HABITAT SUITABILITY MODELLING IN THE NEW FOREST NATIONAL PARK

SARAH JANE DOUGLAS

A thesis submitted in partial fulfilment of the requirements of Bournemouth
University for the degree of Doctor of Philosophy
November 2009

Copyright Statement

Habitat suitability modelling in the New Forest National Park

Sarah Jane Douglas

Abstract

The New Forest National Park is a unique semi-natural landscape which contains many species and habitats which are rare and/or threatened. In order to effectively aid in the conservation of these species, particularly in the face of climate change, there is a requirement to know their habitat requirements and distributions within the New Forest. However, due to limited resources there are gaps in knowledge about this for many of these species.

Habitat suitability modelling was carried out to suggest unsurveyed sites of potentially suitable habitat (and consequently higher likelihood of species occurrence) for selected species of high conservation concern (*Chamaemelum nobile*, *Galium constrictum*, *Gladiolus illyricus*, *Hipparchia semele*, *Nemobius sylvestris*, *Pilularia globulifera*, *Plebejus argus* and *Poronia punctata*). The performance of several modelling approaches was compared. Of the models based on the use of GIS spatial data, an approach requiring only species presence data (Ecological Niche Factor Analysis (ENFA)) was compared to approaches additionally requiring absence or pseudo-absence data (Generalised Linear Models (GLMs) and Generalised Additive Models (GAMs)). An additional approach that did not require GIS data, Bayesian Belief Network (BBN) modelling, was also used to incorporate finer-scale variables not available in GIS format. This relatively new approach to habitat suitability modelling was also used to predict the potential impact of climate change on the suitability of the habitats for the selected species.

The evaluation results showed that the presence-absence GLM and GAM models out-performed the presence-only ENFA method, and that the use of pseudo-absences and automated stepwise variable selection proved effective for developing these models. Species with specialist habitat requirements tended to be modelled more accurately than more generalist species. The BBN models also achieved high evaluation values, and were particularly valuable in being able to provide a

quantitative assessment of the potential impact of climate change on the selected species. Habitat suitability modelling at the scale of an individual predicted area of the size of the New Forest has so far been rare, as have predictions of climate change on specific species at this scale. However, the results of this research show that this can be a valuable approach to aid in management and conservation of species and their habitats in protected areas.

Contents

List of Tables

List of Figures

All photos in this thesis were taken by S. Douglas, except where acknowledged (in Figure A-10).

Acknowledgements

First and foremost I would like to thank my supervisor Adrian Newton for his support and guidance throughout my PhD. I am enormously grateful for the amount of time and advice he has provided. He has helped me to develop into a more confident and able researcher and has shared his enthusiasm for the New Forest and for ecological research, including helping me to search through many a dung pile for nail fungus! Someone else who has lived through this project with me is my husband Euan. As always, he has provided unfaltering love and support every step of the way.

My second supervisor, Anita Diaz, has also been a great help, particularly in starting my fieldwork. There have been many people along the way who helped me with this project. At Bournemouth University, these include Niels Brouwers and Elena Cantarello, as well as Gillian Myers, Sally Keith and Louise Pearson. Valuable help was also provided by Andrew Ford, Kayt Armstrong and Mark Dover in the use of GIS software. I would also like to thank Duncan Golicher, who provided useful discussions and help with R, and Ralph Clarke, who provided statistics guidance.

I could not have done this work without the assistance provided by various individuals and organisations who have been of great help in discussing various aspects of this work and providing information and data. These include Hampshire Biodiversity Information Centre, the Forestry Commission, the New Forest National Park Authority and Hampshire and Isle of Wight Wildlife Trust. I am extremely grateful for the help provided by the species and environmental data providers. I would also like to express my gratitude to the many people who take part in species recording throughout the New Forest. The work that they do is extremely important and this project would not have been possible without them. Further, the individuals who took part in this project as ‘sources of expert knowledge’ in developing some of the models deserve a huge thank-you. They provided rich and informative discussions and the time that they spent in taking part is much appreciated.

Working in somewhere as special and unique as the New Forest has been a privilege and this project has allowed me to experience and appreciate the Forest in a way which I would otherwise have missed. I have really enjoyed my trips into the Forest, especially on the occasions when I was accompanied by my Dad, who had a real knack for spotting things! His help was much appreciated. My family have been hugely supportive throughout this project and Mum and Bella have provided encouragement and support throughout.

I am extremely grateful to Bournemouth University for providing a funded scholarship which allowed me to undertake this PhD. The experience has been a rewarding and enjoyable one.

Author’s Declaration

This thesis comprises only my original work and due acknowledgement has been made in the text to all other material used.

CHAPTER 1

Chapter 1. Introduction

1.1. Background to the biodiversity crisis

There is widespread agreement that biodiversity is being lost globally at an unprecedented rate (Millennium Ecosystem Assessment, 2005). Current rates of extinction are high (at least three orders above average background rates in the fossil record) and accelerating (Pimm *et al.*, 1995). Between 10% and 50% of well-studied higher taxonomic groups (mammals, birds, amphibians, conifers and cyads) are currently threatened with extinction, based on IUCN (World Conservation Union) criteria (Millennium Ecosystem Assessment, 2005).

Conservation of biodiversity is essential as a source of particular biological resources, to maintain different ecosystem services, to maintain the resilience of ecosystems, and to provide options for the future (Millennium Ecosystem Assessment, 2005). Humanity's numbers (and consumption) are increasing, adding to the pressures on biodiversity (Pimm *et al.*, 2001). Changes in biodiversity due to human activities were more rapid in the past fifty years than at any time in human history. The Millennium Ecosystem Assessment (2005) identified the most important (interacting) direct drivers of biodiversity loss and ecosystem service changes as being habitat change (particularly conversion to agriculture), climate change, invasive alien species (and disease), overexploitation of species (especially fishing), and pollution (including nutrient loading).

One of the most important direct drivers of biodiversity loss and ecosystem changes is climate change, which is expected to increasingly affect all aspects of biodiversity (Millennium Ecosystem Assessment, 2005). Although the distribution of ecological systems have repeatedly changed owing to climate change in the past (such as over the Quaternary period), resulting in fluctuations in species ranges through time (Whittaker *et al.*, 2005), the concern with the current climatic change is the rate at which changes are happening, and whether species will be able to respond quickly enough. Observed recent changes in climate, especially warmer regional temperatures, have already had significant impacts on biodiversity and ecosystems, including causing changes in species distributions, population sizes, the timing of reproduction or migration events, and an increase in the frequency of pest and disease outbreaks and the changes are expected to increase (Millennium Ecosystem Assessment, 2005).

The Intergovernmental Panel on Climate Change (IPCC, 2007) project global warming of about 0.2ºC per decade for the next two decades for a range of emissions scenarios. After this the temperature projections increasingly depend on specific emissions scenarios but range from an increase in global average surface temperature between 1.1ºC to 6.4ºC at 2090-2099 relative to 1980-1999. The different emission scenarios project a sea-level rise between 0.18 and 0.59 metres in this time. The IPCC (2007) also suggest that it is very likely that hot extremes, heat waves and heavy precipitation events will become more frequent. Although some ecosystem services in some regions may initially be enhanced by projected changes in climate (such as increases in temperature or precipitation), the scientific evidence suggests there will be a significant net harmful impact on ecosystem services worldwide (Millennium Ecosystem Assessment, 2005).

Some species may be able to cope with landscape transformations and exploit the new habitats (Diack, 1999; Santos *et al.*, 2006). However overall, climate change is projected to exacerbate the loss of biodiversity and increase the risk of extinction for many species, especially those already at risk owing to factors such as low population numbers, restricted or patchy habitats, and limited climatic ranges (Millennium Ecosystem Assessment, 2005). In combination with land-use change and the spread of exotic or alien species, the capability of species to migrate and the ability of species to persist in fragmented or altered habitats are likely to be limited (Millennium Ecosystem Assessment, 2005). Thomas *et al.* (2004) also suggest that many of the most severe impacts of climate change are likely to stem from interactions between threats, rather than from climate acting in isolation.

There are huge inadequacies in our knowledge of the planet's biodiversity and this is hampered largely by the fact that many species have yet to be formally described and catalogued. This knowledge gap is termed the 'Linnean shortfall' (Pimm *et al.*, 2001; Balmford *et al.*, 2005; Whittaker *et al.*, 2005). In addition, there are also inadequacies in our knowledge of the global, regional and local distributions of many species, termed the ‘Wallacean shortfall’ (Whittaker *et al.*, 2005). Therefore, it is uncertain how we will know whether targets such as the Convention on Biological Diversity’s 2010 target (to ‘…achieve by 2010, a significant reduction of the current rate of biodiversity loss at the global, regional and national level…’; (United Nations Environment Programme, 2002) are being met (Balmford *et al.*, 2005).

Measuring and monitoring are required to quantify and assess biodiversity and its loss, particularly where there are gaps and weaknesses, so that habitats and species can be managed effectively. In order to do this it is valuable to know where organisms occur and what their habitat requirements are. However, answers are required quickly and resources are limited; therefore representative sampling, rather than measuring every population or extent of habitat, is likely to be most suitable in the majority of instances (Balmford *et al.*, 2005). Species distribution modelling can provide a useful way of directing such sampling, as well as predicting possible future distributional changes and therefore informing conservation priority setting (Wilson, 2003; Guisan and Thuiller, 2005).

1.2. Species distribution modelling/habitat suitability modelling

Knowledge of biodiversity is incomplete (Wilson, 2001; Balmford *et al.*, 2005), but in order to conserve and manage biodiversity effectively it is important to measure and monitor habitats and species. Given that it is often not possible to know the full distribution of many species (especially those which are difficult to find, or not well known), frequently due to a lack of resources, it is useful to be able apply what is known about their identified occurrence and ecological requirements to map their likely distribution.

The factors affecting where a species occurs are numerous, but the occurrence of a species is ultimately determined by its unique adaptations and how it tolerates physical conditions (e.g. temperature, light levels, soil chemistry) and biotic variables (such as interactions with other species) (Orians and Groom, 2006). However, the availability of habitat is one of the primary factors influencing the distribution and abundance of organisms, so an understanding of distribution patterns through the use of habitat analysis is very useful (Lindenmayer and Franklin, 2002). A simple definition of what constitutes ‘habitat’ is provided by The UK Biodiversity Steering Group (1995), who consider each habitat to be ‘an assemblage of plants and animals found together, as well as the geographical area and features on which they exist.’ However, other factors such as suitable resources necessary for feeding, mating, dispersal and so on, will also be important (Smith *et al.*, 2007).

The ability to reliably predict the potential occurrence of species (i.e. their potential habitat) has a wide range of applications. In addition to suggesting unsurveyed sites of high potential of occurrence for rare species and therefore reducing the time and costs of field surveys, these uses include quantifying the environmental niche of species and assessing the impact of climate, land use and other environmental changes on species distributions, as well as examining the invasive potential of non-native species, aiding in reserve design and identifying areas with a high potential for (re)colonisation (Guisan and Thuiller, 2005; Lütolf *et al.*, 2006; Rodríguez *et al.*, 2007). Such approaches could become increasingly useful for predicting the possible colonisation of new areas by species in response to climate change and for aiding conservation planning for the future (Huntley *et al.*, 2004).

Species distribution models, or habitat suitability models, attempt to provide predictions of distributions by relating presence or abundance of species to environmental predictors (Guisan and Zimmermann, 2000; Elith *et al.*, 2006). Information on species locations and corresponding environmental factors is used to generate statistical functions that can then be projected to areas where environmental factors are known but the species have not been sampled, providing predictions of potentially suitable habitat (Brotons *et al.*, 2004). The development of modelling techniques and the increasing availability of large-scale environmental information in digital format has led to a large number of studies of this nature (Guisan and Zimmermann, 2000; Brotons *et al.*, 2004).

1.2.1. Approaches to habitat suitability modelling

Habitat suitability models can be generated using a wide variety of approaches and these are reviewed in detail elsewhere (e.g. Guisan and Zimmermann, 2000; Elith and Burgman, 2003; Elith *et al.*, 2006; Newton, 2007). Methods can be broadly split into those that require only presence data and those that require both presence and absence data. The main presence-only approaches include: conceptual models based on expert opinion (such as habitat suitability indices (e.g. Yamada *et al*. (2003) and Uhmann *et al*. (2001)) and the relatively new application of Bayesian network models (e.g. Marcot (2006), Smith *et al*. (2007) and Newton (2009a)); environmental (or climatic) envelope models (e.g. Pearson and Dawson (2003) and Huntley *et al.* (2004)), including ecological niche factor analysis (ENFA; e.g. Hirzel *et al*. (2002); see Chapter 2 for more details); and machine-learning methods, which can use presence only data (but employ the background/non-presence as absence) or presence-absence data (such as artificial neural networks, ANN (Lek *et al.*, 1996), genetic algorithm for rule-set prediction, GARP (Stockwell and Peters, 1999) and maximum entropy models, Maxent (Phillips *et al.*, 2006)). The main presence-absence approaches include: regression analysis (usually logistic regression), including generalised linear models, GLMs, and generalised additive models, GAMs (e.g. Pearce and Ferrier (2000b); see Chapter 3 for more details) and tree-based methods (classification and regression trees; e.g. Moisen and Frescino (2002)).

There are also many other modelling approaches in addition to these, including several newer methods that are starting to be used in habitat suitability modelling (e.g. Elith *et al.* (2006) and Austin (2007)). However, Austin (2007) suggests that these newer approaches require further development and investigation before they can be routinely used. The different approaches vary in how they model the distribution of the response, select relevant predictor variables, define fitted functions for each variable, weight variable contributions, allow for interactions, and predict geographic patterns of occurrence (Guisan and Zimmermann, 2000; Elith *et al.*, 2006). Each approach has its advantages and disadvantages and has been shown to work well in some studies, but not as well in others, as can be seen in the examples below.

1.2.2. Model comparison studies

There have been many studies comparing the performance of different species distribution models on a wide variety of species in many different locations. For example, Elith *et al.* (2006) examined some newer modelling approaches in their comparison of 16 modelling methods (including presence-only and presence-absence approaches) applied to 226 species from 6 regions around the globe. They found that the highest performing group of methods were multivariate adaptive regression splines (MARS) community models, boosted regression trees, generalised dissimilarity and maximum entropy models. An intermediate group of methods included most of the standard regression methods (GAM, GLM, BRUTO, OM-GARP, and MARS). The poorest performing group of models included the three methods tested that use only presence data (BIOCLIM, LIVES, DOMAIN) with no inferred absences, as well as desktop GARP and the MARS individual models fitted with interactions. The relative rankings of methods were broadly consistent across the six regions in the study. The greatest variation in the performance of different methods was apparent at the species level, but the same trends in model performance were still found.

Novel modelling methods (such as boosted regression trees and community models) were found by Elith *et al.* (2006) to generally outperform established methods, which they suggest results from their ability to fit complex responses (often including interactions) and select a relevant set of variables. They suggest that it may be the approaches used for building the more established methods, such as model selection, that may mean they do not perform as well as some of the newer methods. Encouragingly, Elith *et al.* (2006) found that predictions based on presence-only data can be sufficiently accurate to be used in conservation planning and in numerous other applications in which estimates of species' distributions are relevant. They also found that predictive performance did not vary consistently with the number of presence records.

A range of modelling methods were also compared by Segurado and Araújo (2004) who examined the performance of the Gower metric, ecological niche factor analysis (ENFA), classification trees, neural networks, generalized linear models, generalized additive models and spatial interpolators, for predicting occurrence of 44 species of amphibians and reptiles in Portugal. They found that artificial neural networks performed generally better, immediately followed by generalized additive models including a covariate term for spatial autocorrelation. The poorest performing methods were DOMAIN and ENFA followed by GLM. However, although some methods generally performed better than others, they noted that no method was superior in all circumstances. Segurado and Araújo (2004) suggest that these apparently divergent patterns of model' performance are likely to be related to variations in the methods' abilities to recover useful relationships between species with different distributions and environmental factors with different strengths and lengths of gradients. They therefore recommend that choice of the appropriate method should be contingent on the goals and kinds of distributions being modelled.

Elith and Burgman (2002) compared four approaches, including those requiring presence-only data and those requiring presence-absence data (the bioclimatic envelope method ANUCLIM, GLMs, GAMs and GARP), to model the distribution of eight plant species in the Central Highlands of Victoria, Australia. Their results indicated no consistent significant difference among the discriminatory performance of predictions from the four different modes, although there was an apparent trend towards better discrimination from the GAMs and GLMs. However, they suggest that, given a more complete and accurate modelling set and more extensive validation data, significant differences between the methods would emerge and modelling success was improved. As found by Elith *et al.* (2006), the difference in modelling success between species tended to be more pronounced than the differences between modelling methods, with some species modelled with sufficient accuracy for the information to be useful in land-use decisions, whereas other species' models were unlikely to be useful.

Logistic regression (GLM), Mahalanobis distance method, classification and regression tree (CART) analysis, and discriminant function analysis were compared by Dettmers *et al.* (2002) for 25 common bird species for an area of over 22 million hectares extending from western Virginia south to northern Georgia in the US. They found that while discriminant function models are likely to distinguish between presence and absence locations within the original data set, such models were unlikely to perform well when tested on new data sets. Both logistic regression and Mahalanobis distance models appeared to be good general methods for predicting probabilities of occurrence for new locations, although they may not perform as well at predicting absolute presence/absence. Dettmers *et al.* (2002) note that logistic regression is a more completely developed technique that has been extensively described by statisticians, whereas the Mahalanobis distance method is not yet a commonly used method and lacks well-defined procedures for model assessment and determination of significant variables in a given model.

Engler *et al.* (2004) compared ecological niche factor analysis (ENFA) and GLMs using pseudo-absences for the endangered plant species *Eryngium alpinum*. Their results showed that GLM models using ENFA-weighted pseudo-absences generally provide better results, which they suggest could be due to the lack of discriminating absences in the ENFA or owing to violation of the assumption of normality of predictors that is required by the ENFA algorithm.

GLMs and ENFA were also compared by Brotons *et al.* (2004) for modelling forest bird species. These authors found that GLMs predicted the distributions with higher accuracy than ENFA. They also found there was a general tendency for ENFA to overestimate the spatial extent of the distributions, especially on the edges of those estimated by GLM, and in some cases areas estimated to have high species' probabilities of occurrence were overlooked by ENFA. Both GLM and ENFA methods seemed to perform equally well on more marginal species (Brotons *et al.*, 2004). Olivier and Wotherspoon (2006) also found that ENFA proved slightly less accurate than GLMs for modelling habitat selection for the snow petrel in Antarctica. However, they suggested that ENFA predictions can provide a useful starting point to categorize habitat suitability prior to the application of other modelling techniques.

Brotons *et al.* (2004) conclude that due to differences in data and application areas it will be impossible to identify one modelling method as universally applicable so the goals and assumptions should be made clear before being applied to particular situations (Segurado and Araújo, 2004). Regarding the use of presence-only or presence-absence data, they suggest that when absence data are not available, or when the main objective of the modelling is to identify overall suitable areas for a given species, methods such as ENFA appear useful, but otherwise presence-absence methods should be used.

This conclusion is echoed by Hirzel *et al*. (2001), who reported equivalent model performance between ENFA and GLM based on a virtual species at equilibrium but found that ENFA could be superior to GLM in a scenario where a species is spreading. They suggest that the robustness of ENFA to data quality and quantity makes it particularly suitable and efficient when the quality of data is either poor (the absence data are unreliable) or unknown, but GLM offers slightly better results when the available presence/absence data are sufficiently good.

Olivier and Wotherspoon (2006) also noted that whilst the GLM approach enabled them to build explanatory models which incorporated the effect of ecological processes (such as snow petrel coloniality and interactions between environmental predictors), the environmental envelopes created with ENFA provided fundamentally different information, potentially expressing results closer to the realised niche of the species.

Similarly, Zaniewski *et al*. (2002), who compared ENFA and GAM, found that while GAM models (using pseudo-absences) appeared to predict species distributions from presence-only data more accurately than ENFA, they appeared less effective than ENFA in highlighting areas of potential biodiversity 'hotspots' from summing of species predictions. Encouragingly, they suggest that in circumstances where species presence/absence data collected from a systematic stratified survey is unavailable and/or unattainable, GAM and ENFA show potential for predicting species spatial distributions from presence-only data sets, such as herbarium or museum collections. Further studies on ENFA and GLMs and GAMs can be found in Chapters 2 and 3.

As these examples illustrate, many modelling techniques are now available for species distribution modelling. There is increased recognition that different techniques yield different results, even when models are calibrated with the same response and predictor variables, and that variability also arises from using different implementations of the same technique (Araújo and Guisan, 2004; Elith *et al.*, 2006). Factors that affect parameterizations include the type of variable selection strategy used, the way absences are estimated, and the way spatial structures are considered (Araújo and Guisan, 2004). It is therefore not surprising, given that comparisons of methods undertaken by different authors usually use different data and predictors and may use different methodology or interpretation (Austin, 2007), that the results of comparisons between approaches are often contradictory.

The fact that there are various sources of error and uncertainty and that different analyses, spatial resolutions, scales, modelling and evaluation methods can produce very different predictions, represents a challenge to modelling (Thuiller, 2004; Guisan and Thuiller, 2005). As discussed by Johnson and Gillingham (2005), it is unlikely that one type of model is best for all situations, so the particular conservation issue and ecology of the focal species should guide the selection of a technique, within the limitations of the data. Austin (2007) notes how, given the importance of knowledge of species distribution for conservation and climate change management, continuous and progressive evaluation of the statistical models predicting species distributions is necessary.

1.2.3. Environmental predictors and scale

Species' distributions are related to needs for food, shelter, reproduction, and biotic interactions such as minimising competition and avoiding predation, and their habitat requirements may vary over their lifecycle, which can make habitat modelling more complex (Mackey and Lindenmayer, 2001). Environmental predictors can exert direct effects (causal or resource) or indirect effects (which have no direct physiological impact on growth or survival processes) on species and are optimally chosen to reflect three main types of influence on the species: limiting factors (or regulators), defined as factors controlling species eco-physiology (e.g. temperature, water, soil composition); disturbances, defined as all types of perturbations affecting environmental systems (natural or human-induced); and resources, defined as compounds that can be assimilated by organisms (e.g. energy and water) (Austin and Meyers, 1996; Guisan and Zimmermann, 2000; Huston, 2002; Guisan and Thuiller, 2005).

Modelling success depends on using appropriate environmental predictors, but prediction becomes increasingly robust and less location-specific as the predictor variables become more process-oriented and relevant to biological processes (Austin and Meyers, 1996). However, the need to use predictors for which estimates are available for unsampled regions may limit the choice to less direct variables (Austin and Meyers, 1996). Indirect, or surrogate, variables, such as altitude are correlated with species occurrence through their correlation with variables such as temperature and rainfall that can have a physiological impact on organisms (Austin, 2007). It is often difficult to incorporate many influential variables because information on them may not be available spatially, for example intraspecific interactions (Wilson, 2003).

In addition, it can be difficult to test the effects of a specific variable as many different environmental conditions can often be correlated with a specific property, making distinguishing the 'causal' predictor from other correlated factors that have no causal role difficult (Huston, 2002). Significantly, it is not possible to infer causation from correlation, owing to the complex and indirect interactions operating, but a lot can still be learnt about ecological relationships, providing modellers are aware of the complexities and limitations (Johnson and Gillingham, 2005).

Model performance generally improves with addition of information, although plateaus exist wherein new information adds little to model performance (Stockwell, 1997; Peterson and Cohoon, 1999; Stockwell and Peterson, 2002b). In fact, in some systems accuracy can be reduced with increasing information (Stockwell, 1997; Stockwell and Peterson, 2002b), usually because the additional variables produce models optimised for a too-specific set of information that performs poorly on new data, a problem recognised as prediction bias or over-fitting (Stockwell and Peterson, 2002b). However, the effect of including too few environmental variables in a modelling effort is that factors critical to limiting species' distributions in space may be omitted, and predicted geographic distributions may be too large (Peterson *et al.*, 2002).

Species distribution models largely rely on the ecological niche concept, clarified by Hutchinson (1957), where the niche of an organism (the *n*-dimensional niche) is described by *n* axes, each representing an environmental variable. A distinction is also made between the fundamental niche, which is primarily a function of physiological performance and ecosystems constraints, and the realised niche (which is actually observed in nature), which additionally includes biotic interactions and competitive exclusion, reducing the realised niche relative to the fundamental niche (Guisan and Zimmermann, 2000; Pulliam, 2000; Guisan and Thuiller, 2005). Species distribution models are generally based on the species' fundamental niche so some areas indicated by models as regions of potential presence may be occupied by closely related species, or may represent suitable areas to which the species has failed to disperse or in which it has gone extinct (Anderson *et al.*, 2003). However, this 'over-prediction' can be advantageous, for example, by suggesting possible areas with high potential for (re)colonisation (Lütolf *et al.*, 2006).

Owing to the complex nature of many species-habitat relations it is often very difficult to disentangle the influencing factors operating, as presence of a species in a certain environment does not necessarily imply conditions suitable for long-term survival and reproduction and suitable habitat does not automatically mean occupation (Schultz *et al.*, 2003). Additionally, probabilities of occurrence may not indicate habitat quality for a species and may not be related to species persistence (Wilson *et al.*, 2005). Species may be absent from suitable habitat and present in unsuitable habitat (Pulliam, 2000). However, Engler *et al.* (2004) suggest that many rare and threatened species will tend to occupy most of their potential habitats, as these have usually been drastically reduced and, as a result, cover only a small proportion of the territory. Alternatively, rare species may be absent from seemingly suitable habitat because they are rare, as organisms may frequently be absent from suitable habitat because of local extinctions and/or dispersal limitation (Pulliam, 2000). However, the absence of species can be difficult to conclude with certainty (Wiens, 2002; and see section 1.2.4).

An important consideration is that species distributions as we observe them today may not be in equilibrium with factors such as the current climate, and the presence of physical barriers to dispersal may prevent some species from occupying their full climatic niche (Pearson and Dawson, 2003). There is increasing evidence which suggests that for many species there is a considerable time lag between changes in climate and changes in distribution and that during much of this time, species may be absent from large portions of their potential geographical ranges (Pulliam, 2000). Further, as noted by Brotons *et al.* (2004), change in habitat may occur rapidly, for example as the result of a management intervention. Such changes may result in species not using areas identified as suitable habitats, or being present in lower suitability areas. Therefore, Brotons *et al.* (2004) recommend caution in the use of habitat suitability methods if there is a strong likelihood of non-equilibrium conditions.

The assumption of equilibrium between the environment and observed species patterns is an important limitation of static distribution models because non-equilibrium is more realistic in ecology (Guisan and Zimmermann, 2000). However, an equilibrium assumption greatly simplifies the modelling approach and allows the production of the quick and accurate predictions needed for management purposes (Lehmann *et al.*, 2002).

Another consideration in developing species distribution models is scale. Two aspects of scale are extent, the area over which the study is carried out, and resolution (or grain), the size of the sampling unit at which the data are recorded (Austin, 2007). There are therefore three different spatial scales that can characterise wildlife habitat relationship models: the grain of the species' distribution data; the grain of the habitat variables; and the extent of the study area; with changes in any of these likely to affect the predictive ability of models (Tobalske, 2002).

Importantly, the resolution and size/extent of the study should be appropriate for the phenomena being studied and the hypotheses being tested (Huston, 2002). However, in the majority of cases the resolution and size are determined by what time and resources are available and the homogeneity of the study area (Elith and Burgman, 2002) and the species of study (Guisan and Thuiller, 2005). The resolution of resulting distribution maps will affect their application, for example, whether at a broad scale or local scale conservation planning and management (Elith and Burgman, 2003), so this needs prior consideration. However, patterns (such as species presence/absence) can be detected at resolutions far coarser than the resolution needed to understand the processes that produce the pattern, and therefore the resolution needed to predict the pattern (Huston, 2002).

1.2.4. Sampling and sample size

The sample size required for developing a species distribution model is an important issue, as decreasing the number of data points required for the model could increase the proportion of species for which species distribution modelling could be carried out (Stockwell and Peterson, 2002b). As noted by Cayuela *et al*. (2009), paradoxically, predictive modelling could be most useful in those cases in which data are fewer and the least reliable.

Pearce and Ferrier (2000b) found that the size of sample used to develop a model had the greatest influence of any factor on the predictive accuracy of their logistic regression models. In general, the larger the sample size the better and there are several reasons why model performance generally decreases with sample size, including increased levels of uncertainty associated with parameter estimates (e.g. means, modes, medians, predicted probabilities of occurrence), with outliers carrying more weight in smaller sample sizes and the fact that large numbers of samples may be needed to allow for accurate description of the range (and complexity) of conditions over which a species occurs (Wisz *et al.*, 2008). For example, McPherson *et al*. (2004) found that increases in training data sample size (through from 50 to 500 points) improved the fit of their models (including logistic regression models) by reducing both false positive and false negative errors.

Vaughan and Omerod (2005) suggest that although sample size guidelines are difficult to formulate, 200–300 sites or more are desirable for a test set, whereas Harrell (2001) suggests that a test set should contain at least 100 sites of the less common event (present/absent). However, Vaughan and Omerod (2005) also note that test set size also relates to the generalisability desired from the model, with greater numbers of test sites likely to be necessary where wider generalisability is needed from a model. This makes daunting demands on field data collection, yet may be essential to provide robust assessments of predictive performance.

For logistic regression models Pearce and Ferrier (2000b) found that a sample size of 50 is too small to allow development of accurate logistic regression models, suggesting that at least 250 sites be available to model the regional distribution of species. However, Stockwell and Peterson (2002b) show that fewer data points than suggested by Pearce and Ferrier (2000b) are required to achieve maximum accuracy in modelling species distributions using logistic regression, instead suggesting a set of 100 data points could be sufficient.

Further, Jimenez-Valverde and Lobo (2007) note that a sample size lower than around 70 observations decreases logistic regression model performance. However, Wisz *et al*. (2008) have shown that the required sample size will be dependent on numerous factors. They found that very different predictions can be obtained depending on region, sample size, and the algorithm used and that no algorithm (from a range of approaches including GAM, Maxent and GARP among others) performed better than all others across a range of sample sizes (10, 30 or 100 training data). Some approaches (such as Maxent) were less sensitive to sample size, and generally outperformed other methods at the smallest sample size. However, predictions made for large sample sizes generally outperformed those at the smallest sample size.

Wisz *et al*. (2008) also note that the degree of predictive accuracy necessary depends on the intended use of the model (Araújo *et al.*, 2005) and that predictions based on small samples are generally unlikely to be suitable for conservation planning and other complex applications. However, they may yield results useful in prioritizing future data collection efforts for rare species. Nonetheless, particular caution should be applied to predictions made from small sample sizes. The sample size also affects the number of predictor variables that can be used, which may also affect model accuracy. A rule of thumb often used in regression models suggests no less than ten observations per number of variables (Harrell *et al.*, 1996; see section 3.2.2.2, Chapter 3).

However, the required sample size will also depend on the variation (and complexity of response) in conditions in which the species occurs (Wisz *et al.*, 2008) and the quality (spatial resolution and locational accuracy) of the data, which Engler *et al.* (2004) found to be more important than quantity (number of occurrences).

Using restricted data (i.e. not capturing the full environmental range of species) strongly reduces the combinations of environmental conditions under which the models are calibrated and reduces the applicability of the models for predictive purposes, such as future projections of species distributions (Pearson and Dawson, 2003; Thuiller *et al.*, 2004). However, the kind of *ad hoc* data that is often available is unlikely to capture the full environmental range of species', although existing data could be supplemented with new data to ‘fill in’ some of the gaps (Araújo and Guisan, 2004). Importantly, models using this kind of restricted data should not be used to project outside of those environmental limits (Thuiller *et al.*, 2004).

There are several issues with using available non-systematic *ad hoc* survey data, such as small sample size (number of sites) and bias. There are many forms of bias that may be introduced into sampling and modelling. Bias can be introduced through the tendency of sampling to be more frequent in certain geographic features, for example, collections along roadsides or the tendency for more inaccessible habitat types to be underrepresented (Stockwell and Peterson, 2002a). Spatial bias of observations to urban areas is often likely in *ad hoc* samples (Araújo and Guisan, 2004) and butterfly distribution maps, for example, have been found to reflect not just the distribution of individual species but also recorder effort (Dennis *et al.*, 1999). However, the ideal of an unbiased stratified survey based on an adequate sampling frame is not usually possible because of cost constraints (Austin and Meyers, 1996). Therefore, although there are issues in using available *ad hoc* survey data, the advantages of the savings in time and resources of carrying out surveying specifically for the modelling far outweigh these (Pearce and Ferrier, 2000b). However, such data must be used critically (Elith *et al.*, 2006).

Another important factor in species distribution modelling is the type of species data available to develop the model, as this will determine which modelling techniques are employed. Species data can be either presence-only data, where only the occurrences of the target species are available, or presence/absence data, where each site is carefully monitored to determine with sufficient certainty whether the species is present or absent. As discussed by Wiens (2002), there is a problem of interpreting ‘presence’ or ‘absence’ in ecologically meaningful terms. For example, if a species is present in an area, does that mean that the habitat is suitable (and suitable long term)? If areas are surveyed only once (as is often the case), records of presence may include transients that do not really ‘belong there’. Conversely, if a species is recorded as absent from a location or habitat, it could be that the species was present but was not detected, or that for historical reasons the species is absent even though the habitat is suitable, or because the habitat is truly unsuitable for the species (Hirzel *et al.*, 2002). The absence of species can be difficult to conclude with certainty and as noted by Hirzel *et al*. (2006b) the reliability of absences will depend on the species characteristics (e.g. biology, behaviour, history), their local abundance, ease of detection, and the survey design.

Reliable absence data takes much longer to obtain because to be confident that a species is absent from a site may require monitoring at different times of day or year and it may still be missed when it was actually present, particularly if it is very inconspicuous, difficult to identify or appears only for a very short time. Therefore, in part due to time and/or financial constraints as well as to data collection strategies aiming at inventories instead of statistical analysis, the majority of data available consists of presence-only data sets collected on an *ad hoc* or non-stratified basis, even though this type of data is the most difficult to successfully incorporate into statistical modelling methods (Zaniewski *et al.*, 2002).

1.2.5. The effect of species' ecological characteristics on the performance of habitat suitability models

Particular ecological characteristics of species may affect the accuracy or efficiency of habitat suitability or species distribution models (Stockwell and Peterson, 2002b). For example, Hepinstall *et al.* (2002) suggest that species' niche breadth is important because generalist species that use many different habitats could be predicted to occur everywhere by habitat association methods, whereas species with narrower niches are more likely to be accurately predicted. Tsoar *et al.* (2007) additionally found that distribution ranges of species with restricted ecological niches can be modelled with higher accuracy than those of more generalist species. Others (e.g. Cowley *et al.*, 2000; Hepinstall *et al.*, 2002; Brotons *et al.*, 2004; Hernandez *et al.*, 2006; Brotons *et al.*, 2007) have also reported that species with restricted ecological niches can be modelled with higher accuracy than more generalist species. Brotons *et al*. (2007) suggested that species which have wider distributions or use a wide range of habitats in one area might not be limited by the measured predictive factors at the scale at which the models are fitted. Cowley *et al*. (2000) also found that the best performing models were for sedentary species that had strong habitat associations and were widespread within those habitats. An effect of species mobility on the performance of species distribution models was also reported by Pearce *et al.* (2001), who found that the discrimination ability of generalised additive models was highest within the less mobile biological groups, particularly vascular plants and small reptiles, whereas performance was poorer for locally nomadic or highly mobile species.

In some cases, reports of range size (or specialist distribution) having an effect on the accuracy of models may potentially be (at least partly) an artefact of sample size and prevalence (Elith *et al.*, 2006; see Appendix 13.1.2). When accounting for this, McPherson and Jetz (2007) found that certain ecological traits, including habitat tolerance and range size, exert real effects on the accuracy of species distribution models. Other traits that decreased model performance included, amongst others, migrant traits or other temporal or spatial variation in habitat associations.

1.2.6. Evaluation of habitat suitability models

Evaluating the predictive performance of habitat suitability models is a vital step in model development as it tests the predictive ability of the models and provides a basis for model comparisons (Pearce and Ferrier, 2000a). Araújo *et al.* (2005) emphasize the need to use independent data for evaluation of species distribution models. They describe three methods of evaluation: resubstitution, where the same data is used to calibrate the model and measure the fit; data-splitting, where the data is split in two at random – a calibration set and an evaluation set; and independent validation, where a totally independent data set from a different region is used (Austin, 2007).

The third alternative does not seem plausible in the current context, as separate regions with the same species complement, ranges and combinations of environmental predictors and ecological history simply do not occur (Austin, 2007). In addition, for the majority of studies (due to time and money constraints) there is only one dataset available (Hirzel *et al.*, 2006b). In fact, Lehmann *et al.* (2002) suggest that it is not clear whether independent datasets are really preferable, even though this is generally claimed. Further, they suggest that by using entirely independent datasets there is a risk of comparing different sampling strategies instead of evaluating a model. Elith *et al.* (2006) adopt a useful compromise by calibrating with one set of data then evaluate the fit with totally separate data collected independently from the same region.

A problem with the resubstitution approach is that models may overfit to the calibration (training) data, leaving users unable to judge whether high accuracy on nonindependent data reflect good predictive accuracy on independent data sets (Araújo *et al.*, 2005). Data-splitting strategies are the current preferred method and have generally been accepted to provide more robust measures of predictive success than resubstitution (Araújo *et al.*, 2005; Austin, 2007). In the data-splitting methods some known occurrences are withheld from model development and accuracy is assessed based on how well models predict the withheld data (Elith *et al.*, 2006). The simplest data-splitting approach is to randomly split the data into a training set and a testing set, of which the former is the largest. There are several alternative techniques including, k-fold partitioning (or grouped cross-validation), leave-one-out and related jack-knifing procedures (where each observation, or partitioned group of observations is omitted in turn and the model is calibrated over all other observations) and bootstrapping (random sampling with replacement) (Fielding, 2002; Araújo *et al.*, 2005). Bootstrapping can be used to repeatedly select a set sample size from a dataset to compute a statistic of interest (such as the area under a receiver operating characteristic curve and assess how it behaves over a certain number of repetitions (Harrell, 2001).

A problem with the data-splitting approach is that the reduction in the size of the training data set will usually produce a corresponding decrease in model accuracy (Fielding and Bell, 1997). There is therefore a trade-off between having a large calibration (training) data set, which is likely to result in a more accurate model, or a large evaluation (testing) set that gives good assessment of model performance (Fielding and Bell, 1997). In cross-validating their GLM and ENFA models, Brotons *et al.* (2004) divided their data into two different sets, by randomly assigning 70% of the data for each species to a calibration data-set and 30% to an independent evaluation data set. A random 70% calibration and 30% evaluation data split was also used by Thuiller (2003). There are numerous measures for assessing the predictive performance of species distribution models on the evaluation data set, dependent on whether presence-only or presence-absence data are used and these are detailed in the following chapters and Appendix 8.

1.3. Project outline and thesis structure

1.3.1. Project objectives

1. To define the potential distribution of selected species of conservation concern in the New Forest by carrying out habitat suitability/species distribution modelling.
2. To compare the use of different approaches for modelling potential habitat suitability (species distribution modelling) for the selected species in the New Forest.
3. To examine the potential impacts of climate change on the selected species and their habitats in the New Forest.

1.3.2. Context

The New Forest is a unique and extremely valuable landscape (see Appendix 1 for further details), which contains many species and habitats that are rare and/or threatened as a result of drivers of biodiversity loss such as habitat change, climate change, invasive alien species (and disease), and pollution, and needs to be conserved appropriately and effectively for the future.

The habitats of the New Forest include ancient pasture woodland, lowland heath, grassland, valley and seepage step mire, or fen. The unenclosed (pasture) woodlands extend to some 4,430 hectares (excluding riverine and bog woodland) and are dominated by oak (*Quercus robur*) and beech (*Fagus sylvatica*), with some trees of early 17th century origin (English Nature, 1996; Wright and Westerhoff, 2001).

The New Forest contains approximately 19,500 hectares of lowland heath, the largest area of this rare habitat remaining in the UK (New Forest National Park Authority, 2006b). The heathlands comprise a series of plant communities, including the dry heath (and associated dry grasslands), which grades into the wetter humid heath (and associated valley mires, streams, ponds, temporary pools and wet grasslands) (English Nature, 1996; Tubbs, 2001; Wright and Westerhoff, 2001).

Within the heathland mosaic, on pockets of richer soils, acid grassland can occur. The more neutral grasslands (or lawns) vary with factors such as soils, topography floodwater nutrient quality and grazing pressure and the species present reflect this (English Nature, 1996). The New Forest also contains ninety separate valley mires; there are no more than twenty in the rest of the English lowlands, and only a handful around the European littoral from Denmark to Spain (Tubbs, 2001). This international importance is reflected in the New Forests' designation as a Ramsar site (a wetland of international importance).

Monitoring of biodiversity in the New Forest is required as a result of its designation as a Special Area of Conservation (SAC) Natura 2000 site, meaning it receives strict protection under the EU Habitats Directive. In order to conserve species for the future, and monitor changes in their distributions and status, it is important to know where they currently occur and the occurrence of potentially suitable habitat.

In the New Forest, as in many other areas, and for many taxonomic groups, the need for distributional information far outstrips the resources available for collection of field data. Collecting distributional data over extensive areas is resource and time intensive (Cowley *et al.*, 2000). Even for well-studied groups, such as butterflies, records are biased towards accessible areas (Dennis *et al.*, 1999; Cowley *et al.*, 2000). Despite being easily accessible and having many conservation designations, there are still many gaps in knowledge about many of the species found in the New Forest and their patterns of distribution. Chatters (2006) discusses how 'the scale and diversity of habitats of the New Forest National Park are still not fully understood and further survey and analysis are needed to gain an adequate understanding of what the National Park contains and whether current designations are adequate. However, such habitat information is far advanced compared with the data on many individual species ... the data sets and analysis of the importance of many of these important species across the National Park are still far from adequate'. Much of the species data are patchy and often recorded on an *ad hoc* basis, with biases towards easily observed species in accessible locations, in large part due to lack of time and resources.

Species distribution (or habitat suitability) modelling can be used to identify unsurveyed sites of high potential occurrence for species, so that time and resources can be directed more efficiently towards these areas. Developments in geographical information systems (GIS) have resulted in the potential for great increases in both the quality and quantity of habitat-level information that can be obtained and analysed (Cowley *et al.*, 2000). This, combined with development of a range of modelling techniques (Guisan and Zimmermann, 2000), suggests that species distribution modelling offers a valuable approach for improving current knowledge of the distribution of species.

Although there are some issues and limitations in their use, species distribution models have been shown to work well in a wide range of applications (e.g. Cowley *et al.*, 2000; Ferrier *et al.*, 2002b; Berry *et al.*, 2005a; Guisan and Thuiller, 2005; Elith *et al.*, 2006; Guisan *et al.*, 2006a; Chefaoui and Lobo, 2007; Matern *et al.*, 2007; Sattler *et al.*, 2007; Wollan *et al.*, 2008). However, species distribution modelling methods have generally been applied to large spatial scales (in both extent and resolution), such as countries (Zaniewski *et al.*, 2002; Guisan and Hofer, 2003; Sérgio *et al.*, 2007; Thomaes *et al.*, 2008; Lachat and Butler, 2009; Puddu *et al.*, 2009) or large regions (Brotons *et al.*, 2004; Santos *et al.*, 2006; Chefaoui and Lobo, 2007; López-López *et al.*, 2007). There have been far fewer examples of smaller scale applications built to address local conservation issues (Seoane *et al.*, 2006), in particular at the scale of an individual protected area or landscape (but see for example Gibson *et al.*, 2004; Seoane *et al.*, 2006; Fei *et al.*, 2007; Podchong *et al.*, 2009). Although larger scale models can be used to predict coarse distributions, it is at local scales where conservation management decisions are typically taken.

In order to test the application of species distribution modelling methods to an individual protected area, this study focuses on species of high conservation value in the New Forest (as listed in the SAC management plan (Wright and Westerhoff, 2001)). This also provides the opportunity to test the application of predictive models to rare and endangered species, of which there have been relatively few studies (Engler *et al.*, 2004; Guisan *et al.*, 2006a; Matern *et al.*, 2007). If models can be shown to work well for the selected species, it is anticipated that it may then be possible to apply them to other species whose distributions are less well understood. Ultimately, it is hoped that the results of these models will be useful to those trying to survey and monitor species and effectively manage the biodiversity of the New Forest. At the same time, it is hoped that the research will provide some findings of general relevance to the implementation of species distribution modelling approaches at the scale of individual protected areas.

In attempting to predict species' potential distributions (or habitat suitability) to aid conservation management, it would be naïve not to consider the impact of (future) environmental change, namely climate change, on those distributions. Climate change and the other pressures facing the New Forest are highlighted in New Forest management plans (e.g. New Forest Committee, 2003; Forestry Commission, 2007; New Forest National Park Authority, 2008), although there has been very little specific research into the impacts of climate change in the New Forest (except for an investigation into temporary pond macroinvertebrate communities (Ewald, 2008) and some general analyses undertaken for the whole of Hampshire e.g. (Berry *et al.*, 2005a; Hossell and Rowe, 2006) and the UK (Hossell *et al.*, 2000; Berry *et al.*, 2007b)). Therefore, a review of the impact of climate change on the selected species and their habitats was also undertaken during this research to suggest potential impacts, using the Bayesian belief network models.

1.3.3. Choice of modelling methods

The use of different techniques, scales and analyses and availability of suitable data, can produce very different modelling results (Thuiller, 2004; Guisan and Thuiller, 2005). Given that every study area is different and the use of habitats by species likely to be different, results from different modelling studies frequently cannot be applied specifically to other areas. Therefore, it is beneficial to examine the performance of different approaches in the New Forest to determine the most suitable methods for predicting species' distributions in this area.

As can be seen from the literature (sections 1.2.1 and 1.2.2), there are a wide array of methods employed in species distribution (or habitat suitability) modelling, with varying success in different applications. Ecological niche factor analysis (ENFA, implemented in the Biomapper (Hirzel *et al.*, 2002) package) is a widely used method that has a key advantage of requiring only species' presence data, which is frequently (as in the case of New Forest species) the only type of data available. It is freely downloadable and easy to use, and is compatible with software packages such as IDRISI (Clark Labs, 2006). By using ENFA it is also possible to assess the contribution of each environmental variable to the final model, allowing an extra validation of the results by an ecologist; this is not possible, at least in a straightforward manner, with other ecological niche modelling methods, such as GARP and Maxent (Sérgio *et al.*, 2007).

GLMs and GAMs (logistic regression) are well established methods, which have been extensively used in species distribution modelling, partly because they are so well documented (Guisan, 2002). They are also relatively simple to implement within a geographical information system (Guisan *et al.*, 1998). Although GLMs are the most commonly used modelling technique in species distribution modelling (Rushton *et al.*, 2004) an extension of these, GAMs, are increasingly being used (Thuiller, 2003). GAMs have an additional advantage that the exact shape of a species response to an environmental predictor does not have to be specified prior to fitting the model, which is extremely useful for investigating unknown relationships (Austin, 2002a). Logistic regression is a powerful technique, in part due to its use of presence-absence data. However, this approach is frequently used with pseudo-absence data (see Appendix 13.1) if true absence data are unavailable. A further advantage of ENFA and logistic regression models is that they both have easily-implementable ways of selecting from a large number of potentially important spatial variables.

Bayesian belief networks (BBNs) are a relatively new approach to modelling habitat suitability for species, but have been reported to be useful approach in the application of environmental problems (Newton, 2009a). They are particularly appropriate for modelling finer scale variables (which are not available as spatial GIS data, such as aspects of habitat structure) that may be important for more accurate prediction of habitat suitability for many species. This is mainly owing to their ability to combine empirical data with expert judgement, which makes them an extremely flexible and useful modelling tool (Smith *et al*., 2007). They are also a powerful tool for assisting in the elicitation, integration and analysis of expert knowledge (Newton, 2009a), which is an important source of knowledge for many species (particularly those for which there is very little available literature). BBNs are also easy to create and amend (McCann *et al*., 2006) and extensive computer programming or modelling expertise is not required to develop and update models (Smith *et al.*, 2007).

Each of these modelling methods was applied to examine the potential distribution of selected species of conservation concern in the New Forest, as detailed in the following chapters.

1.3.4. Thesis structure

Chapter 2 examines the performance of a broad-scale modelling approach (using spatial GIS data) requiring only species' presence data for predicting potential habitat suitability for selected species, addressing the following aim:

1. To examine the performance of the ecological niche factor analysis method (implemented in the Biomapper software) for predicting potential habitat suitability for selected species.

Chapter 3 examines the performance of broad-scale modelling approaches (also using spatial GIS data) requiring both species' presence and absence data for predicting potential habitat suitability/distribution for selected species, addressing the following aims:

1. To examine the performance of species distribution models requiring presence-absence data; mainly Generalised Linear Models (GLMs) as well as Generalised Additive Models (GAMs).
2. To compare the selection of variables based on expert knowledge and the literature to automated stepwise selection.

Chapter 4 examines the performance of Bayesian Belief Network (BBN) models in predicting potential habitat suitability for finer-scale variables based on site data from fieldwork, addressing the following aims:

1. To identify factors influencing the suitability of habitat for species at a finer scale/in more detail using available literature and expert knowledge.
2. To incorporate this information into BBN models to predict the potential habitat suitability of sites.
3. To test the BBN models using fieldwork data.

Chapter 5 presents a review of the literature to assess the potential impact of climate change on the selected species and their habitats in the New Forest, addressing the following aims:

1. To examine the potential impacts of climate change on the study species and their habitat in the New Forest.
2. To apply this to the Bayesian Belief Network (BBN) model variables (from Chapter 4) for the study species and examine the potential change in habitat suitability.

Chapter 6 integrates and discusses the results of Chapters 2 – 5, addressing the following aims:

1. To compare and discuss the performance of the different modelling approaches used in Chapters 2 – 4.
2. To examine sources of error and uncertainty in the use of the models.
3. To discuss the use of species distribution models in the New Forest.
4. To examine the broader implications of the research, in terms of modelling species distributions at the scale of individual protected areas.

CHAPTER 2

Chapter 2. Presence-only habitat suitability models

2.1. Introduction

Numerous species distribution modelling approaches have been developed that use presence-only data (see section 1.2.1, Chapter 1). A widely used approach is Ecological Niche Factor Analysis (ENFA). Hirzel *et al.* (2002; 2006a) developed Biomapper, a statistical tool based on ENFA, which can build habitat suitability models and maps for any kind or animal or plant species. ENFA quantifies the niche that the species occupies by comparing its distribution in an ecological space, defined by one or more variables (known as ecogeographical variables, EGVs), with the distribution of all cells in that space, focusing on the marginality (the species niche position) and specialisation of the species (Hirzel *et al.*, 2002; Elith and Burgman, 2003).

Species are expected to be non-randomly distributed regarding ecogeographical variables. A species with an optimum temperature preference, for instance, is expected to occur preferentially in grid cells lying within its optimal range (Hirzel *et al.*, 2002). In Biomapper, environmental variables are summarised into two types of uncorrelated factors: marginality (representing the deviation, + or –, of a species' mean distribution from the global mean) and specialisation (a ratio comparing the range of the global distribution to that of the species'). Together these define a hyper-volume of space corresponding to the ecological niche of the species (Hirzel *et al.*, 2002; Zaniewski *et al.*, 2002). Habitat suitability indexes (scaled to range between 0 and 100) are computed by comparing these factors for observed species distribution to the distribution of the environmental variables in the whole area (Hirzel *et al.*, 2002; Zaniewski *et al.*, 2002). Habitat suitability maps produced in Biomapper are based on an environmental envelope algorithm; these envelopes are fitted to the observed distribution in the niche space (Hirzel, 2008). ENFA predicts habitat suitability rather than probability of presence (Zaniewski *et al.*, 2002).

In Biomapper the ecological niche refers to the subset of cells in the ecogeographical space where the focal species has a reasonable probability of occurring (Hirzel *et al.*, 2002). This multivariate niche can be quantified on any of its axes by an index of marginality and specialization. Some of these axes are more influential than others, which is why a factor analysis is introduced (Hirzel *et al.*, 2002). Since ecological variables are not independent, multicollinearity and redundancy arise as more and more are introduced into the description. Factor analyses aim to transform *V* correlated variables into the same number of uncorrelated factors. As these factors explain the same amount of total variance, subsequent analyses may be restricted to the few important factors (e.g., those explaining the largest part of the variance) without losing too much information (Hirzel *et al.*, 2002).

As many environmental variables as available can be inputted into the Biomapper model, as the factor analysis does not reject any of them but instead weights them (Hirzel *et al.*, 2002). This has the advantage that the subjective components and *a priori* knowledge required are kept minimal, and correlations among variables and axes are immediately visible and interpretable (Hirzel *et al.*, 2002). Hirzel *et al*. (2002) suggests that this approach is an advantage over the stepwise variable selection procedures used in logistic regression.

ENFA was used in the MONARCH 3 UK climate change analysis (O'Hanley, 2005b), as a relatively simple method for modelling presence-only datasets. Others, such as Zaniewski *et al*. (2002) and Olivier and Wotherspoon (2006) (see section 1.2.2, Chapter 1) have also reported on the beneficial use of ENFA. Santos *et al*. (2006) found ENFA (developed in Biomapper) to be an ‘outstanding’ method to evaluate the factors that limit the distribution range of secretive and widespread species (such as *Vipera latastei*), contributing to the evaluation of their conservation status. Sattler *et al*. (2007) used ENFA to suggest reclassification the conservation status of a cryptic bat species.

However, the majority of ENFA studies have been undertaken at a relatively large scale, such as the extent of large regions (Brotons *et al.*, 2004; Santos *et al.*, 2006) or whole countries (Zaniewski *et al.*, 2002; Sérgio *et al.*, 2007; Lachat and Butler, 2009; Puddu *et al.*, 2009), with very few studies at a smaller scale (Olivier and Wotherspoon, 2006), particularly for relatively small protected areas but see (Fei *et al.*, 2007; Soares and Brito, 2007; Podchong *et al.*, 2009). Further, although there are a number of studies demonstrating the use of the ENFA approach for modelling bird (e.g. Brotons *et al.*, 2004; Hirzel *et al.*, 2004; Olivier and Wotherspoon, 2006) and mammal (e.g. Hirzel *et al.*, 2002; Sattler *et al.*, 2007; Puddu *et al.*, 2009) species, application to insect and fungi species has also been limited (Chefaoui *et al.*, 2005; Lachat and Butler, 2009), as has comparison between species of different functional groups. The resulting habitat suitability maps can also be used to indicate sites with the highest indices of 'habitat potential' for concentrating habitat improvement measures, leading to ecological and economical optimisation of conservation efforts (Braunisch and Suchant, 2005).

The main advantage of ENFA is that it uses presence-only data, although this means it has a tendency to over-predict species distribution due to the lack of discriminating absences, a common problem with presence-only methods (Engler *et al.*, 2004). However, for purposes such as predicting potentially suitable sites of occurrence or for suggesting possible areas with high potential for (re)colonisation (Lütolf *et al.*, 2006), it is better to over-predict than under-predict species distributions. The ENFA method also allows for relatively easy assessment of the contribution of different environmental variables to the final model (see section 2.3.2.2.1.2), allowing an extra validation of the results (Sérgio *et al.*, 2007). A further advantage of ENFA is that because it is fundamentally a descriptive analysis, it does not rely on any underlying hypothesis for the data, in particular spatial autocorrelation (see Appendix 13.3) is not a problem as such (Basille *et al.*, 2008).

For these reasons, ENFA (implemented using the Biomapper software) was chosen for modelling potential habitat suitability for selected species in the New Forest. The fact that the method is designed for use with presence-only data (the only data available for this study, and for many other applications) is a key advantage and avoids some of the issues that arise with absence data (as discussed in section 1.2.4, Chapter 1).

Objectives:

- To examine the performance of ecological niche factor analysis (implemented in Biomapper) for predicting potential habitat suitability for selected species of conservation concern in the New Forest.

2.2. Methods

2.2.1. Selection of species for modelling and preparation of species occurrence and environmental variable layers

Biomapper requires spatial (GIS) layers of environmental variables and species occurrence to generate models.

2.2.1.1. Species data

It was decided to focus on species of conservation concern (Wright and Westerhoff, 2001) for the modelling, as these are the species for which it would be most useful to know the occurrences of for effective management and conservation. In addition, the locations in which they are found tend to be more accurately recorded because they are rare, and therefore attract particular attention from recorders. Additionally, species of conservation concern may be more sensitive to environmental change, particularly where they tend to be more specialised (see Appendix 33). Many studies (e.g. Hepinstall *et al.* (2002), Brotons *et al*. (2004), Hernandez *et al*. (2006) and Tsoar *et al.* (2007)) have suggested that species with restricted ecological niches can be modelled with greater accuracy than more generalist species (see section 1.2.5, Chapter 1). However, some of the species of conservation concern that are listed in the New Forest SAC Management Plan (Wright and Westerhoff, 2001) may not necessarily be rare in the New Forest, as for several species, the New Forest is their stronghold in the UK. Although rarer species have been selected as part of this study, more common species should not be overlooked, as they could also become rare and may suffer declines as a result of future environmental changes (Gaston and Fuller, 2007). They may also have significant roles for ecosystem structure and functioning (Gaston and Fuller, 2007) and important roles in the habitats of rarer species (e.g. be a foodplant for a rare butterfly species).

Over 600 Species of Conservation Concern are listed in the New Forest Special Area of Conservation (SAC) Management Plan (Wright and Westerhoff, 2001) and the lists of these species for each terrestrial (marine species were not included) species group (plants, butterflies etc.) were extracted from the plan. It was intended that examples of species from a range of species groups could be used to test the models because, as Tsoar *et al*. (2007) found, the characteristics of the species (e.g. niche width, dispersal) may have a greater impact on predictive accuracy than differences in modelling techniques.

Experts on each of the species groups, including local recorders and organisations, were then approached with the appropriate lists and where possible they provided data for species for which there were sufficient records (approaching 100 unique 6-figure grid reference records, post 1990). Data were only available for presences of species, rather than any absences. Occurrence data was used rather than abundance data mainly because the vast majority of records were of this form and because binary information is sufficient for, and less misleading in, determining whether a habitat is suitable for a species, whereas species abundance is influenced by many factors other than habitat quality (Matern *et al*., 2007).

Most of the records of species occurrence were collected in an *ad hoc* manner rather than by comprehensive surveys. In that respect, there may be biases in density of recordings, for example towards areas that are easily accessible. However, this is often the case with the type of data used to develop species distribution models (see section 1.2.4, Chapter 1). It is assumed that these data are reliable (even though not complete) as it is typically collected by experienced recorders, although any records can be prone to mis-identification, particularly for species that are difficult to identify.

The extent of the National Park was chosen as the area of study rather than the Special Area of Conservation (SAC) because the National Park is the latest conservation designation and management will now be focussed more at this extent. This is the conservation designation covering the largest area of the New Forest (and includes all of the other designated areas). A Management Plan for the National Park is currently being developed (New Forest National Park Authority, 2008). Species records therefore had to occur within this boundary to be included.

A cut-off of 1990 onwards was used to select occurrence records, so that they would reflect the current distribution. However, where a larger number of records were available a cut-off of 1995 was used (see Appendix 3). Although ideally it would be better to use a more recent date as a cut-off, such as 2000 onwards, this would greatly reduce the number of records available for many of the species. The habitat layer employed here was based on aerial photos from 1995/1996 and subsequently revised, based on ground truthing (see Appendix 5). A resolution of 6-figure grid references (i.e. 100 m x 100 m resolution) was chosen because this was generally the most accurate resolution for which species occurrence records were available.

A minimum of approximately 100 unique (i.e. different six-figure grid references) records was defined as a criterion for potential selection of a species for study. Although there are mixed suggestions on an appropriate sample size (see section 1.2.4, Chapter 1), Stockwell and Peterson (2002b) suggest 100 data points could be sufficient for logistic regression models and Jimenez-Valverde and Lobo (2007) note that a sample size lower than around 70 observations decreases logistic regression model performance. The required number of presence data for Biomapper has also not been established (see section 2.2.2.2 below). The required sample size is also dependent to some extent on the individual species and their range of habitat requirements (see section 1.2.4, Chapter 1). However, using a larger minimum sample size would reduce the number of species with sufficient records to include in the modelling.

From a list of over 600 candidate species, there were very few with sufficient data, according to the selection criteria defined above. Based on these criteria, eight species fulfilled the requirements at the time, and were selected. These were four plant species: wild chamomile (*Chamaemelum nobile*), slender marsh-bedstraw (*Galium constrictum*), wild gladiolus (*Gladiolus illyricus*) and pillwort (*Pilularia globulifera*); two butterfly species: silver-studded blue (*Plebeius argus*), grayling (*Hipparchia semele*); one Orthoptera species: wood cricket (*Nemobius sylvestis*); and one fungus species: nail fungus (*Poronia punctata*).

C. nobile is a pleasantly aromatic flowering perennial herb which favours mown or grazed grassland in open places (Sterry, 2006; Plantlife, 2007a). *G. constrictum* is a trailing or scrambling perennial herb with small white flowers, which occurs in marshy places, ditches and pond-sides (Stace, 1997; Meek, 2002). *G. illyricus* is a cormous perennial with erect stems (25-50 cm) with reddish-purple flowers (Sell and

Murrell, 1996). In the New Forest it is strongly associated with bracken on acid grass heath, frequently close to woodland edges (Stokes, 1987). *P. globulifera* is a perennial, grass-like rhizomatous aquatic fern, typical of the seasonally flooded margins of mildly acid ponds, pools or slow-flowing rivers (Preston, 2002; Plantlife, 2006). It is not a competitive species and requires bare substrate to colonise, with suitable conditions maintained at many of its sites by the trampling of cattle and horses (Preston and Croft, 1997).

H. semele is a relatively large butterfly with cryptic colouring which provides it with excellent camouflage (Bailey *et al.*, 1989; Asher *et al.*, 2001). It is a butterfly of arid places, frequently occurring in acid grass heaths, especially where heathers are regenerating after burns (Oates *et al*., 2000). *P. argus* is a smaller silvery-blue butterfly which breeds exclusively on heathland, including the edges of wet heathland and pockets of heathland in coniferous plantations (Oates *et al*., 2000). The species requires managed heathland and colonies die out on neglected heaths as mature heaths are unsuitable (Oates *et al.*, 2000).

N. sylvestris is a small dark brown non-flying cricket, occurring in deep, frequently broadleaved, leaf-litter in woodland clearings or along woodland edges (Richards, 1952; Marshall and Haes, 1988). *P. punctata* is a saprotrophic ascomycete fungus which forms its nail-like fruiting bodies on the dung of horses and ponies which have grazed on unimproved acid-loving vegetation (Webster, 1999; Poland, 2004). It therefore tends to occur in open acid heathland and grassland (Cox and Pickess, 1999; Poland, 2004; Cox *et al.*, 2005).

Further information about each of these species can be found in Appendix 2 and the individuals and organisations that kindly provided data are listed in Appendix 3. It should be noted here that *H. semele* was not included in the SAC Management Plan (Wright and Westerhoff, 2001) as a species of conservation concern but has since been made a UK Biodiversity Action Plan species. Bird species were not included as although sufficient data was available for three species, their distributions were already very well studied in the New Forest. Another species for which there was sufficient data was the stag beetle (*Lucanus cervus*) but the majority of records for this species were from people’s gardens. Marine species were not included as the focus of this study was on the terrestrial habitats.

The occurrence data for each species were divided into a set for ‘training’ and developing the models, and a set for ‘testing’ and evaluating the models. There are various ways of doing this, each with advantages and disadvantages (see section 1.2.6, Chapter 1) but in this case a random split was used, with the data randomly assigned into the training set (70% of the occurrence values) or the testing set (30% of the occurrence values) (such as used by Brotons *et al*. (2004)). The number of records allocated to each of these datasets is listed in Appendix 3. This approach was chosen because it was felt to be the closest approximation to using an entirely independent data set, which although is the preferred option, was not available in this case. The main disadvantage of this method is that the quantity of data used to train the model is reduced.

Raster layers of the species locations were created for each species using Database Workshop in IDRISI Andes (Clark Labs, 2006). Pixels were assigned values of 1 (species presence/occurrence record) or 0 (no occurrence record).

2.2.1.2. Environmental data

All spatial environmental data that were available for the extent of the New Forest National Park were obtained. The environmental GIS data obtained included a habitat type layer, a soil type layer, a digital terrain model layer and climate data, and these were used to derive the following spatial data layers:

- Habitat type/percentage cover
- Habitat type patch area
- Habitat type patch compactness
- Habitat type edge density
- Euclidean distance to habitat type
- Soil type
- DTM/Elevation
- Slope
- Apect

- Maximum, minimum and average temperature
- Maximum, minimum and everage precipitation

Further information about each of these layers, including how they were derived and used can be found in Appendices 4 - 7.

Many of the environmental variables may not directly affect species occurrences, but may act as a surrogate for another variable(s). The different types of variables are discussed in section 1.2.3 (Chapter 1). For example, although soil type may have a direct effect on plant species occurrence, it may be indirectly related to butterfly occurrence because preferred food-plants only occur on certain soil types. Slope and aspect are also often included in habitat suitability modelling because of their relationship to local temperature and soil moisture, but are a lot easier to incorporate spatially than other variables associated with them.

Although the available habitat layer was fairly detailed, even using many different habitat types is unlikely to incorporate enough information to fully predict the habitat requirements and occurrences of the species in detail. For example, components of the habitat that vary within habitat types, such as habitat structure or condition, may be important for certain species, but unfortunately this information was not available spatially for the New Forest and it would be extremely time consuming to obtain. Therefore, it was decided first to develop and test the models with available data then examine what further, if any, information would be useful in improving the performance of the models.

2.2.2. Ecological Niche Factor Analysis (ENFA): Biomapper

The free Biomapper software (version 3.2., available from http://www2.unil.ch/biomapper/; Hirzel *et al.*, 2002; Hirzel *et al.*, 2006a) was used to develop the models in this chapter.

2.2.2.1. Preparation of the species and environmental layers

The Biomapper models were developed using the ‘training’ presence layers (Boolean layers, with a value of 1 indicating presence of the species and a value of 0 indicating background/no presence record). Biomapper provides an option to normalise the EGV (environmental layer) maps using the Box-Cox algorithm, which aims to optimise the normality of the variable distribution. However, Hirzel (2008) reports that empirically, normality was not found to be a crucial factor and this step could as well be ignored. Normalisation was carried out for one of the habitat cover maps, but created a nearly boolean map (with almost all cells assigned a small range of values) so as suggested by Hirzel (2008) in this case, the original maps were used.

2.2.2.2. Ecological Niche Factor Analysis

ENFA is not sensitive to irrelevant data or to its order, as all useful information is extracted and summarised into the ecological niche factors, so although non-relevant data will increase the computation time and memory needs, it will not significantly influence the accuracy of the result (Hirzel, 2004). However, mathematically, ENFA cannot be computed with more environmental variables (termed ecogeographical variables, EGVs, in Biomapper) than species presence points, and practically it is best to have at least three times more presences than EGVs (Hirzel, 2008). Therefore, when there are only a low number of presences and/or a high number of potential EGVs and it is unknown which EGVs are most important for a given species, Hirzel (2008) suggests grouping the EGVs into classes (e.g. habitat type, topography), ideally so that the number of EGVs in each group is less than one third of the number of presence points, and then computing an ENFA separately with each group. The ‘best’ EGVs (those that have the highest coefficients on the marginality and specialisation factors; see below) from the ENFA outputs for each group should then be kept and pooled together and a final ENFA computed and used for the habitat suitability computations.

Therefore, as the number of training records available for the species ranged from 70-180, but there were 98 environmental (EGV) layers available, this procedure was followed. As noted by Hirzel (2008), there is no automated variable selection function in Biomapper, in the sense of the regression stepwise methods (see section 3.2.2.3, Chapter 3). Instead, there is factor interpretation, through the information provided by the ENFA (see section 2.3.2.2.1.2).

The number of presence locations required for use in Biomapper is not clear, as it depends on several factors such as the variance of the study area, the specialisation of the focal species, the design and accuracy of the sampling (Hirzel, 2008). Although Hirzel (2008) has generally used several hundreds of points, he also notes that far fewer points could have been used without significantly decreasing the accuracy of the model, perhaps even as few as 20 or 30 points. The environmental layers (EGV's) were split into 6 groups (habitat cover, patch area, patch compactness, edge density, Euclidean distance, and soil, terrain and climate (see Appendix 7 for explanation)) and the ENFA's run for each species with each of the groups. Negative or very large eigenvalue warnings were ignored at this stage, as this was not going to be the final selection of variables for the ENFA.

The most relevant environmental variables (EGVs) were selected by examining the ENFA score matrix (which indicates how the factors are correlated with the variables), in particular the marginality factor (column 1). In ENFA, the first axis is chosen so as to account for all the marginality of the species, and the following axes so as to maximize specialization, i.e. the ratio of the variance in the global distribution to that in the species distribution (Hirzel *et al.*, 2002). The marginality maximizes the multivariate distance of the EGVs between the cells occupied by the species and the cells within the whole reference area (Sattler *et al.*, 2007). The coefficients *mi* of the marginality factor express the marginality of the focal species on each EGV, in units of standards deviations of the global distribution (Hirzel *et al.*, 2002). The marginality is defined as the absolute difference between the global mean and the species mean, divided by 1.96 standard deviations of the global distribution (to remove any bias introduced by the variance of the global distribution) (Hirzel *et al.*, 2002). Further explanation of marginality can be found in Hirzel *et al*. (2002). The higher the absolute value of a coefficient (close to 1 or -1), the further the species departs from the mean available habitat regarding the corresponding variable (i.e. the more particular the habitat relative to the global habitat). Negative coefficients indicate that the focal species prefers values that are lower than the mean with respect to the study area, while positive coefficients indicate preference for higher-than-mean values (Hirzel *et al.*, 2002). A low marginality value (close to 0) indicates that the species tends to live in average conditions throughout the study area (Hirzel, 2008).

The rows of the score matrix are the EGV contributions to each factor and the other columns are the *V*-1 specialisation factors (*V* is the number of variables). Specialization is defined as the ratio of the standard deviation of the global distribution to that of the focal species (Hirzel *et al.*, 2002). The specialisation factors account for the decreasing residual variance after removal of upper-ranked explanatory factors (and therefore most of the variance is explained by a few of the first factors), and denote to what extent the species' EGVs distribution is narrow with respect to the overall distribution of the EGVs in the whole reference area (Sattler *et al*., 2007). The inverse of specialization is therefore a measure of species' tolerance (Sattler *et al.*, 2007). For the specialisation factors it is only the absolute value that is important (the signs are arbitrary), so EGVs that had particularly high values (>0.5 or <-0.5) on the first few specialisation factors were also considered (as long as the sign of the coefficient of the marginality factor for that variable was appropriate (see below)), as the higher the absolute value, the more restricted is the range of the focal species on the corresponding variable (Hirzel *et al.*, 2002).

When selecting EGVs for the final pool of 'best' variables, only the EGVs with positive coefficients (positive marginality values) for habitat percentage cover, patch area, patch compactness, edge density, soil type percentage cover and aspect layers, were selected. This was because if, for example, percentage cover of coniferous woodland had a strong negative marginality value, it may end up being selected as one of the most important variables (and therefore used as a factor to develop the HS map). However, this indicates that the species tends to occur where there is less coniferous woodland cover than globally (throughout the site) available. Consequently, anywhere that did not have coniferous woodland (regardless of what else the habitat type was) may be more likely to have a high HS value. This is also an issue in stepwise regression (see section 3.2.2.3, Chapter 3). However, for Euclidean distance to habitat type variables, it is negative coefficient (marginality) values that are relevant, because they indicate that the species tends to occur in sites that are closer to that habitat type. For the elevation, slope and climate variables, the coefficient may be positive or negative as the species occurrence may be associated with more positive or more negative values than the global mean.

There was generally a clear group of variables within the groups with more positive or negative marginality values (which tended to be <-0.1 or >0.1). If the species correlation tree showed a pair or group of variables to be highly correlated (>0.5), only the variable(s) of the pair or group with the highest marginality value were selected.

2.2.2.2.1. ENFA with selected 'best' variables from each of the groups

An ENFA was run with the selected EGVs from the six groups of EGVs for each species. If there was no eigenvalue warning then a habitat suitability (HS) map was created (see section 2.2.2.3). If a negative or very large eigenvalue warning was received then the species and global correlations trees were examined as negative or very large eigenvalues can be caused by highly correlated variables (Hirzel, 2008). Where required, one out of each of the pairs of highly correlated (>0.5) variables was removed, based on which had the lower marginality value. Loss of information through the removal of these variables was not a problem as the removed maps contain mostly redundant information (Hirzel, 2008). The ENFA was then re-run and the eigenvalues checked again. If there was still a warning, then the process was repeated until a warning message was no longer received.

It was also important to ensure that the number of EGVs used for the final ENFA did not include more than the suggested number of variables (i.e. no more than one third of the total number of presence data in the training data set; see Appendix 9 for details). If the number of selected variables from the groups was greater than suggested, the number of variables could be reduced by removal of correlated variables (as above) or those with the lowest marginality and specialisation factors. A list of the final set of variables used in the final ENFA for each species can be found in Appendix 9.

Biomapper does not produce a formula for the habitat suitability, as the maps are produced based on an environmental envelope algorithm fitting to the observed distribution in the niche space (Hirzel, 2008). Therefore, for interpreting the ecological requirements of the study species, Hirzel (2008) suggests examining the coefficients of the ecological niche factors from the ENFA score matrix, which indicate how marginal and specialised the species is on the various relevant environmental variables. This was carried out and the results are displayed in section 2.3.2.2.1.2.

2.2.2.3. Habitat suitability computation

The species distribution on the marginality and specialisation factors allows the computation of a habitat suitability (HS) index for any set of EGV values and a HS map to be created (Hirzel *et al.*, 2002). Biomapper provides a choice of four different algorithms to compute the habitat suitability (see Hirzel (2004)). However, owing to the type of non-systematic species data used in these models, the medians algorithm had to be used because the three other options provided (distance geometric mean, distance harmonic mean and minimum distance algorithms) take into account the distance between observations (species occurrences) so the density of observations must be representative of the species niche (Hirzel, 2004). The distribution data available for this project were not, as they were collected *ad hoc*.

The median algorithm has been reported to give good results in most situations and be the quickest of the available algorithms (Hirzel, 2004). To compute the median algorithm, the species range on each factor is divided into 25 classes, in such a way that the median would exactly separate two classes. For every point (cell) in the environmental space, the number of observations (from the species distribution) are counted that are in either the same class or in any class further apart from the median (Hirzel, 2004). Normalization is achieved by dividing twice this number by the total number of cells in the species distribution (total number of observations) (Hirzel *et al.*, 2002). Thus, a cell laying in one of the two classes directly adjacent to the median would score one, and a cell laying outside the observed species distribution would score zero (Hirzel *et al.*, 2002).

An overall suitability index for the focal cell is computed by the weighted average of its scores on each factor, the weights being given by the amount of information explained by each factor (indicated by the eigenvalues; see Appendix 11 for explanation) (Hirzel, 2004). Repeating this procedure for each cell allows the production of a habitat-suitability map, where suitability values range from 0 (least suitable) to 100 (Hirzel *et al.*, 2002), where cells with values closer to the medians are deemed more suitable (O'Hanley, 2005b).

Habitat suitability computation was carried out using the 'final' ENFA for each species. The number of factors used to create the HS map (i.e. the number of retained factors) was determined by the MacArthur's broken stick advice (except for *C. nobile*; see section 2.3.2.3), incorporated as part of the habitat suitability computation within Biomapper. This is based on the expected distribution when breaking a stick randomly. Therefore, the eigenvalues that are larger than what would have been obtained randomly may be considered 'significant' (Hirzel, 2008). The eigenvalues give an indication of how much variance is explained by the factors: the larger they are, the more information each factor is conveying (Hirzel, 2008). Eigenvalues usually rapidly decrease from the second factor to the last one, so that only the first four or five axes are useful to compute habitat suitability (Hirzel *et al.*, 2002). The appropriate factor maps were computed and the weights assigned to each factor (see Appendix 11) left as the default, where the weights are computed from the eigenvalues and represent the amount of information explained by each factor. The isopleth option was selected for scaling. The HS maps can be found in section 2.3.2.3.

2.2.2.4. Evaluation of the habitat suitability maps

The habitat suitability maps were evaluated using the 'testing' set of presence data for each species. (Information about the number of records for each species can be found in Appendix 3). The evaluation tools included within Biomapper for evaluating models with separate presence 'testing' data were used. These were:

- Box plots (histograms) and accompanying statistics comparing the habitat suitability values from the predicted map for the global (all cells in the study area) and the validation/testing presence data;
- Graphs of the distribution (frequency) of global and validation cells in 50 habitat suitability classes;
- Continuous Boyce Index – using the evaluation/training data to calculate the expected frequency of evaluation points, to compare to an expected random distribution for different habitat suitability values (as described in Appendix 8.1).

Presence-only models are more difficult to evaluate than presence-absence models, but among the presence-only evaluators, the continuous Boyce index has been found to be the most accurate for characterising predictive capability (Hirzel *et al.*, 2006b). However, it is useful to make use of a range of different measures to obtain an overall assessment of the performance of the models. The evaluation outputs can be found in section 2.3.2.4.

2.3. Results

2.3.1. Selected species and environmental data

Information on the species and environmental data can be found in Appendices 3 – 7.

2.3.2. Biomapper results

2.3.2.1. Preparation of the species and environmental layers

When the environmental layers were verified in Biomapper several of the maps were reported as not being continuous enough (for example the habitat types percentage cover; 0-100). However, Hirzel (2008) reports that while such layers may cause negative or very large eigenvalues, they should not prevent the analysis from being performed, and therefore they were retained.

2.3.2.2. Ecological Niche Factor Analysis results

When running the initial ENFA's for *G. illyricus*, an error message of 'floating point division by zero' arose for the habitat percentage cover, patch area, patch compactness and edge density groups. This was because some of the variables had only values of zero at the presence locations, as *G. illyricus* has a small and restricted distribution (i.e. did not occur at all in some habitat types). The values for each of the environmental layers for the *G. illyricus* training data were examined to determine for which variables this was the case and these were excluded from the groups for the *G. illyricus* ENFA's. (These were AQ6, GL3, GL11, GL13, ST3).

2.3.2.2.1. Final ENFAs (with 'best' variables from each of the groups)

Final ENFAs for all of the species were run without producing an eigenvalue warning, except for *P. punctata*, which even after removal of all the highly correlated variables, still produced a 'very large eigenvalue' warning. Low sample sizes (*P. punctata* had the second lowest sample size) are more prone to getting very large or negative eigenvalues because of the correlation among EGVs, owing to low variance among species sites (Hirzel, 2008). However, all of the highly correlated variables had been removed. Negative or low eigenvalues may also be caused by the EGV maps not being continuous enough (Hirzel, 2008), although this was not a problem for the other species. Therefore, the HS map was computed anyway. A list of the variables included in the final ENFAs for each species can be found in Appendix 9.

2.3.2.2.1.1. Score matrices for final ENFA's

The output ENFA score matrices for each species can be found in Appendix 10 and were used to determine the most important ecological variables for each species (see section 2.3.2.2.1.2 below). Overall marginality, specialisation and tolerance values (see Table 1 below) were also provided with the ENFA score matrix for each species, and allow for among-species comparisons within a given area (Hirzel *et al.*, 2002). The global marginality takes into account all the EGVs and gives a summary of how much the species habitat differs from the available conditions (Hirzel, 2008). The global specialisation is the inverse of global tolerance, but as it varies between 1 and infinity, it is less easy to interpret (Hirzel, 2008).

Species	**Marginality**	**Specialisation**	**Tolerance**
C. nobile	0.74	1.08	0.926
G. constrictum	1.38	2.15	0.464
G. illyricus	2.57	3.16	0.317
H. semele	1.62	1.48	0.674
N. sylvestris	0.982	3.35	0.299
P. argus	1.66	1.46	0.684
P. globulifera	1.53	2.20	0.454
P. punctata	1.14	3570000	0.000

Table 1. Overall marginality, specialisation and tolerance values to accompany the score matrices obtained by ENFA analysis using Biomapper. For explanations of each of these, see text.

None of the species had low (close to 0) marginality values, suggesting that all of the species tend to live in particular habitats relative to the mean (i.e. the available conditions in the study area) (Sattler *et al.*, 2007). The lowest marginality value (at 0.744) was for *C. nobile*, which of all the species is probably one of the most widespread in different habitats within the New Forest, and the highest marginality value was for *G. illyricus* (2.565), which occurs in a fairly particular bracken-dominated habitat.

All of the tolerance values were below 1, indicating that the species all have some form of specialisation (a randomly chosen set of cells is expected to have a tolerance of 1 (Sattler *et al.*, 2007)). However, *C. nobile* had the highest tolerance value, closest to 1 (0.926), so is the least specialised/more generalist in its living environment (Hirzel, 2008). This fits with it being the least marginal as well. *H. semele* (0.674) and *P. argus* (0.684) also had fairly high specialisation values, although both had high marginality values (1.620 and 1.664 respectively). The lowest tolerance value was for *P. punctata* (0), indicating that it is a very specialist species living in a very narrow range of conditions (Hirzel, 2008). However, *P. punctata* did not have the highest marginality value, although it was still high.

2.3.2.2.1.2. Interpretation of species' ecological requirements from score matrix

Further information about the species' relationships with the environmental variables can be obtained from the score matrix by examining the marginality values (column 1) and the specialisation values (subsequent columns) (see Appendix 10 for the score matrices). The most important variables (EGVs) in terms of the highest (absolute) marginality values and specialisation values for each species from the score matrixes are shown in Tables 2 to 9 below. (Note that positive marginality values mean the species prefers locations with higher values on the corresponding EGV than the mean location in the New Forest study area, whereas negative values on this factor mean the species prefers locations with lower values. Signs of coefficient have no meaning on the specialisation values, where a higher absolute value means the range of the species is more restricted on that variable).

C. nobile

EGV	**Marginality value**	**Highest specialisation value**
Edge density of dry heath/acid grassland mosaic (HL3)	0.640	0.189
Edge density of dense scrub (ST1)	0.345	0.064
% cover of soil type 64303	0.305	0.503
Euclidean distance to HL3	-0.293	0.719
Slope	-0.287	-0.643
Euclidean distance ponds >0.5 ha (AQ6)	-0.198	0.556
Edge density of dry heath (HL1)	0.195	0.470

Table 2. Marginality and specialisation values for *C. nobile* resulting from ENFA analysis. EGVs are sorted by decreasing absolute value of coefficients on the marginality factor. Highest specialisation value indicates the highest (absolute) specialisation value (signs of coefficient have no meaning on the specialisation values) for that variable from only the number of factors retained to calculate the habitat suitability map (i.e. the first few columns) (see Table 10, section 2.3.2.3).

The highest marginality value for *C. nobile* was edge density of dry heath/acid grassland mosaic, demonstrating that *C. nobile* is found in locations with higher edge density HL3 values than the mean HL3 edge density of the whole national park. This means that it tends to be found in sites (pixels) containing dry heath/acid grassland mosaic with neighbouring pixels that also contain this habitat type, which is consistent with it being a species of moderately acidic grassland (see Appendix 2.1). However, the highest specialisation value was not particularly high, but was still greater than 0, indicating that *C. nobile* was found to occupy a narrower range of HL3 edge density values that the available range across the whole of the national park study area. Low Euclidean distance (i.e. closer) to this habitat type was also important. The next highest marginality value was for edge density of dense scrub, although the specialisation value for this was very low. Soil type 64303 is associated with lowland heath habitats, so it is not surprising that high marginality and specialisation values were obtained for *C. nobile* for this variable. *C. nobile* was also associated with lower slope values (i.e. flatter), which are perhaps more likely to be wetter, particularly in winter, which *C. nobile* favours (see Appendix 2.1). This association with wetter areas of habitats was also suggested by the preference for lower (closer) Euclidean distance values to larger ponds (>0.5 hectares).

None of the factors accounted for a large proportion of the specialisation (see Appendix 10), with a low value for the first (marginality) factor of 11%. This species also had relatively low global marginality and high tolerance values (see section 2.3.2.2.1.1), so appears to occur in a more generalist range of habitats.

G. constrictum

EGV	**Marginality value**	**Highest specialisation value**
% cover of soil type 84102	0.366	0.069
% cover of ponds <0.5 ha (AQ5)	0.319	0.014
Edge density of dry heath/acid grassland mosaic (HL3)	0.301	-0.059
Patch area of HL3	0.294	0.085
% cover of HL3	0.271	-0.257
Slope	-0.234	0.346
Euclidean distance to dry heath (HL1)	-0.228	-0.244
Edge density of HL1	0.219	0.067
Annual mean temperature (Bio1)	0.215	0.365
Elevation	-0.214	0.531
Euclidean distance to unimproved acid grassland (GL11)	-0.212	0.391
Euclidean distance to HL3	-0.211	0.403

Table 3. Marginality and specialisation values for *G. constrictum* resulting from ENFA analysis. EGVs are sorted by decreasing absolute value of coefficients on the marginality factor. Highest specialisation value indicates the highest (absolute) specialisation value (signs of coefficient have no meaning on the specialisation values) for that variable from only the number of factors retained to calculate the habitat suitability map (i.e. the first few columns) (see Table 10, section 2.3.2.3).

The results showed that *G. constrictum* tends to occur on sites with higher percentage cover of soil type 84102 than the global mean. This is a seasonally wet deep loam (see Appendix 6) associated with permanent grassland and deciduous woodland, so is consistent with *G. constrictum* being a plant of wet habitats (see Appendix 2.2), as is the association with ponds less than 0.5 ha in size, suggesting that smaller ponds are preferable to larger (>0.5 ha – AQ6) ponds. Edge density, patch area and percentage cover of dry heath/acid grassland mosaic were also important, which although it is a dry habitat, may contain ponds or other wetland features. The specialisation values for all these variables were fairly low. Euclidean distance to this habitat type was also important, but had a lower marginality value but a higher specialisation value. Euclidean distance to unimproved acid grassland and to dry heath also featured in this table of most important variables, so the cover of *Calluna* (see Appendix 5) does not seem to be too relevant, just the presence of the acidic habitat, consistent with the literature (see Appendix 2.2). The association with lower values for slope (i.e. flatter) and elevation also indicate sites where ponds or other wetland habitats are more likely to occur. The first two factors (marginality and the first specialisation factor) accounted for 42% of the specialisation (see Appendix 10).

G. illyricus

EGV	**Marginality value**	**Highest specialisation value**
Patch area of continuous bracken cover (GL8)	0.540	-0.006
% cover of GL8	0.444	0.016
Patch compactness of GL8	0.349	-0.012
Edge density of GL8	0.299	-0.042
Patch compactness of dry heath/acid grassland mosaic (HL3)	0.196	-0.025
Edge density of merged dry heath and dry heath/acid grassland mosaic (HL1_HL3)	0.189	-0.025
% cover of ponds <0.5 ha (AQ5)	0.178	0.004
Euclidean distance to acid fen/flush/valley mire in heathland situations (AQ1)	-0.167	-0.666
Euclidean distance wet heath (HL2)	-0.164	-0.807
% cover HL3	0.147	0.038

Table 4. Marginality and specialisation values for *G. illyricus* resulting from ENFA analysis. EGVs are sorted by decreasing absolute value of coefficients on the marginality factor. Highest specialisation value indicates the highest (absolute) specialisation value (signs of coefficient have no meaning on the specialisation values) for that variable from only the number of factors retained to calculate the habitat suitability map (i.e. the first few columns) (see Table 10, section 2.3.2.3).

Consistent with the literature (see Appendix 2.3), *G. illyricus* showed a strong preference for continuous bracken habitats (GL8), although the size of the patch of that habitat type was the most important, with the highest marginality value (0.540), indicating a preference for larger patches than the average. Association with dry heath and dry heath/acid grassland mosaic can also be seen from the table (patch compactness, edge density and Euclidean distance), which is again consistent with the suggestion in the literature of its occurrence on moderately acid soils. Bracken is also often found in these habitats. The association with small ponds may also confirm the findings of Stokes (1987) (see Appendix 2.3) that most New Forest *G. illyricus* sites were within 100 m of water. The preference of closer Euclidean distances to the valley mire (AQ1) and wet heath (HL2) habitats also suggest this association with water or wetter areas. Interestingly, Euclidean distance to continuous bracken (GL8) and Euclidean distance to coniferous woodland edge were highly correlated, but the latter, although it made the final ENFA selection, did not show a high marginality value.

The specialisation values were all low except for Euclidean distance to AQ1 and to HL2, for which *G. illyricus* occupied a narrower range of values than available, confirming the message of the high marginality values, demonstrating that it tends to occur at sites close to these habitat types (i.e. low Euclidean distance values). However, 50% of the specialisation was accounted for by the first (marginality factor) and 23% percent by the second factor, with 88% of the specialisation accounted for by the first four factors (see Appendix 10).

H. semele

EGV	**Marginality value**	**Highest specialisation value**
Patch area of merged dry heath and dry heath/acid grassland mosaic (HL1_HL3)	0.414	0
% cover of soil type 64303	0.334	0
Edge density of dry heath (HL1)	0.316	0
% cover of HL1	0.296	0
Euclidean distance to HL1	-0.24	0
Euclidean distance to wet heath (HL2)	-0.233	0
Euclidean distance to acid fen/flush/valley mire in heathland situations (AQ1)	-0.222	0
Patch compactness of HL1	0.216	0
Edge density of dense scrub (ST1)	0.213	0
Patch compactness of ST1	0.197	0

Table 5. Marginality and specialisation values for *H. semele* resulting from ENFA analysis. EGVs are sorted by decreasing absolute value of coefficients on the marginality factor. Highest specialisation value indicates the highest (absolute) specialisation value (signs of coefficient have no meaning on the specialisation values) for that variable from only the number of factors retained to calculate the habitat suitability map (i.e. the first few columns) (see Table 10, section 2.3.2.3).

The preference of *H. semele* for larger patches of dry heath or dry heath/acid grassland mosaic is consistent with the literature describing this butterfly as being associated with arid acidic grassy habitats (see Appendix 2.4). These habitats are also associated with soil type 64303, on which the species mean for *H. semele* is higher than the global mean. *H. semele* also shows a preference for sites with lower (i.e closer) Euclidean distance values to wet heath and valley mire (AQ1), which is not consistent with the literature. However, Euclidean distances to all the heathland habitats (HL1, HL2, HL3 and AQ1) were all highly correlated as they tend to occur close to each other, often in a mosaic, so occurrence near one type is likely to be close to another. In addition, *H. semele* is a fairly mobile species so could be recorded as it moves through them.

H. semele also showed higher than the global mean edge density and patch compactness values of dense scrub. This may not seem like preferable habitat, but this association could be because *H. semele* is often found in areas where heathers are regenerating after burns (see Appendix 2.4) and areas of dense scrub are often cleared as part of heathland management. As the habitat layer is based on the habitat types at a fixed point in time and the species observations span several years, the dense scrub habitat type in particular is likely to have changed over that time.

All the specialisation values were 0 for the variables included in Table 5, indicating that *H. semele* can occur across a wide range of values on those variables, but (as shown by the marginality values) is most likely to occur on sites that are closer to the Euclidean distance variables, or sites that have a higher than the global average value for the other variables. None of the factors individually accounted for a large proportion of the specialisation (see Appendix 10).

N. sylvestris

EGV	**Marginality value**	**Highest specialisation value**
Patch area of deciduous woodland (W1_W2)	0.430	-0.085
% cover of deciduous woodland (W1_W2)	0.373	-0.212
Euclidean distance to mixed woodland (W7_W8)	-0.290	0.190
Edge density of mixed woodland (W7_W8)	0.259	-0.042
Edge density of dry heath/acid grassland mosaic (HL3)	0.204	0.059
Patch compactness W7_W8	0.203	0.144
Euclidean distance W1_W2	-0.198	-0.229

Table 6. Marginality and specialisation values for *N. sylvestris* resulting from ENFA analysis. EGVs are sorted by decreasing absolute value of coefficients on the marginality factor. Highest specialisation value indicates the highest (absolute) specialisation value (signs of coefficient have no meaning on the specialisation values) for that variable from only the number of factors retained to calculate the habitat suitability map (i.e. the first few columns) (see Table 10, section 2.3.2.3).

The highest marginality value for *N. sylvestris* was for patch area of deciduous woodland, which fits with its requirements for deciduous leaf litter, and confirming the conclusions of Brouwers (2008), that the likelihood of *N. sylvestris* presence increased with patch size (see Appendix 2.5). The high marginality value for percentage cover of that habitat type within the 100 m x 100 m pixel also suggests that relationship. However, it was Euclidean distance to mixed (deciduous and coniferous) woodland that had the next highest marginality value (ahead of Euclidean distance to broadleaved woodland), with a preference for sites closer to mixed woodland. Higher edge density and patch compactness values of mixed woodland were also favoured. This is particularly interesting because mixed woodland has a lower proportion of deciduous trees, which are favoured by *N. sylvestris*. However, as long as some deciduous trees, and the resulting leaf litter, are present it seems suitable. It may just be that the records happened to be more from those types of woodlands.

A preference for higher edge density values of dry heath/acid grassland mosaic habitat is also interesting and unexpected, as *N. sylvestris* is a woodland species. However, perhaps this is something to do with woodland edge meeting heathland, as when *N. sylvestris* is recorded at a woodland edge, then it is a location where the woodland meets another habitat type which, in the New Forest, is frequently heathland. Therefore, within a 100 m x 100 m pixel containing woodland edge, there will be another habitat, which has a high likelihood of being heathland, present. Another explanation is that *N. sylvestris* has been recorded on heathland sites under bracken (S. Douglas, personal observation), and is known to venture several hundred metres from a woodland edge in hot weather (N. Brouwers, personal communication, January 9, 2009).

The specialisation values were generally low, with the highest values for Euclidean distance to broadleaved woodland (-0.229) and to mixed woodland (0.190), indicating, as for the marginality values, that *N. sylvestris* tends to occur on sites with a lower value for the Euclidean distance to these habitats. However, most of the specialisation (74%) was accounted for by the first (marginality) factor (see Appendix 10), indicating that these values are the most important. *N. sylvestris* also had the second lowest overall tolerance value (see section 2.3.2.2.1.1).

P. argus

EGV	**Marginality value**	**Highest specialisation value**
Edge density of dry heath (HL1)	0.304	0.003
% cover of merged dry heath and dry heath/acid grassland mosaic (HL1_HL3)	0.302	-0.640
Patch area of HL1_HL3	0.298	-0.006
% cover of HL1	0.251	0.524
Patch compactness of HL1	0.236	0.014
Euclidean distance to HL1	-0.232	0.060
Euclidean distance to wet heath (HL2)	-0.223	0.064
Euclidean distance to acid fen/flush/valley mire in heathland situations (AQ1)	-0.215	0.073

Table 7. Marginality and specialisation values for *P. argu s* resulting from ENFA analysis. EGVs are sorted by decreasing absolute value of coefficients on the marginality factor. Highest specialisation value indicates the highest (absolute) specialisation value (signs of coefficient have no meaning on the specialisation values) for that variable from only the number of factors retained to calculate the habitat suitability map (i.e. the first few columns) (see Table 10, section 2.3.2.3).

As documented in the literature (see Appendix 2.6), *P. argus* had a strong association with dry heath (HL1) and dry heath/acid grassland mosaic (HL3), with high marginality values for edge density, percentage cover, patch area and Euclidean distance, as well as high specialisation values for percentage cover of dry heath and dry heath/acid grass mosaic. This indicates that *P. argus* occurs on a more restricted range of values (i.e. higher percentage cover values) than the range of values available for sites across the study area. *P. argus* is also known to use wet heath, which is demonstrated by the marginality values for Euclidean distance to wet heath (HL2) and valley mire (AQ1), indicting a preference for sites closer to these habitat types. None of the factors individually accounted for a large amount of the specialisation (with the first factor explaining 17%) and the total amount of specialisation accounted for by the 9 retained factors was only 68% (see Appendix 10).

P. globulifera

EGV	**Marginality value**	**Highest specialisation value**
% cover of ponds (AQ5_AQ6)	0.358	0
Edge density of merged wet heath and acid fen/flush/ valley mire in heathland situations (HL2_AQ1)	0.234	0
Edge density of dry heath/acid grassland mosaic (HL3)	0.287	0
% cover of HL2_AQ1	0.234	0
% cover of HL3	0.233	0
Euclidean distance to wet heath (HL2)	-0.232	0
Euclidean distance to acid fen/flush/valley mire in heathland situations (AQ1)	-0.230	0
Patch area of HL2_AQ1	0.219	0
Euclidean distance to HL3	-0.212	0
Euclidean distance to dry heath (HL1)	-0.208	0
Patch compactness of HL2_AQ1	0.205	0
Edge density of HL1	0.186	0
% cover of Aspect - Flat	0.179	-0.485

Table 8. Marginality and specialisation values for *P. globulifera* resulting from ENFA analysis. EGVs are sorted by decreasing absolute value of coefficients on the marginality factor. Highest specialisation value indicates the highest (absolute) specialisation value (signs of coefficient have no meaning on the specialisation values) for that variable from only the number of factors retained to calculate the habitat suitability map (i.e. the first few columns) (see Table 10, section 2.3.2.3).

The percentage cover of ponds had the highest marginality value for *P. globulifera*, known as a wetland plant (see Appendix 2.7). This was also reflected by the high marginality values for edge density, percentage cover, patch area and patch compactness of wet heath and valley mire habitats combined (HL2_AQ1), as well as occurrence on sites with lower Euclidean distances to these habitat types individually. However, *P. globulifera* also showed a preference for high percentage cover of dry heath/acid grassland mosaic (HL3) and low Euclidean distance to this habitat, but this could be because water bodies such as ponds can occur within this habitat. This may also explain the preference for higher edge density of dry heath (HL1). The association with sites with a higher percentage of flat ground may reflect the fact that ponds, rivers and other water bodies may be more likely to be found in areas with flat ground, rather than sloping, where the water is more likely to lie.

The specialisation values were all zero except for this last variable, suggesting that *P. globulifera* tends to occupy a narrower range of values (i.e. higher percentage of flat aspect) than the available range across the study area. The rest of the specialisation values were zero, indicating that *P. globulifera* occurs across a range of values on those variables (for example, *P. globulifera* can occur in sites with a range of percentage cover of AQ5_AQ6 ponds, but tends to occur at sites with a higher than average percentage cover than the global average, as indicated by the marginality value). Further, a large proportion (58%) of the specialisation was accounted for by the first (marginality) factor, with none of the subsequent factors accounting for much of the specialisation (between 3% and 8%) (see Appendix 10).

P. punctata

EGV	**Marginality value**	**Highest specialisation value**
% cover of dry heath (HL1)	0.623	0.004
% cover of soil type 64303	0.517	0.001
Euclidean distance to HL1	-0.368	0.066
Edge density of dry heath/acid grassland mosaic (HL3)	0.252	-0.001
Average July precipitation (Prec7)	0.246	0
Euclidean distance to ponds >0.5 ha (AQ6)	-0.199	0.004
Edge density of dense scrub (ST1)	0.152	-0.001
Euclidean distance to ponds <0.5 ha (AQ5)	-0.129	-0.003

Table 9. Marginality and specialisation values for *P. punctata* resulting from ENFA analysis. EGVs are sorted by decreasing absolute value of coefficients on the marginality factor. Highest specialisation value indicates the highest (absolute) specialisation value (signs of coefficient have no meaning on the specialisation values) for that variable from only the number of factors retained to calculate the habitat suitability map (i.e. the first few columns) (see Table 10, section 2.3.2.3).

The highest marginality values indicated that *P. punctata* tends to be found at sites with a higher percentage cover of dry heath (HL1) or close (low Euclidean distance) to this habitat type, which is also reflected by an association with higher percentage cover of soil type 64303, which is associated with lowland heath habitats. A relation with higher edge density of dry heath with lower cover of *Calluna* and higher cover of grass (HL3) was also shown and may reflect the fact that the ponies on which *P. punctata* depends are more likely to be found where there is a greater cover of grass for grazing.

P. punctata also showed a preference for sites that receive higher July precipitation than the average July precipitation across the New Forest. This may reflect a requirement for some moisture leading up to the fruiting time (generally from early Autumn), which is common for fungi, although the moisture requirements of *P. punctata* are not clear (see Appendix 34.4). It may also be that grass growth is better where the rainfall is higher, and therefore better for ponies. However, there was only a difference of 3mm between the highest and lowest values for this variable, so the apparent relationship may just reflect the concentration of records to the areas of the study site where rainfall is higher (e.g. the south-east and the north-west of the National Park). However, a preference for slightly wetter sites than the average was also reflected in the preference for sites with a lower Euclidean distance to ponds (small or large, although preferentially large). This may also potentially reflect the behaviour of ponies, which may gather fairly close to ponds, for drinking, or because the vegetation is more lush.

A preference for sites with higher edge density values for dense scrub (ST1) was also shown but this could reflect recordings of *P. punctata* at locations where scrub was present at the time that the habitat layer was devised but has since been cleared, as part of management practices. Cleared scrub often provides areas of young growth where ponies like to graze, and consequently excrete dung (S. Douglas, personal observation).

The specialisation values for *P. punctata* were all low, indicating that it does not occur on a particularly restricted range of values for the variables, although these are not really relevant as 100% of the specialisation was explained by the first (marginality) factor (see Appendix 10). This corresponds with the very high overall specialisation value and very low tolerance value (see section 2.3.2.2.1.1).

2.3.2.3. Habitat suitability computation

The broken stick advice was followed for all the species regarding the number of factors to retain for the habitat suitability computation, except for *C. nobile*, where the suggested number of factors to retain was two, but this accounted for only 64.8% of the specialisation. Therefore, it was decided to increase the number of factors until the explained information was over 80% (which all the other species were), which took the number of retained factors to six (within the range of the number of factors retained for other species maps) and the explained information to 82.8%. The number of factors used to compute each of the HS maps and the amount of information explained by these are given in Table 10 below (with the values for the weighting of each factor, based on the explained marginality and specialisation for each species, provided in Appendix 11). The explained information value takes into account the marginality and specialisation and is explained in more detail in Appendix 11.

Species	**Number of factors**	**Explained information (%)**
C. nobile	6	82.8
G. constrictum	6	86.8
G. illyricus	4	94.3
H. semele	7	86.8
N. sylvestris	4	94.1
P. argus	9	84.2
P. globulifera	6	92.2
P. punctata	2	100

Table 10. Number of factors used to compute habitat suitability maps for each species and the amount of information explained by the total number of factors.

The resulting habitat suitability (HS) maps for the New Forest National Park study area are shown in Figures 1 to 8 below. The HS maps are also shown alongside maps of the variables with the highest marginality values for each species (see Tables 2 to 9 in section 2.3.2.2.1.2) in Appendix 12.

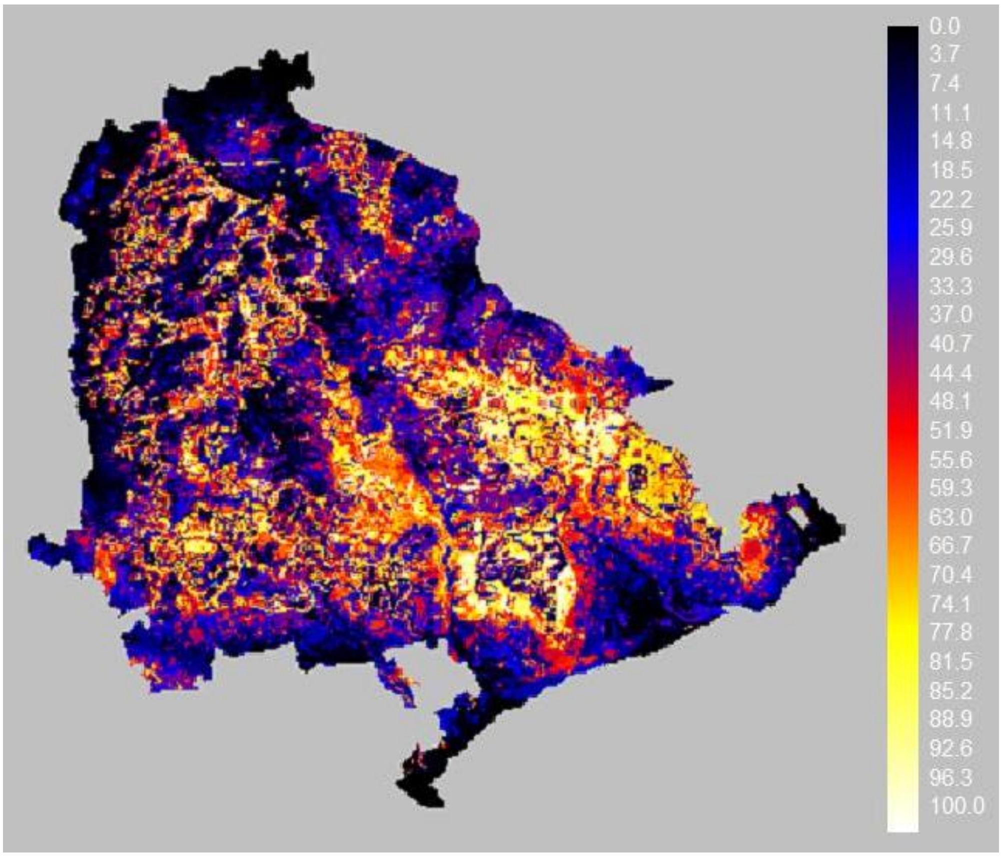


Figure 1. Biomapper habitat suitability map for *C. nobile*.

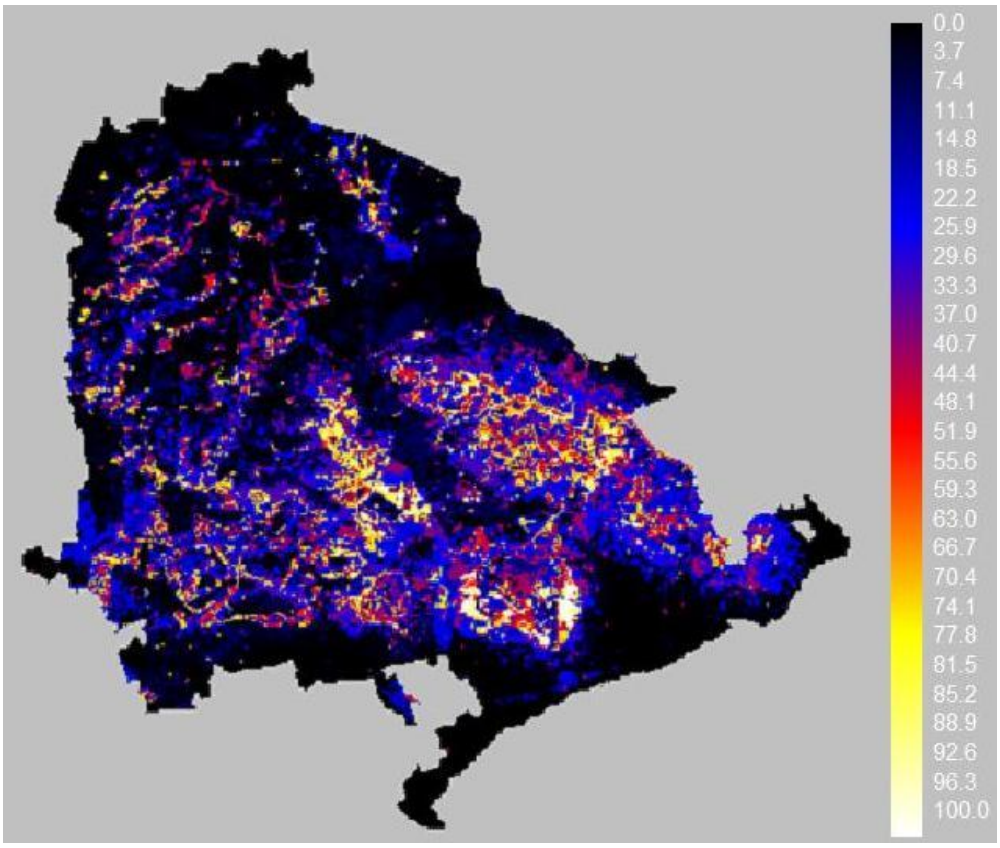


Figure 2. Biomapper habitat suitability map for *G. constrictum*.

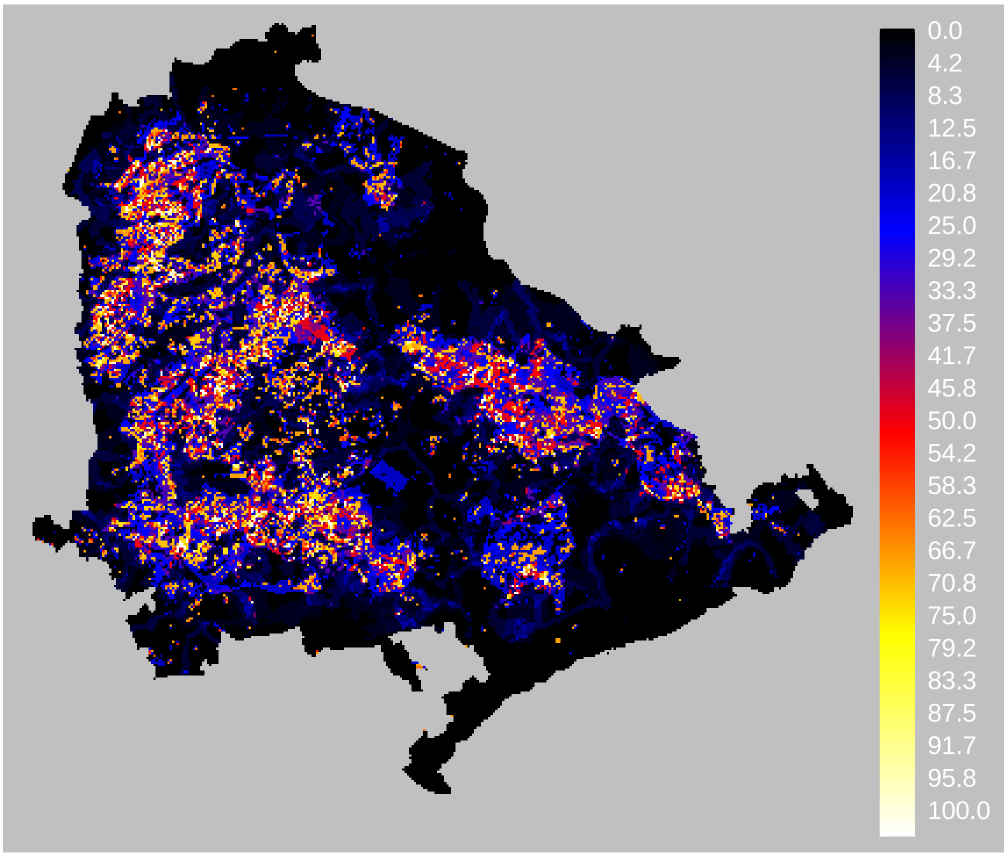


Figure 3. Biomapper habitat suitability map for *G. illyricus*.

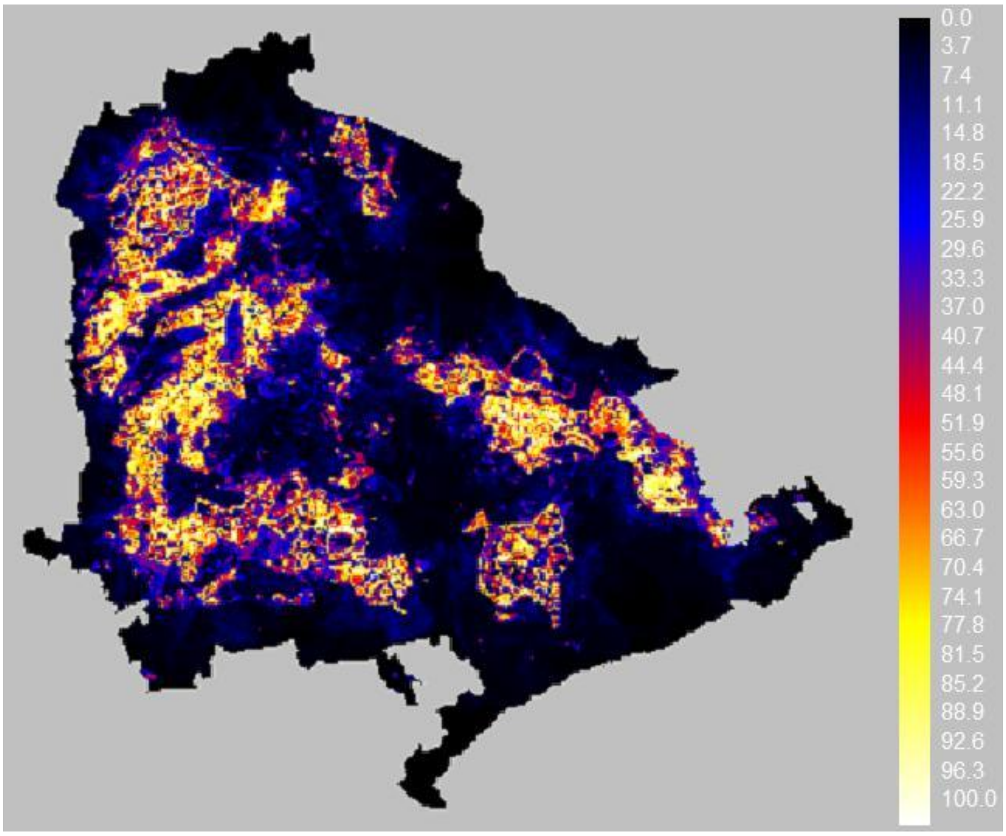


Figure 4. Biomapper habitat suitability map for *H. semele*.

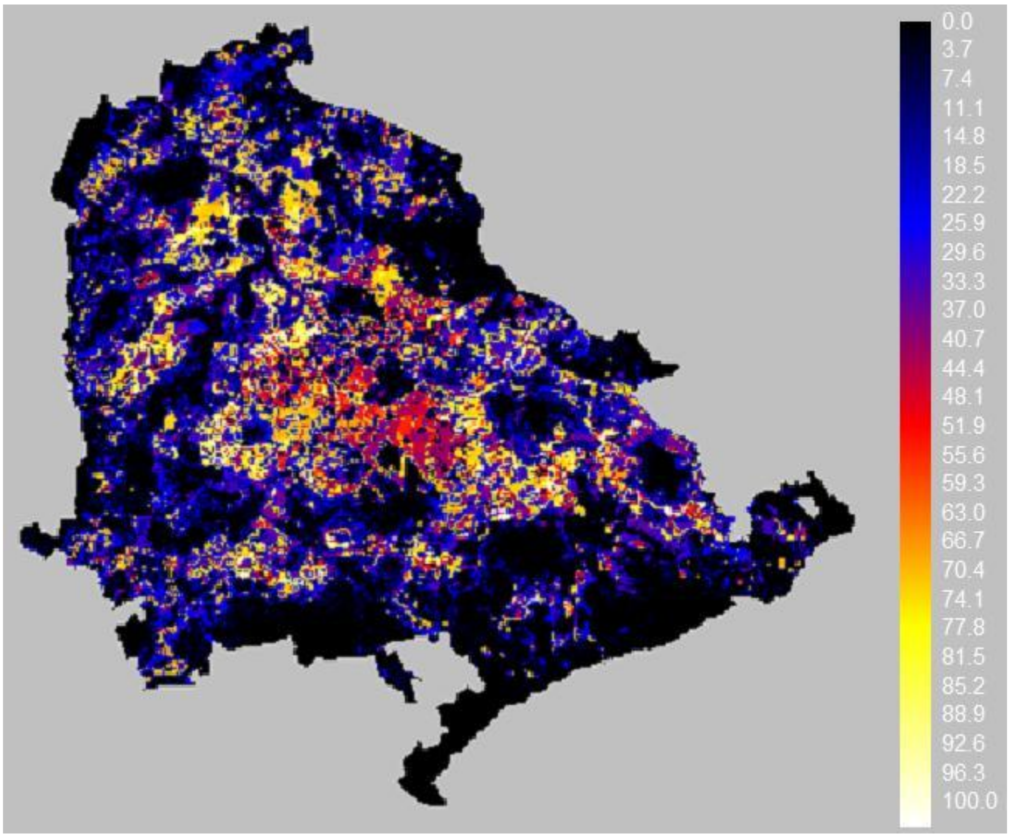


Figure 5. Biomapper habitat suitability map for *N. sylvestris*.

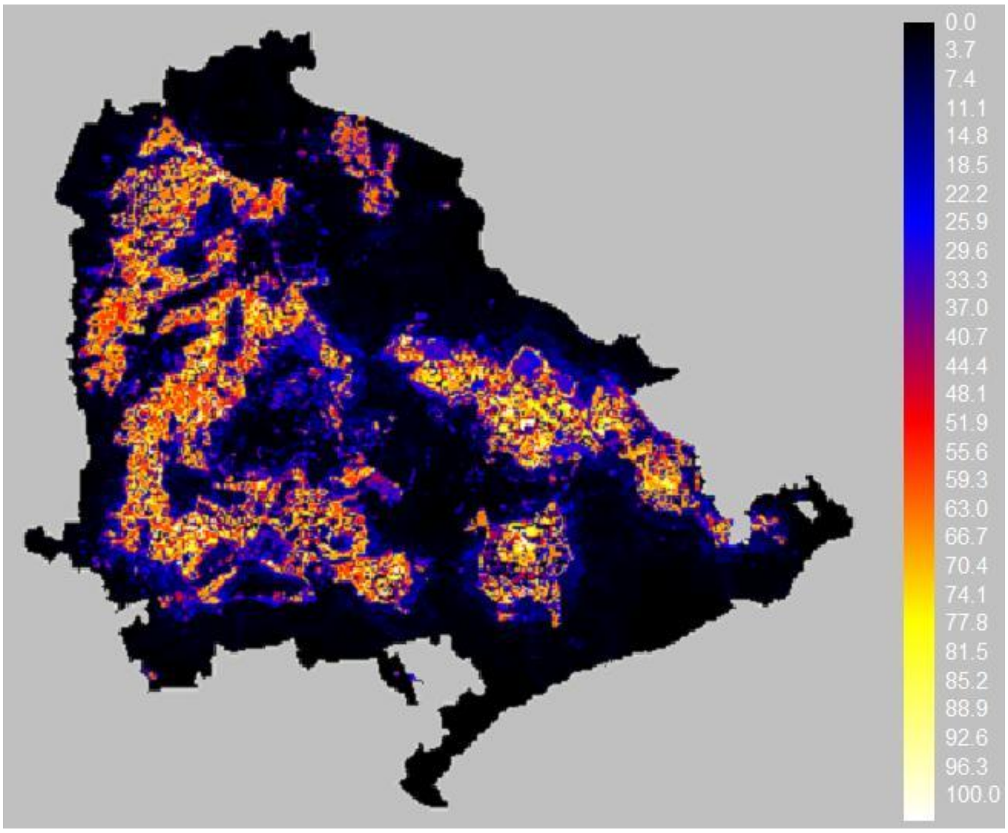


Figure 6. Biomapper habitat suitability map for *P. argus*.

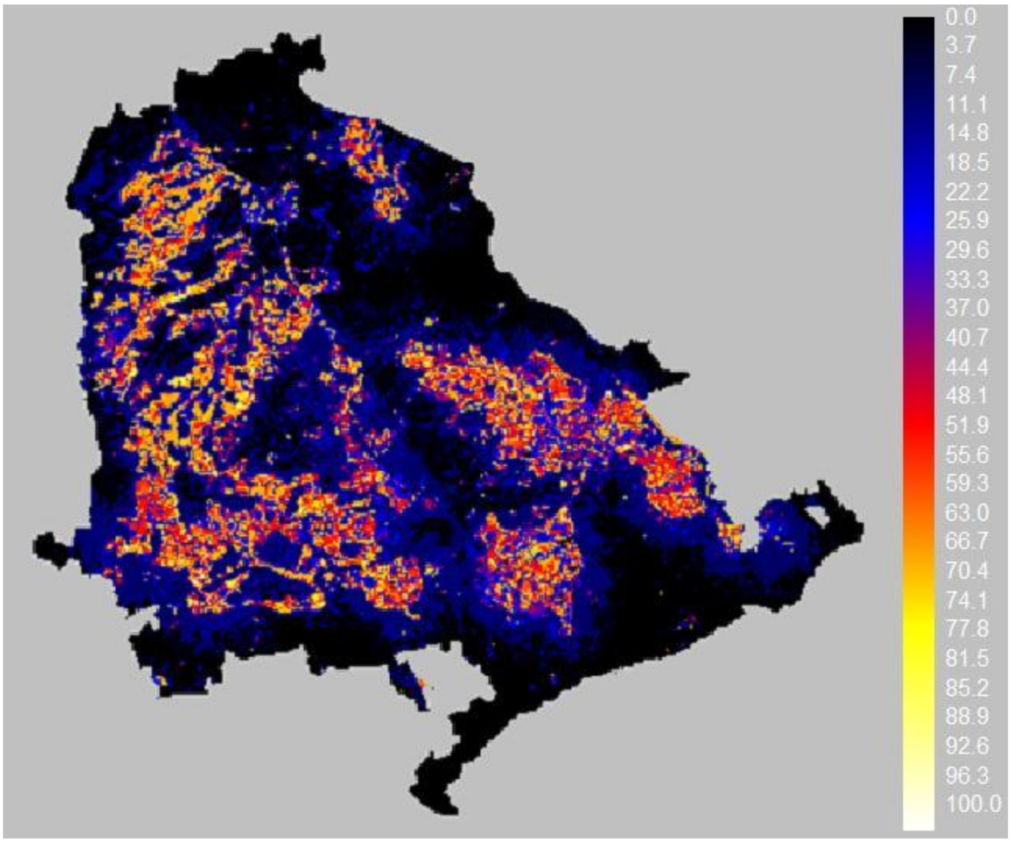


Figure 7. Biomapper habitat suitability map for *P. globulifera*.

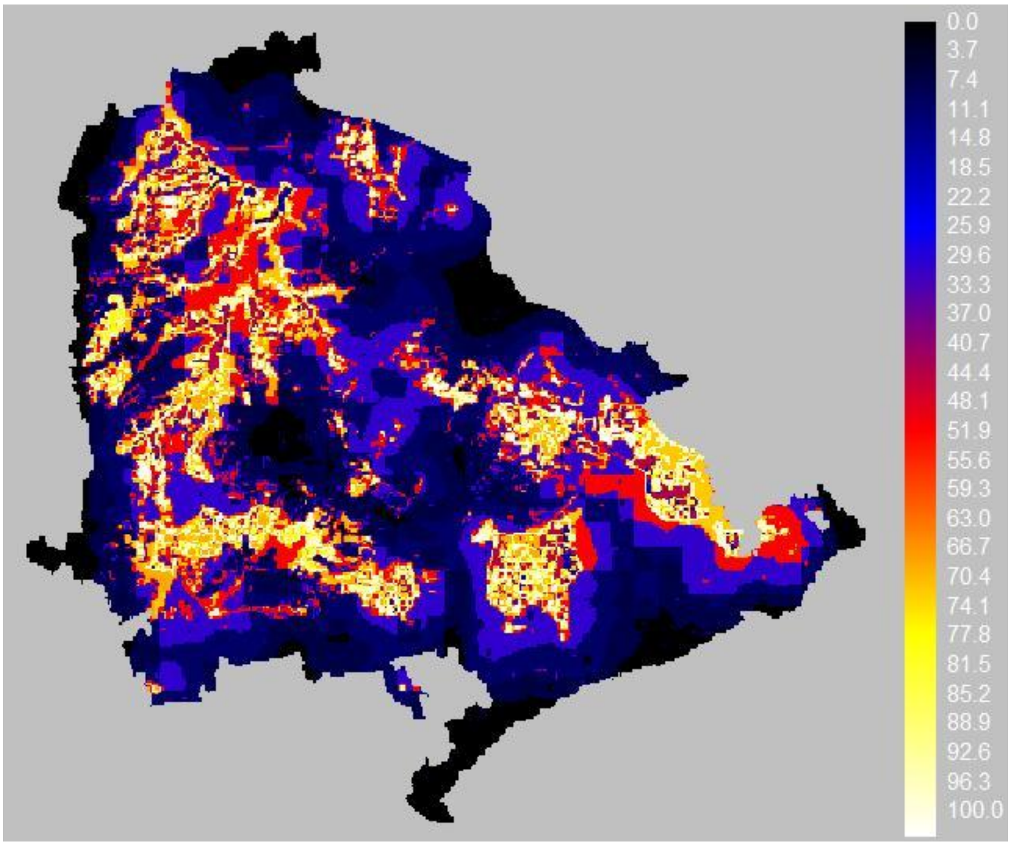


Figure 8. Biomapper habitat suitability map for *P. punctata*.

There is a visible difference between the HS maps for different species in the amount of habitat with higher HS values (i.e. the lighter colours). For example, the map for *C. nobile* has a greater area of higher habitat suitability than the map for *G. constrictum*. This may reflect less specialist habitat requirements and/or availability of suitable habitat meeting those requirements. It is important to note that the maps do not indicate the quality (i.e. condition) of the habitat, as this was not provided as an EGV.

2.3.2.4. Evaluation of the habitat suitability maps

The evaluation methods within Biomapper were used to evaluate the HS maps which, when no reliable absence data are available, involve computing statistics on the predicted habitat suitability map, comparing the whole study (global) area with the validation (testing) species presence data (Hirzel, 2008).

The box plots of the distributions of habitat suitability values for the global and species validation sets are displayed in Figures 9, 11, 13, 15, 17, 19, 21 and 23 below. A good model should produce high (80-100) HS values for the species validation set. The global box-plot gives a comparison to indicate how marginal the species is in the study area and therefore how likely the results could have been obtained by chance only (Hirzel, 2008). The box plots represent the distributions of the habitat suitability values (y axis) for the whole (global) set of cells (left box) and the validation (species) subset (right box). The boxes indicate the interquartile range, the middle line indicates the median and the whiskers encompass the 90% confidence interval. Accompanying statistics can be found in Table 11. The graphs (Figures 10, 12, 14, 16, 18, 20, 22 and 24 below) show the distribution (frequency) of global and validation cells in 50 habitat suitability classes for each of the species; better models should have a higher frequency of cells in the higher HS classes than the global cells and lower frequency in the lower HS classes.

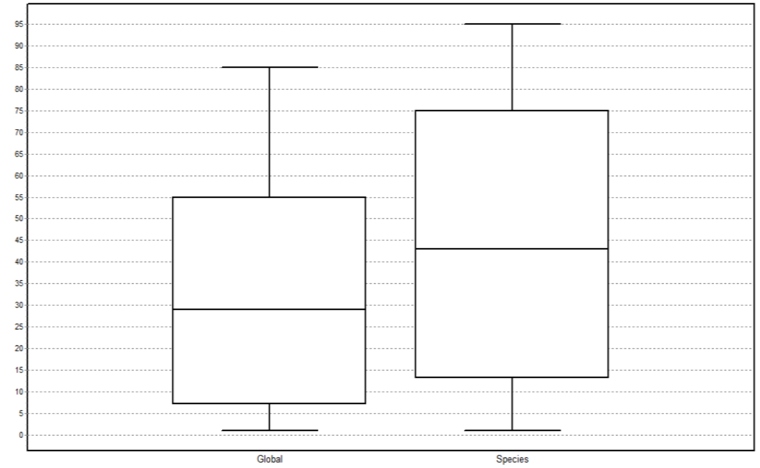


Figure 9. Box plots for *C. nobile*. Median: Species = 43, Global = 29; 1st quartile: Species = 13, Global = 7; 3rd quartile: Species = 75, Global = 55.

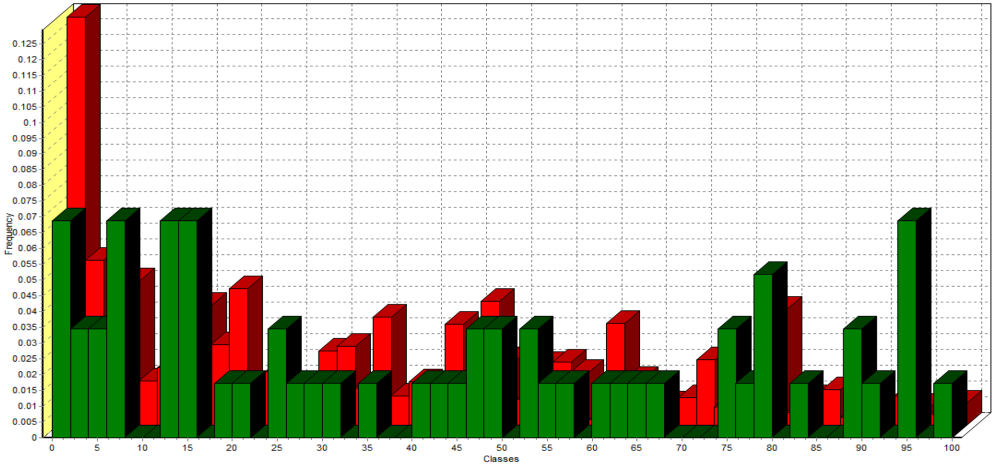


Figure 10. Biomapper graph of the distribution (frequency) of global (red) and validation (green) cells in 50 habitat suitability classes for *C. nobile*.

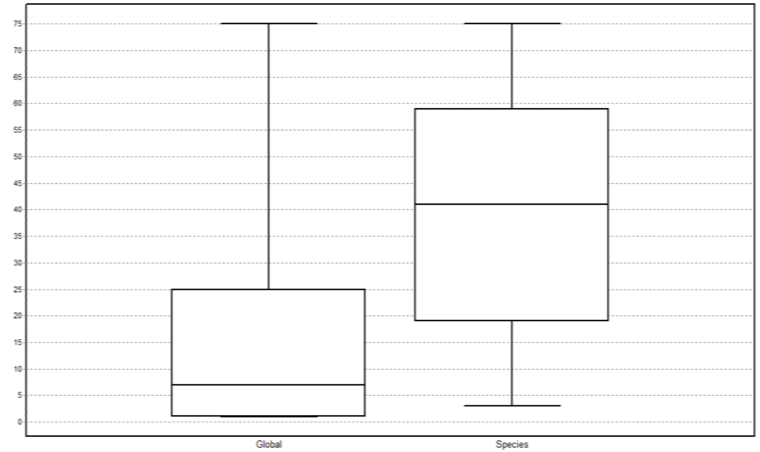


Figure 11. Box plots for *G. constrictum*. Median: Species = 41, Global = 7; 1st quartile: Species = 19, Global = 1; 3rd quartile: Species = 59, Global = 25.

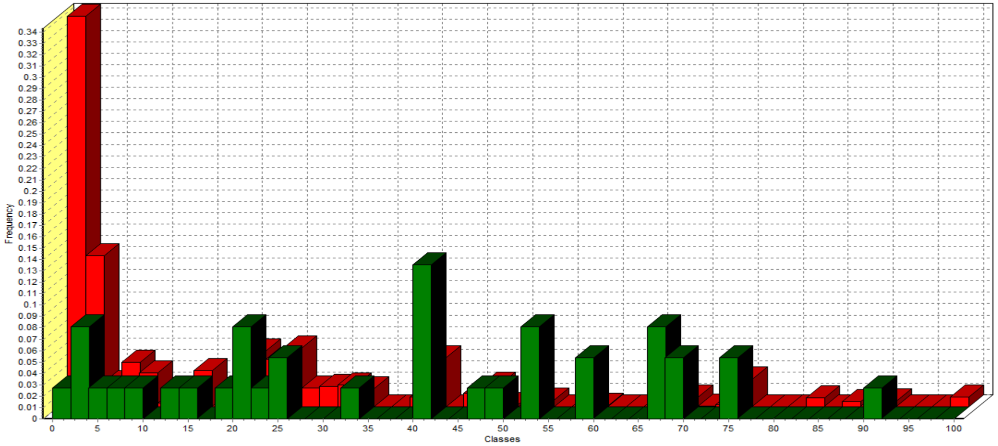


Figure 12. Biomapper graph of the distribution (frequency) of global (red) and validation (green) cells in 50 habitat suitability classes for *G. constrictum*.

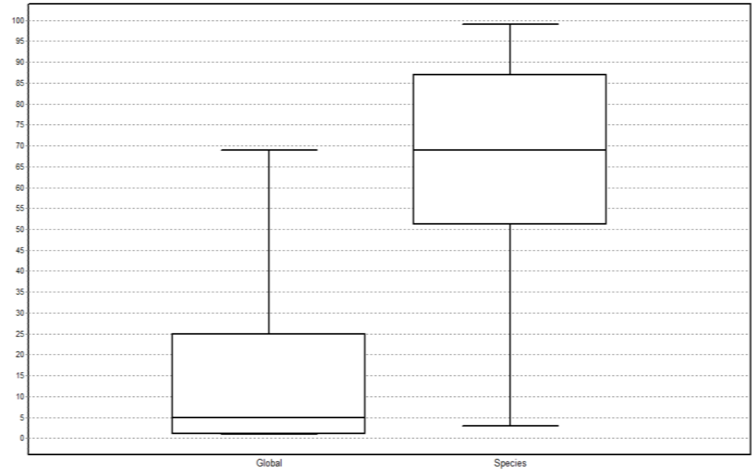


Figure 13. Box plots for *G. illyricus*. Median: Species = 69, Global = 5; 1st quartile: Species = 51, Global = 1; 3rd quartile: Species = 87, Global = 25.

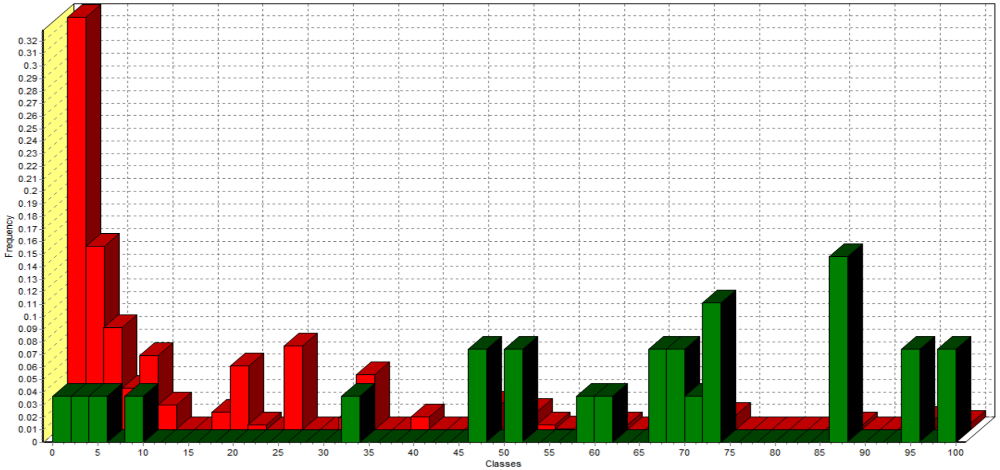


Figure 14. Biomapper graph of the distribution (frequency) of global (red) and validation (green) cells in 50 habitat suitability classes for *G. illyricus*.

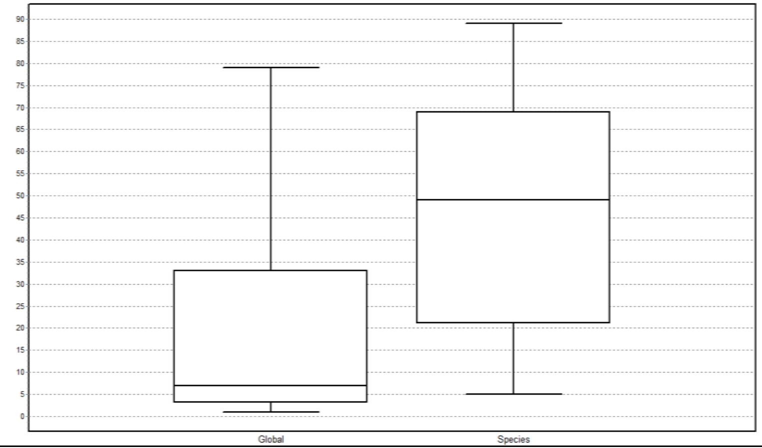


Figure 15. Box plots for *H. semele*. Median: Species = 49, Global = 7; 1st quartile: Species = 21, Global = 3; 3rd quartile: Species = 69, Global = 33.

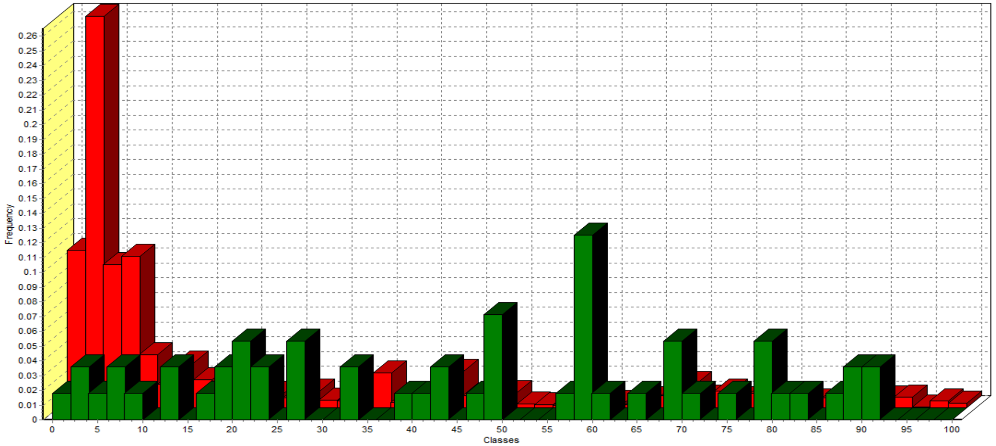


Figure 16. Biomapper graph of the distribution (frequency) of global (red) and validation (green) cells in 50 habitat suitability classes for *H. semele*.

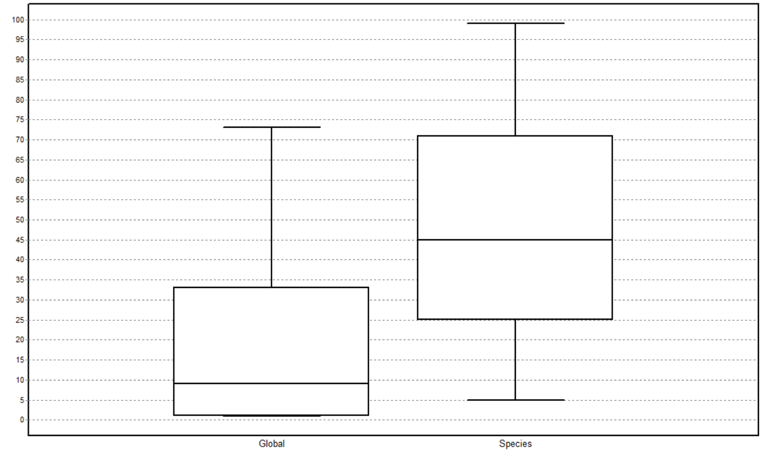


Figure 17. Box plots for *N. sylvestris*. Median: Species = 45, Global = 9; 1st quartile: Species = 25, Global = 1; 3rd quartile: Species = 71, Global = 33.

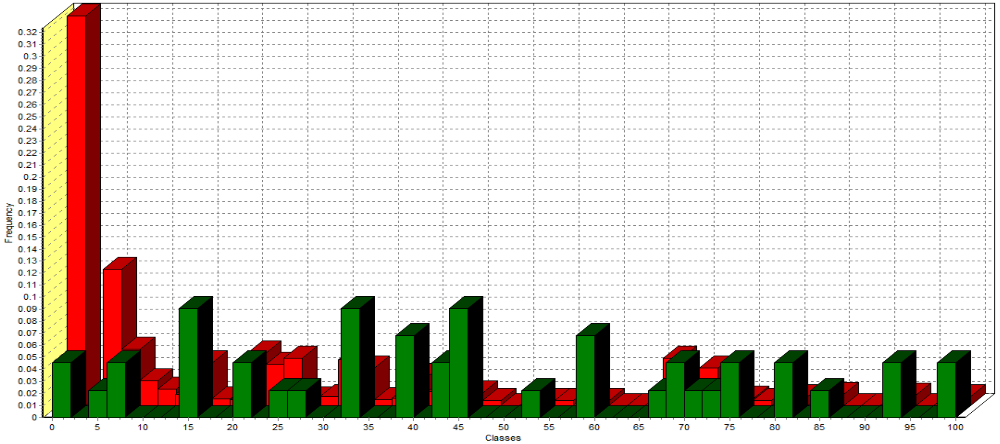


Figure 18. Biomapper graph of the distribution (frequency) of global (red) and validation (green) cells in 50 habitat suitability classes for *N. sylvestris*.

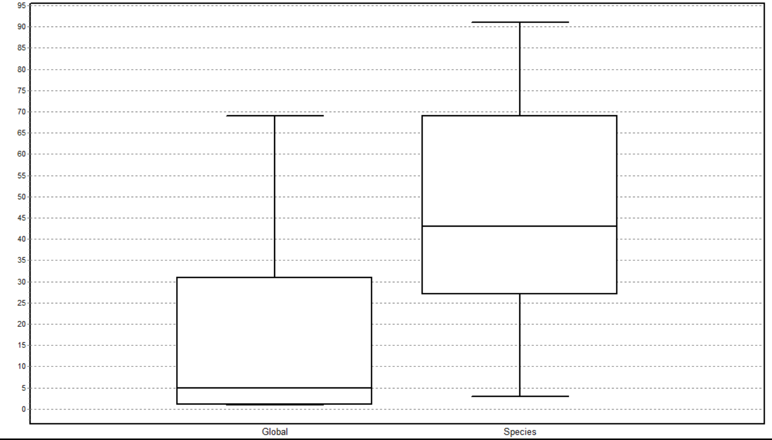


Figure 19. Box plots for *P. argus*. Median: Species = 43, Global = 5; 1st quartile: Species = 27, Global = 1; 3rd quartile: Species = 69, Global = 31.

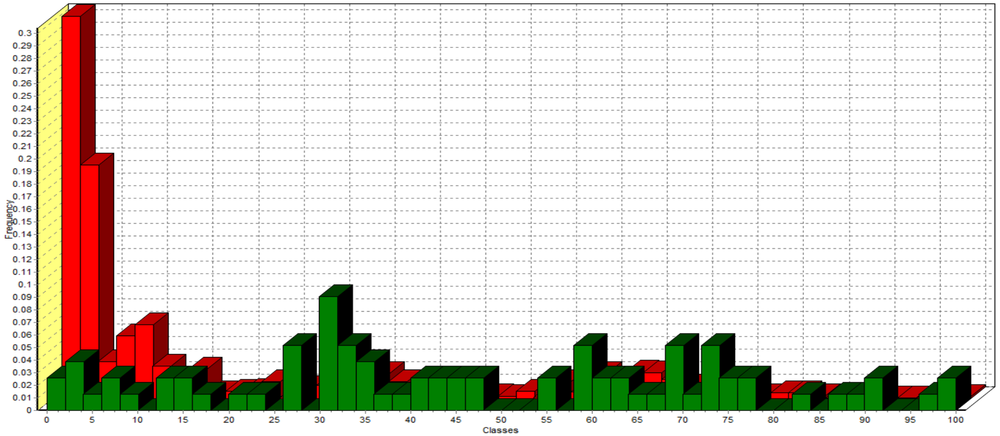


Figure 20. Biomapper graph of the distribution (frequency) of global (red) and validation (green) cells in 50 habitat suitability classes for *P. argus*.

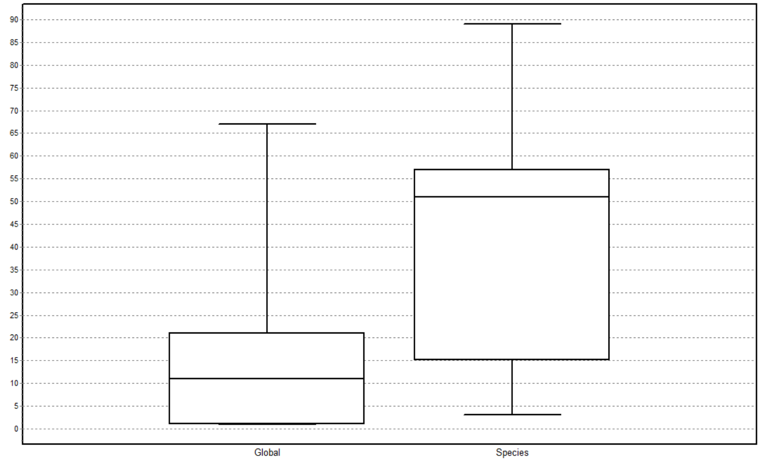


Figure 21. Box plots for *P. globulifera*. Median: Species = 51, Global = 11; 1st quartile: Species = 15, Global = 1; 3rd quartile: Species = 57, Global = 21.

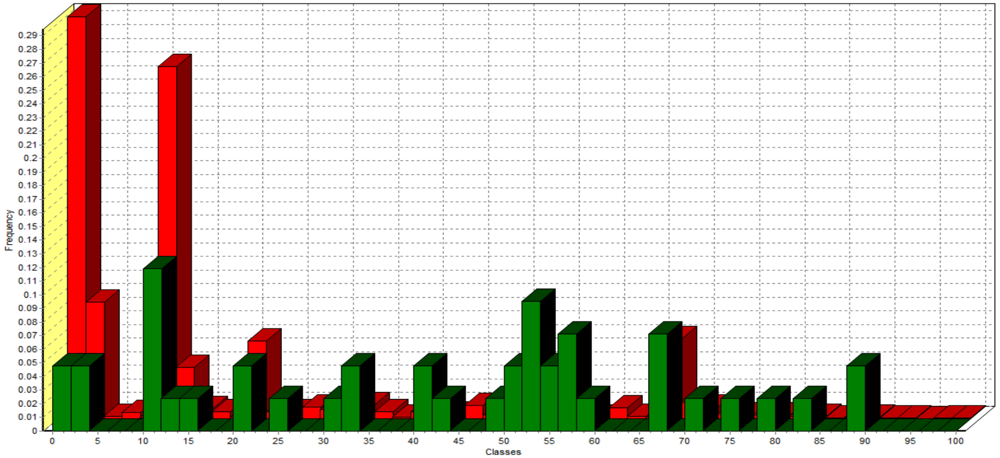


Figure 22. Biomapper graph of the distribution (frequency) of global (red) and validation (green) cells in 50 habitat suitability classes for *P. globulifera*.

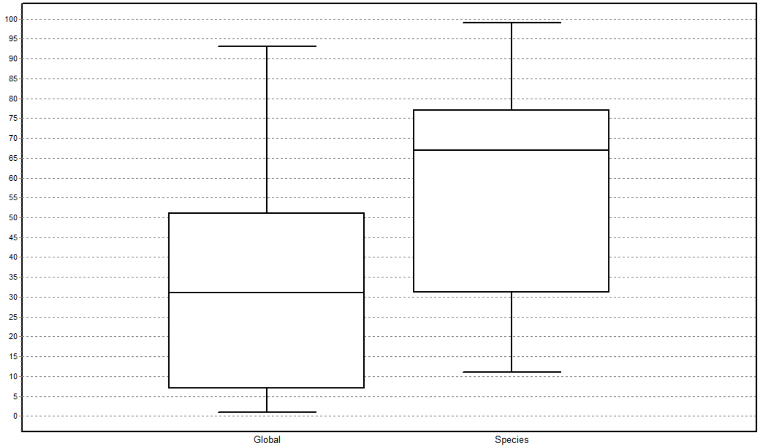


Figure 23. Box plots for *P. punctata*. Median: Species = 43, Global = 5; 1st quartile: Species = 27, Global = 1; 3rd quartile: Species = 69, Global = 31.

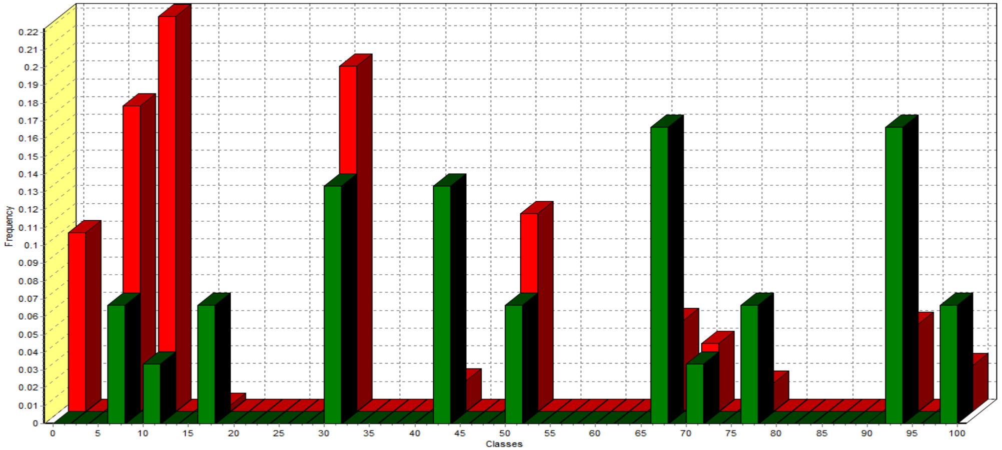


Figure 24. Biomapper graph of the distribution (frequency) of global (red) and validation (green) cells in 50 habitat suitability classes for *P. punctata*.

Biomapper also produces several statistics to accompany the box-plots and provide further information. These are provided in Table 11 below.

Species		***C. no***	***G. co***	***G. il***	***H. se***	***N. sy***	***P. ar***	***P. gl***	***P. pu***
Mean HS value	V	41.9	37.5	61.2	45.5	46.2	45.6	41.7	55.8
	G	33.1	17.0	15.5	19.9	21.2	17.7	18.1	30.0
Standard deviation	V	31.7	24.7	29.1	27.1	28.3	26.9	25.6	29.3
	G	27.8	22.6	22.4	26.1	25.7	24.2	22.6	28.3
Minimum HS value	V	1	1	0	1	0	1	0	7
	G	0	0	0	0	0	0	0	0
1st quartile	V	13	19	51	21	25	27	15	31
	G	7	1	1	3	1	1	1	7
Median HS value	V	43	41	69	49	45	43	51	67
	G	29	7	5	7	9	5	11	31
3rd quartile	V	75	59	87	69	71	69	57	77
	G	55	25	25	33	33	31	21	51
Maximum HS value	V	99	91	100	91	100	100	88	100
	G	100	100	100	100	100	100	100	100
Interquartile range	V	62	40	36	48	46	42	42	46
	G	48	24	24	30	32	30	20	44
Mode HS value	V	1	41	87	59	15	31	11	67
	G	1	1	1	3	1	1	1	11
Proportion of validation cells with HS >50 (AVI)	V	0.397	0.351	0.741	0.446	0.409	0.429	0.500	0.567
	G	0.284	0.085	0.114	0.146	0.156	0.162	0.167	0.291
Probability of getting this value by chance	V	0.588	0.928	0	0	0.918	0.753	0.097	0
	G	0.406	0.428	0.427	0.252	0.480	0.460	0.429	0.281
Value that 90% of validation cells are above	V	5	1	9	9	7	9	11	17
	G	1	1	1	1	1	1	1	1
Value that 95% of validation cells are above	V	1	3	3	5	5	3	3	11
	G	1	1	1	1	1	1	1	1
Comparisons between validation and global/all cells									
Continuous Boyce index		0.744	0.567	0.959	0.680	0.673	0.603	0.707	0.910
Standard deviation		0	0	0	0	0	0	0	0

Table 11. Validation statistics for HS maps, including statistics to accompany box plots. C. no = *C. nobile*, G. co = *G. constrictum*, G. il = *G. illyricus*, H. se = *H. semele*, N. sy = *N. sylvestris*, P. ar = *P. argus*, P. gl = *P. globulifera*, P. pu = *P. punctata.* V = Validation (species testing data) results, G = global/all cells results.

Statistic	**Description**
Proportion of validation cells >50 (Absolute Validation Index, AVI; see Appendix 8.1)	The proportion of validation points that have a predicted habitat suitability (HS) value over 50. The higher this value, the better the model. This is the most useful statistic (Hirzel, 2008).
Probability of getting this value (above) by chance	This statistic uses a bootstrap procedure to assess how likely it is that this value (proportion of validation cells with a HS value >50) could have been obtained by chance (Hirzel, 2008). The global distribution of HS values is bootstrapped using the procedure described in Hirzel (2008).
Comparisons between validation and all cells:	
Continuous Boyce Index	Boyce index values range from –1 to 1 with positive values indicating a model whose predictions are consistent with the presences distribution in the evaluation dataset, values close to zero indicating that the model is not different

	from a chance model, and negative values indicating an incorrect model (Hirzel *et al.*, 2006b). (Further information on the Boyce Index can be found in section Appendix 8.1).

Table 12. Descriptions of the statistics shown in Table 11 (from Hirzel (2008)).

Hirzel (2008) cautions that the comparisons between the validation and all-cells values are useful to assess how the model is different from what could be achieved by a random model but say nothing about the absolute quality of the model. This is because they are highly related to the global HS of the study area and if the study species is not very marginal nor very specialised, the model could be very effective but produce a very low 'far from random' score.

The best validation results were obtained for *G. illyricus*. The HS map had the highest mean HS value (61.2 (standard deviation, SD = 29.1)) for the validation data set, which was considerably higher than the global mean (15.5, SD = 22.4). It also had the highest median, first and third quartile values and the lowest interquartile range, as well as the highest mode HS value and the highest proportion of validation cells with HS>50. However, the mode statistic may not be as useful, because with a small validation sample size and a large range of possible values (0-100), the most common value may only occur one or two times more than the least common value.

The mode HS value for *C. nobile* was 1 which, without examining the other statistics, would suggest that the model is very poor. However, the other measures were also not as good. The largest interquartile range was obtained for *C. nobile*, with a low first quartile value, but a fairly high third quartile value. The difference (14) between the median HS value validation cells (43) and the global cells (29) was the smallest for *C. nobile*. The next smallest difference in medians was for *G. constrictum*, of 34, and the largest difference (of 64) was between the global median (5) and the validation median (69) for *G. illyricus*.

The lowest mean HS value was for the *G. constrictum* map, which also had the lowest median and first and third quartile values, although these values were considerably different from the values for the global cells. The mean HS values for the rest of the species' maps were over a similar range (all between 41.7 and 46.2), except for *P. punctata*, which was slightly higher at 55.8. The difference between the mean HS value for the validation cells and the global cells was generally between 20.5 and 27.9, with two exceptions: the difference was much greater for *G. illyricus* (a difference of 45.7), which confirms the good performance results described above, and a much smaller difference (8.8) for *C. nobile*, which suggests that the validation cells were not as different from the global distribution cells as for the other species. However, this could be because *C. nobile* is not as specialised or marginal (as shown by the marginality and specialisation values in section 2.3.2.2.1.1) or restricted to certain habitat types (that are not as widespread) as the other species. Therefore, the model could be very good even though it does not achieve a good 'difference from random' score (Hirzel, 2008). On the contrary, for species which are very specialised and occur in a more restricted niche, the values are likely to be more different to the global values. This is why the models are very dependent on the study area (Hirzel, 2008).

For some of the species, none of the validation cells had a HS value of 100, and for *P. globulifera*, the maximum HS value was 88 for a validation cell. This could in part be due to small sample sizes or because there just were not many sites predicted at the highest suitability values for those species. This can be seen for the HS map for *P. globulifera* (Figure 7, section 2.3.2.3).

All of the validation data sets had low minimum HS values of mostly 0 or 1, except for a value of 7 for *P. punctata*. This is a problem because, if a threshold of 50, or even much lower, was used to categorise sites into potentially suitable and unsuitable habitat, these sites would have been classified as unsuitable and would have been far less likely to be visited when trying to record new locations of presences for the species. So although the site would have been classified as unsuitable, it is a known presence site. This could be due to the species occurring at a site that is not as similar to the majority of the sites that it is found at, and it may not be as suitable and the species may have only been there short-term. It may also suggest that the model is not performing well. However, it may only be one validation record site with a low HS value, and this is where using a range of statistics (such as the interquartile range) can help.

The graphs of the distribution of global and validation cells in the HS classes should show a tendency for the validation cells to be distributed less in the low HS classes, with more in the higher HS classes and also showing the same relationship compared to the global cells (i.e. lower frequency of validation cells in the lower HS classes compared to the frequency of global cells in these classes and a higher frequency of validation cells in the higher HS classes than frequency of global cells in these higher classes). The graphs for all the species generally showed this relationship with the validation and global cells. However, they did not necessarily show a clearly higher frequency of validation cells in the higher HS value classes than the lower HS classes. The best graph for displaying these relationships was for *G. illyricus*, as it had a higher frequency of validation cells in the higher HS categories (above 50) than in the lower HS classes (below 50) and had a higher frequency of validation cells in the higher HS classes than the global cells and the opposite for the lower HS classes.

The statistic of ‘proportion of validation cells with HS>50’ also provides information about the distribution of the validation cells in lower and higher HS classes. For all but three of the species (*G. illyricus*, *P. punctata* and *P. globulifera*), the proportion of validation cells with a HS value greater than 50 was less than 0.5, which seems low, although these proportions were greater than the proportion for the global cells for all the species. However, the probabilities of getting the validation proportion values by chance were high for several of the species, in particular *G. constrictum* (0.928) and *N. sylvestris* (0.918), as well as *P. argus* (0.753) and *C. nobile* (0.588), suggesting that the HS maps were not predicting HS values greater than 50 for the validation cells any better than chance, which is poor. These results were also reflected by the ‘probability of a cell being over 50 by chance’ values, which should be lower than the proportion of validation cells with HS>50 and higher than the global cells proportion. This was the case for *G. illyricus*, *P. punctata*, *H. semele* and *P. globulifera* (although only just), which was good but was not the case for the other four species listed above. However, as discussed above, species that are not as specialised or marginal may not achieve good ‘difference from random’ scores even if the models are good. However, those species are not generally any less specialised than the other species, which all achieved low (0 – 0.097) probability values for this statistic. However, they still had low proportions of HS values greater than 50.

The *G. illyricus* map scored the best on the ‘proportion of validation cells with HS>50’ statistic, with the highest value of 0.741 and a probability of 0 of getting this value by chance. This value was also much higher than the proportion of global cells with HS values greater than 50 (the difference between the proportion of validation and global cells given by the ‘proportion of cells significantly above 50’ statistic). The *G. illyricus* map also obtained the highest value for this statistic (0.627), with the next highest value for the *P. globulifera* map (0.333). The values for the rest of the species’ maps were similar to this (down to 0.253 for *N. sylvestris*), apart from for *C. nobile* with the lowest value of 0.112.

The value that 90% and 95% of cells are above also provides information about the distribution. The values for these were generally very low, although the *P. punctata* map achieved the highest value of 17 for the 90% of validation cells statistic, compared to a value of 1 for the global cells.

The Boyce index values were quite high (generally above 0.6) for all the species, in particular for the species that scored well on the other statistics, with the highest values for *G. illyricus* (0.959) and *P. punctata* (0.910), both very high. These values indicate that the model predictions were consistent with the presences distribution in the evaluation dataset (Hirzel *et al.*, 2006b). The lowest Boyce index value was for the *G. constrictum* map (0.567), which also generally scored the least well on the other statistics. This was also the species with the lowest predicted area of suitable habitat (see HS map Figure 2, section 2.3.2.3). The other species map which scored fairly low on the other statistics, *C. nobile* (in particular the proportion of cells with HS>50 = 0.397), achieved the third highest Boyce index value (0.744).

2.4. Discussion

The validation statistics indicated that two of the species’ Biomapper models (*G. illyricus* and *P. punctata*) performed substantially better than the other models. Of the rest of the models, those for *H. semele* and *P. globulifera*, had the next highest validation statistics and the remaining four models performed relatively poorly (*G. constrictum*, *C. nobile*, *N. sylvestris* and *P. argus*, in order of poorest first). The confidence in using the HS maps for these last four species would be lower, although the HS maps for the other of species, particularly *H. semele* and *P. globulifera*, should still be used with caution. For example, lots of validation points (for all the HS maps) had low HS values and consequently these sites may have been overlooked by someone using the maps as a tool for targeting recording efforts. False negatives (omission errors) are more of a problem when trying to identify potential new sites for recording species occurrence. However, too many false positives (commission errors) are also undesirable because they would waste valuable time, and it should be noted that the false positive rate of these models has not been tested, due to lack of absence data. Therefore, although the *G. illyricus* and *P. punctata* models appear to predict suitable habitat relatively well, their ability to predict unsuitable habitat is not known. Presence-only models are known to have a tendency to over-predict species distribution owing to the lack of discriminating absences (Engler *et al.*, 2004).

Although the Boyce index values for the species' models were generally quite high, some of the other validation measure results were not so high. The Boyce index values (all with standard deviations of 0) ranged from 0.567 for *G. constrictum* to 0.910 for *P. punctata* and 0.959 for *G. illyricus*, with the remaining values between 0.603 (*P. argus*) and 0.744 (*C. nobile*). However, the proportion of validation cells with HS values greater than 50 (the Absolute Validation Index; AVI) ranged from 0.351 for *G. constrictum* to 0.567 for *P. punctata* and 0.741 for *G. illyricus*. It has been suggested that a good model should have an AVI value greater than 0.75 (Podchong *et al.*, 2009), which none of the models achieve, although the *G. illyricus* model value was only slightly lower than 0.75. Nonetheless, this value for the *G. illyricus* model was substantially higher than values achieved for threatened lichen species in Spain, where Martinez *et al.* (2006) reported AVI values ranged from 0.486 to 0.567, which they suggested were 'accurate enough'. The models for the lichen species were carried out over a much larger extent (the whole of Peninsular Spain) and at a much coarser resolution (10 km x 10 km), which could result in less accurate predictions. However, these values were still higher than the majority of AVI values obtained in this study at a much finer scale.

The AVI and Boyce index values for *G. illyricus* and *P. punctata* were better than the AVI value of 0.52 and the Boyce index value of 0.83 (SD 0.23) achieved for woodland grouse in the Black Forest in Germany (Braunisch and Suchant, 2005). However, performance of the woodland grouse model was better than the models for the other species in this study. Although the woodland grouse study area was much larger than the New Forest, the species was reported to have a relatively narrow niche breadth. Similar AVI (0.5 and 0.52) and Boyce index (0.84 (SD 0.34) and 0.81 (SD 0.27)) values to Braunisch and Suchant (2005) were obtained by Qi *et al*. (2009) for giant and red pandas in China, although similarly the standard deviations for the Boyce index were quite high, whereas all the standard deviation values for the Boyce index values for the species in this study were 0. The panda species were also highly specialised bamboo feeders and although the study area was of a much larger extent than the New Forest, Qi *et al*. (2009) used a smaller resolution of 30 m x 30 m.

None of the AVI values obtained in for the models in this study were as high as the AVI value of 0.830 obtained by Hirzel *et al*. (2002) for Alpine ibex in Switzerland, which used the same 100 m x 100 m resolution as used in this study, although across the whole of Switzerland. However, Hirzel *et al*. (2002) reported that the Alpine ibex showed very specific habitat requirements. Apart from the model for *G. illyricus*, the AVI values for the other species were all considerably lower than AVI values (0.686, 0.756, 0.715) obtained for models of three mice species developed by Reutter *et al*. (2003) at a 100 m x 100 m resolution across the whole of Switzerland. Likewise, models for cryptic bat species in Switzerland (Sattler *et al*., 2007), achieved AVI values of 0.68 and 0.77, and Boyce Index values of 0.84 and 0.98. However the *G. illyricus* and *P. punctata* models achieved higher Boyce index values than one of the bat models (for the more generalist species. A very high Boyce index value (0.97), higher than the Boyce index values achieved by any of the models in this study (although only just higher than the value for the *G. illyricus* model), was reported by Fei *et al*. (2007) for American chestnut. Of the reported studies, this was one of the most similar in extent to the New Forest study area (which is just over double the size of the Mammoth Cave National Park (USA) study area of the Fei *et al*. (2007) study. That study also used a much finer resolution (10 m x 10 m), as well a much larger sample size (2156), all factors which may have contributed to the superior model performance.

It is difficult to make meaningful comparisons between different studies which use different species (with different levels of specialism), different study area extents and resolutions, different numbers of records and different environmental variables. One of the most important factors may be whether the models include the appropriate environmental variables which are important for predicting species distributions at the scales used in the studies. Related to this is the suggestion that more specialist species may be modelled better than more generalist species (Sattler *et al.*, 2007; see below). According to the global marginality and specialisation/tolerance values (see Table 1, section 2.3.2.2.1.1), *P. punctata* had the highest specialism/lowest tolerance, but this was followed by both *N. sylvestris* and *G. illyricus*, although *N. sylvestris* had one of the lower marginality values, which is why it may not have been one of the better performing models. However, Hirzel (2008) suggests that too much significance should not be placed on comparisons of the global values between species, particularly if different variables have been used (see Hirzel (2008) for further explanation). In comparison, the results of Reutter *et al*. (2003) showed that the most specialist species mouse species in their study achieved a slightly lower AVI value (0.686) than the two far more generalist species (0.715 and 0.756), although this difference was relatively small. Although there have been a limited number of ENFA studies comparing the performance between species with different levels of specialism, higher performance of models for more specialist species has also been reported in other modelling approaches (see section 1.2.5, Chapter 1).

The use of a threshold for suitable/unsuitable of 50 is arbitrary and not necessarily ideal, but still gives an indication of model performance (alongside other measures that do not require the use of a threshold). This threshold could be lowered to reduce the number of false negatives, but this would also increase the number of false positives. However, in predicting new potentially suitable habitat, false positives are less costly than false negatives (see Appendix 8.2).

However, all of the models worked better than random, as for all of the species the proportion of validation cells with a HS value greater than 50 was greater than for the global distribution (although not all had significant P values). This was found in particular for *G. illyricus*, where the predicted suitability exceeded 50 in 74% of validation cells, which differed highly significantly ($P < 0.000$, bootstrap test) from the value of 11% expected if cells were randomly chosen from the global distribution. The HS map for *G. illyricus* could be used with most confidence as this achieved relatively high results on all of the validation statistics. The Biomapper models also helped to identify important EGVs, through the score matrix output. They also highlighted unexpected relationships, such as a preference of *N. sylvestris* for higher edge density values of dry heath/acid grassland mosaic habitat, although the occurrence of woodland edge next to heathland is the likely explanation for this (as discussed in section 2.3.2.2.1.2).

The weak performance of some of the models may in part reflect the problem of less specialised species (in comparison to the study area) perhaps getting poor ‘difference from global’ results (see section 2.3.2.4). Poorer performance of more generalist species in comparison to more specialised species was also reported for ENFA modelling of two bat species by Sattler *et al*. (2007). Although the species may seem to be relatively specialised, some of them can occur on a wide range of the available habitats within the New Forest. For example, *C. nobile* occurs in many different grassland and heath areas and *P. argus* uses both wet and dry heath which, apart from woodland, makes up a large proportion of the habitat in the New Forest. Therefore, for some of the models that appeared to perform badly, it may just be that the distributions of these species were more similar to the global distribution.

The issue of comparing the validation cells to the global cells cannot really be improved other than by using true absence data, which would require additional field survey. However, the results of the models can be used to direct field-testing. It is poor practice to just examine performance of (presence-only) validation data because a model could predict high habitat suitability everywhere and have a high rate of true positives (sensitivity), but still be a very poor predictive model if it did not distinguish between suitable and unsuitable sites (Fielding and Bell, 1997). However good a HS model is, it is only predicting habitat suitability, based on characteristics of the sites at which a species is already known. Even with the best model, it would be difficult for it to always predict suitable habitat perfectly because species do not always occur in optimal habitat. In addition, the presence of a larger proportion of the training records from this habitat may bias the relationships of species occurrence with the environmental variables towards the slightly less suitable habitat, rather than towards the more suitable habitat. Furthermore, important variables may not have been included in the model, and there are also many reasons why a site may be suitable but the species not present at that site (see section 1.2.3).

The models are also only as good as the data used to develop them and the outputs should not be used without critical ecological appraisal. For example, the score matrix showed that *P. punctata* and *H. semele* were found at sites with higher edge density values for dense scrub (ST1). However, this does not seem to fit the published information about this species. This finding may reflect recordings of these species at locations where scrub was present at the time that the habitat layer was devised but has since been cleared, as part of management practices. Cleared scrub often provides areas of young growth where ponies like to graze, and consequently excrete dung (S. Douglas, personal observation). This is one of the problems of using a habitat layer that is correct at a snapshot in time, but for some habitats in particular (such as scrub) may vary quite a bit over the period of several years of species records (see section 6.2.1, Chapter 6, for a discussion). In addition, although suggestions may be made about the ecological requirements of species based on the score matrix, these associations should not be over-interpreted, as they may arise for other reasons.

As with any inferential modelling approach, there are limitations to the Biomapper method. ENFA is a purely descriptive method and cannot extract causality relations. In particular, a variable might turn out to correlate with one of the main axes not because of its intrinsic importance, but because it correlates strongly with another crucially important variable (Hirzel *et al.*, 2002). However, this may still be useful in prediction of potentially suitable sites. Biomapper also only allows for linear dependencies within the species niche; an assumption that may be unrealistic for some variables; although this can be circumvented to some extent by transforming the environmental (EGV) data (Hirzel *et al.*, 2002; Elith and Burgman, 2003).

A further reason for poor model performance may result from the use of the median algorithm. An issue with the median algorithm is that it makes the assumption that the best habitat is at the median of the species distributions on each factor, and that these distributions are symmetric. Although this is often true, in some cases it is not:

for example when the distributions is bimodal (Hirzel *et al.*, 2002). The median algorithm was the most suitable of those available at the time.

Although the models are not without their flaws, as commented by Hirzel *et al.* (2002), they provide (at worst) important clues about preferential conditions, and are a powerful tool to draw potential habitat maps. Significantly, the models can be developed using presence-only data, which is frequently the only type of data available. These maps can be developed easily within the (free) Biomapper package, and even without outstanding performance, may still help to identify previously 'unknown' sites of high habitat suitability and potential species occurrence. Importantly, this helps to focus survey efforts and reduce time and resources in establishing a fuller understanding of species' distributions, for example as suggested by Braunisch and Suchant (2005). Further, if only small areas of potentially suitable habitat are shown to be available within the study area, the species could be deemed more vulnerable. For example, the areas of high predicted habitat suitability for *G. constrictum* and *G. illyricus* were fewer than for the other species. ENFA has been used in a similar way to evaluate the conservation status of species. For example, Santos *et al*. (2006) used ENFA to evaluate the factors that limited the distribution of a snake species, and consequently update evaluation of its conservation status, and Sattler *et al*. (2007) were able to recommend reassessment of the conservation status of a patchily distributed bat species.

2.5. Conclusion

The Biomapper habitat suitability maps varied in their predictive ability, so that some could be used more confidently than others. Apart from the models for *G. illyricus* and *P. punctata*, the results were generally poorer than those reported from other studies. However, the models still performed better than random, and although they did not show excellent performance, they could be used to help identify important ecological variables and provide an indication of sites of potentially high habitat suitability for the species, aiding in a better understanding of those species' distributions in the New Forest.

CHAPTER 3

Chapter 3. Presence-absence habitat suitability models

3.1. Introduction

Several approaches have been developed for modelling species distributions using presence-only data (see section 1.2.1, Chapter 1), such as ecological niche factor analysis (ENFA). However, Zaniewski *et al.* (2002) suggest that such approaches have generally not been found to surpass the performance of statistical modelling techniques that require systematic presence and absence data. Presence-only techniques tend, in general, to overestimate species distributions owing to the lack of absence data, which would otherwise restrict predictions where needed (Zaniewski *et al.*, 2002; Engler *et al.*, 2004; Jimenez-Valverde and Lobo, 2007). Brotons *et al.* (2004) suggest that if absence data are available, methods using this information should be preferably used in most situations. However, absence data are more difficult to obtain (see section 1.2.4, Chapter 1) and are frequently unavailable so 'pseudo-absence' data may be used instead (see Appendix 13.1).

Generalised Linear Models (GLMs) are the most commonly used technique in species distribution modelling (Rushton *et al.*, 2004). An extension of these, Generalised Additive Models (GAMs), are also increasingly being used (Thuiller, 2003). Both of these regression analysis approaches require presence and absence data and estimate the probability of presence or the abundance of a species (Elith and Burgman, 2003). Regression relates a response variable (e.g. species presence-absence) to a single (simple regression) or a combination (multiple regression) of environmental predictors (explanatory variables) (Guisan and Zimmermann, 2000). Guisan *et al.* (1998) note how the GLM approach is particularly attractive because it is very simple to implement within a geographical information system (GIS). GLMs and GAMs are extensively used in species distribution modelling (for example, see Guisan and Zimmermann (2000) and Scott *et al.* (2002b)), partly because they are so well documented and because they are particularly appropriate for such modelling (Guisan, 2002).

GLMs have three important properties: the error structure, the linear predictor, and the link function. The error structure is defined by means of the family directive.

Selection of the binomial family allows for the response (i.e. the distribution of Y) to be binary (i.e. 1/0, presence/absence) and for the model to have binomial errors (Crawley, 2007). The structure of the model relates each observed y value to a predicted value and the predicted value is obtained by transformation of the value emerging from the linear predictor. The transformation to be employed is specified in the link function, which relates the mean value of y to its linear predictor (Crawley, 2007). The most appropriate link function is the one that produces the minimum residual deviance, which for a binomial model is the logit link (Crawley, 2007). GLMs using the binomial distribution with the logit link are known as logistic regressions and are the most appropriate GLMs for presence-absence data.

GAMs are powerful nonparametric extensions of GLMs, where the assumption of a linear relationship between the response and each explanatory variable is relaxed (Pearce *et al*., 2001). One or more of the GLM functions is replaced with a smoothed data-dependent function (such as splines), which enables the fitted response surface to be a more realistic representation of the true response shape (Elith and Burgman, 2003). The great advantage of this is that the exact shape of a species response to an environmental predictor does not have to be specified prior to fitting the model, which is important because a great variety of shapes are possible (Austin, 2002a; b). This is particularly useful when there is no *a priori* reason to choose one parametric form over another for describing the shape of the relationship between the response variable and the explanatory variables (Crawley, 2007).

GLMs and GAMs are used extensively in species' distribution modelling (Elith *et al.*, 2006) and there are a large number of studies illustrating their use (also see section 1.2.2, Chapter 1, for example). They are a particularly valuable tool in a conservation context and have been used to identify important environmental attributes defining the occurrence of cryptic species (e.g. Gibson *et al.*, 2007), increase sampling efficiency for rare species, saving up to 70% of time spent in the field (e.g. Guisan *et al.*, 2006a), and aid in species' management and conservation (e.g. Pearce *et al.*, 2001; López-López *et al.*, 2007).

To predict the distribution of twenty shrub species in southern California, Franklin (1998) used GLMs, GAMs and classification tree models incorporating climate and terrain-derived variables. GAMs proved useful for exploring the shape of the response functions and evaluating them for ecological rationality, and GLMs allowed those response functions to be parameterised and their significance tested. However, classification trees were difficult to interpret in terms of the ecological response functions they approximate and hierarchical interactions they describe, but they did yield the lowest prediction errors. Meynard and Quinn (2007) found that in general GAMs and GLMs outperformed classification trees and GARP (genetic algorithm for rule-set production) when using artificial species and therefore recommend the use of GAM or GLM over the other two methods. Pearce and Ferrier (2000b) found models fitted using GAMs to provide a slightly greater level of accuracy than those derived using GLMs across a range of plant and animal species.

GLMs and GAMs have frequently been used to model species distributions at large spatial scales, such as for whole countries (e.g. Zaniewski *et al.*, 2002; Guisan and Hofer, 2003; Thomaes *et al.*, 2008) or large regions (e.g. Brotons *et al.*, 2004; Brotons *et al.*, 2007; Chefaoui and Lobo, 2007; López-López *et al.*, 2007). However, although they have been applied to a lesser extent at relatively small scales (e.g. Gibson *et al.*, 2004; Seoane *et al.*, 2006), they do not appear to have been applied to individual protected areas, the scale at which management decisions are frequently made. There have also been very few applications of GLM or GAM to modelling species' distributions in the UK (although see Eyre *et al.*, 2004; Eyre *et al.*, 2005; Newton-Cross *et al.*, 2007).

Although only presence data were available for the study species, pseudo-absences can be generated. While there are some issues with using pseudo-absences (see Appendix 13.1) they are frequently used to successfully generate GLM and GAM models (e.g. Ferrier *et al.*, 2002b; Zaniewski *et al.*, 2002; Elith *et al.*, 2006; Gibson *et al.*, 2007). Stepwise variable selection is a frequently used approach for selecting the most important variables to include in the models and is readily implemented. However, this approach has also received some criticism (see Appendix 13.2), which is why the selection of variables based on expert knowledge (from the literature, and from fieldwork experience) was tested as a comparison. Testing for spatial autocorrelation was also carried out as suggested in the literature (see Appendix 13.3). Finally, the models were tested using a range of measures, rather than relying on a single measure such as AUC (see Appendix 8.2) to test the performance of the models in predicting species potential distributions in the New Forest, to evaluate the effectiveness of the approach.

Objectives:

- To examine the performance of species distribution models requiring presence-absence data; focusing on Generalised Linear Models (GLMs) and Generalised Additive Models (GAMs).
- To compare the selection of variables based on expert knowledge and the literature to automated stepwise selection.
- To evaluate the effectiveness of this approach.

3.2. Methods

3.2.1. Species and environmental data preparation

The environmental data raster layers, as described for the presence-only models in Chapter 2 (section 2.2.1.2) were used to extract values for the environmental variables at the locations for the species data as described below (section 3.2.1.2). The presence data (including the training and testing data split) for the same eight species as used for the presence-only models described in Chapter 2 (see section 2.2.1.1, Chapter 2) were used for the presence-absence models. However, owing to the requirement of absence data for GLMs and GAMs, pseudo-absences had to be generated, as sufficient true absence data were not available. There are several options for creating pseudo-absences (see Appendix 13.1). None of these are as effective as using true absence data, but pseudo-absences were employed here because of the lack of systematically collected absence data.

It was decided to select the pseudo-absences at random from the background (i.e. excluding known presence records, of any date), which is the most common way of generating pseudo-absences (Gibson *et al.*, 2007). The number of pseudo-absences generated was the same as the number of presence data for each species (for example as used by Engler *et al.* (2004) and Hirzel *et al*. (2006b)), so that the issues of prevalence would be avoided (see Appendix 13.1.2). As absence data were not already available, this did not constitute a loss of information. However, the main problem with selecting pseudo-absences at random is that sites might be selected that are actually suitable habitat or are unknown presence locations (Engler *et al.*, 2004). Another option (which is similar to using ENFA-aided pseudo-absences (Engler *et al.*, 2004), depending on the threshold used) is to select pseudo-absences at random only from areas that are known to be unsuitable (e.g. woodland locations for a heathland species). However, as it is known that it is extremely unlikely that the species will occur in those locations, this will only indicate that the species is not present where it was always unlikely to have been in the first place (D. Golicher, personal communication, April 11, 2008).

The Hawth's Analysis Tools extension (Beyer, 2006) for ArcGIS was used within ArcMap (ESRI, 2005) to generate random points from within the study area of the New Forest National Park to use as pseudo-absences. The presence layers for each species were used to ensure that none of the pseudo-absence points were selected from known presence sites (100 m x 100 m area). Coordinates were generated for the pseudo-absence points using Hawth's Analysis Tools (Beyer, 2006). The pseudo-absence data were then randomly split into 70% training and 30% testing, as carried out for the presence data (see section 2.2.1.1, Chapter 2) and raster layers of the new pseudo-absence data created in IDRISI Andes (Clark Labs, 2006) (as for the presence data; see section 2.2.1.1, Chapter 2).

3.2.1.1. Extraction of environmental data values for species presence and absence locations

GLMs and GAMs were run outside of IDRISI Andes, so values for each of the environmental layers at each of the presence and pseudo-absence locations for each species were extracted (for use in Microsoft Excel for importing into R) using the SPLUSIDRIS function in IDRISI Andes. Any environmental variables with all values of 0 (or less than 10, or 10% non-zero values, whichever was lower) for a species' presence training data were not included in the analyses, as they can cause errors when the models are run. This was done by checking the Excel spreadsheets of the SPLUSIDRIS extractions for each species. Values for each of the environmental layers for 100 m x 100 m pixels for the whole of the New Forest

National Park study area were also extracted for model predictions (see section 3.2.2.7).

3.2.2. Generalised Linear Models (GLMs)

The free open-source software environment for statistical computing and graphics, ‘R’ (version 2.9.0) (R Development Core Team, 2009) was used for running the regression models. The training data set for each of the species was then read into R using the `read.table` function. The following steps were carried out for each species in turn.

3.2.2.1. Checking for collinearity between environmental variables

It is important to check for collinearity between environmental variables when running GLMs and GAMs. When at least one of the predictors can be predicted well from the other predictors, the standard error of the regression coefficients can be inflated and corresponding tests have reduced power (Harrell, 2001). Also, a common observation is that two highly correlated predictors can both appear non-significant even though each would explain a significant proportion of the deviance if considered individually (Guisan *et al.*, 2002). In stepwise variable selection, collinearity can cause predictors to compete and make the selection of ‘important’ variables arbitrary (Harrell, 2001) and fewer authentic variables may gain entry into the final model (Derksen and Keselman, 1992). Tabachnick and Fidell (2001) suggest a value of 0.90 and above as indicating high correlation but suggest that careful consideration should be given before including two variables with a bivariate correlation of 0.70. Therefore, in general, if the correlation between predictor variables is greater than 0.70, one of the variables should be removed from the analyses (for example: Tobalske (2002); Strauss and Biedermann (2005) and Matern *et al.* (2007)).

The `cor` function in R was used to test for high correlations between all of the environmental variables for each of the species. If there were any pairs of variables that showed correlation values equal or greater than 0.70 or equal to or more negative than –0.70, one of the pairs was removed, based on which had the least significant *P*-value from univariate GLMs of those variables (univariate GLMs were run using the function `glm` in R with family=binomial and link=logit), or if a variable had a negative estimate/coefficient value for any of the habitat percentage cover, patch area, patch compactness, edge density, soil type percentage cover or aspect variables, or a positive value for Euclidean distance to habitat type variables (as described in section 2.2.2.2, Chapter 2, for the presence-only models). Using the univariate GLMs provided an objective way of removing the correlated variables. Although variables with high *P*-values may be removed, when the variables are highly correlated, the removed variables will contain mostly redundant information.

3.2.2.2. Selection of variables for GLMs based on literature review/expert-selection

An alternative to automated stepwise variable selection is to select variables based on available literature or expert knowledge (Derksen and Keselman, 1992; Harrell *et al.*, 1996; Steyerberg *et al.*, 1999; also see Appendix 13.2). Therefore, this approach was used, and compared to stepwise variable selection (see section 3.2.2.3). Variables were selected before running any stepwise selections, so that the choice of variables was not influenced by those selections.

A set of relevant variables for each species were selected based on a review of the literature (see Appendices 2 and 20 (excluding information from species experts used to develop models for Chapter 4)) and on personal experience from carrying out fieldwork as part of this project. The maximum number of variables selected for each species was determined by the rule of thumb suggested by Harrell *et al.* (1996), which is often used in species distribution modelling studies to determine the number of observations required in relation to the number of environmental predictors (Guisan and Zimmermann, 2000). The rule suggests that no more than *m*/10 predictors should be included in the final model, where *m* is the total number of observations or the number of observations in the least represented category in the case of a binary response (see Appendix 15 for this number for each species).

The variables were selected independently of which variables were kept or discarded when they were checked for collinearity (section 3.2.2.1). However, the selected variables were checked for collinearity (as in section 3.2.2.1) and if any variables were highly correlated only the most relevant of the pair was kept. The variables selected for each species can be found in Tables 13 to 20 below and justification for these choices in Appendix 14. More information about these variables can be found in Appendices 4 to 7.

GLMs were run using the selected variables using the function `glm` in R with family=binomial and link=logit (for more information on the structure of GLMs see section 3.1). Any variables with 'wrong' relationships (see section 3.2.2.1 above) were removed (except for Eden.W1.W2 for *N. sylvestris*; see section 3.3.1.2) after the initial GLM, and the GLM re-run. (The variables that were removed from the GLMs are indicated by * in tables 13 to 20 below).

C. nobile:

Variables categories	**Variable names**
Terrain	Apect.flat*, Slope
Soil type	S64301, S64303, S71107, S84102
Habitat type (% cover in pixel)	HL2.cover*, HL3.cover
Patch area of habitat type	-
Patch compactness of habitat type	-
Edge density of habitat type	Eden.HL1.HL3
Euclidean distance to habitat type	ED.HL1.HL3, ED.AQ5.AQ6
Total number of variables selected:	11 → 9

Table 13. Expert-selected variables for *C. nobile*. * indicates variables that were removed after initial GLM due to wrong slope/estimate value. The first number of the total number of variables selected is the number of variables in the initial set of variables and the second number is the number of variables after the 'wrong relationship' variables were removed.

G. constrictum:

Variables categories	**Variable names**
Terrain	Aspect.flat*, Slope
Soil type	S71107*, S84102
Habitat type (% cover in pixel)	AQ5.cover, HL1.HL3.cover, HL2.AQ1.cover
Patch area of habitat type	-
Patch compactness of habitat type	-
Edge density of habitat type	-
Euclidean distance to habitat type	ED.HL2.AQ1, ED.AQ5.AQ6/ED.AQ5*
Total number of variables selected:	9 → 6

Table 14. Expert-selected variables for *G. constrictum*. * indicates variables that were removed after initial GLM due to wrong slope/estimate value. The first number of the total number of variables selected is the number of variables in the initial set of variables and the second number is the number of variables after the 'wrong relationship' variables were removed.

G. illyricus:

Variables categories	**Variable names**
Terrain	Slope
Soil type	S64301
Habitat type (% cover in pixel)	GL8.cover, HL3.cover
Patch area of habitat type	-
Patch compactness of habitat type	-
Edge density of habitat type	-
Euclidean distance to habitat type	ED.GL8, ED.wood.edge
Total number of variables selected:	6

Table 15. Expert-selected variables for *G. illyricus*. * indicates variables that were removed after initial GLM due to wrong slope/estimate value.

H. semele:

Variables categories	**Variable names**
Terrain	-
Soil type	S64303
Habitat type (% cover in pixel)	ST1.cover
Patch area of habitat type	PA.HL3
Patch compactness of habitat type	-
Edge density of habitat type	-
Euclidean distance to habitat type	ED.HL1.HL3
Total number of variables selected:	4

Table 16. Expert-selected variables for *H. semele*. * indicates variables that were removed after initial GLM due to wrong slope/estimate value.

N. sylvestris:

Variables categories	**Variable names**
Terrain	Aspect.south*
Soil type	S64303
Habitat type (% cover in pixel)	-
Patch area of habitat type	PA.W1.W2, PA.W7.W8
Patch compactness of habitat type	-
Edge density of habitat type	Eden.W1.W2, Eden.W5.W6, Eden.W7.W8
Euclidean distance to habitat type	ED.wood.edge
Total number of variables selected:	8 → 7

Table 17. Expert-selected variables for *N. sylvestris*. * indicates variables that were removed after initial GLM due to wrong slope/estimate value. The first number of the total number of variables selected is the number of variables in the initial set of variables and the second number is the number of variables after the ‘wrong relationship’ variables were removed.

P. argus:

Variables categories	**Variable names**
Terrain	Aspect.flat, Aspect.south, Slope
Soil type	S64301, S64303, S71107
Habitat type (% cover in pixel)	-
Patch area of habitat type	PA.HL1.HL3, PA.HL2
Patch compactness of habitat type	-
Edge density of habitat type	Eden.HL1.HL3, Eden.HL2
Euclidean distance to habitat type	ED.HL2
Total number of variables selected:	11

Table 18. Expert-selected variables for *P. argus*. * indicates variables that were removed after initial GLM due to wrong slope/estimate value.

P. globulifera:

Variables categories	**Variable names**
Terrain	Aspect.flat*, Slope
Soil type	64301*, 64303*, S71107* (+ S84102)
Habitat type (% cover in pixel)	HL1.HL3.cover, HL2.AQ1.cover, AQ5.AQ6.cover
Patch area of habitat type	-
Patch compactness of habitat type	-
Edge density of habitat type	-
Euclidean distance to habitat type	ED.AQ5.AQ6, ED.HL2.AQ1
Total number of variables selected:	10 → 7

Table 19. Expert-selected variables for *P. globulifera*. * indicates variables that were removed after initial GLM due to wrong slope/estimate value. The first number of the total number of variables selected is the number of variables in the initial set of variables and the second number is the number of variables after the 'wrong relationship' variables were removed.

P. punctata:

Variables categories	**Variable names**
Terrain	-
Soil type	S64301, S64303
Habitat type (% cover in pixel)	HL1.HL3.cover, HL2.cover*
Patch area of habitat type	-
Patch compactness of habitat type	-
Edge density of habitat type	-
Euclidean distance to habitat type	ED.HL1.HL3
Total number of variables selected:	5 → 4

Table 20. Expert-selected variables for *P. punctata*. * indicates variables that were removed after initial GLM due to wrong slope/estimate value. The first number of the total number of variables selected is the number of variables in the initial set of variables and the second number is the number of variables after the 'wrong relationship' variables were removed.

3.2.2.3. Stepwise (and full) GLMs

Independently of the expert-selected variables, GLMs were run in R for each species with the remaining variables after exclusion of correlated variables (see section 3.2.2.1) and those with no, or very few, non-zero values (see section 3.2.1.1), so that stepwise variable selection could be run. Stepwise selection was run as well as expert-variable selection to provide a comparison, and to benefit from the advantages of automated stepwise selection as discussed in Appendix 13.2.

Although there are several drawbacks of stepwise selection (see Appendix 13.2), this method of variable selection was used so that models could be run objectively without too much prior knowledge of the species, bearing in mind that the outcome of the project is that these models could be applied for large numbers of species, quickly and at low cost. In addition, the requirements of many species may not be well known and new associations (or surrogate variables) may become apparent (Hosmer and Lemeshow, 2000). In particular, stepwise variable selection was used to help identify which of the variables of percentage cover, patch area, edge density, patch compactness and Euclidean distance to habitat types had the best predictive ability for each species.

As one of the criticisms of stepwise selection procedures is that variables that are negatively related to species distribution may be selected above those with positive relationships (R. Clarke, personal communication, February 19, 2008; see Appendix 13.2), those with negative coefficients (or positive coefficients for Euclidean distance variables), except for the elevation, slope and climate variables (where the coefficient may be positive or negative), were removed (as explained in section 3.2.2.1) and the GLMs re-run. A list of the set of variables used in the full GLMs for potential selection by stepwise variable selection for each species can be found in Appendix 15.

Stepwise GLMs were fitted using the `step.AIC` function from the `MASS` library (Venables and Ripley, 2002) in R, with Akaike's Information Criterion (AIC; see Appendix 13.2) as the stopping rule. AIC is a widely used stopping rule and is preferable, particularly when modelling species' distributions that may exhibit spatial autocorrelation (see Appendix 13.3), because it does not fully rely on significance thresholds (Segurado *et al.*, 2006). Tests of forwards selection were carried out but it was found to be highly dependent on the order of the variables entered into the model (as reported by Pearce and Ferrier (2000b)) and both-directional stepwise selection was found to produce the same results as backwards selection, so consequently it was decided to just use backwards selection. Backwards stepwise selection is also reported to usually perform better than forward stepwise methods (see Appendix 13.2), and it requires examination of a full model fit, which is the only fit providing accurate standard errors, error mean square, and *P*-values (Harrell, 2001).

3.2.2.4. Final GLMs (of combined stepwise and expert selected variables)

To create a final 'best' model for each species, sets of variables were determined based on the most significant variables selected by both stepwise and by expert selection. These models should then include the most statistically significant variables, as well as being checked that they include variables that are known to be of ecological importance (Hosmer and Lemeshow, 2000). Variables were included in the final set of variables based on a significance value of $P<0.1$, rather than the more stringent $P<0.05$, so that important variables were not omitted, which might be more significant when part of a different set of variables. The number of variables included within the final set was kept within the maximum suggested number for each species (see Appendix 15), with the most significant being selected first. Descriptions of which variables were selected can be found in section 3.3.1.4 and Appendix 16.

3.2.2.5. Testing for spatial autocorrelation in final GLMs

As discussed in Appendix 13.3, species distribution models should be checked for spatial autocorrelation (SAC) in the model residuals. Moran's *I* test is a frequently used measure for this purpose (see Appendix 13.3) and was used to check for SAC in the residuals of the final 'best' GLM models for each species.

To calculate Moran's *I* in R, a 'weights list' object is required, which is created most simply from a neighbour object. Following the guidelines of Crawley (2007), the `knearneigh` function (`spdep` library (Bivand, 2009)) was used to convert the list of coordinates for the training presence and absence locations (which contains no information about neighbours) into an object of class `knn` (asking for the four nearest neighbours). This `knn` object was then converted into a neighbour object using the `knn2nb` function and this neighbour object then converted into a weights object using the `nb2listw` (`spdep` library) function. Moran's *I* was then calculated using the `lm.morantest` function in the R library `spdep` (Bivand, 2009).

3.2.2.6. Generalised Linear Mixed Models (GLMMs)

As the majority of the GLMs showed the presence of spatial autocorrelation in the residuals, Generalised Linear Mixed Models (GLMMs; see Appendix 13.3), which allow for within-group errors to be spatially autocorrelated, were run for each species using the training data and the 'full' set of variables used for entry into stepwise GLMs (see section 3.2.2.3 and Appendix 15). The `glmer` function in the `lme4` library (Bates *et al.*, 2008) was used and the GLMMs were fit by the Laplace approximation, which appears to be the current best method (Bolker *et al.*, 2009). For each species, all data were assigned to the same group using a grouping factor in R (after Dormann *et al.*, 2007a), as within-group errors of GLMMs may be spatially autocorrelated (Venables and Ripley, 2002; Dormann *et al.*, 2007b). This formed the random effects part of the GLMM model. The environmental variables made up the fixed effects part of the model.

GLMMs were also run for the expert set of selected variables and the best/final set. In all cases the coefficient and *P*-values were almost identical to the values obtained by the GLMs (see Appendix 17), so this was not taken any further as it appeared that the same variables would be selected whether GLM or GLMM was used.

3.2.2.7. Predictions for testing data for final GLMs

The `predict.glm` function in R was used to make predictions for the testing data (the 30% split) using the 'final/best' model (best expert and backwards stepwise variables combined) for each species so that the models could be evaluated. `Type=response` in this function was selected to ensure that the returned values were back transformed so that they were on the scale of the response variables (i.e. on the probability scale) rather than on the logit scale (Fox, 2002; Crawley, 2007).

Predictions were also made for the whole of the New Forest study area so that predicted values were available for any site and, if required, maps of the predicted values could be created in IDRISI.

3.2.3. Generalised Additive Models (GAMs)

In order to ascertain whether better models could be developed for *C. nobile* and *P. punctata*, the only two GLM models that did not have AUC evaluation values of 0.8 or above (see section 3.3.3), GAMs were run to determine whether any of the variable responses showed non-linear relationships (which would not have been selected for by the GLMs) that may improve model performance. The `gam` package (Hastie, 2009a) was used to run the GAM models in R as it was developed to match the `glm` function in R and has a similar stepwise selection procedure (Hastie, 2009b). It uses smoothing splines to produce a smooth fit to the data.

3.2.3.1. Checking for collinearity between environmental variables (GAMs)

The same correlated variables were removed as for the GLMs (based on the coefficient values of the GLMs; see section 3.2.2.1).

3.2.3.2. Selection of variables for GAMs based on literature review/expert-selection

The full set of expert-selected variables as used for the GLMs (see section 3.2.2.3) were used for the GAMs, including the variables with ‘wrong’ relationships that were removed in the expert GLMs (marked with * in Tables 13 to 20 (section 3.2.2.2); Aspect.flat and HL2.cover for the *C. nobile* model and HL2.cover for the *P. punctata* model).

3.2.3.3. Stepwise (and full) GAMs

Univariate GAMs were run using the set of variables used for the GLMs after (one of the pair of) the correlated variables and the variables with all or mostly zero variables had been removed (before removal of variables with ‘wrong’ relationships, as with non-linear fitting, there is not necessarily a straightforward positive or negative relationship, so this was not a factor for the GAMs). This was carried out to determine which variables were more significant to reduce the number of variables entered into the stepwise GAM later on, because as each variable is entered into the model linearly, smoothed or not at all (see below), the model call can become very long and may not run if there are too many variables. This is an issue for running GAMs.

Full GAMs were run using the variables from the univariate GAMs with *P*-values less than or equal to 0.1 (erring on the side of generosity, as the significance may change when in combination with other variables). A list of these variables (15 for *C. nobile* and 10 for *P. punctata*) can be found in Appendix 15. The R function `gam` (in the `gam` package (Hastie, 2009a)) with family=binomial and link=logit was used to run the univariate and 'full' GAMs. All of the variables were entered into the model as non-parametrically smoothed functions, s(Variable).

Stepwise GAMs were run using the `step.gam` function in the `gam` package (Hastie, 2009a). The function requires a scope argument that is a list of formula corresponding to a term in the model. Each of these formulae specifies a 'regimen' of candidate forms in which the particular term may enter the model (Hastie, 2009b). For example, a term formula might be ~1 + Variable + s(Variable), meaning that the variable could either appear not at all, linearly or as a smoothed function estimated non-parametrically. The stepwise selection was run in the backwards direction as well as in the 'both' direction for comparison.

Final 'best' models (as created for the GLMs; see section 3.2.2.4) were not created for the GAMs as the two significant expert-selected variables for *C. nobile* were entered into the stepwise models and none of the *P. punctata* expert-selected variables were significant. However, non-significant variables were removed from the *C. nobile* model and the GAM re-run.

3.2.3.4. GAM plots of variables selected for final GAMs

GAM plots can be used to examine the shape of the smoothed relationship for model variables, so were generated for the stepwise-selected variables using the `plot.gam` function in the `gam` package (Hastie, 2009a).

3.2.3.5. Predictions for testing data for GAMs

The `predict.gam` function (in the `gam` package (Hastie, 2009a)) was used to make predictions for the testing data using the stepwise GAM models for each species so that the models could be evaluated. As for the GLMs (see section 3.2.2.7), predictions were also made for the whole of the New Forest study area.

3.2.4. Evaluation of GLMs and GAMs

A range of evaluation measures and thresholds were used to examine the predictive performance of the models, rather than relying on a single measure. Despite AUC being widely used in species distribution modelling studies as a threshold independent measure, and being the current best practice for assessing model success for presence-absence data (Austin, 2007), there are several issues with it (e.g. Lobo *et al.*, 2008), as there are with alternative methods. Therefore, as there does not appear to be an ideal measure, it is best to report on the results of a range of measures, including sensitivity and specificity in particular (Lobo *et al.*, 2008).

The free Delphi program ROC_AUC (Schröder, 2006) was used for calculation of all of the evaluation measures (using the data from the GLM or GAM (testing data) predictions (section 3.2.2.7)) as it provides a very quick and easy-to-use interface. It provides calculation of the area under Receiver Operating Characteristic (ROC) curves as well as a range of threshold-dependent criteria with optimisation of a range of cut-off values (see Table 21 below). It also allows for different weightings to be assigned to the cost of omission and commission errors (see below). This software was used by Matern *et al.* (2007) for example, who used it to validate their logistic regression models for predicting habitat suitability for a rare ground beetle species.

Cut-off value	Description
P_opt	The cut-off value which maximises the proportion of correct classifications (Schröder, 2008).
P_fair	The cut-off value at which sensitivity and specificity are the same. It has the advantage that both misclassifications – false presence predictions and false absence predictions – have the same probability to occur (Schröder and Richter, 1999/2000).
P_Kappa	The cut-off value at which Kappa is maximised.
P=0.5	A commonly used cut-off threshold.
P_crit	The applied cut-off value.

Table 21. Cut-off values provided for the calculation of threshold-dependent performance criteria; sensitivity, specificity, correct classification rate and Kappa in ROC_AUC software.

There are often situations when the assumption of equivalent false negative (commission) errors and false positive (omission) errors used to derive the measures from the confusion matrix can be questioned (Fielding and Bell, 1997; see Appendix 8.2). In this application, false negative (FN) errors would be more costly (in conservation terms) than false positive (FP) errors, because FN errors may mean that a new occurrence (or potential occurrence) location for a species is overlooked. Although too many FP errors would present a cost in time and resources, this cost is still felt to be less than the FP cost. In addition, Lobo *et al*. (2008) suggest that because absences have a higher degree of uncertainty than presences when recording species presence-absence, owing to low detectability or non-sampled areas, false absences are more likely to occur than false presences. This is particularly relevant when pseudo-absences are randomly selected from the background, as this procedure inflates the number of false absences. Therefore, Lobo *et al*. (2008) suggest that commission errors (FP) should not weigh as much as omission errors (FN).

These inequalities can be compensated for partly by the choice of error measure and threshold, for example if FN errors are more serious than FP errors the threshold can be adjusted to decrease the FN rate at the expense of an increased FP error rate (Fielding and Bell, 1997). However, different weights can be easily allocated to the FP cost and the FN cost within the ROC_AUC, so this was carried out as a comparison to the equal weighting of errors.

As noted by Fielding and Bell (1997), in the absence of clear economic gains and losses, the allocation of weights must be subjective. Therefore, the choice of cost weightings for this application was subjective, but based on the cost of false negatives being greater, but not too much greater, than the cost of false positives (for the reasons discussed above). The false positive cost was set at 0.4 and the false negative cost at 0.6.

3.3. Results

3.3.1. Generalised Linear Models (GLMs) results

3.3.1.1. Checking for collinearity between environmental variables

Unsurprisingly, the merged habitat type layers and the individual habitat layers of the same habitat type were frequently correlated and the percentage habitat type, patch area, patch compactness and edge density variables for a habitat type were also frequently found to be correlated for some of the habitat types. The DTM (elevation) layer was also found to be correlated with the climate variables, particularly the temperature-related variables.

Of further note was the fact that the Euclidean distance layers to all the heathland types (ED.HL1, ED.HL2, ED.HL3, ED.AQ1, ED.HL1.HL3 and ED.HL2.AQ1) were always correlated, suggesting that these habitat types all tend to occur near each other. In addition, for *G. illyricus N. sylvestris* and *P. argus*, ED.HL8 (continuous bracken) was also correlated with the distance to the heathland habitat types. Euclidean distance to wood edge (ED.wood.edge) was always correlated with distance to broadleaved edge (ED.Bld.edge) and to broadleaved woodland (ED.W1.W2), suggesting that the woodland edge was frequently broadleaved woodland. Finally, unsurprisingly, Euclidean distance to improved grassland (ED.GL3) was frequently correlated with Euclidean distance to residential areas (ED.UR.res).

3.3.1.2. Results of GLMs using expert-selected variables

The outputs for the GLMs run using the expert-selected variables (after removal of variables with ‘wrong’ coefficient values are provided in Tables 22 and 23 below.

Values for the null deviance and the residual deviance give an indication of the fit of the data to the model (Fox, 2002) and the AIC values can be used to compare models (see section Appendix 13.2), where smaller values indicate a better fit of the model to the data (Fox, 2002). Table 22 below shows the results of the GLMs run with the expert selected variables and Table 23 shows the coefficient (estimate) values and their corresponding standard errors, z-values and *P*-values, from which the most significant variables in that model can be determined.

Species	**No. of variables**	**Null deviance**	**No. of degrees of freedom**	**Residual deviance**	**No. of degrees of freedom**	**AIC**
C. nobile	9	337	271	307	262	327
G. constrictum	6	236	169	163	163	177
G. illyricus	6	180	129	84.1	123	100
H. semele	4	360	259	305	255	315
N. sylvestris	7	282	202	236	195	252
P. argus	11	498	358	359	347	383
P. globulifera	7	274	197	200	190	216
P. punctata	4	191	137	106	133	116

Table 22. Results of GLMs using expert-selected variables. No. of variables is the number of selected variables. Null deviance is the deviance for the model with only an intercept and the residual deviance is the deviance for the fitted model. The numbers of degrees of freedom are provided for each. Akaike's Information Criterion (AIC) values are also shown.

Species	**Selected variables**	**Estimate**	**Standard error**	**Z value**	**Pr(>|z|)**
C. nobile (9)	(Intercept)	-0.205	0.471	-0.435	0.664
	S71107	0.00580	0.00432	1.34	0.179
	S84102	0.0155	0.00529	2.92	0.00350
	S64301	0.00688	0.00550	1.25	0.211
	S64303	0.00686	0.00473	1.45	0.147
	Slope	-0.331	0.0893	-3.71	0.000207
	HL3.cover	0.00539	0.00652	0.827	0.408
	Eden.HL1.HL3	0.114	0.0321	3.55	0.000385
	ED.HL1.HL3	0.000104	0.000279	-0.373	0.709
	ED.AQ5.AQ6	0.000193	0.000230	-0.840	0.401
G. constrictum (6)	(Intercept)	0.985	0.532	1.85	0.0642
	S84102	0.0170	0.00510	3.35	0.000817
	AQ5.cover	0.867	0.928	0.934	0.350
	Slope	-0.426	0.142	-2.99	0.00276
	HL2.AQ1.cover	0.00904	0.0143	0.632	0.527
	HL1.HL3.cover	0.00559	0.00603	0.926	0.354
	ED.HL2.AQ1	-0.00154	0.000500	-3.10	0.00193
G. illyricus (6)	(Intercept)	0.166	0.791	0.210	0.834
	S64301	0.0249	0.00928	2.69	0.00721
	Slope	0.0663	0.144	0.460	0.645
	GL8.cover	0.0362	0.0186	1.95	0.0510
	HL3.cover	0.0354	0.0120	2.95	0.00314
	ED.GL8	-0.00650	0.00215	-3.02	0.00250
	ED.wood.edge	-0.000555	0.00275	-0.202	0.840

H. semele (4)	(Intercept)	-0.151	0.201	-0.748	0.455
	S64303	0.0115	0.00334	3.44	0.000581
	ST1.cover	0.00900	0.0118	0.760	0.447
	PA.HL3	0.0333	0.0251	1.32	0.185
	ED.HL1.HL3	-0.00156	0.000517	-3.02	0.00251
N. sylvestris (7)	(Intercept)	0.316	0.489	0.646	0.518
	S64303	0.00681	0.00478	1.43	0.154
	Eden.W1.W2	-0.00204	0.0351	-0.0580	0.953
	Eden.W5.W6	0.0203	0.0414	0.492	0.623
	Eden.W7.W8	0.0134	0.0461	0.293	0.770
	PA.W1.W2	0.000961	0.000296	3.25	0.00115
	PA.W7.W8	0.00389	0.0224	0.174	0.862
	ED.wood.edge	-0.0153	0.00563	-2.71	0.00671
P. argus (11)	(Intercept)	-1.18	0.524	-2.25	0.027
	S64301	0.0113	0.00446	2.54	0.0112
	S64303	0.00588	0.00408	1.44	0.150
	S71107	0.00151	0.00344	0.440	0.660
	Aspect.flat	0.0109	0.00632	1.73	0.0840
	Aspect.south	0.00657	0.00406	1.61	0.106
	Slope	0.0368	0.0692	0.531	0.595
	Eden.HL1.HL3	0.0838	0.0262	3.20	0.00136
	Eden.HL2	0.0491	0.0347	1.41	0.157
	PA.HL1.HL3	0.00251	0.000738	3.40	0.00136
	PA.HL2	0.00323	0.0192	0.168	0.867
	ED.HL2	-0.000803	0.000255	-3.15	0.00163
P. globulifera (7)	(Intercept)	1.10	0.566	1.96	0.0500
	S84102	0.0142	0.00553	2.56	0.0105
	Slope	-0.373	0.105	-3.54	0.000398
	HL1.HL3.cover	0.00794	0.00540	1.47	0.141
	HL2.AQ1.cover	0.00754	0.00677	1.11	0.265
	AQ5.AQ6.cover	0.156	0.0906	1.72	0.0847
	ED.AQ5.AQ6	0.000241	0.000328	-0.734	0.463
	ED.HL2.AQ1	-0.00185	0.000497	-3.72	0.000203
P. punctata (4)	(Intercept)	-0.142	0.657	-0.217	0.829
	S64301	0.00588	0.00801	0.734	0.463
	S64303	0.0193	0.00618	3.12	0.00181
	HL1.HL3.cover	0.00277	0.00865	0.320	0.749
	ED.HL1.HL3	-0.00696	0.00298	-2.34	0.0192

Table 23. Expert-selected GLM variables results. The estimate column shows the regression coefficients for the intercept and slope (for the selected variables) and their standard errors, which are both in logits. The Z-value (Wald statistic) is the estimate value divided by the standard error (the ratio of the coefficient to its standard error) and tests the hypothesis that the regression coefficient is zero (Fox, 2002), and the *P*-value indicates whether the Z-value is significantly different from zero. The numbers in brackets next to the species name indicate the number of selected variables.

For all of the species there were several expert-selected variables that were not significant (even at *P*= 0.10). Although some of these variables may be important, they may not be the dominant factors influencing the species' distributions or the variable may not be adequately represented by the available data, or have a linear relationship with species occurrence (see section 3.4). The selected variables are discussed in more detail in Appendix 16 (and section 3.3.1.4), where they are compared to the stepwise-selected variables.

For the *N. sylvestris* model, the values for Eden.W1.W2 are shown (and the variable kept in the model), even though it had a negative coefficient value, as this species is associated with woodland edge. This may mean that a negative coefficient value for this variable would be more appropriate. However, it was not statistically significant.

3.3.1.3. Results of stepwise (and full) GLMs

Table 24 below shows the deviance and AIC values for the full and backwards stepwise GLMs for each species. (The final variables entered into the stepwise GLMs are listed in Appendix 15).

Model/Species	**No. of variables**	**Null deviance**	**No. of degrees of freedom**	**Residual deviance**	**No. of degrees of freedom**	**AIC**
C. nobile	9	377	271	284	262	304
G. constrictum	8	236	169	124	161	142
G. illyricus	8	180	129	70.38	121	88.4
H. semele	10	360	259	256	249	278
N. sylvestris	11	281	202	171	191	195
P. argus	18	498	358	323	340	361
P. globulifera	14	274	197	153	183	183
P. punctata	10	191	137	76.3	127	98.3

Table 24. Results of the stepwise (backwards) GLMs. No. of variables is the number of variables entered into the stepwise selection (full) or the number of variables selected by stepwise selection (step). Null deviance is the deviance for a model with only an intercept and the residual deviance is the deviance for the fitted model. The numbers of degrees of freedom are provided for each. Akaike's Information Criterion (AIC) values are also shown.

Table 25 below shows the variables selected by backwards stepwise selection for the GLMs for all eight species and the associated results.

Species	**Selected variables**	**Estimate**	**Standard error**	**Z value**	**Pr(>|z|)**
C. nobile (9)	(Intercept)	1.25	0.500	2.51	0.0123
	S84102	0.0112	0.00464	2.42	0.0155
	S64303	0.00735	0.00420	1.75	0.0802
	Slope	-0.472	0.104	-4.54	0.00000562
	Eden.HL1.HL3	0.128	0.0306	4.19	0.0000283
	Eden.W1.W2	0.0760	0.0290	2.62	0.00884
	ED.Bld.edge	-0.00147	0.000983	-1.49	0.136

	ED.UR.nonres	-0.000485	0.000194	-2.51	0.0123
	ED.UR.res	-0.000643	0.000389	-1.65	0.0982
	ED.HL2.AQ1	-0.000492	0.000249	-1.98	0.0478
G. constrictum (8)	(Intercept)	-3.57	1.60	-2.23	0.0259
	S84102	0.0166	0.00630	2.64	0.00841
	Aspect.south	0.0156	0.00662	2.36	0.0186
	Slope	-0.473	0.187	-2.53	0.0115
	Tmin2	3.55	1.03	3.46	0.000549
	Eden.HL2.AQ1	0.0891	0.0567	1.57	0.116
	PC.W1.W2	0.0640	0.0398	1.61	0.108
	ED.UR.nonres	-0.000776	0.000359	-2.16	0.0307
	ED.HL1.HL3	-0.00461	0.00106	-4.34	0.0000146
G. illyricus (8)	(Intercept)	-1.88	1.24	-1.52	0.128
	S71107	0.0335	0.0127	2.63	0.00858
	S64301	0.0537	0.0179	3.00	0.00270
	S64303	0.0229	0.0122	1.88	0.0598
	Eden.HL1.HL3	0.122	0.0611	2.00	0.0456
	PC.HL1.HL3	0.138	0.0489	2.83	0.00462
	PA.W1.W2	0.00239	0.000995	2.40	0.0164
	PA.W7.W8	0.212	0.0906	2.33	0.0196
	ED.GL8	-0.0148	0.00354	-4.18	0.0000287
H. semele (10)	(Intercept)	21.0	12.0	1.75	0.0794
	S71107	0.00963	0.00442	2.18	0.0293
	S64303	0.0139	0.00469	2.97	0.00301
	DTM	-0.0155	0.00687	-2.25	0.0243
	Slope	0.265	0.0840	3.15	0.00162
	HL1.HL3.cover	0.0121	0.00510	2.36	0.0182
	Bio12	-0.0272	0.0152	-1.78	0.0747
	Eden.HL2	0.0610	0.0381	1.60	0.109
	Eden.ST1	0.120	0.0464	2.58	0.00981
	ED.AQ5	-0.000538	0.000275	-1.96	0.0504
	ED.HL1	-0.00142	0.000466	-3.05	0.00231
N. sylvestris (11)	(Intercept)	3.01	0.661	4.56	0.00000512
	Eden.HL3	0.0600	0.0394	1.53	0.127
	PC.W1.W2	0.0909	0.0252	3.61	0.000302
	PA.GL8	0.863	0.524	1.65	0.100
	PA.HL2.AQ1	0.0832	0.0275	3.02	0.00252
	PA.W1.W2	0.00133	0.000347	3.84	0.000124
	PA.W5.W6	0.0222	0.00780	2.85	0.00436
	ED.AQ6	-0.000618	0.000177	-3.49	0.000479
	ED.HL3	-0.00176	0.000542	-3.24	0.00121
	ED.mix.edge	-0.00342	0.000766	-4.46	0.00000817
	ED.UR.res	-0.00135	0.000534	-2.53	0.0115
	ED.wood.edge	-0.00857	0.00521	-1.65	0.100
P. argus (18)	(Intercept)	-1.80	0.738	-2.44	0.0146
	S71107	0.00815	0.00495	1.65	0.100
	S84102	0.0104	0.00612	1.70	0.0996
	S64301	0.0171	0.00584	2.93	0.00338
	S64303	0.00950	0.00564	1.69	0.0920
	Aspect.flat	0.0157	0.00670	2.34	0.0192
	Aspect.south	0.00645	0.00442	1.46	0.144
	Slope	0.124	0.0782	1.59	0.112
	HL1.HL3.cover	0.0112	0.00560	2.00	0.0460

	W1.W2.cover	0.00957	0.00524	1.83	0.0681
	Eden.AQ1	0.07002	0.0421	1.67	0.0952
	Eden.GL8	0.0672	0.0338	1.99	0.0469
	Eden.HL1.HL3	0.0626	0.0308	2.03	0.0423
	Eden.ST1	0.128	0.0565	2.26	0.0241
	PA.AQ1	0.0900	0.0611	1.47	0.141
	PA.HL2	0.0338	0.0209	1.62	0.105
	PA.HL1.HL3	0.00177	0.000922	1.93	0.0541
	ED.AQ6	-0.000516	0.000135	-3.83	0.000130
	ED.HL1	-0.000824	0.000364	-2.26	0.0236
P. globulifera (14)	(Intercept)	31.8	15.3	2.08	0.0372
	S71107	0.00835	0.00569	1.47	0.142
	S84102	0.0215	0.00787	2.73	0.00629
	S64303	0.0114	0.00685	1.70	0.0950
	Aspect.south	0.0143	0.00685	2.09	0.0370
	Slope	-0.663	0.162	-4.09	0.0000432
	AQ5.AQ6.cover	0.204	0.113	1.80	0.0721
	Bio12	-0.0395	0.0194	-2.03	0.0425
	Eden.HL1.HL3	0.0682	0.0380	1.79	0.0728
	Eden.HL2.AQ1	0.188	0.0514	3.66	0.000256
	Eden.W5.W6	0.185	0.0876	2.11	0.0350
	PC.HL3	0.0684	0.0408	1.68	0.0939
	ED.con.edge	-0.00172	0.000993	-1.73	0.0832
	ED.HL3	-0.00310	0.00110	-2.83	0.00470
	ED.AQ5.AQ6	-0.000783	0.000440	-1.78	0.0748
P. punctata (10)	(Intercept)	-55.5	24.0	-2.31	0.0209
	S64303	0.0128	0.00839	1.53	0.127
	Aspect.north	0.0324	0.0117	2.77	0.00556
	Slope	-0.246	0.140	-1.75	0.0797
	Bio12	0.0721	0.0307	2.35	0.0188
	Eden.HL1	0.115	0.0619	1.85	0.0639
	PA.HL3	0.0718	0.0880	0.816	0.415
	PA.HL2.AQ1	0.0370	0.0250	1.48	0.138
	ED.AQ6	-0.000391	0.000235	-1.66	0.0968
	ED.UR.nonres	-0.00124	0.000429	-2.89	0.00388
	ED.HL1.HL3	-0.00662	0.00242	-2.74	0.00620

Table 25. Stepwise-selected GLM variables results. The estimate column shows the regression coefficients for the intercept and slope (for the selected variables) and their standard errors (Std. Error column), which are both in logits. The Z-value (Wald statistic) is the estimate value divided by the standard error (the ratio of the coefficient to its standard error) and tests the hypothesis that the regression coefficient is zero (Fox, 2002), and the *P*-value indicates whether the Z-value is significantly different from zero. The numbers in brackets next to the species name indicate the number of selected variables.

One or more soil variables were selected for all the species, except for *N. sylvestris*. Although soil type may not be directly important for the occurrence of the butterfly species, or *P. punctata*, it is the vegetation type (including food plants) associated with those soil types that influence the species' occurrence. This provides an example of how variables, such as soil type, can act as surrogate variables for other factors. Slope was selected for all the species except for *G. illyricus* and *N. sylvestris* and all of the sets of variables contained at least one Euclidean distance variable. The selected variables are discussed in more detail in Appendix 16 (and section 3.3.1.4).

A large number of variables were selected for *P. argus* and *P. globulifera* and although this was within the permitted maximum number of variables for *P. argus*, there were four too many variables for *P. globulifera*. Only the models for *C. nobile*, *G. constrictum*, *H. semele* and *P. argus* were within the permitted maximum number of variables for the sample size (see Appendix 15). The number of variables over the maximum ranged from 1 (for *G. illyricus* and *N. sylvestris*) to 3 for *P. punctata* and 4 for *P. globulifera*. However, several of the variables selected by stepwise selection did not have significant *P*-values, even at *P*=0.1. This is because stepwise selection of terms on the basis of AIC can be somewhat permissive in its choice of terms (Venables and Ripley, 2002), tending to err on the side of generosity, which is preferable for an automated procedure (Crawley, 2007). Therefore following stepwise selection, variable selection can proceed by manually dropping the less significant terms (Venables and Ripley, 2002; Crawley, 2007) and reducing the number of variables. This was carried out when selecting the final 'best' models (section 3.3.1.4).

3.3.1.4. Results of final GLMs (combined 'best' models of stepwise and expert-selected variables)

The selection of variables for the final 'best' GLMs are discussed in Appendix 16. Although the models are referred to as the 'best' models they may not actually be the 'best' model, as there are often many subsets of predictors of a given size that are nearly equally good (Fox, 2002). In addition, variables may become more or less significant depending on which other variables are present (as demonstrated in the discussions below). The final set of selected variables for each species (which ended up being based on the stepwise-selected set of variables, but with some of the less significant variables removed) are shown in Table 27 below (and results also in Table 26 below).

Species	**No. of variables**	**Null deviance**	**No. of degrees of freedom**	**Residual deviance**	**No. of degrees of freedom**	**AIC**
C. nobile	7	377	271	288	264	304
G. constrictum	6	235	169	125	162	141
G. illyricus	7	180	129	74.3	122	90.3
H. semele	9	360	259	259	250	279
N. sylvestris	9	281	202	177	193	197
N. sylvestris	7	281	202	198	195	214
P. argus	7	498	358	344	351	360
P. globulifera	7	274	197	169	190	185
P. globulifera	6	274	197	175	191	189
P. punctata	7	191	137	83.7	130	99.7

Table 26. Results of final GLMs. Where two different versions of a model were run (i.e. with different numbers of variables; see Appendix 16 for details) these are shown. No. of variables is the number of variables included in the final selection. Null deviance is the deviance for a model with only an intercept and the residual deviance is the deviance for the fitted model. The numbers of degrees of freedom are provided for each. Akaike's Information Criterion (AIC) values are also shown.

Species	**Selected variables**	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**
C. nobile (7)	(Intercept)	1.25	0.469	2.66	0.00792
	S84102	0.00948	0.00454	2.09	0.0367
	Slope	-0.445	0.0991	-4.49	0.00000713
	Eden.HL1.HL3	0.125	0.0292	4.28	0.0000191
	Eden.W1.W2	0.0851	0.0268	3.18	0.00147
	ED.UR.nonres	0.000484	0.000191	-2.53	0.0113
	ED.UR.res	-0.000669	0.000368	-1.82	0.0694
	ED.HL2.AQ1	-0.000608	0.000244	-2.49	0.0128
G. constrictum (6)	(Intercept)	-3.12	1.58	-1.98	0.0480
	S84102	0.0138	0.00608	2.27	0.0235
	Aspect.south	0.0158	0.00642	2.46	0.0140
	Slope	-0.454	0.173	-2.62	0.00877
	Tmin2	3.60	1.01	3.51	0.000453
	ED.UR.nonres	-0.000739	0.000352	-2.10	0.0357
	ED.HL1.HL3	-0.00493	0.00105	-4.69	0.00000271
G. illyricus (7)	(Intercept)	0.0751	0.589	0.128	0.899
	S71107	0.0147	0.00676	2.17	0.0301
	S64301	0.0347	0.0133	2.61	0.00902
	Eden.HL1.HL3	0.117	0.0591	1.98	0.0473
	PC.HL1.HL3	0.126	0.0449	2.81	0.00497
	PA.W1.W2	0.00177	0.000843	2.10	0.0357
	PA.W7.W8	0.160	0.0785	2.03	0.0421
	ED.GL8	-0.0142	0.00332	-4.28	0.0000188
H. semele (9)	(Intercept)	21.9	11.9	1.84	0.0661
	S71107	0.0105	0.00439	2.40	0.0165
	S64303	0.0146	0.00462	3.17	0.00151
	DTM	-0.0168	0.00677	-2.48	0.0131
	Slope	0.271	0.0830	3.26	0.00112
	HL1.HL3.cover	0.0109	0.00505	2.17	0.0304
	Bio12	-0.0281	0.0152	-1.85	0.0646
	Eden.ST1	0.116	0.0462	2.51	0.0121
	ED.AQ5	-0.000516	0.000275	-1.88	0.0608

	ED.HL1	-0.00164	0.000465	-3.53	0.000421
N. sylvestris (9)	(Intercept)	3.12	0.635	4.92	0.000000863
	PC.W1.W2	0.0863	0.0247	3.50	0.000472
	PA.HL2.AQ1	0.0736	0.0267	2.75	0.00591
	PA.W1.W2	0.00121	0.000338	3.57	0.000359
	PA.W5.W6	0.0193	0.00746	2.59	0.00974
	ED.AQ6	-0.000587	0.000170	-3.45	0.000553
	ED.HL3	-0.00201	0.00201	-3.85	0.000117
	ED.mix.edge	-0.00299	0.00299	-4.35	0.0000135
	ED.UR.res	-0.00113	0.00113	-2.21	0.0273
	ED.wood.edge	-0.00917	0.00917	-1.82	0.0692
N. sylvestris (7)	(Intercept)	1.58	0.429	3.67	0.000242
	PC.W1.W2	0.0617	0.0224	2.76	0.00580
	PA.W1.W2	0.00126	0.000330	3.81	0.000138
	PA.W5.W6	0.0155	0.00655	2.37	0.0177
	ED.HL3	-0.00165	0.000474	-3.48	0.000495
	ED.mix.edge	-0.00192	0.000562	-3.43	0.000612
	ED.UR.res	-0.000900	0.000455	-1.98	0.0480
	ED.wood.edge	-0.0103	0.00461	-2.24	0.0249
P. argus (7)	(Intercept)	0.614	0.352	1.75	0.0811
	S64301	0.00745	0.00388	1.92	0.0550
	Eden.AQ1	0.0699	0.0381	1.83	0.0669
	Eden.HL1.HL3	0.0816	0.0285	2.86	0.00425
	Eden.ST1	0.0913	0.0525	1.74	0.0820
	PA.HL1.HL3	0.00224	0.000739	3.03	0.00245
	ED.AQ6	-0.000436	0.000120	-3.64	0.000273
	ED.HL1	-0.00139	0.000342	-4.07	0.0000476
P. globulifera (7)	(Intercept)	0.979	0.419	2.34	0.0194
	S84102	0.0168	0.00654	2.56	0.0103
	Aspect.south	0.0117	0.00607	1.92	0.0545
	Slope	-0.584	0.135	-4.32	0.0000154
	AQ5.AQ6.cover	0.178	0.0993	1.79	0.0735
	Eden.HL2.AQ1	0.143	0.0436	3.27	0.00108
	Eden.W5.W6	0.183	0.0741	2.47	0.0135
	ED.HL3	-0.00470	0.00102	-4.60	0.00000432
P. globulifera (6)	(Intercept)	1.05	0.412	2.54	0.0111
	S84102	0.0151	0.00629	2.40	0.0165
	Aspect.south	0.0114	0.00593	1.93	0.0537
	Slope	-0.499	0.127	-3.94	0.0000815
	AQ5.AQ6.cover	0.178	0.104	1.72	0.0861
	Eden.HL2.AQ1	0.124	0.0414	3.01	0.00263
	ED.HL3	-0.00463	0.000984	-4.70	0.00000258
P. punctata (7)	(Intercept)	-57.3	21.4	-2.68	0.00735
	Aspect.north	0.0228	0.00999	2.28	0.0226
	Slope	-0.281	0.137	-2.05	0.0406
	Bio12	0.0763	0.0272	2.81	0.00500
	Eden.HL1	0.100	0.0560	1.79	0.0736
	ED.AQ6	-0.000384	0.000218	-1.76	0.0781
	ED.UR.nonres	-0.00121	0.000397	-3.05	0.00230
	ED.HL1.HL3	-0.00836	0.00232	-3.60	0.00032

Table 27. Final model selected GLM variables. Where two different versions of a model were run (i.e. with different numbers of variables; see Appendix 16 for details) these are shown. The numbers in brackets next to the species name indicate the number of selected variables. The estimate column shows the regression coefficients for the intercept and slope (for the selected variables) and their standard errors, which are both in logits. The Z-value (Wald statistic) is the estimate value divided by the standard error (the ratio of the coefficient to its standard error) and tests the hypothesis that the regression coefficient is zero (Fox, 2002), and the *P*-value indicates whether the Z-value is significantly different from zero.

3.3.1.5. Moran's *I* test for spatial autocorrelation in residuals of final GLMs

The results of the Moran's *I* tests for spatial autocorrelation in the residuals of the final GLMs (from section 3.3.1.4) are shown in Table 28 below.

Species	**Moran's *I***	**Standard deviate**	***P*-value**
C. nobile	0.122	3.89	0.000
G. constrictum	0.178	2.54	0.006
G. illyricus	0.310	0.914	0.180
H. semele	0.145	3.58	0.000
N. sylvestris	9= 0.047, 7= 0.113	2.10, 2.69	0.018, 0.004
P. argus	0.201	4.54	0.000
P. globulifera	7= 0.080, 6= 0.093	1.80, 1.80	0.036, 0.036
P. punctata	0.090	0.974	0.165

Table 28. Moran's *I* results for final GLMs. Standard deviate is the standard deviation of Moran's *I*. *P*-value is the *P*-value of the test. Where two different models were run for a species (see section 3.3.1.4. and Appendix 16) values are shown for both models.

A non-significant *P*-value means that the null hypothesis (of a Moran's *I* value of 0, i.e. no spatial autocorrelation) cannot be rejected. Therefore, from Table 28 it can be seen that the GLM residuals for *G. illyricus* and *P. punctata* did not exhibit spatial autocorrelation. The Moran's *I* value and standard deviation values for *P. punctata* were the lowest. The other six species' GLMs all had highly significant *P*-values, meaning that the alternative hypothesis, of the presence of spatial autocorrelation (Moran's *I* value not 0) was accepted. Although the Moran's *I* values were mostly below 0.2, the standard deviation values were quite high, which is why the null hypotheses for these tests were rejected.

3.3.1.6. Results of Generalised Linear Mixed Models (GLMMs)

The results for the GLMMs for the final set of variables are shown in Appendix 17 as an example of how similar they were to the GLM results. The GLMM results for the other sets of variables are not shown, as they were also as similar and it was decided to use the GLMs as they were so similar to the GLMMs.

The residual deviance and AIC values for the GLMMs were very slightly higher for the GLMMs (Table A-24, Appendix 17) than the GLMs (Table 26, section 3.3.1.4) for the final set of variables, but the coefficient and *P*-values for the variables were identical at three decimal places except for the intercept values for the *C. nobile*, *G. constrictum*, *P. argus* and *P. globulifera* models. This suggests that the effects of the environmental variables were not overestimated and the *P*-values were not inflated, so variables were not wrongly selected during stepwise selection (see Appendix 13.3). This indicates that using GLMMs would select the same variables as using GLMs, so the GLMs can be used.

3.3.1.7. Predictions for testing data for final GLMs

Using the intercept and variable coefficient values from Table 27 (section 3.3.1.4), the equation below can be used to work out the predicted probability value (between 0 and 1) for a particular site:

$$Y = \frac{1}{1 + e^{-x}}$$

Where *x* is:

> Intercept + (coefficient of variable × value for that variable at a site) + (coefficient of variable × value for that variable at a site) + ...

The intercept is the value of *x* when the value of all variables is zero. Note that numbers in tables are rounded, and for the calculations the full numbers should be used. However, this is all done automatically within R using the `predict.glm` function (see section 3.2.2.7). The predicted values for the testing data set were used to evaluate the models.

3.3.2. Generalised Additive Models (GAMs) results

3.3.2.1. Results of GAMs using expert-selected variables

Tables 29 and 30 below show the results of the GAMs run with the expert-selected variables (including variables removed with ‘wrong’ relationships from the expert-selected GLMs). Instead of Z scores and P-values for the coefficients as in the GLM output, the GAM output provides Chi-square values and corresponding *P*-values for each of the terms. Chi-square values are provided for the non-parametric effects to indicate the importance of the smooth for each term in the model (Insightful Corporation, 2001). Therefore, only the values for the smoothed terms are provided.

Species	**No. of variables**	**Null deviance**	**No. of degrees of freedom**	**Residual deviance**	**No. of degrees of freedom**	**AIC**
C. nobile	9	337	271	241	243	300
P. punctata	4	191	137	87.4	117	129

Table 29. Results of GAMs using expert-selected variables. No. of variables is the number of variables in the model. Null deviance is the deviance for a model with only an intercept and the residual deviance is the deviance for the fitted model. The numbers of degrees of freedom are provided for each. Akaike's Information Criterion (AIC) values are also shown.

Species	**Selected variables**	**Df Npar**	**Chi square**	**P(Chi)**
C. nobile (11)	s(S71107)	3	2.59	0.459
	s(S84102)	3	5.72	0.126
	s(S64301)	3	2.47	0.481
	s(S64303)	3	1.54	0.674
	s(Aspect.flat)	3	1.91	0.591
	s(Slope)	3	3.77	0.288
	s(HL2.cover)	3	2.79	0.425
	s(HL3.cover)	3	10.7	0.0136
	s(Eden.HL1.HL3)	3	8.78	0.0324
	s(ED.HL1.HL3)	5.7	9.02	0.149
	s(ED.AQ5.AQ6)	3	3.08	0.380
P. punctata (5)	s(S64301)	3	0.00000266	1.00
	s(S64303)	3	0.991	0.803
	s(HL1.HL3.cover)	3	2.95	0.399
	s(HL2.cover)	3	4.86	0.182
	s(ED.HL1.HL3)	3	4.33	0.229

Table 30. Expert-selected GAM variables results. s(Variable) indicates smoothed terms. Df Npar is the non-parametric degrees of freedom for each variable in the model. Chi square is the Chi-square value for the non-parametric effects. P(Chi) is the *P*-value for the Chi-square.

Of the *C. nobile* expert-selected variables, only two showed evidence of significant non-linearity (HL3.cover and Eden.HL1.HL3). Eden.HL1.HL3 was also significant in the GLMs, suggesting that the response may be partly linear and partly non-linear, and HL3.cover was not. Two other variables also showed significant linear relationships in the GAMs. None of the *P. punctata* expert-selected variables showed evidence of significant non-linearity in the GAMs, whereas two showed significant linearity in the GLMs. As for the expert-selected GLMs, the majority of the expert-selected variables did not show significant relationships using the GAMs either.

3.3.2.2. Results of stepwise (and full) GAMs

Tables 31 and 32 below show the results for the backwards stepwise GAMs (the both directional stepwise GAMs produced the same results) and the selected variables with the values for the nonparametric effects of the models. After running the stepwise GAM for *C. nobile*, three of the variables (S84102, PA.HL1 and ED.AQ6) were not significant (at *P*=0.1) so were removed and a GAM run just with the remaining significant variables. Two more variables (Slope and ED.HL2.AQ1) were then not significant, so the procedure repeated. The remaining seven variables were all significant at *P*=0.05, but the AIC value was higher than for the model with 12 variables. All of the selected smoothed variables for the *P. punctata* model were significant at *P*=0.05 and two variables were selected in their linear form. These models were then the final GAM models as the significant expert-selected variables were entered into the stepwise selection.

	No. of variables	**Null deviance**	**No. of degrees of freedom**	**Residual deviance**	**No. of degrees of freedom**	**AIC**
C. nobile	12	377	271.00	181	221	282
C. nobile	7	377	271.00	248	243	306
P. punctata	5	191	137.00	56.5	123	86.5

Table 31. Results of stepwise (backwards) GAMs. No. of variables is the number of variables in the model. Null deviance is the deviance for a model with only an intercept and the residual deviance is the deviance for the fitted model. The numbers of degrees of freedom are provided for each. Akaike's Information Criterion (AIC) values are also shown.

Species	**Selected variables**	**Df Npar**	**Chi square**	**P(Chi)**
C. nobile (12)	s(S84102)	3	5.35	0.148
	s(Slope)	3	7.68	0.0531
	s(HL3.cover)	3	16.5	0.000900
	s(Tmin2)	3	10.9	0.0125
	s(PC.GL3)	3	7.70	0.0527
	s(PC.HL1.HL3)	3	7.64	0.0540
	s(PA.HL1)	4.2	5.61	0.251
	s(ED.AQ6)	3	5.81	0.121
	s(ED.Bld.edge)	3	20.8	0.000100
	s(ED.UR.nonres)	3	9.07	0.0284
	s(ED.UR.res)	3	7.60	0.0551
	s(ED.HL2.AQ1)	3.3	8.27	0.0528
C. nobile (7)	s(HL3.cover)	3	11.8	0.00820
	s(Tmin2)	3	9.27	0.0260
	s(PC.GL3)	3	8.03	0.0453
	s(PC.HL1.HL3)	3	14.8	0.00200
	s(ED.Bld.edge)	3	15.4	0.00150
	s(ED.UR.nonres)	3	13.1	0.00450
	s(ED.UR.res)	3	11.5	0.00940

P. punctata (5)	Bio12			
	s(Eden.HL3)	3	8.97	0.0297
	s(ED.AQ6)	3	12.5	0.00590
	s(ED.UR.nonres)	3	10.0	0.0185
	ED.HL1.HL3			

Table 32. Stepwise-selected GAM variables results. The numbers in brackets next to the species name indicate the number of selected variables. s(Variable) indicates smoothed terms. Df Npar is the non-parametric degrees of freedom for each variable in the model. Chi square is the Chi-square value for the non-parametric effects. P(Chi) is the *P*-value for the Chi-square.

Of the *C. nobile* GAM model containing seven variables (the same number as for the final GLM), two of these were also selected for the final GLM (ED.UR.nonres and ED.UR.res) and it can be seen from Figures A-44 and A-45 in Appendix 18 that these responses contain linear and non-linear elements. Two of the other GAM variables (PC.HL1.HL3 and ED.Bld.edge) were partly similar to two of the final GLM variables (Eden.HL1.HL3 and Eden.W1.W2). An additional three smoothed GAM-selected variables in the *C. nobile* GAM models containing twelve variables also occurred in the final GLM.

Four of the five stepwise *P. punctata* GAM variables were also part of the final GLM model. Two of these four were selected by the GAM in their linear form (Bio12 and ED.HL1.HL3) and two were selected in their smoothed form (ED.AQ6 and ED.UR.nonres), suggesting that the responses of these two variables had linear and non-linear components. This can be seen from Figures A-48 and A-49 in Appendix 18, where the overall relationship is of a linear trend. The other GAM-selected smoothed variable (Eden.HL3) was not selected by the GLM (although Eden.HL1, which is similar, was) and shows a non-linear response (see Figure A-47 in Appendix 18). A greater number of variables were included in the final GLM (7) than the final GAM (5) for *P. punctata*.

3.3.2.3. GAM plots of variables selected for final GAMs

GAM plots and discussion of the variables selected by the stepwise GAMs (section 3.3.2.2) can be found in Appendix 18. The GAM plots additionally allow checking of observation density along each environmental predictor (i.e. the rugs; see Appendix 18), which can also be used to improve the sampling of unevenly sampled predictors during further field campaigns, if required (Guisan *et al.*, 2006a).

3.3.2.4. Predictions for testing data for GAMs

Predictions for the testing data for each of the species were generated automatically by `predict.gam`, which gets `predict.glm` to produce the parametric part of the predictions and for each nonparametric term, `predict.gam` reconstructs the partial residuals and weights from the final iteration of the local scoring algorithm, with the appropriate smoother called for each term, and the prediction for that term produced (Hastie, 2009b).

3.3.3. Results of evaluation of GLMs and GAMs

The results of the threshold-dependent and threshold-independent (AUC) evaluation measures for the predictions using the final GLMs and GAMs for each species for their testing data are presented in Tables 33, 34 and 35 below. Details on the different cut-off values are provided in Table 21 (section 3.2.4) and information about the different measures is provided in Appendix 8.2.

Species	***C. nobile***		***G constrictum***		***G. illyricus***		***H. semele***		***N. sylvestris -7***		***N. sylvestris -9***		***P. argus***		***P. globulifera-6***		***P. globulifera-7***		***P. punctata***	
Weighting	1:1	W	1:1	W	1:1	W	1:1	W	1:1	W	1:1	W	1:1	W	1:1	W	1:1	W	1:1	W
P_opt																				
P_crit	0.638	0.195	0.365	0.305	0.370	0.190	0.458	0.373	0.535	0.260	0.200	0.200	0.483	0.408	0.495	0.495	0.438	0.438	0.230	0.073
Sensitivity	0.569	0.966	0.838	0.892	0.889	0.926	0.786	0.875	0.705	0.909	0.932	0.932	0.844	0.922	0.857	0.857	0.857	0.857	0.867	0.967
Specificity	0.845	0.293	0.703	0.676	0.714	0.679	0.714	0.607	0.773	0.523	0.614	0.614	0.714	0.649	0.714	0.714	0.714	0.714	0.533	0.433
False +ve	0.155	0.707	0.297	0.324	0.286	0.321	0.286	0.393	0.227	0.477	0.386	0.386	0.286	0.351	0.286	0.286	0.286	0.286	0.467	0.567
False -ve	0.431	0.035	0.162	0.108	0.111	0.074	0.214	0.125	0.296	0.091	0.068	0.068	0.156	0.078	0.143	0.143	0.143	0.143	0.133	0.033
CCR	0.707	0.629	0.770	0.784	0.800	0.800	0.750	0.741	0.739	0.716	0.773	0.773	0.779	0.786	0.786	0.786	0.786	0.786	0.700	0.700
Kappa	0.414	0.259	0.541	0.568	0.601	0.602	0.500	0.482	0.477	0.432	0.546	0.546	0.558	0.571	0.571	0.571	0.571	0.571	0.400	0.400
P_fair																				
P_crit	0.535	0.535	0.420	0.420	0.800	0.800	0.470	0.470	0.495	0.495	0.488	0.488	0.558	0.558	0.568	0.568	0.538	0.538	0.655	0.655
Sensitivity	0.638	0.638	0.784	0.784	0.704	0.704	0.750	0.750	0.750	0.750	0.727	0.727	0.753	0.753	0.738	0.738	0.738	0.738	0.633	0.633
Specificity	0.638	0.638	0.784	0.784	0.714	0.714	0.750	0.750	0.750	0.750	0.727	0.727	0.753	0.753	0.738	0.738	0.738	0.738	0.600	0.600
False +ve	0.362	0.362	0.216	0.216	0.286	0.286	0.250	0.250	0.250	0.250	0.273	0.273	0.247	0.247	0.262	0.262	0.262	0.262	0.400	0.400
False -ve	0.362	0.362	0.216	0.216	0.296	0.296	0.250	0.250	0.250	0.250	0.273	0.273	0.247	0.247	0.262	0.262	0.262	0.262	0.367	0.367
CCR	0.638	0.638	0.784	0.784	0.709	0.709	0.750	0.750	0.750	0.750	0.727	0.727	0.753	0.753	0.738	0.738	0.738	0.738	0.617	0.617
Kappa	0.276	0.276	0.568	0.568	0.418	0.418	0.500	0.500	0.500	0.500	0.455	0.455	0.507	0.507	0.476	0.476	0.476	0.476	0.233	0.233
P_Kappa																				
P_crit	0.638	0.638	0.365	0.365	0.190	0.190	0.458	0.458	0.535	0.535	0.200	0.200	0.483	0.483	0.495	0.495	0.438	0.438	0.230	0.230
Sensitivity	0.569	0.569	0.838	0.838	0.926	0.926	0.786	0.786	0.705	0.705	0.932	0.932	0.844	0.844	0.857	0.857	0.857	0.857	0.867	0.867
Specificity	0.845	0.845	0.703	0.703	0.679	0.679	0.714	0.714	0.773	0.773	0.614	0.614	0.714	0.714	0.714	0.714	0.714	0.714	0.533	0.533
False +ve	0.155	0.155	0.297	0.297	0.321	0.321	0.286	0.286	0.227	0.227	0.386	0.386	0.286	0.286	0.286	0.286	0.286	0.286	0.467	0.467
False -ve	0.431	0.431	0.162	0.162	0.074	0.074	0.214	0.214	0.296	0.296	0.068	0.068	0.156	0.156	0.143	0.143	0.143	0.143	0.133	0.133
CCR	0.707	0.707	0.770	0.770	0.800	0.800	0.750	0.750	0.739	0.739	0.773	0.773	0.779	0.779	0.786	0.786	0.786	0.786	0.700	0.700
Kappa	0.414	0.414	0.541	0.541	0.602	0.602	0.500	0.500	0.477	0.477	0.546	0.546	0.558	0.558	0.571	0.571	0.571	0.571	0.400	0.400
P=0.5																				
P_crit	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500
Sensitivity	0.638	0.638	0.703	0.703	0.889	0.889	0.661	0.661	0.727	0.727	0.727	0.727	0.844	0.844	0.833	0.833	0.786	0.786	0.733	0.733
Specificity	0.569	0.569	0.784	0.784	0.714	0.714	0.770	0.770	0.750	0.750	0.727	0.727	0.727	0.727	0.714	0.714	0.738	0.738	0.600	0.600
False +ve	0.431	0.431	0.216	0.216	0.286	0.286	0.232	0.232	0.250	0.250	0.273	0.273	0.273	0.273	0.286	0.286	0.262	0.262	0.400	0.400
False -ve	0.362	0.362	0.297	0.297	0.111	0.111	0.339	0.339	0.273	0.273	0.273	0.273	0.156	0.156	0.167	0.167	0.214	0.214	0.267	0.267
CCR	0.603	0.603	0.743	0.743	0.800	0.800	0.714	0.714	0.739	0.739	0.727	0.727	0.786	0.786	0.774	0.774	0.762	0.762	0.667	0.667
Kappa	0.207	0.207	0.487	0.487	0.601	0.601	0.429	0.429	0.477	0.477	0.455	0.455	0.571	0.571	0.548	0.548	0.524	0.524	0.333	0.333

Table 33. Threshold-dependent evaluation criteria with different cut-off values calculated for testing dataset in ROC_AUC software for final GLMs. Values were calculated under equal weighting of false positive costs and false negative costs (1:1) and unequal weighting of 0.4:0.6 (W) for the weight of false positive costs: weight of false negative costs (where false negatives are more costly than false positives). P_crit is the applied cut-off value and CCR is the correct classification rate.

Species	***C. nobile - 7***		***C. nobile - 12***		***P. punctata***	
Weighting	1:1	W	1:1	W	1:1	W
P_opt						
P_crit	0.460	0.335	0.555	0.555	0.205	0.205
Sensitivity	0.759	0.897	0.776	0.776	0.867	0.867
Specificity	0.638	0.586	0.741	0.741	0.600	0.600
False +ve	0.362	0.414	0.259	0.259	0.400	0.400
False -ve	0.241	0.103	0.224	0.224	0.133	0.133
CCR	0.698	0.741	0.759	0.759	0.733	0.733
Kappa	0.397	0.483	0.517	0.517	0.467	0.467
P_fair						
P_crit	0.565	0.565	0.588	0.588	0.743	0.743
Sensitivity	0.724	0.724	0.741	0.741	0.667	0.667
Specificity	0.741	0.741	0.741	0.741	0.667	0.667
False +ve	0.259	0.259	0.259	0.259	0.333	0.333
False -ve	0.276	0.276	0.259	0.259	0.333	0.333
CCR	0.733	0.733	0.714	0.714	0.667	0.667
Kappa	0.466	0.466	0.483	0.483	0.333	0.333
P_Kappa						
P_crit	0.460	0.460	0.555	0.555	0.205	0.205
Sensitivity	0.759	0.759	0.776	0.776	0.867	0.867
Specificity	0.638	0.638	0.741	0.741	0.600	0.600
False +ve	0.362	0.362	0.259	0.259	0.400	0.400
False -ve	0.241	0.241	0.224	0.224	0.133	0.133
CCR	0.698	0.698	0.759	0.759	0.733	0.733
Kappa	0.397	0.397	0.517	0.517	0.467	0.467
P=0.5						
P_crit	0.500	0.500	0.500	0.500	0.500	0.500
Sensitivity	0.759	0.759	0.776	0.776	0.733	0.733
Specificity	0.655	0.655	0.724	0.724	0.600	0.600
False +ve	0.345	0.345	0.276	0.276	0.400	0.400
False -ve	0.241	0.241	0.224	0.224	0.267	0.267
CCR	0.707	0.707	0.750	0.750	0.667	0.667
Kappa	0.414	0.414	0.500	0.500	0.333	0.333

Table 34. Threshold-dependent evaluation criteria with different cut-off values calculated for testing dataset in ROC_AUC software for final GAMs. Values were calculated under equal weighting of false positive costs and false negative costs (1:1) and unequal weighting of 0.4:0.6 (W) for the weight of false positive costs: weight of false negative costs (where false negatives are more costly than false positives). P_crit is the applied cut-off value and CCR is the correct classification rate.

Species	**AUC**
C. nobile	0.700
C. nobile - GAM - 7	0.792
C. nobile - GAM - 12	0.799
G. constrictum	0.836
G. illyricus	0.828
H. semele	0.812
N. sylvestris - 7	0.811
N. sylvestris - 9	0.798
P. argus	0.824
P. globulifera - 6	0.815
P. globulifera - 7	0.809
P. punctata	0.721
P. punctata - GAM	0.723

Table 35. AUC (Area Under the ROC Curve) results calculated on testing datasets in ROC_AUC software for final GLMs and GAMs. (Models are GLMs unless stated otherwise). The ROC plots are shown in Appendix 19.

The area under the ROC curve plots the probability of detecting the true signal (sensitivity; plotted on the *y*-axis against) and false signal (1 – specificity; plotted on the *x*-axis) for an entire range of possible cutpoints, where the values can range from 0 to 1 (Hosmer and Lemeshow, 2000). A good model will achieve a high true positive rate while the false positive rate is still relatively small; thus the ROC plot will rise steeply at the origin, then level off at a value near the maximum of 1 (Moisen *et al.*, 2006). Hosmer and Lemeshow (2000) suggest as a general rule for the area under the ROC curve, if:

AUC = 0.5 – this suggests no discrimination
0.7 ≤ AUC < 0.8 – this is considered acceptable discrimination
0.8 ≤ AUC < 0.9 – this is considered excellent discrimination
AUC ≥ 0.9 – this is considered outstanding discrimination

Based on this guide all of the GLM models achieved at least acceptable discrimination (*C. nobile* and *P. punctata*). Apart from those two models, the rest of the GLM models all achieved excellent discrimination, although at the lower end of this category. The higher AUC values were generally obtained from ROC plots where the sensitivity (true positive rate) increases quite steeply while the 1-specificity (false positive rate) is still fairly low (see Appendix 19). The correct classification rates were all reasonably high at above 0.7, apart from for the *C. nobile* model and the *P. punctata* model (for P=0.5 only). The Kappa results were all fairly low, with the highest Kappa value of 0.602 for the *G. illyricus* model with the P_opt or P_Kappa cut-off. Most of the Kappa values fell within the 0.4 – 0.6 category of moderate performance (Landis and Koch, 1977; Fielding and Bell, 1997; see Appendix 8.2), except for the P_fair and P=0.5 cut-offs for the *C. nobile* and *P. punctata* models, which fell within the 'fair' category.

Applying the unequal cost weighting only affected the values using the P_opt cut-off and had the effect of decreasing the P_crit threshold values for all of the species except for *N. sylvestris* (9) and both of the *P. globulifera* models, as false negatives were assigned as more costly, so decreasing the cut-off meant that more values were likely to be assigned as positives (presences). This also had the effect of increasing the sensitivity (true positive) rate and decreasing the specificity (true negative) rate, apart from for the models where P_crit was unchanged. The correct classification rate either decreased (*C. nobile*, *H. semele* and *N. sylvestris* (7)), increased (*G. constrictum*, *P. argus*) or stayed the same (*G. illyricus*, *N. sylvestris* (9), both *P. globulifera* models and *P. punctata*) with the unequal weighting. When equal weights were applied all of the values for each species for P_opt and P_Kappa were the same except for the *G. illyricus* model.

The highest sensitivity rate over all of the thresholds was achieved for the *P. punctata* model with unequal weightings (at 0.967, which was very high) and for *N. sylvestris* (9) for the equal and unequal weighting (at 0.932) using the P_opt cut-off. The false positive rate was not too high for the *N. sylvestris* model, considering how high the sensitivity rate was, but the false positive rate for the *P. punctata* was consequently one of the highest. All of the sensitivity rates were high for the P_opt cut-off, except for the equal weighting for the *C. nobile* model (0.569), which appeared to predict more lower values/absences. High sensitivity is important for the application of finding new sites of potential occurrence for species as a high true positive rate means that the model is predicting presence locations well and not missing too many (i.e. not many false negatives). The true negative rate is less important in that respect but is important for checking that the model is actually discriminating between presences and absences and not just predicting presences everywhere (which would achieve a very high sensitivity rate but a low specificity rate). The specificity rates also remained fairly high, apart from for the unequal weightings for the *C. nobile* model and both weightings for the *P. punctata* model, suggesting that the models are discriminatory. However, when the unequal weighting was applied, although the *C. nobile* model achieves a high sensitivity value, the false positive rate was also very high.

However, the P_crit values were generally very low for the P_opt cut-off (particularly with unequal weights for the *P. punctata* model), suggesting that the arbitrary cut-off of 0.5 would not be as good. This was further reinforced by the generally lower sensitivity values obtained with the P=0.5 threshold, although the specificity values were generally higher with the P=0.5 threshold. The correct classification rate was also lower for all of the species' models using the P=0.5 cut-off compared to the P_opt cut-off, apart from for *P. argus*. However, when the P=0.5 cut-off was used the sensitivity and specificity values were all above 0.7, except for sensitivity for the *C. nobile* and *H. semele* models (but which were still above 0.6) and for specificity for *P. punctata* (0.600) and *C. nobile* (0.569).

Taking into account all of the evaluation measures, the *C. nobile* and *P. punctata* models appeared to be the least effective of the GLM models as they achieved the lowest AUC values (although still above 0.7, which is considered acceptable discrimination) and generally the lowest sensitivity, specificity, correct classification and Kappa values, whichever cut-off was used. The best performing models were those for *G. constrictum* and *G. illyricus*, which had the highest AUC scores and generally among the highest values for the appropriate threshold-dependent measures, with high sensitivity and specificity values, whilst still maintaining low false positive and false negative rates. The *G. constrictum* model had the highest sensitivity, specificity, correct classification and Kappa values for P_fair, but not for the other cut-offs. The *G. illyricus* model had the highest Kappa and correct classification rate values for all the cut-offs except for P_fair and the highest sensitivity value for the P=0.5 cut-off and the second-highest for the P_Kappa cut-off. The values were not exactly in the same order (of best to worst) for the different measures, but in general, if a model performed poorly with one of the measures it tended to perform poorly with all of them.

In terms of different GLM models run for two of the species, there was very little difference in performance between the two *N. sylvestris* models. The model with 7 variables (with the two variables which did not make ecological sense, ED.AQ6 and PA.HL2.AQ1, removed; see Appendix 16) performed slightly better. It achieved a slightly higher AUC value than the 9-variable model and higher correct classification rates, Kappa values and higher sensitivity and specificity values (except for a lower sensitivity value for the P=0.5 threshold) for the different thresholds, except for a lower correct classification rate and Kappa value for the P_opt and P_Kappa thresholds. This was despite the fact that the residual deviance and AIC values were higher for the 7-variable model than for the 9-variable model.

For the *P. globulifera* models there was also very little difference in performance between the two different versions. Indeed, the values for all of the thresholds were the same (excluding the P_crit values) except for the P=0.5 threshold. With the P=0.5 threshold the correct classification rate was slightly higher for the *P. globulifera* model with 6 variables, and the AUC value was slightly higher for this model. This suggests that the removal of the Eden.W5.W6 variable which did not make ecological sense (see Appendix 16) was beneficial.

When GAMs were run for the two species for which the AUC values for the GLMs were the lowest, *C. nobile* and *P. punctata*, the AUC value was significantly improved for both *C. nobile* GAMs, suggesting that allowing for non-linear relationships improved model performance in this case, although the AUC value was still lower than for any of the other GLM models (except for the *P. punctata* and *N. sylvestris* - 9 models). The *C. nobile* GAM containing 12 variables (i.e. before removal of non-significant variables after the stepwise selection) had a very slightly higher AUC value with better performance on most of the threshold-dependent measures with equal and unequal weighting (i.e. Kappa, higher sensitivity and specificity values and lower false positive and false negative values) than the model containing 7 variables. However, there was no significant improvement in the AUC value for the *P. punctata* GAM model, with an AUC value of 0.723, compared to a value of 0.721 for the GLM. However, the performance based on the threshold-dependent measures was slightly improved, particularly for the P_fair threshold, so that GAM model was very slightly better than the GLM model for this species.

The unequal weighting was applied to decrease the false negative rate, as this would be more of a problem in the context of aiming to find new potentially suitable sites or populations of a species. Obviously, as the false negative rate decreases, the true negative (specificity) rate decreases and the false positive rate increases (although this is less of a problem), as does the true positive (sensitivity) rate. The Kappa values and correct classification rates either did not change, increased or decreased for different species. If a threshold was required to convert predictions for the whole of the New Forest into potential presence or absence sites, then the P_crit cut-off from the P_opt threshold for the weighting could be applied.

Finally, Chefaoui and Lobo (2008) provide a word of caution in interpreting the results of models using pseudo-absences. As both specificity and AUC scores estimate the degree of accuracy of the absence information used in the model training process, a high specificity score only implies that most of the data considered as absence data are correctly predicted and does not imply a high performance in the prediction of the unknown true absences.

3.4. Discussion

The GLMs generally performed well, all achieving AUC values of at least 0.7 (considered acceptable discrimination), and apart from the models for *C. nobile* and *P. punctata*, all of the GLM models achieved AUC values of at least 0.8 (excellent discrimination). These values were higher than the AUC value of 0.66 achieved by Seone *et al*. (2006) who modelled the distribution of potentially suitable habitat (at the same 100 m x 100 m resolution) for the endangered Dupont's lark over a much smaller study area (~42 km^2) in north-east Spain. However, the data used to test the models was collected several years later than the data used to test the models, which may have been a factor in this lower value. However, higher AUC values (0.879 – 0.972) were obtained, for example, by López-López *et al*. (2007) for GLM models to predict potential golden eagle habitat for application to Important Bird Areas design, and by Thomaes *et al*. (2008) for GLM and GAM models (AUC values between 0.86 and 0.90) developed to predict potential stag beetle distribution across Belgium.

Brotons *et al*. (2004) reported AUC values ranging from 0.76 to 1 for GLM models of 30 forest bird species in the Catalan region (northeastern Iberian Peninsula). This was a similar range of values to those reported by Guisan and Hofer (2003) for 13 reptile species across the whole of Switzerland. Further, Pearce *et al*. (2001) found that of GAM models developed for 153 reptile, marsupial, bird and vascular plant species in north-east New South Wales, Australia, 70% yielded a discrimination index better than 0.7. There therefore appears to be a range of AUC values reported by GLM and GAM studies, dependent on a range of factors, such as the characteristics of the study species (see below), the study area and available data. The results of the models reported in this study fall within this range, although perhaps towards the lower end of this range. Nonetheless, all the AUC values were above 0.7.

The *C. nobile* GAM model performed considerably better than the GLM model. Pearce and Ferrier (2000b) also found that models fitted using GAMs provided a slightly greater level of accuracy than those derived using GLMs across a range of plant and animal species. However, using GAM for fitting a *P. punctata* model did not result in a significant improvement in model performance. Therefore, it seems that in certain cases, better models may be obtained by using GAMs as opposed to GLMs, but in some cases it may just be difficult to fit a good model for a species. However, GAMs may overfit the data, particularly with small data sets (Gibson *et al.*, 2007). Further, with small data sets aspects of the shape of the smoothed response may be influenced by a small number of data that may not fit with the general trend and with a larger sample size the trend may be simpler (which may have been the case for the minimum February temperature variable for the *C. nobile* GAM). However, a large sample size is rarely available, and was not for any of the selected species. Olivier and Wotherspoon (2006) also suggest that GLMs may be a more robust tool than GAMs when analysing presence-only data with computer-generated pseudo-absences, as they may not 'overreact' to potentially mis-selected pseudo-absences points, as GAMs may do, by over-fitting the data. Therefore it may be preferable to fit GLMs where possible and GLMs are also favourable because the results are easier to interpret by way of a retrievable model formula (Gibson *et al.*, 2007).

However, in some cases the response may not be linear and it would not be appropriate to allow only a linear relationship (Guisan and Thuiller, 2005), although stepwise GAM can be used to select linear and non-linear forms of variables. An alternative could be to initially fit (univariate) GAMs and inspect the shape of the response curves of the GAM plots before fitting GLMs, for example as carried out by Gibson *et al*. (2007). If there was then evidence of non-linearity GAMs could be fit or transformations (such as logarithmic) of the data could be carried out and GLMs run (Gibson *et al.*, 2007). However, this would be more time consuming, particularly if this had to be carried out for a large number of potential variables. It may also be beneficial to include interaction terms, but with a large number of variables there would be a very large number of possible interactions. Further, the number of candidate models would be greatly increased, as would the complexity, which may not necessarily be supported by relatively small data sets (Gibson *et al.*, 2004).

Stepwise variable selection appeared to work better, by selecting more significant variables, than expert-variable selection, and included significant expert-selected variables, as well as additional variables that had not been expert-selected. For example for *C. nobile*, stepwise selected more significant variables than the expert variable selection, as well as including the significant expert-selected variables, so worked very well. Indeed, all of the final models were based on the stepwise-selected set of variables, but with some of the less significant variables removed. Therefore, despite some of the criticisms of stepwise variable selection (see Appendix 13.2), in this case it can be used fairly confidently. This is encouraging because it means that the approach can be effectively used to quickly run large numbers of models, particularly for species where expert knowledge is lacking. The performance of stepwise variable selection was likely improved by removal of variables with coefficient (estimate) values with the ‘wrong’ relationship.

However, some of the variables selected by stepwise did not appear to make sense ecologically (for example some of the stepwise-selected variables for *N. sylvestris*). Therefore, it is important that stepwise variable selection is not solely relied upon to always select the best variables and where possible, the selected variables should be checked for ecological plausibility (Hosmer and Lemeshow, 2000). In addition, it is important not to over-interpret the results, as the final set of predictors may be just one of several best subsets of predictors (Fox, 2002) and as demonstrated, they may not make sense ecologically. However, this may be due to identification of new associations that were not previously known from the literature. Or even if they do not appear to make sense ecologically, they may be acting as a surrogate for another factor (or having an indirect influence, such as slope for laying water) and as long as they predict well, then this should not be too much of a problem, particularly because the models will not be applied outside of the study area used to train the models.

However, it may also be the case that associations arise due to recording artefacts, such as may be the case for *N. sylvestris*, which is why it is important to be critical of the results. When the ED.AQ6 variable was removed from the *N. sylvestris* model, the AUC value for the model was higher. Similarly, the removal of the Eden.W5.W6 variable from the *P. globulifera* model also resulted in an improved AUC value. Some of the associations that do not appear to make sense ecologically may be due to the influence of the randomly selected pseudo-absences.

The size of the training data set did not appear to be a factor in model performance as the *C. nobile* model had one of the larger sample sets (136) and the *P. punctata* model had one of the smaller data sets (100). In addition, one of the best performing models was for *G. illyricus*, which had the lowest number of training data (99). However, the difference in these sample sizes was not very large and the effective sample size also depends on how clustered the records are and whether they capture the full range of conditions in which the species occurs. Nonetheless, a larger sample size is likely to be beneficial, but is rarely available, so the fact that in general good models were achieved with the available records is promising.

Another factor affecting model performance may be how restricted the habitat is in which the species occurs. In general, there did not appear to be a huge difference between the species in this respect. However, *C. nobile* may occur on a wider range of habitats as it occurs on grassy heaths, on New Forest lawns, in woodland glades, on cricket pitches and in both wetter and drier areas (see Appendix 2.1), whereas *G. illyricus* for example, occurs in a more specific bracken-heath habitat, often close to woodland. Further, *P. punctata* occurs in a comparatively wider range of habitats and it may be that this makes it difficult for the model to discriminate between potential presence sites and absence sites if a site could just as likely be a presence or an absence site. Indeed, others (e.g. Pearce *et al.*, 2001; Hepinstall *et al.*, 2002; Brotons *et al.*, 2004; Hernandez *et al.*, 2006; Brotons *et al.*, 2007; Tsoar *et al.*, 2007; also see section 1.2.5, Chapter 1) have suggested that species with restricted ecological niches can be modelled with higher accuracy than more generalist species. Wollan *et al*. (2008) also reported that better models were achieved for fungi species with restricted distributions compared to more widespread distributions. This was because their distributions were characterised more precisely by environmental variables than are the distributions of more widespread species, and the areas in which the species were present were better contrasted with absence areas.

Cowley *et al*. (2000) found that the lowest performing logistic regression models in their butterfly and moth study were for species that were either very rare, had had habitat requirements that cut across the coarse habitat categories used, or were ubiquitous. For this last case they note that although their models lacked statistical significance they were realistic in that the species were extremely widespread and were predicted to occupy most sites. The best performing models were for sedentary species, which had strong habitat associations and were widespread within those habitats. Pearce *et al*. (2001) also found that model discrimination ability was higher for less mobile biological groups, such as vascular plants. This trend was not seen in the species used in this study, although this could be because none of the species were very mobile (even the butterfly species were fairly sedentary) and because the effect of species' habitat specialism may have had a greater affect.

Fielding and Bell (1997) also suggest that the greatest difficulty that ecological processes can create for classifiers is that some of the negative locations may be similar, and possibly identical, to positive locations (particularly for generalist species), which will degrade the performance of the classifier and/or result in too many false positives. This is particularly likely when pseudo-absences selected at random are used as, by chance, a pseudo-absence may be selected in area that is suitable for the species. To help reduced this in part, it may be preferable to create a buffer around known presence sites where pseudo-absence points will not be selected (Olivier and Wotherspoon, 2006; Gibson *et al.*, 2007). Owing to the likelihood of pseudo-absences being selected in suitable presence locations, it will generally be very difficult for any of the models to achieve a very high discriminatory performance. In addition, owing to limited occurrence data, the full range of conditions in which the species occur may not have been captured.

Further, in conservation-based studies it is almost inevitable that the species will be restricted to few locations, thus only a small proportion of the potentially positive cases will be occupied, as many suitable sites are not saturated (Fielding and Bell, 1997). In addition, organisms, particularly sessile organisms such as plants, may be in their current locations because of past rather than current events, although it would be very difficult to incorporate this ecologically important information into a classifier (Fielding and Bell, 1997). The models also do not include the effect of inter- or intra-specific competition or ecological processes, which are again difficult to incorporate (Fielding and Bell, 1997).

If a species is naturally rare within the study area (for example *G. illyricus*), the model is likely to predict lots of false presences just because the species is so rare and does not occur at many seemingly suitable sites (just because, for example, it has not spread there). Additionally, although a site may appear suitable at a coarse level, there may be more subtle factors that influence the occurrence of the species. That the models also do not incorporate finer scale habitat variables that may be important is another factor which is likely to be responsible for reducing model performance. For example, *P. argus* is a heathland species, but has a preference for heathland in the pioneer growth stage, which will therefore exclude its occurrence on large areas of heathland. However, this sort of more detailed habitat information is very rarely available spatially over large areas for use in species distribution models. It can also be very difficult to incorporate spatially other potentially important factors, such as biotic interactions, which means that these sorts of models tend to over-predict species' distributions, although not as much as presence-only models (Zaniewski *et al.*, 2002). Therefore, a common cause of poorer model performance is that not all of the ecologically-relevant processes were specified in the model because they were not available (Fielding and Bell, 1997).

Nonetheless, in the context of identifying potential ‘new’ sites of species occurrence, it is better for the models to over-predict than to under-predict and ‘over-prediction’ can also be advantageous, for example, by suggesting possible areas with high potential for (re)colonisation (Lütolf *et al.*, 2006). Further, over-prediction at a coarse-scale can be used to identify broadly suitable habitat (for example presence of heathland for *P. argus*) and then finer-scale models (such as the Bayesian models in Chapter 4) can be used to further calculate the suitability of the site (such as presence of pioneer heath, and other more detailed variables). Therefore, a high sensitivity (true positive) rate is important and a high false negative rate would be a particular problem. However, too many false positives could waste time.

The choice of a threshold for defining presence-absence (if required) is not necessarily straightforward. However, when using the output probability values for focusing surveying of new populations, rather than defining an absolute threshold, the best approach may be to start with areas with the highest probability values and work from there. In particular because the predicted values should be seen as relative (ranked), rather than absolute values as the calibration of the models has not been checked (Vaughan and Ormerod, 2005; and see discussion below). Using the predictions in this way, rather than dichotomising them into presence-absence, also conveys more information rather than potentially disguising a wide range of habitat variation (Vaughan and Ormerod, 2005).

It should be noted that the discrimination ability (i.e. how well a model distinguishes occupied from unoccupied sites (Vaughan and Ormerod, 2005)) and not the calibration of the models (i.e. the numerical accuracy of predictions, e.g. whether 40% of sites with predicted probabilities of 0·40 are occupied (Vaughan and Ormerod, 2005)) was tested during the evaluation. In the context of identifying potentially suitable habitat and most likely new occurrence locations, the models will only really be required to rank sites according to their relative suitability/likelihood of being occupied (Pearce and Ferrier, 2000a; Vaughan and Ormerod, 2005). Pearce and Ferrier (2000b) note that good discrimination ability implies that a model can differentiate between occupied and unoccupied sites and that predictions from the model thereby act as a good index of likely species occurrence, even if the actual predicted values do not represent true probability of occurrence. Lack of discrimination ability usually arises because the explanatory variables in the model are not strongly associated with the presence of the species (Pearce and Ferrier, 2000a).

As noted in section 3.3.3, the evaluation measures do not provide an indication of how well the models perform in the prediction of true absences, only using the pseudo-absences (Chefaoui and Lobo, 2008). One of the main issues with the development of presence-absence models with presence-only data is the use of pseudo-absences. A problem with selecting pseudo-absences at random from non-presence locations is that they may be selected at sites that are suitable for the species. Also, when only a small number are selected (i.e. to match the number of presence records, to ensure equal prevalence) they could occur in very different (in terms of habitat type etc.) conditions every time an alternative selection was made, having a large influence on the characteristics of the pseudo-absences, and resulting in different model predictions. An alternative, as used by Gibson *et al*. (2007), is to sample, once only, a larger number of pseudo-absences (say 5000) randomly over the entire area of concern. Then a weight is assigned to each of *n*/5000, where *n* is the number of presence points, so that the sum of the weight of all pseudo-absences adds up to give the number of presences, simulating a prevalence of 0.5.

A further way of generating pseudo-absences is to use the predictions from presence-only models, such as Biomapper (see Appendix 13.1) as used by Engler *et al*. (2004), although this relies on reliable presence-only models. A similar approach to selecting pseudo-absences from ENFA habitat suitability maps may be to stratify their distribution along a suitability gradient, such as mean annual temperature (Engler *et al.*, 2004) or slope, for example.

However, the best way is to use true absence data, which would involve recorders having to also record when they had visited sites but not seen the species. However, this is much more time consuming in terms of being certain that the species was not present (see section 1.2.4, Chapter 1) and is less attractive (in terms of interest and involvement) than recording presences. Further, given limited time and resources, it is surely more important to record presences of species. However, a wider range of data would be advantageous, for example, data from a stratified survey to incorporate the full variability of conditions in which the species occur in the study area (Gibson *et al.*, 2007). Nonetheless, models can be used to suggest priority areas for surveys, the results of which can be incorporated as new data into an enlarged data set, which in turn can be used to make new, and probably better, models (Wollan *et al.*, 2008).

The data used for evaluation of models is also an issue and although it is generally claimed that independent datasets are preferable, Lehmann *et al.* (2002) suggest that it is not clear whether they really are. In addition, as these models of the current investigation are not going to be used outside of the New Forest National Park, where the data was collected, truly independent testing data is less important. An alternative to a single data split is to use statistical resampling techniques such as cross-validation and jackknifing (Pearce and Ferrier, 2000a), or to average the results from several divisions of the data (into training and testing sets) so that the assessment of model performance is less dependent on a single partition (Fielding and Bell, 1997; Harrell, 2001), although this would potentially involve running a large number of models.

As the available data was collected *ad hoc* without the intention for use in modelling, and it was not possible to collect the amount of additional data ideally required (as is frequently the case), the potential errors associated with such data should be recognised. For example, sampling biases in the data mean that modelled relationships can be dominated by the patterns at the sampled sites rather than the patterns across the entire study area (Barry and Elith, 2006). For example, this may be the case for the *N. sylvestris* GLM model, which appeared to occur close to large ponds (see section 3.3.1.4). However, this may have been an artefact of the data, in that the species was recorded whilst looking for wetland species as part of the study and the model did perform very slightly better when this (and another questionable variable) were removed (see section 3.3.3).

This example shows the important of assessing model results for ecological plausibility (Hosmer and Lemeshow, 2000), although this may be more difficult for running a large number of models, particularly if many of the species are not well known. As the difference in performance between models with the stepwise-selected variables and those with the variables removed from those selections that did not make ecological sense, removal of these variables may not necessarily be required. It was surprising that many of the expert-selected variables were not significant either for the GLMs or the GAMs. This may just be due to the limited number of records so that a clear relationship was not apparent, or that the relationship was not a linear one (see section 3.3.1.2), but shows that the use of stepwise variable selection provided a useful way of identifying suitable variables. This is an advantage because it makes the models quicker to develop.

There are also issues to consider with the environmental data used. For example, the use of a habitat layer which is correct at a fixed point in time, but using species records spread over several years may also lead to discrepancies. This is unavoidable, as only using species records from the same year as the habitat layer would drastically reduce the number of available records. Most of the habitat types will change very little, but for habitat types such as dense scrub, these may change within that time frame, for example through being cleared as part of management. Although the cover of dense scrub appeared to be associated with *H. semele* presence-absence, it may not be advisable to use variables that are more prone to change radically. The habitat data also does not give an indication of the quality of the habitat, although this may be difficult to assess as the ‘quality’ may differ for different species. Measures such as the condition assessments used by Natural England (Natural England, 2009) could potentially be incorporated as an indication of habitat quality.

It is important not to over-interpret the results of the models as they are not necessarily the ‘best’ models (as different models could be obtained, for example, by using different combinations of variables, using different selection rules or using a different selection of pseudo-absences) and are based on suboptimal data. Further, Rushton *et al*. (2004) note that the ‘best’ model identified may depend critically on the route taken to find it. Ecological scrutiny is therefore required (as discussed above) and it should be remembered that measures such as AUC assesses model predictive success, not explanatory value (Austin, 2007). However, the models did provide clear indications of potentially important associations between species occurrence and environmental variables and discriminatory predictions. Despite the warnings about the use of the available (and commonly used) evaluation measures, in the absences of alternatives, the use of a range of measures provided a more detailed insight into the performance of the models, which could be used with more confidence rather than reliance on a single value. It is also important that the models are provided for use with an indication of their sensitivity and specificity and other evaluation measures so that users are aware of the potential weaknesses of the models, such as over-prediction (Pearce and Ferrier, 2000a).

Despite the limitations of the models, they can still be used to aid surveying for new occurrences of species and identify potentially suitable areas for species to expand into (as long as users are aware of the limitations and recognise that they are not a panacea). Although the models are fairly straightforward to develop and to use, there are numerous issues and factors that need to be taken into account and that may affect model performance as discussed above, such as generation of pseudo-absences, adequate sample size and selection of variables, particularly when there are a large number available. The significance of variables is also dependent on which other variables are present, which means that different models could be obtained just based on which variables were initially entered into the model. The models are also sensitive to too many zero values.

Another issue is spatial autocorrelation (SAC), which may lead to an increase in false positives (see Appendix 13.3), although his would not be as much of a problem in this application. Using a method that allowed for SAC did not affect which variables were selected, although other methods are available which work in different ways, including incorporating an additional function to capture the spatial configuration (rather than just allowing for within-group errors to be spatially autocorrelated) (see Appendix 13.3), which may produce different results.

Some SAC may be caused by important autocorrelated variables being omitted from the model (Lichstein *et al.*, 2002; Guisan and Thuiller, 2005; Dormann *et al.*, 2007b). This was a likely reason in this case as variables such as abundances of competitors and habitat quality were not available (Dormann *et al.*, 2007b). SAC in model residuals may also be caused by mis-specifying the functional relationship of a variable (or variables) with the response (Legendre, 1993; Dormann *et al.*, 2007b).

Another likely cause of SAC in these models was from observer and sampling bias (Dormann *et al.*, 2007b), although it will also exist just as a result of environmental factors being more similar in neighbouring sites and due to intrinsic factors such as dispersal (Dormann, 2007; and see Appendix 13.3). Many of the species records are clustered just because species tend to occur in patches, particularly if they do not disperse very far, as is generally the case for the study species, although this is also true for *G. illyricus* and *P. punctata*, which did not exhibit SAC. This clustering is an inherent property of species distributions and by including all of the available data at a useful scale (i.e. 100 m x 100 m sites) such clustering is likely to arise. SAC may be reduced to some extent by thinning out data points (Segurado *et al.*, 2006), which to a certain degree is done by splitting the presence-absence data into training and testing sets. However, at least for the six species' models exhibiting SAC in their residuals, this obviously did not thin out the data enough and did not work. However, for this application, where the models will only be used to predict within the area used to train the models, as long as the models predict well, which the evaluation results show that they do, then that is the most important thing.

3.5. Conclusion

All of the models showed at least acceptable (and the majority showed excellent) discrimination ability, as determined by AUC values, which was within the range of values reported by other studies. In general, as reported by others, more specialist species appeared to be better modelled than more generalist species. The results show that despite some of the issues arising in the development of presence-absence models, such as in the quality of the data, the models can provide a useful tool in identifying areas of potentially suitable habitat for the study species. Further, this was an effective approach for modelling rare species at the scale of individual protected areas. However, users of the models should be aware of their limitations.

CHAPTER 4

Chapter 4. Bayesian network models of habitat suitability

4.1. Introduction

4.1.1. Introduction to Bayesian Belief Networks

A Bayesian Belief Network (BBN) is essentially a set of variables, represented as a network of nodes that are linked by probabilities, that affect some outcome(s) of interest (Marcot, 2006). A BBN can be represented in the form of a network diagram, known as a directed acyclic graph (DAG), to provide a visual representation of the components and dependencies of a domain (Newton, 2009a).

In a Bayesian network diagram, variables (known as nodes), data and parameters are represented by different shapes (such as ellipses and rectangles), which are connected by arrows (known as directed links) to indicate conditional dependencies (Newton, 2009a). A link between two nodes, from node A (a ‘parent’ node of B) to node B (a ‘child’ node of A), indicates that A and B are functionally related, or that A and B are statistically correlated (Newton, 2009a). Variables without parents are known as input nodes. Directed links are representations of conditional dependence and represent influence rather than causality, as there is no requirement that that links represent causal impact (Lauritzen and Spiegelhalter, 1988; Jensen and Nielsen, 2007; Newton, 2009a). The graph links have directions but no directed cycles, or closed loops, are permitted (Newton, 2009a). Propagations can be made through the links in either direction of the network, enabling the model to be explored in reverse: the BBN can be used to infer the most likely set of conditions for a given outcome, in contrast to most other modelling approaches (McCann *et al.*, 2006).

Each child node (i.e. a node linked to one or more parents) contains a conditional probability table (CPT) which gives the conditional probability for the node being in a specific state given the configuration of the states of its parent nodes (Newton, 2009a). When networks are compiled, Bayes’ theorem is applied according to the values in the CPT, so that changes in the probability distribution for the states at node A are reflected in changes in the probability distribution for the states at node B (Jensen and Nielsen, 2007; Newton, 2009a). A BBN can be explored by changing the states of the nodes (or variables) incorporated within the model (Newton, 2009a) and when the state of a node is known, that variable is said to be *instantiated* (Jensen and Nielsen, 2007). Once a node has been instantiated, this will then influence the probabilities associated with the states of other nodes to which it is linked, according to the values in the CPTs (Newton, 2009a).

There is general agreement that it is advantageous to have a relatively simple network structure with the minimum number of nodes (preferably three or fewer parent nodes) and no more states per node (preferably five or fewer) than are necessary (Marcot *et al.*, 2006). This keeps the associated conditional-probability table (CPT) small enough to be tractable and understandable (Marcot *et al*., 2006). In addition, Marcot *et al*. (2006) suggest that the depth of the model – the number of layers of nodes – should be kept to four or fewer, to reduce propagation of unnecessary uncertainty from input nodes to output nodes and to prevent the sensitivity of the output node to input nodes being swamped and dampened by intermediate nodes.

When developing BBNs, it will usually be necessary to discretise variables by defining a number of discrete states for each variable (this reflects the computational difficulties of performing Bayesian inference with continuous variables) (Newton, 2009a). Although algorithms are available for discretising continuous variables (for example, Clarke and Barton (2000)), these are aimed for use with machine-learning of (generally extensive) data and not for models based on expert knowledge. Discretisation may not necessarily be straightforward and some degree of interpretation and subjective judgement can be used to a certain extent (Newton, 2009a). Pollino *et al*. (2007) provide a useful description of how they achieved this, by establishing states, where possible using recognised classifications, management thresholds or guidelines. Where these were not available, sub-ranges were specified with the guidance of experts (Pollino *et al.*, 2007). The number of ‘states’ or ‘classes’ assigned to each variable were not pre-determined, but evaluated and assigned on an individual basis. Pollino *et al*. (2007) note that while discretisation of continuous variables is not desirable, it facilitates parameterisation process by simplifying expert elicitation, and it acknowledges that understanding of many parameters, and the data available to support such relationships, is often quite rudimentary.

It is widely recognised that obtaining values to populate the CPTs is one of the main challenges to modelling with BBNs as obtaining appropriate values is often difficult because of a lack of appropriate information, yet the entered values will have a major influence on the performance of the model (Newton, 2009a). The problem increases with the number of directed links associated with each node (Newton, 2009a). It can also be a particularly difficult task for rare events and when the number of probabilities to be estimated is large (McCann *et al*., 2006). The sources of information that are used (often in combination) are expert knowledge, observational or experimental evidence or data which is either available directly or extracted from the scientific literature, outputs of other empirical, mechanistic or stochastic models, or stakeholder consultation (Newton, 2009a). Different information sources, such as data and expert estimations, can be combined and weighted, which is a key advantage of BBNs (Pollino *et al*., 2007).

BBNs are able to learn CPT values directly from a data set, although this is rarely possible in investigations relating to conservation management, where available datasets are often limited. Therefore, they are frequently completed using expert knowledge, but in situations where information is lacking conditional probability values may be based on very restricted information, something which should be borne in mind when interpreting results (Marcot *et al.*, 2001; Newton *et al.*, 2007). In addition, BBN models can easily be built that reflect personal biases, although peer review can help to prevent this (Marcot *et al*., 2001).

The creation of a network diagram, representing the domain of interest (i.e. identification of relevant variables and the relationships between them) is the first stage of BBN development, followed by assignment of states and probabilities to each variable (Newton, 2009a). This can itself be a useful way of eliciting information from experts and structuring the information available and the visual nature of the network can foster communication between interested parties (Newton, 2009a). This participatory modelling process of BBNs can also help to document and communicate current understanding and identify key uncertainties or gaps in knowledge and also identify suitable indicators to provide a basis for monitoring and adaptive mangement (Nyberg *et al.*, 2006; Smith *et al.*, 2007; Newton, 2009a). This is seen as a key advantage of BBNs.

Although BBNs are somewhat similar to decision trees and other decision models, their interactive and graphical representation is a great advantage, particularly in permitting more effective communication of cumulative effects and outcomes of alternative conditions and decisions than do more static models such as decision trees and other traditional statistical approaches like classification or regression trees (McCann *et al*., 2006). They are also more readily understood by non-modellers (McCann *et al*., 2006), which is an important advantage, particularly when they often rely on expert knowledge, and may well be used by non-modellers.

The ease with which BBNs can be created and amended is an advantage over other modelling approaches (McCann *et al*., 2006). Different model structures may be explored and simulations can be run very rapidly at relatively low cost as extensive computer programming or modelling expertise is not required to develop and update models (Smith *et al*., 2007). By being able to instantly recalculate and display probabilities of conditions and outcomes as alternative decisions are specified (for example by comparing the probability of different outcomes arising from alternative management decisions), McCann *et al*. (2006) suggest that BBNs offer a uniquely valuable tool for supporting decision-making. BBNs can also be used to infer the most likely set of causal conditions for a given outcome by solving the models conditional probabilities backward through the model structure, which is something that many other models, such as decision trees, cannot provide (McCann *et al*., 2006).

The fact that BBNs use probabilistic, rather than deterministic, expressions to describe the relationships among variables also makes them different from most other environmental modelling approaches and makes them particularly useful in the context of risk assessment and for supporting decision making (Borsuk *et al.*, 2004; Newton, 2009a). This use of BBNs can also be enhanced by incorporating decision nodes and utility nodes to create influence diagrams (related to decision trees) (Newton, 2009a). Decision nodes represent two or more choices or decisions (made by a user of the model) that influence the values of other response nodes and do not have CPTs associated with them (Nyberg *et al.*, 2006; Newton, 2009a). A utility node represents some measure that can be used to assess the success or failure of a decision and is associated with a utility table, which specifies the utility of each configuration of the parents of the utility node (which may be influenced by a decision, through a link with a decision node) (Nyberg *et al.*, 2006; Newton, 2009a). Once parameterised, such a model can be explored to identify the choice in each decision node that minimises the costs or maximises the benefits or values considered (Nyberg *et al.*, 2006).

BBNs are particularly useful when there are uncertainties in the available information used to construct a model, something typically associated with environmental modelling (Newton, 2009a). For example, Newton (2009c) used a Bayesian Network to produce IUCN Red List (www.redlist.org) classifications for taxa in situations where the input data were uncertain. Newton *et al*. (2007) suggest that a key feature of BBNs is that the results are presented as probability distributions or relative likelihoods of different outcomes, which provides a highly visual means of representing the uncertainty surrounding the potential outcomes of, for example, conservation management interventions. Smith *et al*. (2007) also highlight the suitability of BBNs for species habitat modelling in data and information poor environments, and for accounting for uncertainty in data and knowledge. These features, as well as the ability of BBNs to combine empirical data with expert judgement, makes them an extremely flexible and useful modelling tool (Smith *et al*., 2007).

However, despite their advantages, BBNs are prone to many of the general limitations common to other modelling approaches, including the difficulty of incorporating all sources of causality, uncertainty and variability in a model without errors and inaccuracies (McCann *et al*., 2006). In addition, in some environmental domains, the interactions among variables may be highly complex and difficult to quantify (Gu *et al.*, 1996). BBNs, like other decision models, may oversimplify criteria affecting a decision and fail to depict subtle variations of decisions and changing conditions that so often occur in real-world situations (Nyberg *et al*., 2006).

A further shortcoming of BBNs is that they do not strictly permit feedback functions either within a node or from response (output) variables back to predictor (input) variables. Feedback can be important in many systems such as density-dependent survivorship and reproduction in wildlife population models and consumer performance in economic models (Nyberg *et al.*, 2006). They are also poorly suited to examine dynamics over time, although there are some approaches that can be used to overcome this (Newton, 2009a). Another issue is that discretising continuous-variable distributions, as is necessary in most BBNs, might oversimplify state responses (Nyberg *et al*., 2006). Some of these drawbacks will be more of an issue in certain applications than others, but Nyberg *et al*. (2006) suggest that BBNs be viewed as decision-aiding tools to help inform and advise the decision-maker who, ultimately, must weigh the ramifications of decisions that can be far more subtle and complex than any model can depict. As with any modelling approach, it is important that BBNs are used with appropriate knowledge of their strengths and weaknesses (Nyberg *et al*., 2006).

4.1.2. Evaluation of BBNs

An important aspect of developing models is model evaluation. Marcot *et al*. (2006) emphasize the importance of testing and validation of BBN models to ensure reliability and accuracy, as otherwise models built solely on the basis of expert judgement represent nothing more than unconfirmed belief structures. Statistical procedures typically used to test environmental models include using confusion matrices and the kappa statistic and receiver–operator characteristic (ROC) curves (Marcot *et al.*, 2006; see Appendix 8.2), but their use to test BBNs has been rare to date (Newton, 2009a). BBNs are commonly evaluated by measuring predictive accuracy; the frequency with which the predicted node state (that with the highest probability) is observed, relative to the actual value (Pollino *et al*., 2007).

Evaluation using experts is also valuable and can be done via a structured review of the model (Pollino *et al*., 2007). Another important aspect of evaluating BBNs is sensitivity analysis. Sensitivity analysis involves running the model with a range of parameter estimates and inputs and observing how responsive the probabilities of query nodes (outputs) are to these changes outputs (Pollino *et al*., 2007). Each of the uncertain parameters can be varied in turn (such as values in the CPTs), recording the response of the model, while holding all other parameters constant at their most likely values (Newton, 2009a). Sensitivity analysis can examine ‘sensitivity to findings’, which considers how the BBNs posterior distributions change under different conditions, as well as, ‘sensitivity to parameters’, which considers how the BBNs posterior distributions change when parameters are altered (Pollino *et al.*, 2007). Sensitivity analysis can help experts determine whether the model is performing according to their beliefs, and can be used to help guide the collection of field data for model validation (Marcot, 2006).

4.1.3. The use of expert knowledge in BBNs

BBNs are generally considered to be particularly useful when expert knowledge is an important part of the knowledge available for the chosen problem (Uusitalo *et al.*, 2005), and they offer powerful tools for assisting in the elicitation, integration and analysis of expert knowledge (Newton, 2009a). Indeed, in many cases the models, in particular the use of probabilities to populate the CPTs, can only be developed using expert knowledge.

Expert knowledge (also called expert opinion or expert judgement) can be considered as a form of data given by an expert in response to a technical problem, and is an informed opinion based on the expert’s experience (Meyer and Booker, 1991). An expert is the person whose knowledge is sought; the term ‘expert’ does not necessarily signify any more than that (Garthwaite *et al.*, 2005), although they will generally be someone who has background in the subject area, and be recognised by his or her peers or those conducting the study as qualified to answer questions (Meyer and Booker, 1991).

The problems of using expert knowledge are widely acknowledged, giving rise to potential biases and error (Morgan and Henrion, 1990; Meyer and Booker, 1991). Such errors may arise because individual ‘experts’ may have an imperfect knowledge of a domain, they may lack the skills to assess probabilities accurately, or may be biased in their beliefs (Newton, 2009a). However, there are often no alternative source of information available (Newton, 2009a). For example, the data available for a particular domain will often be partial, and not presented in the form of probabilities (required for populating the CPTs). In such cases, some degree of interpretation and subjective judgement will be necessary, for example in discretising variables (Newton, 2009a).

Established techniques are available for eliciting expert knowledge, including techniques for eliciting probabilities (for example see Meyer and Booker, 1991), but these have not been widely employed in BBN investigations to date (Newton, 2009a). As Reckhow (1999) notes, there is considerable evidence that people are generally not particularly good at making judgements and representing their knowledge (such as in the form of probabilities) under uncertainty, indicating the importance of using appropriate and effective elicitation techniques (Newton, 2009a).

Pollino *et al*. (2007) note that elicitation of parameters, particularly for complex BBNs, can be a difficult and time-consuming task, and often the knowledge of experts is incomplete. However, although expert judgement is not a substitute for definitive scientific research, it can provide useful insights for policy makers and research planners while research to produce more definitive results is ongoing (Morgan *et al.*, 2001) or it may be the only source of detailed information regarding particular problems.

4.1.3.1. Elicitation of expert knowledge

Expert knowledge may be used in all stages, or just for part of BBN model development, but a key aspect generally involves populating CPTs with probability values. Elicitation is the process of gathering expert knowledge through specially designed methods of verbal or written communication (Meyer and Booker, 1991). The method used will depend on factors such as the type of information the experts must provide, the number of experts available and the amount of time and resources available (Meyer and Booker, 1991). However, there are three basic elicitation situations: individual interviews, interactive groups and Delphi situations, of which further details are provided in Table 36 below.

Elicitation situation	**Description**	**Advantages**	**Disadvantages**
Individual interview	Experts are interviewed alone, usually in a face-to-face situation with the interviewer. This can be structured to any degree.	Best method for obtaining detailed, in-depth data, especially about the expert's problem-solving processes. Avoids potential bias from group dynamics, and data can be combined later, usually by mathematical aggregation.	Time consuming. No synergistic effects from inter-expert discussion.
Delphi situation	Experts, in isolation from one another, give their judgements to a moderator. These anonymous judgements are then redistributed to the experts, allowing them to revise their previous judgements. These iterations can be continued until consensus, if it is desired, is achieved.	Designed to avoid biases arising from group dynamics (although some doubts exist about whether this is accomplished).	Limited in the amount of data that can be gathered. Less synergism than in the interactive group.
Interactive group	The experts meet in a face-to-face situation with one another and a session moderator (data gatherer). The expert's interactions with one another can be structured to any degree.	Generates more accurate data, particularly for predictions, and a greater quantity of ideas than the other two situations.	Possesses the potential for group-think bias. Poses logistical problems in scheduling and handling multiple interacting experts, particularly if there are more than four to seven experts per moderator. Possibility of strong personalities in the group having too much weight in the discussion. It may also be true that the pressure to reach consensus leads to the experts suppressing dissenting views, or alternatively it may not be possible to reach consensus (Garthwaite *et al*., 2005).

Table 36. Descriptions and advantages and disadvantages of the three main situations for eliciting expert knowledge (developed from Meyer and Booker (1991)).

O’Hagan *et al*. (2006) report that there are few studies which have compared these methods for eliciting expert judgements and that the results have been inconsistent. However, it is widely agreed (for example see Meyer and Booker (1991) and O’Hagan *et al*. (2006)) that face to face contact (rather than phone or mail) is best for obtaining detailed data, although it is the most time consuming and expensive. A combination of methods may also be used. For example, in the development of BBNs by Uusitalo *et al*. (2005), discussions on the model structure were discussed as groups and the estimated probabilities for the final model were then carried out by each expert individually, via a questionnaire form, so that estimations were based on individuals own judgement and not affected by the other experts opinions and judgements.

4.1.3.2. Aggregation of expert knowledge

When eliciting expert knowledge from multiple experts, some form of aggregation is required to combine the answers from the experts. There are two main types of aggregation; behavioural and mathematical. Behavioural aggregation occurs during, rather than after the elicitation session and attempts to generate agreement among the experts by having them interact in some way (Clemen and Winkler, 1999). It is therefore only suitable for interactive groups or Delphi situations, through which a consensus is elicited from the group as a whole (O'Hagan *et al.*, 2006).

Behavioural aggregation has the advantage of an aggregated result being produced during the session, which is based on a consensus (if experts are able and willing to do so), instead of requiring separately elicited beliefs to be combined afterwards (O'Hagan *et al.*, 2006). However, it can foster a group-think situation in which no-one truly thinks but unconsciously complies and can be very time consuming if group-think does not facilitate unconscious agreement (Meyer and Booker, 1991). It can also suppress differences between the experts answers and the reasons for the differences, both of which can be critical to the understanding, analysis, and use of these data (Meyer and Booker, 1991).

Mathematical aggregation involves the elicitation of answers from each individual individually and independently of the others and then mathematically combining the results into a single estimate (O'Hagan *et al.*, 2006). This form of aggregation has the advantage of not having to be planned as early or as closely in conjunction with the elicitation situations as behavioural aggregation and means that different mathematical schemes can be applied in succession to the individual's data, whereas with the behavioural aggregation the process can usually only be done once (Meyer and Booker, 1991). However, like any type of aggregation, mathematical aggregation obscures the differences between the expert's answers (Meyer and Booker, 1991). Meyer and Booker (1991) also caution that it is easy to perform mathematical aggregation incorrectly, such as by combining the estimates from experts who have made such different assumptions in answering the question that they have essentially solved different questions. In addition, they suggest that mathematical aggregation can lead to the creation of a single answer that all of the experts would reject.

In a review of the literature on combining probability distributions from experts in risk analysis, Clemen and Winkler (1999) suggest that mathematical aggregation outperforms intuitive aggregation (based on the decision-makers assessment, rather than formal aggregation) and that mathematical and behavioural approaches tend to be similar in performance, with mathematical rules having a slight edge. O'Hagan *et al*. (2006) also report that generally, the group often performs less well than individual elicitations followed by a simple mathematical averaging of the individual judgements.

4.1.3.2.1. Aggregation estimators

The most commonly used method of combining a set of answers is to calculate a single summary value (estimator) based on all the values in a data set, with the most popular estimators for central tendency being the mean, median and geometric mean (Meyer and Booker, 1991). The mean is the average of the values and gives equal weight to each datum. This has a serious implication: if only a few experts provide answers and one expert gives an answer that is far away in value from the rest, then that extreme value will greatly influence the mean value, which may not be desirable, especially if the extreme value appears questionable or seems unreasonable (Meyer and Booker, 1991). However it is possible to run different versions of the models, with and without the extreme values.

To overcome the influence of extreme values when forming an aggregation estimate, Meyer and Booker (1991) suggest using the median or geometric mean (the average of the data values based on a logarithmic scale) as the central values of the data set tend to influence both of these estimators while the extreme values do not. The median is the 50th percentile value: half of the data is larger than the median and half of the data is smaller than the median. If the data set has an odd sample size, the median is calculated by finding the central value of the ordered data points and if the data set has an even sample size, then the median is the average or halfway point between the two centre values (Meyer and Booker, 1991).

Another alternative is to use a weighted average or mean where each datum (expert answer) is given its own individual weight. The advantage of this aggregation method is that the analyst can control which values (or experts) have the most influence on the estimator (Meyer and Booker, 1991). However, the biggest disadvantage is that the weights must be determined for each expert, which can be difficult (Meyer and Booker, 1991). There are several different ways of determining weights but Meyer and Booker (1991) suggest using equal weights unless some unusual circumstances indicate the use of some different weights. O'Hagan *et al.* (2006) also suggest that the general message of the literature is that simple aggregation methods (e.g. a simple average) work well in comparison with more complex methods. This is also suggested by Clemen and Winkler (1999), in a review of the literature, who advocate that simple rules are also practical because of their ease of use and robust performance.

There is limited coverage in the literature on the aggregation of probability values from experts for ecological BBNs. However, Uusitalo *et al.* (2005) combined the probability distributions of experts in their BBN for estimation of salmon smolt capacity of rivers using a simple average, based on evidence that simple combinational models outperform group judgements compared with more complex combinational rules. They also considered their experts to be exchangeable in the sense that their probabilities were treated equally and symmetrically. Martin *et al.* (2005) also used an unweighted average to determine the mean response elicited from their experts (on the impact of grazing on birds). By taking an unweighted average they avoided difficulties concerned with rating the comparative 'accuracy' of each expert's opinion. They also note that to some extent, the expert's ability to provide this measure was contained within the expert data, since experts only provided responses for which they were confident.

4.1.3.3. Evaluation of expert knowledge elicitation

It is important that the elicited knowledge is verified. Verification is the process of checking whether the probabilities provided by the expert are well calibrated (conform to observed frequencies), obey the laws of probability (are coherent) and are reliable (Renooij, 2001). However, checking whether the assessments conform to 'reality' is often impossible, since the events for which the probabilities are assessed are often unobservable (Renooij, 2001). Coherence can be checked during the elicitation by checking whether all probabilities sum to 1 where required (Renooij, 2001).

The reliability of elicited probabilities could be checked by re-eliciting the probabilities from the experts or testing whether the experts agree with their own assessments, that is, whether they would provide the same estimates when asked for the same probabilities again (Renooij, 2001). However, when dealing with belief networks, the number of probabilities to be assessed is often so large that it is infeasible to assess them more than once (Renooij, 2001). An indication of the validity of the assessments can also be obtained by entering observations into the belief network and computing the effect of the observations on the probabilities for certain variables of interest (Renooij, 2001), a form of sensitivity analysis.

4.1.3.4. Bias in expert elicitation

An elicitation is done well if the distribution that is derived accurately represents the expert's knowledge, regardless of how good that knowledge is (Garthwaite *et al.*, 2005). However, achieving accurate elicitation is not straightforward. For example, the expert may be unfamiliar with the meaning of probabilities, but even when the expert is familiar with probabilities and their meaning it is not often easy to assess a probability value for an event accurately (Garthwaite *et al.*, 2005).

Bias can be introduced into the elicitation by both the experts and the elicitor. For example, in face-to-face situations, data gatherers can intentionally or unintentionally influence the expert through their body language, facial expression and wording of questions (Meyer and Booker, 1991). People are limited in the number of things that they can mentally juggle, so when experts are asked to make complicated judgements such as probability assessments, they often subconsciously use heuristics, or rules of thumb, to simplify the task (Renooij, 2001). In general these heuristics are quite effective (Garthwaite *et al.*, 2005), but their use can introduce cognitive biases in probability assessments. Cognitive bias occurs when the expert's predictions do not match what can be observed and measured, although it is often not possible to check (Meyer and Booker, 1991). Commonly used heuristics include judgement by availability, representativeness and anchoring and adjustment. Further information on biases and heuristics can be found in Appendix 22.

As well as cognitive bias, another form of bias, motivational bias, can also occur. This can happen, for example, when an expert consciously or unconsciously makes accommodations to please the interviewer, something which is assumed to be driven by our human needs, such as for approval (Meyer and Booker, 1991). Meyer and Booker (1991) report that experts may not report their actual solutions or thought processes, for example because of factors such as social acceptability, impression management bias and wishful thinking (see Appendix 22 for further details).

It is important to account for bias in the elicitation of expert knowledge as it can degrade the quality of the data and may impact on the credibility of the project (Meyer and Booker, 1991). Meyer and Booker (1991) hold the view that humans cannot directly perceive reality because it is interpreted through the filter of their mental models. Similarly, it is not possible to determine objectively what is biased and what is not; it is only possible to make value judgements as to what is considered desirable (e.g. more objective) in the data (Meyer and Booker, 1991).

4.1.4. Use of BBNs in habitat suitability modelling and ecology

Newton (2009a) reports that although interest in the use of BBNs for environmental modelling has increased in recent years, there have still only been relatively few publications on conservation management (including habitat suitability analysis) or studies addressing spatial management or planning. However, in a recent review of studies describing the application of BBNs to environmental problems, Newton (2009a) found that the vast majority of authors considered that the BBN models they developed were in some way useful.

Smith *et al*. (2007) developed a BBN model and linked it to a GIS to map habitat suitability for the endangered Julia Creek dunnart in Australia, at the regional scale (10,000 km^2). Expert knowledge (elicited using a process similar to the Delphi technique), supported by field data, was used to develop the models and they found that BBNs provided a cost-effective tool for prioritising where animals were most likely to occur and hence help guide targeted field surveys for locating and monitoring the species. A BBN model was also developed by Rowland *et al*. (2003) to provide spatially explicit estimates of relative habitat capability for an elusive wolverine in the Columbia River Basin in the United Sates. Their results could be used to guide regional conservation planning for the species.

A series of BBN models were developed by Marcot (2006) to predict suitability of sites for, and presence of, rare and little-known species associated with late-successional and old growth (LSOG) forests. The models were built at two spatial scales: a broad eco-province scale, and a site-specific scale. The broad-scale models were built in a geographic information system (GIS) to predict potential environmental suitability of selected species based on climatic and geophysical variables and potential natural vegetation across landscapes. The site-specific models were developed using BBNs and could be run independently of the broad-scale models. Alternatively, the broad-scale models could be used first to map large polygon areas of potentially suitable environments for a species, and then the site-specific models applied within those polygons to refine the predictions, because many of the species modelled responded as much to very fine grained environmental and habitat features that were not represented in GIS, as they did to broader-scale features. The example model presented by Marcot (2006) accurately predicted species presence, but incurred a fairly high error rate in predicting species absence (false positives).

Effective BBNs were developed by McNay *et al*. (2006) in evaluating habitat for woodland caribou in British Columbia. These authors reported that BBNs and the use of expert knowledge were useful in that the approach: supported the development of common understanding across disciplines and among stakeholders (including learning where understanding differed or was most lacking); allowed for consistent applications of general ecological understanding across variable situations; allowed for sources and implications of uncertainty to be transparent within a decision context; and allowed forecasts of alternative management scenarios prior to implementation, thereby providing decision-makers with relative assessments of risk and probabilities of success. BBNs have also been used in a similar way, to evaluate the ecosystem management alternatives on elk and deer (Lehmkuhl *et al*., 2001) and to investigate the responses of terrestrial lichens to forest management in British Columbia (Nyberg *et al*., 2006). Raphael *et al*. (2001) also used BBNs to analyse the effects of three land management alternatives on 31 terrestrial vertebrates of conservation concern in the interior Columbia river basin.

From their experiences of using expert knowledge in developing BBNs as a tool for estimating Atlantic salmon smolt carrying capacity of rivers, Uusitalo *et al*. (2005) report that BBNs are useful tools when expert knowledge is the most important part of the knowledge available for some part of a management problem and the ease with which the probabilities in those parts of the problem where data exist can be updated is advantageous. Based on their use of BBNs to assess the potential impact of different management interventions in four case studies of conservation interest, Newton *et al*. (2007) also suggest that BBNs have outstanding potential for supporting evidence-based approaches to conservation management.

4.1.5. Context

The above studies show that although application of this approach to habitat suitability modelling has so far been very limited (and with no application across a range of species), BBNs have effectively been used to model habitat suitability and have a number of advantages that make them particularly suitable for this application. BBN models can be used to capture finer-scale variables that are important for species habitat suitability, but are not available as spatial GIS layers for inclusion in the models developed in Chapters 2 and 3. The ability of BBN models to quantitatively incorporate expert knowledge is a key advantage in this respect, as is their ability to easily incorporate any relationship (i.e. not necessarily linear). Expert knowledge is an important source of information for many species (particularly very rare species, for which there are very few records or for species where there is limited information available in the literature) and BBNs provide an excellent way of capturing that knowledge, which may otherwise be unrecorded. The nature of the models, including the visual representation, can also be used to encourage deeper thinking about the habitat requirements of the species and also to identify gaps in knowledge of the species.

The fact that BBNs are easy to use and also to update is an important factor, as the aim is to develop models that non-modellers (i.e. those recording in the field) can use and develop. The ability of BBNs to be used to examine the impact of different management decisions or other factors (such as environmental change) is also important and the fact that the uncertainty often associated with these can be incorporated is a big advantage. BBNs therefore provide a suitable tool for developing models of habitat suitability for the selected species of conservation concern in the New Forest.

The BBN models were applied in a different way to the ENFA and logistic regression models from Chapters 2 and 3, in that they could not be used to predict potential distribution across the whole of the New Forest National Park study area using spatial GIS data, as this was only available for fairly coarse habitat and soil types and topography and not available for finer-scale variables. The BBN models were developed (based on fine-scale habitat variables measured on the ground) for use at site-level, at focused sites (such as determined, for example, by the coarser-scale models from Chapters 2 and 3), to determine habitat suitability in more detail. For example, although a site containing a high cover of heathland may be predicted as being highly suitable for a heathland-specialist species, at a finer-scale, the site may not actually be suitable because the structure of the heathland (such as the growth phase) may not be suitable.

Objectives:

- To identify factors influencing the suitability of habitat for species using available literature and expert knowledge.

- To incorporate this information into BBN models to predict the potential habitat suitability of sites.
- To test the BBN models using field survey data.

4.2. Methods

4.2.1. Identification of relevant variables

A literature search was carried out for each of the species to construct tables of variables important for habitat suitability (see Tables A-26 to A-33 in Appendix 20) and to identify areas requiring further clarification from experts. Tools such as ISI Web of Knowledge (www.isiknowledge.com) and Google (www.google.com) were used to search (using the species' names as search terms) for studies and other information on each of the species. In addition, relevant books and other 'grey' literature, such as unpublished reports, were searched for from locations such as the New Forest Library and Hampshire and Isle of Wight Wildlife Trust. Where possible, literature specific to the New Forest was used, but this had to be widened to the UK in order to provide sufficient information for model parameterisation.

Importantly, species may have slightly differing requirements in different geographical areas. For example, *P. argus* has slightly different foodplant preferences in different areas of the UK, such as for bell heather, *Erica cinerea* (common to dry heathland) in Suffolk and Surrey and for cross-leaved heath, *Erica tetralix* (more common to humid and wet heaths) in Dorset (Ravenscroft and Warren, 1996). Therefore, it was important to try and include information specific to the New Forest study area, although this was not always possible. This was why expert knowledge was particularly important, to verify where information was lacking.

4.2.2. Development of network diagrams based on the literature

Influence diagrams were constructed in Hugin Researcher 6.9 (Hugin Expert A/S, 2007) using the information from the variables tables created using the literature (see Appendix 20). The models were also developed and checked in an iterative process in consultation with an experienced BBN modeller (A. Newton), which also formed part of the model evaluation (see section 4.2.11.1). Although the development of the diagrams was based primarily on the accumulated literature, they also reflected the knowledge of the modellers. Therefore, they represented a particular viewpoint about an appropriate structure of the domain, based on the personal beliefs of the modellers, but underpinned by a comprehensive assessment of relevant literature.

Not all variables identified in the literature were used to build the models; for example, variables were rejected if they over-complicated the model (see notes in Appendix 25). ‘Proxy’ nodes (which are not measured directly (Newton *et al.*, 2007)) were used where required to reduce the number of parent nodes into a child node. Where possible, i.e. where information was available, values for the states of each node (variable) were also included, to begin construction of the initial BBN. However, for the majority of nodes this information was not available from the literature, but construction of preliminary influence diagrams helped to determined gaps in knowledge and focus questions for expert consultation.

The guidelines provided by Marcot *et al.* (2006) for constructing BBNs were followed as far as possible, such as keeping the maximum number of states to five, where appropriate. Exceptions were for ‘habitat type’ nodes where more than five habitat types were required. In this case, for example as used for the *P. punctata* network, a proxy node was used with ‘habitat type’ as the only parent node, so that the CPT was not too complicated.

The structure of the influence diagrams was kept as simple as possible, so that they would be easy to use. In this sense the models did not have to be ‘complete’ as long as they worked sufficiently well. This meant that even if a relationship existed between two variables (nodes), there did not necessarily need to be a link between them, unless it improved model effectiveness. For each network diagram the terminal node was designed to infer the probability of a particular 1 ha site (the same size site as used for the models in Chapters 2 and 3) being suitable for the species. The final BBN diagrams can be found in section 4.3.7 and notes about their construction are provided in Appendix 25.

4.2.3. Selection of and contact with experts

Selection of experts was carried out by referring to the sources of the literature on each of the species, particularly the literature directly relevant to the New Forest and by contacting the local experts and organisations that initially provided species records for this project. Taxonomic specialists, natural historians, and field ecologists with direct field experience in the New Forest were targeted for inclusion. Selection was aimed at experts who had knowledge of the species in the New Forest, if possible. Selected experts were also asked to suggest other experts in their field. The only problem with recruitment of experts through other experts is that the experts may only recommend other experts with similar opinions or views to themselves (Renooij, 2001), which may bias the results. However, as this was not the only method used to select experts, and the knowledge of the species was likely to be similar anyway, it was judged unlikely to be a problem.

It has been suggested in the risk analysis literature (Clemen and Winkler, 1999) that experts who are very similar (in modelling style, philosophy, access to data and so on) tend to provide redundant information, so heterogeneity among experts is highly desirable. However, in this application, where different outcomes were not being investigated, it was likely that there would be some homogeneity between experts. It has also been found that the marginal utility of information decreases as the number of experts increases, and using 3 to 5 experts is suggested (Clemen and Winkler, 1999). Particularly in the case where the experts are likely to have similar knowledge, it is unnecessary to have a large number of experts, which would also be very time-consuming to interview. Further, for some poorly known species there may not be many experts available. However, it is still beneficial to include more than one expert per species as a set of experts can provide more information than a single expert (Clemen and Winkler, 1999). It was therefore decided to use a minimum of three experts for each species, although initial contact was made to more than this number as several people were unable to help (owing to time commitments) or did not respond to the request for assistance.

When selecting experts the following criteria were borne in mind: tangible evidence of expertise; reputation; availability and willingness to participate; understanding of the species' habitat requirements; and impartiality (Hora and VonWinterfeldt, 1997;

O'Hagan *et al.*, 2006). Letters or emails were sent to the selected experts explaining the study and asking whether they would be interested in participating. A small sum of money was offered as a contribution to cover the experts' time, to encourage participation. Experts were also offered the option to remain anonymous in the thesis or any resulting publications.

Experts were asked to take part in both the development of the conceptual models and elicitation of probabilities for the CPTs as Renooij (2001) suggests that when eliciting probabilities for belief networks, it is best to select an expert who has also been involved in building the structure for the network, to prevent errors due to the possible existence of different definitions for certain variables. A list of the experts that took part in the development of the BBNs can be found in Appendix 21.

4.2.4. Assessment of potential bias

A number of different types of bias may be present when eliciting expert judgements, which require analysis. However, studying or trying to counter bias is complicated by not having a readily available reference point from which to determine the direction and magnitude of bias; for most questions asked of experts, there are no known single correct answers or empirical data (Meyer and Booker, 1991). Meyer and Booker (1991) suggest a program for handling bias in which the first step is to anticipate which biases are likely to occur in the planned elicitation and then to redesign the planned elicitation to make it less prone to the anticipated biases. They also suggest that the elicitation be monitored by the data gatherer for the occurrence of bias. Further, they recommend that the data gatherer should study the signs that warn of possible bias and watch for these and then adjust, in real time, to counter the occurrence of bias, such as stopping for a break if the expert appears to becoming fatigued. An assessment of the main sources of bias likely to occur in the expert elicitations and how to minimise them is shown in Appendix 22.

4.2.5. Pilot interview

As suggested by Meyer and Booker (1991), a pilot interview with a *Poronia punctata* expert (A. Newton) was carried out to determine whether the chosen procedure was suitable (such as the experts' understanding of the questions and use of the response mode), to identify any potential problems with the procedure and to provide an indication of how long the elicitation would take. The procedure followed was as planned for the elicitation with experts (see section 4.2.6 below). Although Meyer and Booker (1991) advise against using the advisory experts in the pilot sample as they will not approach the test materials from a fresh perspective if they helped develop them, it was decided as the procedure was not altered, it would be a wasted opportunity not to make use of the knowledge of the expert.

4.2.6. First consultation with experts: discussion of relevant variables and suitable states

The individual interview was chosen for expert elicitation in both stages of BBN model development because it allows for a richer, more in-depth discussion and areas of uncertainty or new knowledge can be more easily elaborated on and probed further (see section 4.1.3.1). In addition, many experts are extremely busy and it would be difficult to arrange for them all to meet at the same convenient time and location on two different occasions. Further, group meetings would likely take longer for the experts than individual meetings due to the interaction and discussion between experts (O'Hagan *et al.*, 2006), although individual meetings take longer for the facilitator. The data from the interviews can easily be combined through mathematical aggregation afterwards. Individuals meetings also avoid potential bias from group dynamics, although this is at the expense of the potential generation of new ideas through expert interactions.

Discussion with experts was carried out at informal, face-to-face meetings, except for discussions with two experts which were carried out over the telephone, owing to distance and time constraints. Although telephone discussions are generally not considered as good as face-to-face discussions (Meyer and Booker, 1991), it was more important to ensure meetings with the experts and for the first stage this was not as much of a problem.

The initial meetings with the experts were mainly to identify which variables they thought were most important in determining habitat suitability for the selected species. An important aspect of this was deciding on suitable states for each of the nodes. Discretisation is an important issue in the development of BBNs, and as information about suitable states was not generally something available in the literature, the consultation with experts was key for choosing appropriate states. This also meant that the choice of states reflected those which would be most useful and useable for the people involved in surveying the species.

The project was explained to the experts and several points were clarified before proceeding. It was explained that the information from all of the experts would be pooled together to developed a ‘best’ model of current knowledge. It was also made clear that the assessments provided did not necessarily need to be ‘correct’, but just accurate in the sense that they should represent the (current) knowledge and judgement of the experts: there are no right and wrong answers (Renooij, 2001). In addition, the point was made that the models do not even need to be ‘complete’, in the sense of including every important factor or link, as long as they work and are easy to use, and that the models can easily be updated with new knowledge at a later date.

The experts were also asked to provide as much detail as possible on their reasoning (justification, evidence, assumptions) for providing the given answers and asked to try to ‘think aloud’ as much as possible. They were also asked to report if they were not sure of an answer and were given the option to provide a range of values in which they thought the most likely probability would lie, if they were particularly unsure. They were also informed that they could ask questions at any time, particularly if something was not clear or required recapping and that they could take a break at any time.

For each variable suggested, where appropriate, the following themes were discussed:

- A description of the variable, such as quantity;
- How the variable might be measured;
- Why the variable is important;
- The minimum, maximum and ideal tolerances of the species for that variable and the difference in suitability within that range (in order to help determine suitable states);
- What factors might affect that variable.

Experience of the species gained during fieldwork carried out as part of this project, and knowledge from the literature helped to facilitate discussion about the species and important variables. If experts did not mention a variable that had already been identified in the literature, then that variable was suggested to the expert for their comments and elaboration. This was also used, where necessary, to prompt suggestion of other variables. Experts were monitored for signs of bias, such as inconsistency (as described in section 4.3.4) and where appropriate, measures were taken to reduce it, such as asking a question in a different way, or asking for further clarification.

4.2.7. Adjustment of network diagrams and assignment of states

Suitable states were assigned, and where appropriate, adjustments to the influence diagrams were made, based on the pooled information derived from the expert interviews. Nodes and states were kept as simple as possible, so that they were easy to use and assign, but still had the desired precision of estimates (i.e. were still useful). Marcot *et al*. (2006) note that this is a balance between parsimony and precision, but that increased precision is not necessarily equivalent to increased accuracy. The final structures are shown in section 4.3.7.

4.2.8. Creation of additional information/survey sheets to accompany BBNs

Information obtained from the literature and from the experts was used to create additional information sheets to accompany the BBNs (see Appendix 26) so that they were clear to use. These included a definition of the variable/node, why it was important for habitat suitability, definitions of each of the states and how to measure the variable. This could then be used to explain exactly what was meant by each of the variables/nodes and their states so that experts could assign the probabilities to the CPTs. Additionally it provided details on how to survey sites in the field to instantiate the BBN model for the appropriate species to assess the potential habitat suitability of that site for that species. Providing such information also meant that the surveys could be carried out by others and would not be limited to those who were familiar with the model. The ways of measuring the variables were kept as straightforward as possible so that the surveys could be carried out quickly and without a specialist. This was influenced by experiences from fieldwork carried out as part of this project.

4.2.9. Second consultation with experts: probability elicitation for CPTs

Information about habitat suitability in the form of probabilities of each of the states was not available in the literature (and rarely is (Newton, 2009a)), so was based entirely on expert knowledge. I also filled in CPTs as an expert, based on my ‘expert’ knowledge, based on the literature and my fieldwork experiences. This was carried out before probability elicitation from the experts so that I was not influenced by their answers. It also provided a way of checking that the models made sense (particularly in terms of appropriate links), before presenting them to the experts.

Experts were all met at face-to-face interviews for the probability elicitations as filling in the CPTs would be easier in this situation and inconsistencies could be more easily identified. It was also easier to have a print-out of the (blank) CPTs for each node, which could be filled in with the expert. Experts were reminded again of the points outlined in section 4.2.6 (not necessarily needing to provide ‘correct’ answers, ‘thinking aloud’ etc.). A simple example was initially used to explain the procedure to the experts. They were also informed that the columns of the tables should sum to 1, but the rows did not have to, and that cells could be denoted as representing impossible combinations (Marcot *et al*., 2006).

It has been reported that people often find it difficult to provide their answers in numerical form, i.e. probability values (Garthwaite *et al.*, 2005; see section 4.1.3.4). It was therefore decided to aid the elicitation of probabilities using a probability scale with both verbal and numerical anchors, based on a scale developed by Renooij and Witteman (1999) and used by van der Gaag *et al.* (1999), Renooij (2001) and Witteman and Renooij (2003). Experts were provided with print-outs of this probability scale, which they could use as much or a little as they preferred. A copy of the scale can be found in Appendix 24 and details of its use and modifications made to the scale are provided in Appendix 23.

Elicitation of probabilities involved asking questions such as, “What is the probability that variable A takes state X, if Y is ... and Z is ...?” (after Pollino *et al.* (2007)). Experts were asked to provide ‘best estimate’ probability values, preferably to the nearest 0.05, as it would be very difficult to be more accurate than that and smaller increments may become meaningless. The additional information/survey sheets (from section 4.2.8) were used to clarify the meaning of the nodes and states. Additional information, such as definitions of the habitat types and soil types were also provided.

Experts were told that if they were unsure of a value they could give a range of possible values, as it has been suggested that many experts prefer to give a range of possible values instead of a single point estimate to reflect their uncertainty in providing a single value (Meyer and Booker, 1991). Pollino *et al*. (2007) also suggest that eliciting a range of values can be used to later identify other parameters needing attention. This option was not offered for every parameter as it was identified in the pilot interview that eliciting too many values may over-complicate the process, and experts generally found it relatively easy to provide a single best estimate value. In addition, ranges would be more difficult to aggregate. If experts preferred to provide a single value, but were less confident than for some of the other values, this was recorded.

The probabilities for the most suitable (e.g. 1) and the least suitable scenarios were generally assigned first for each node and the rest filled in from there. This was based on a suggestion by Marcot *et al*. (2006) of setting all the CPTs to a uniform value then 'pegging the corners' by setting the extreme cases to 0 or 1, adjusting the middle or most moderate case and then back-interpolating the other entries. A similar approach was used by Smith *et al*. (2007) who elicited properties for selected scenarios first: (a) the best-case scenario where all of the parent nodes (input variables) are in the best state; (b) the worst-case scenario where all of the parent nodes are in the worst state, and; (c) scenarios where only one parent node is not in the best state. These scenarios were then used as reference points for eliciting probabilities for the remaining scenarios in a conditional probability table. As suggested by Marcot *et al*. (2006), the values entered in the CPTs were cross-checked by scanning down each column and checking whether the entries with the highest (and lowest) values really represented the most (and least) likely causal conditions for that state. Experts were also monitored for signs of bias as described in section 4.3.4.

4.2.10. Aggregation of expert probability results

As more than one expert provided probability values for the CPTs for each of the models, the values had to be combined. Mathematical aggregation was required due to the individual interviews method of elicitation chosen. The unweighted mean was used as a simple aggregation estimator, as simple methods have been suggested to work well in comparison with more complex methods (Meyer and Booker, 1991; Clemen and Winkler, 1999; O'Hagan *et al.*, 2006) and it avoided the difficulty of rating experts. For example, experts may be more confident about certain aspects of the model than others, and they may only be unsure about a few values, such as those at the extremes of the tolerances of the species where experts are less likely to have encountered them. Further, it is difficult to weight experts in an unbiased way, as it would be tempting to rate those experts higher whose opinions appeared to fit more closely with personal expectations. On the rare occasion where an expert had been particularly unsure about a value for a CPT or had left it blank, these were excluded from the aggregation. In this way, as suggested by Martin *et al*. (2005), the expert's ability (i.e. a weighting) was partly contained within the data, since experts did not provide values if they were uncertain. The aggregated values are displayed in the CPTs in Appendix 27.

4.2.11. Evaluation of BBNs

Evaluation of the BBNs is an important part of testing their ability to predict potential habitat suitability. Several evaluation measures were implemented, as described below.

4.2.11.1. Model checking

The models were checked for consistency, in particular the CPT values were checked to make sure they made ecological sense and reflected the qualitative discussions from the experts (i.e. the information from the first meetings) and, where available, the information from the literature. For example, if it was known that a site would become less suitable the greater the canopy cover, then the values in the CPTs were checked that they reflected that. Checking for consistency of answers was also carried out during the expert elicitation (sections 4.2.6 and 4.2.9). Coherence was checked for during the elicitation by checking whether all probabilities summed to 1 (Renooij, 2001). Although the reliability of elicited probabilities can be checked by re-eliciting the probabilities from the experts or testing whether the experts agree with their own assessments (Renooij, 2001), this was not carried out due to the large number of probabilities to be assessed and the amount of time required. The model structures were also checked (before probability elicitation) by another subject-matter expert, as suggested by Marcot *et al*. (2006).

4.2.11.2. Model evaluation using field survey data

In order to evaluate the BBNs (i.e. to test whether they realistically predicted the potential suitability of sites for the species), data were collected on the variables for each of the species to instantiate the BBNs. An initial field survey was carried out during May to October of 2008 recording relevant habitat suitability variables for each species, determined from the literature, and as incorporated in the BBNs. This generally involved measuring variables using continuous scales (e.g. the depth of deciduous leaf litter in centimetres), where appropriate, rather than categorically (so that states could be assigned retrospectively). A second field survey was carried out during June and July of 2009 to record any additional variables that had arisen from discussion with experts, or to assign states that had not been accounted for, at the same sites as the 2008 fieldwork.

Twenty recent known sites (each 1 ha in area) for each of the species were visited to record variables for the presence sites (where the species could still be found at the time of the survey). The sites with the most recent records of the species were selected, as these were more likely to be currently suitable for the species. Where neighbouring sites were identified, only one was visited. A hand-held GPS was used to locate the centre of the site so that the area of the extent of the site could be established. Survey sheets of the habitat suitability variables were completed by walking throughout the site to obtain good coverage. Searches for all of the selected species were carried out whilst visiting each site and presences or absences recorded as well as recording of variables for each species. Approximately 45 minutes to 1 hour was spent at each site, on average.

Absences for testing the BBNs were also obtained from the field survey. Whilst visiting known presence sites for the study species, absences were recorded for any of the non-focus species if it could be determined with some certainty that the species was absent from the site. Absences were only obtained from surveys that were carried out during the time when the species should be present (e.g. flight time for the butterflies), because as the fieldwork was carried out over a period of six months, most of the species would only occur during a restricted length of time within that. Some sites were revisited at an appropriate time of year to confirm presences or absences. However, the majority of sites were only visited once, and ideally they should be visited several times (one of the problems with obtaining absence data), but nevertheless, due to the measures taken as described above, they were likely to be true absences.

The absence sites were not random locations (due to time constraints, which prohibited survey of further sites), because they were determined by visiting a known location for another of the study species, and were therefore subject to a degree of bias. Where more than 20 absence records were obtained for a species, sites were selected by which were likely to be the most certain of absence (e.g. suitable weather conditions at the time of the survey for the butterfly species). A total of 118 sites were surveyed (as some sites were used as presence or absence sites for more than one species).

BBNs for each species were instantiated using the results from the field surveys and run for each of the 20 presence locations and 20 absence locations for each species. Where sites contained more than one state of a variable (for example habitat type), the percentage cover of each state was entered as a likelihood value in the model.

As suggested by Marcot *et al*. (2006) the classification power of the BBNs were depicted using a ROC curve. The AUC (Area Under the ROC Curve) value was also calculated. However, as discussed in Appendix 8.2, it is useful to provide a range of measures of model performance. Therefore, the ROC_AUC program (Schröder, 2006) was used for calculation of a range of evaluation measures, as applied for the models in Chapter 3 (see section 3.2.4).

4.2.11.3. Sensitivity analysis

Some of the models were run with different nodes instantiated (i.e. at different levels of the model (in terms of input nodes or just their child nodes) in the cases of *C.*

nobile, *H. semele*, *P. argus* and *P. globulifera*), or with different CPT values where there was a clear difference in opinion between different experts (for *P. punctata*). Details of these models are provided in section 4.3.11.2. As the models were run they were checked that they were exhibiting realistic behaviour (as suggested by Marcot *et al*. (2006)).

4.3. Results

4.3.1. Identification of relevant variables

For some species there was quite a lot of information available in the literature on habitat requirements, such as for *C. nobile*, *P. globulifera* and the two butterfly species. In the case of *G. illyricus*, a detailed report on its habitat requirements in the New Forest was available (Stokes, 1987). However, there was very little detailed information available for *G. constrictum*. For *N. sylvestris*, there was an in-depth thesis (Brouwers, 2008) available on its requirements on the Isle of Wight (although nothing specific to the New Forest), so this was the main literature source for the development of the BBN for this species. It was assumed that the requirements would be largely similar as the locations are close geographically. Although there were several reports about *P. punctata*, its detailed ecological requirements are still relatively unknown. Tables of relevant variables for each species can be found in Appendix 20.

4.3.2. Development of network diagrams based on the literature

Initial influence diagrams were created based on information from the literature. In general there was sufficient information available to develop complete diagrams, except for *G. constrictum*. The most difficult part was deciding on states for the variables/nodes, as information to support this was not generally available, so this particularly required incorporation of expert knowledge (see section 4.2.6).

4.3.3. Selection of and contact with experts

Several of the experts that were contacted did not respond, or responded but were unable to help, for example owing to time constraints. However, at least three experts were able to help with each of the species. A list of the experts who contributed to the development of the BBN models can be found in Appendix 21.

4.3.4. Assessment of bias

An assessment of the main sources of bias likely to occur in the expert elicitations and how to minimise them are shown in Appendix 22. From this assessment there were several approaches to reducing bias that frequently arose, which were incorporated into the elicitation procedure. The main one was asking the experts to ‘think aloud’ as much as possible, verbalising their thoughts and answers and explaining (justifying) their answers in detail. Another important aspect was that the questions posed to the experts were made as clear, unambiguous and simple as possible (for example, not asking too much at once, by breaking the question up) and were not leading. Structuring of the questions so that they flowed from the general to the specific and in a logical manner was also found to help.

Recapping of definitions, assumptions and background to the questions and previous answers to related questions was also deemed beneficial as well as generally monitoring for signs of bias, in particular inconsistencies or contradictions or signs of anchoring. Asking questions in a different way also helped to clarify inconsistencies. Finally it was noted that the interview should not last for more than two hours and a break should be provided if required. These points were important to follow, whether accounting for bias or not, as they help to improve the quality of the experts’ answers.

One aspect of the program for handling bias suggested by Meyer and Booker (1991) and other authors (including O’Hagan *et al.* (2006)) is to make the experts aware of the potential intrusion of particular biases (particularly the definitions and causes of these biases) and familiarise them with the elicitation procedure. However, after discussion during the pilot interview (see section 4.3.5), it was decided that this may cause the experts to feel that they could not provide unbiased answers and they may therefore be less willing to participate. Therefore, this approach was not pursued.

4.3.5. Pilot interview

The main conclusions of the pilot interview were that it was best not to show the expert the preliminary BBN structure in the first consultation as they may become too focused on the structure, rather than the identification of important variables, and would also take more time. Another conclusion was that asking for upper and lower values (i.e. a range) or probability values might over-complicate the probability elicitation. Instead, a ‘best guess’ should be given, and only a range of values provided if the expert was particularly unsure about what probability value to assign. This would also give an indication of the level of uncertainty. An additional result of the pilot interview was the decision not to make the experts aware of all the potential biases (see section 4.3.4). Apart from these points, the elicitation procedure and modifications of the probability scale were deemed appropriate.

4.3.6. First consultation with experts: discussion of relevant variables and suitable states

The experts generally did not suggest any additional variables (apart from bare ground/disturbance and water pH for *G. constrictum*, ground moisture for *N. sylvestris*, and speed of water flow for *P. globulifera*), but were able to provide more detail on the variables and therefore the potential states. The additional information derived from the expert interviews was added to the tables of variables in Appendix 20. No contradictions in the relevant variables arose, and in general, contradictions were also not encountered in the potential range of suitability and states, although some experts were able to provide more detail than others for certain variables. The experts generally knew the optimum level of the variable in which the species most often occurred, but did not necessarily know the limits of the species tolerance, as these were less frequently encountered.

4.3.7. Adjustment of network diagrams and assignment of states: final structures

After discussion with experts, the structures for the influence diagrams were finalised and are shown in Figures 25 to 32 below. The most difficult aspect of developing the models was deciding on appropriate states for each of the nodes. The states were based on a pooling of knowledge from the literature, expert discussion and experience from fieldwork to produce categories that were relatively easy to assign in a survey but detailed enough to be useful and discriminatory. The number of states was kept to five or fewer, in line with the recommendations of Marcot *et al*. (2006), to keep the associated CPT tractable and understandable. Where the same variables were used for more than one species, the same states were used for each species (as long as this was sensible), for example for ground moisture, so that variables could be surveyed more easily for several species at the same time.

The states used for habitat type and soil type were based on the GIS habitat layers used in the initial statistical models (chapters 2 and 3), so that the models could be more easily used in conjunction with each other. In addition, by using this approach, the soil type of a site could be established using the GIS layer, rather than having to take samples, which would be more resource-intensive.

Figures of the final BBN structures are shown in Figures 25 to 32 below, and their CPTs can be found in Appendix 27. Notes of the development of the models can be found in Appendix 25 and accompanying sheets describing the variables and their states (section 4.2.8) can be found in Appendix 26.

The models for *C. nobile* and *H. semele* include rectangular ‘decision’ nodes to indicate the importance and influence of those variables, which relate to management interventions. However, as currently configured, the decision nodes do not have an effect on the values of their child nodes. They could potentially be incorporated as ‘discrete chance’ nodes in the same way as the other nodes, but it was felt that it would be better (and more accurate) to just instantiate the child nodes (‘bare ground’ and ‘ground vegetation height’ for *C. nobile* and ‘pioneer heath’ and ‘bare ground’ for *H. semele*).

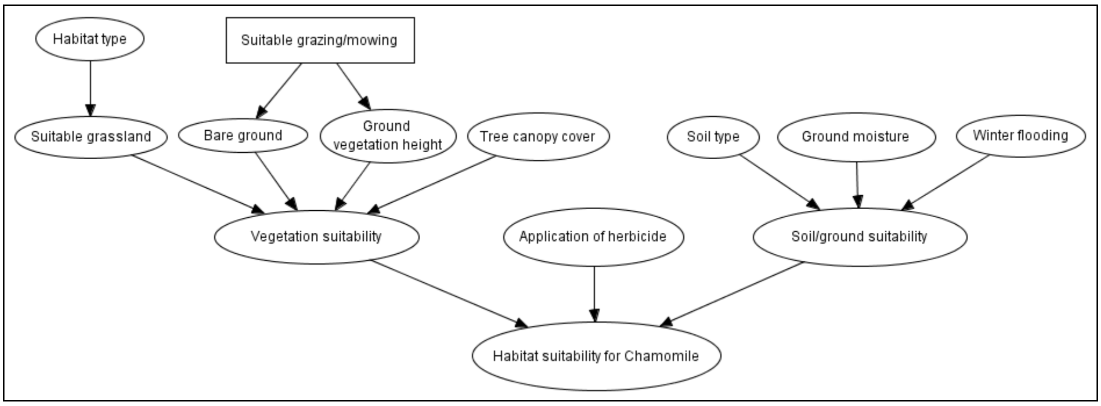


Figure 25. BBN network structure for *C. nobile*.

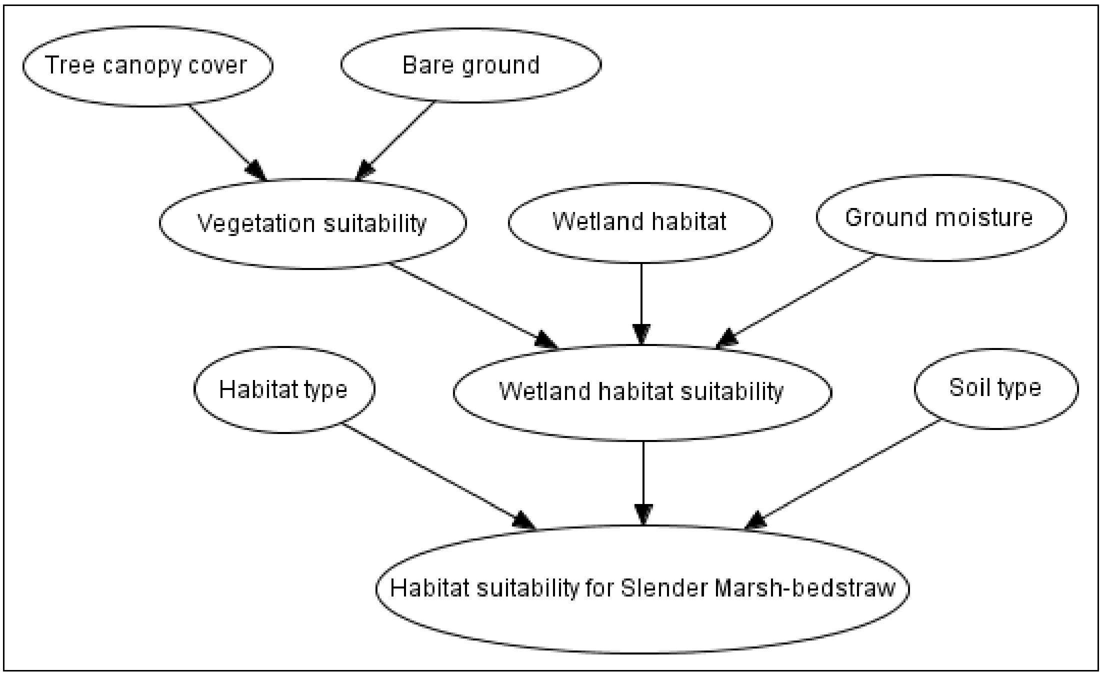


Figure 26. BBN network structure for *G. constrictum*.

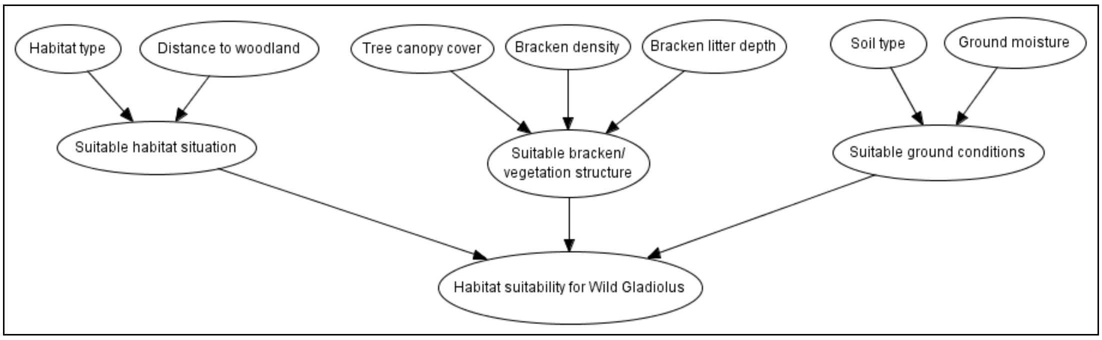


Figure 27. BBN network structure for *G. illyricus*.

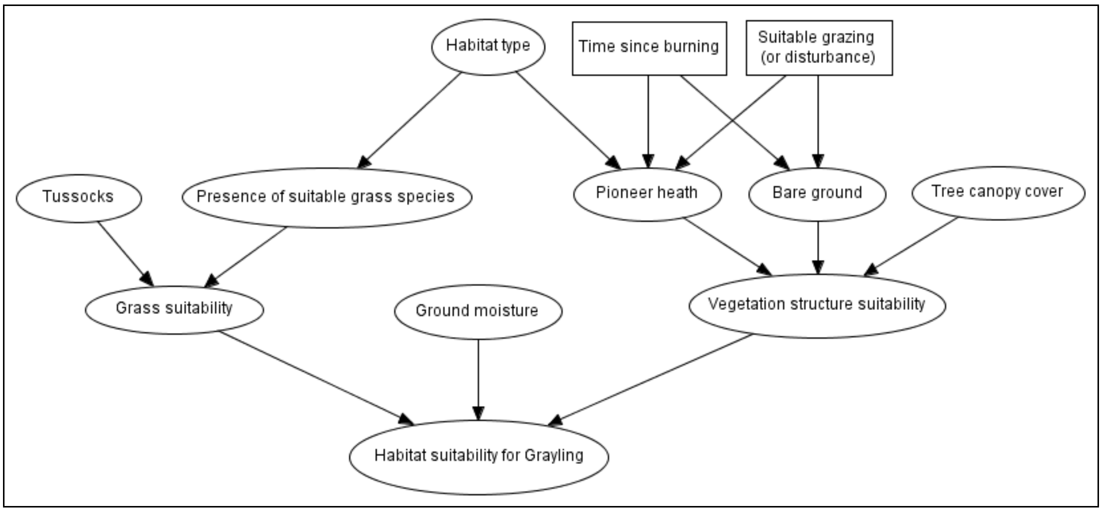


Figure 28. BBN network structure for *H. semele*.

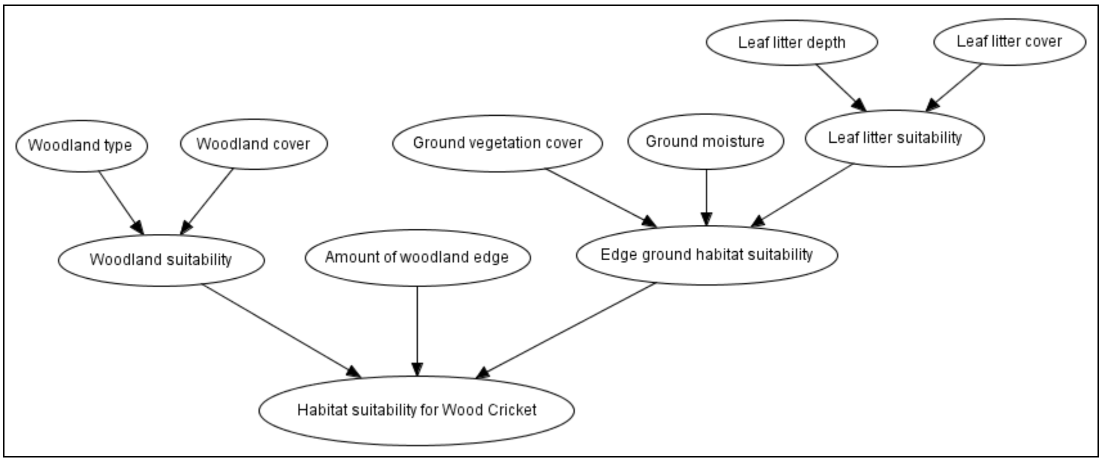


Figure 29. BBN network structure for *N. sylvestris*.

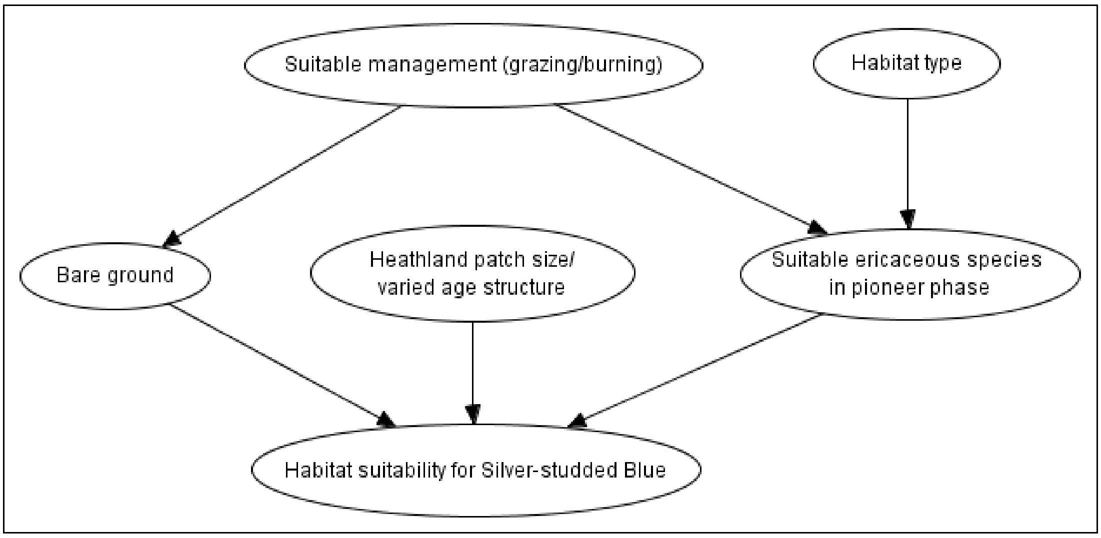


Figure 30. BBN network structure for *P. argus*.

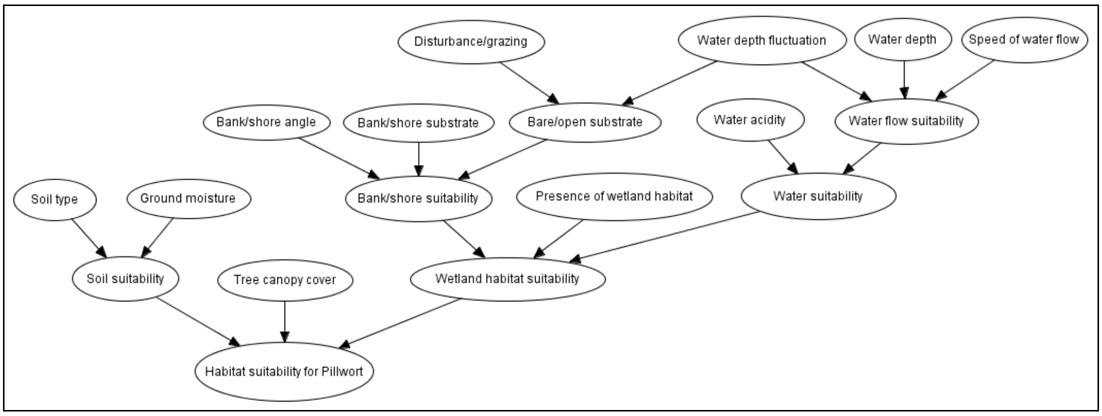


Figure 31. BBN network structure for *P. globulifera*.

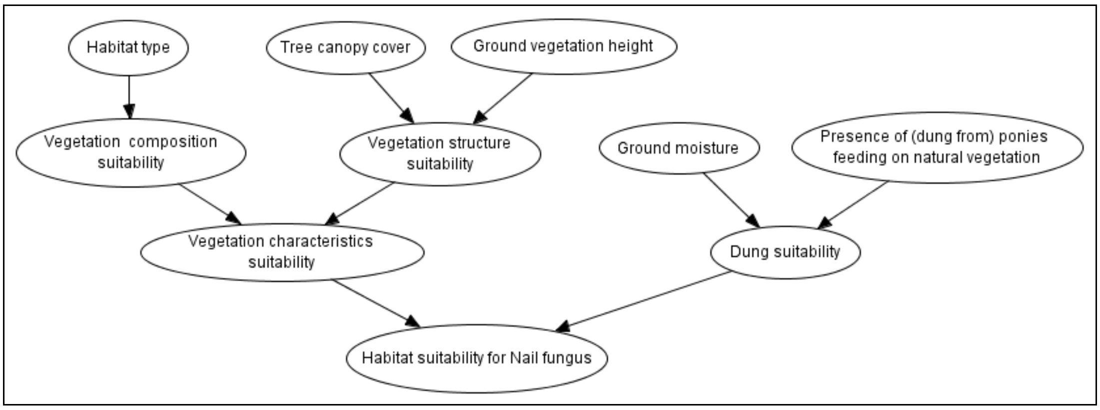


Figure 32. BBN network structure for *P. punctata*.

4.3.8. Creation of additional information/survey sheets to accompany BBNs

The information sheets developed to accompany the BBN models for each species can be found in Appendix 26.

4.3.9. Second consultation with experts: probability elicitation for CPTs

Experts generally found it relatively easy to complete the CPTs with single values. Only three of the experts chose to use the probability scale and they did not use it for every value, but just as a guide for the first CPT. One expert commented that he found it easier to think in probabilities, rather than in verbal terms such as likely and unlikely, because they more precise and it was easier to make steps from the highest probabilities to decreasing probability values with a drop in suitability for a less suitable combination. All of the experts picked up the procedure very quickly after the initial explanation. Although experts were offered the option of providing a range of values, rather than an individual ‘best guess’ value for each cell in the CPTs, none of them did this and preferred to just comment if they were unsure and leave those cells blank. The aggregated results of the expert probability elicitations can be found in Appendix 27.

4.3.10. Aggregation of expert probability results

The CPTs showing the final aggregated results can be found in Appendix 27. In one of the CPTs for the *P. punctata* BBN there was a particular difference of opinion over which states were the most suitable, assigning the highest probability to different combinations of states. The experts differed over whether they thought that dry or moist soil was more suitable, when filling in the ‘dung suitability’ CPT. This meant that when the results were aggregated, none of the states were assigned a probability value greater than 0.86 (see Appendix 27). As it was not a single expert whose opinion differed and none of the experts expressed uncertainty over this aspect of the model, the model was run with the CPT based on an average of only the values provided by the experts who thought that dry soil was more suitable, and also run with only the values provided by the experts who thought that moist soil was more suitable. (The experts who thought that moist soil was more suitable also thought that wet could be quite suitable (average 0.58), whereas the experts who thought that dry soil was more suitable also thought that wet soil was unsuitable (average 0.03)). This also formed part of the sensitivity analysis (see section 4.3.11.3). The model was also run using an average of the values from all of the experts.

There was also a slight difference of opinion in the *P. globulifera* ‘soil suitability’ CPT over which soil type was the most suitable, but the experts rated two of the soil types highly and did not assign a large drop in value for the soil type that they thought was the second most suitable, so the highest value was still very high at 0.97. Therefore, this model was not run with different experts’ values. The possible reasons for differences in opinion between experts are discussed in section 4.4.

4.3.11. Evaluation of BBNs

4.3.11.1. Model checking

Exploration of each of the models confirmed that their general behaviour was intuitive and conformed with expectations based on the literature review, and the knowledge provided by experts. The structures of the models were also approved by an experienced BBN modeller, as described in section 4.2.2.

4.3.11.2. Model evaluation using field survey data

Predicted habitat suitability values for each site for each of the species' models can be found in Appendix 28 and the averages and range of suitability values are shown in Table 37 below. AUC values and other evaluation measures can be found in Tables 38 and 39 below. Some of the species have more than one set of results. This is because the models were run with different variables/nodes instantiated (i.e. at nodes of different levels of the models, where the model allowed for this), or where they were run with different CPT values owing to differences in experts opinion of which states were the most suitable (see section 4.3.10). These were:

C. nobile

C. nobile – 1: With 'suitable grass' instantiated.

C. nobile – 2: Without 'suitable grass' instantiated, but 'habitat type' instantiated.

H. semele

H. semele – 1: With 'presence of suitable grass species' instantiated.

H. semele – 2: Without 'presence of suitable grass species' instantiated, but 'habitat type' (and 'pioneer heath') instantiated.

P. argus

P. argus – 1: With 'bare ground' and 'suitable ericaceous species in pioneer phase' instantiated.

P. argus – 2: Without 'bare ground' and 'suitable ericaceous species in pioneer phase' instantiated, but with 'suitable grazing' and 'habitat type' instantiated.

P. globulifera

P. globulifera – 1: With ‘bare/open substrate’ instantiated.

P. globulifera – 2: Without ‘bare/open substrate’ instantiated, but with ‘disturbance/ grazing’ (and ‘water depth fluctuation’) instantiated.

P. punctata

P. punctata – all: With the ‘dung suitability’ CPT values based on all of the experts values.

P. punctata – dry: With the ‘dung suitability’ CPT values based only on the values provided by the experts who rated dry ground as more suitable.

P. punctata – moist: With the ‘dung suitability’ CPT values based only on the values provided by the experts who rated moist ground as more suitable.

Where there was a difference in opinion of experts (as discussed above) it was not possible to get a maximum overall habitat suitability of 100. This was the case for the *G. illyricus* model, where the highest overall value that could be obtained was 63.06 (with the highest value in the ‘suitable habitat situation’ CPT of 0.8 (for habitat type of bracken or unimproved acidic grassland) and the highest value in the ‘suitable ground conditions’ CPT of 0.72 (for soil type 64303 or 71107)). The highest overall habitat suitability obtainable for the *P. globulifera* models was 97.69 due to the highest value in the ‘soil suitability’ CPT of 0.97. For the ‘*P. punctata* – all’ model the highest overall habitat suitability that could be obtained was 86.28 (due to the highest value in the ‘dung suitability’ CPT of 0.86), but with both of the other *P. punctata* models it was possible to get an overall habitat suitability of 100.

Species	**Presence sites**		**Absence sites**	
	Average	**Range**	**Average**	**Range**
C. nobile – 1	59.62	7.08 – 97.84	17.52	0.45 – 49.56
C. nobile – 2	37.18	1.62 – 82.02	8.55	0.45 – 37.53
G. constrictum	41.55	11.40 – 92.62	9.62	0 – 62.00
G. illyricus	25.80	2.66 – 60.33	9.20	0 – 42.00
H. semele – 1	61.59	36.20 – 100.00	11.07	0 – 65.00
H. semele – 2	53.53	30.58 – 100.00	6.51	0.02 – 21.60
N. sylvestris	53.59	2.28 – 84.60	4.38	0 – 22.36
P. argus – 1	95.56	26.17 – 100.00	11.43	0 – 80.00
P. argus – 2	71.17	29.86 – 100.00	20.94	0 – 76.64
P. globulifera – 1	57.24	14.34 – 89.22	5.42	0 – 50. 72
P. globulifera – 2	57.24	14.34 – 89.22	9.52	0 – 50. 72
P. punctata – all	61.01	32.10 – 74.60	45.05	15.25 – 67.66
P. punctata – dry	77.42	34.08 – 98.81	44.88	9.12 – 81.44
P. punctata – moist	53.66	32.50 – 65.29	46.58	19.03 – 74.19

Table 37. Average (mean) and range of values determined by the BBN models of probability of habitat being suitable for focus species for known presence sites and known absences sites from fieldwork data. Values presented are percentages, representing likelihood of the habitat being suitable.

As shown in Table 37, the *P. argus* (1) model with ‘bare ground’ and ‘suitable ericaceous species in pioneer phase’ instantiated had the highest average value for the presence sites, as well as a fairly low average for the absence sites, giving it a very high AUC value of 0.999 (see Table 38 below). The *P. argus* (2) model was not as discriminating (as shown by the lower, although still very high, AUC value of 0.91; see Table 38 below), with higher averages for both the presence and absence sites, suggesting that it is better to instantiate the bare ground and pioneer heath nodes. All of the sites had suitable management so ‘bare ground’ was always present in terms of the *P. argus* (2) model. However, the habitat types differed, so the suitability values varied for the ‘suitable ericaceous species in pioneer phase’ variable. Instantiating ‘suitable management’ and ‘habitat type’ generally gave a more conservative overall habitat suitability value for the presence sites and either the same or a higher value for the absence sites. Interestingly, only two of the presence sites were composed of mainly wet heath, which was identified as the most suitable habitat type in the CPT for ‘suitable ericaceous species in pioneer phase’, which was the main reason why the values for the presence sites for the *P. argus* (2) model were lower.

The model with the next highest presence site average was the *P. puncata* model based on the average values for the 'dung suitability' CPT of just the values from the experts that thought dry soil was more suitable than moist soil. The overall habitat suitability values for the presence sites were all higher for the 'dry' model than the model with all of the experts' values combined and approximately half of the absence sites values were higher and half were lower. In comparison, apart from four sites, the values for the presence sites for the 'moist' model averages for the presence sites were lower than the averages for the model using all of the experts' values for the 'dung suitability' CPT, and all of the values were lower than for the 'dry' model.

The absence site values for the 'moist' averages were higher than the averages for the 'all' model, except for six sites. Therefore, the model based on the assumption that dry sites are better than moist sites performed better and showed more of a difference in habitat suitability values between presence sites and absent sites (as confirmed by the higher AUC value; see Table 38 below). This suggests that if by chance, only the 'moist better' experts had participated in the model building then the model would not have been as effective. However, it could just be that the sites used to test the models happened to be drier sites and were not representative of the range of sites in which *P. punctata* occurs. In addition, the surveys were carried out during the summer and autumn when the ground would generally not have been as wet as when experts may normally record the species (although the ground moisture was supposed to be an average moisture throughout the year).

None of the absence sites for any of the *P. punctata* models had an overall habitat suitability value of 0 because all of the sites had dung present and contained a habitat type that did not have a suitability value of 0. The values for the absence sites were therefore generally quite high, as shown by the high average values for the absence sites. It is likely that there are not that many places in the New Forest (outside of urban areas) which are completely unsuitable (apart from in some closed canopy woodlands), as long as pony dung is present, which is why it is hard for a model to discriminate between presence and absence sites.

The *C. nobile* (1) model also had quite a high average for the absence sites, although the highest value for the absence sites was much lower (approximately half) than the highest value for the presence sites. Some of the presence sites for *C. nobile* (both models) had very low overall habitat suitability values, suggesting that *C. nobile* can occur in habitats that seem less suitable (although the lowest value of 7.08/1.62 was for a site with no bare ground and 'other' soil type). This made it more difficult for the model to discriminate between presence and absence suitability and consequently the AUC value (see Table 38 below) was one of the lower ones, although still high.

The values for the *C. nobile* (2) model without 'suitable grassland' instantiated were mostly lower than (with a few the same as, and only two (for absence sites) higher than) the values for model 1 with 'suitable grassland' instantiated (and this model achieved a lower AUC value). This suggests that using 'habitat type' to estimate the likelihood of suitable grassland for *C. nobile* underestimated that likelihood. In some cases (such as for presence sites numbers 2 and 11) this was by a large amount. Therefore, where possible it would be advantageous to instantiate 'suitable grassland'. However, habitat type could still be used as a rough guide to indicate potentially suitable grassland (one of the most important variables) using GIS spatial data before visiting a site.

As with *C. nobile*, the *N. sylvestris* model also had low values for some of the presence sites, although only one really low value (2.28; this was for a site with no woodland present within the site, but just outside of the site). Therefore, the model may be improved by including distance to a larger woodland as a variable (distance to an occupied woodland was identified as a variable (see Table A-30, Appendix 20) but was not included; see Appendix 25). This shows how applying and testing the model with fieldwork data can help to identify the importance of particular variables. Low values may also be due to the ability of *N. sylvestris* to occur in habitats that are significantly less than optimal, particularly as it is so widely distributed within the New Forest (S. Douglas, personal observation), and it may be that the experts underestimated this. Although a site may have a low habitat suitability, it does not mean that the site is not tolerable or suitable to a certain extent, although *N. sylvestris* may be slightly less likely to occur there. However, the average for the absence sites was very low, as was the range, suggesting that the model is good at discriminating between potential presence and absence locations. This was confirmed by the very high AUC score of 0.963 (see Table 38 below). None of the presence sites achieved the highest possible suitability value, with the highest value of only 84.60. This was mainly because none of the sites had the most suitable states for all of the variables. In particular, only two of the sites had high leaf litter cover and only one had a leaf litter depth of more than 5cm. It may be that the classifications for the states need to be changed, or the CPT values changed. Or it may just be that the limited number of testing sites did not encompass the more suitable sites.

The average habitat suitability value for the presence sites for the *H. semele* (1) model (with ‘presence of suitable grass species’ instantiated) was only 61.59. The cause of low values for some of the presence sites was largely due to the absence of tussocks at these sites, which dropped the overall habitat suitability values from 100 to 40. It may therefore be that the presence of tussocks is not as important as envisaged from the CPT values and should be given a lower weighting in the ‘grass suitability’ CPT. However, the model appeared to provide good discriminatory ability (as confirmed by the very high AUC value of 0.965; see Table 38 below), with an average of 11.07 for the absence sites, although model 2 provided a very low absence sites average (6.51) and the range of overall habitat suitability values for the presence sites did not overlap with the range of values for the absence sites. This meant that it achieved the highest AUC score of 1, indicating perfect discrimination.

When ‘presence of suitable grass species’ was not instantiated (model 2) the values for the presence sites were either lower or stayed the same, showing that the use of habitat type to indicate suitable grassland provided a more conservative estimate of the likelihood of suitable grass species being present. Therefore, although the AUC value for this model was higher, it is better to instantiate presence of suitable grass species if possible. However, although slightly lower, the values are still in a similar range, so could be used to provide an indication of potentially suitable sites just by using habitat data, which is available as a GIS layer, so could be determined without visiting the site.

The next highest presence sites averages were for the *P. globulifera* models (which also had low average values for the absence sites). The values did not change for the presence sites whether ‘bare/open substrate’ was instantiated or not, and only four values differed for the absence sites. This suggests that ‘disturbance/grazing’ and ‘water depth fluctuation’ were good predictors of ‘bare/open substrate’. Although the *P. globulifera* BBN included a large number of variables in comparison to some of the other species’ models, the only things that differed between the presence sites were the states for ground moisture, soil type and tree canopy cover. The main reason for low values for the absence sites was absence of wetland habitat.

In terms of ‘ground moisture’, ‘wet’ was assigned the highest suitability value in the ‘soil suitability’ CPT, with ‘dry’ given quite low values (see Appendix 27). However, most of the presence sites were moist with a few dry sites but none of the sites were all wet, although four sites were half wet and half moist. This could suggest that sites do not need to be as wet in general as long as a suitable wetland habitat is present, or it could be that the testing sites under-represented ‘wet’ sites in which *P. globulifera* occurred (as they were only a limited sample of known locations). It may also be a factor that there are not as many records for *P. globulifera* in very wet sites, despite it being known to favour wet conditions, because such sites are more difficult to access. Although the lowest suitability values were generally given for soil type 84102, five of the presence sites were of that soil type, suggesting that experts may sometimes not be as familiar with soil type.

As for the *P. globulifera* model, the main reason for low values for the absence sites for *G. constrictum* was the absence of a wetland habitat. The average for the absence sites was substantially lower than for the presence sites, mainly due to this factor, but this shows that the model was capable of discriminating between more and less suitable sites (as confirmed by the high AUC value of 0.886; see Table 38 below). However, some of the values for the presence sites were quite low. This was partly because several of the presence sites occurred where there was woodland or where there was no bare ground, which were both assigned lower suitability values in the CPTs. *G. constrictum* was one the species for which there was the least knowledge, so it may just be that the experts were not as familiar with the species in these situations.

Even though the highest overall habitat suitability value that could be achieved for the *G. illyricus* model was 63.06, only one of the sites achieved a value close to that, with the majority scoring less than two thirds of that value. The low values for many of the presence sites were mainly due to the soil type. The soil type assigned the highest suitability was 64303 (see the ‘suitable ground conditions’ CPT, Appendix 27) but only four of the presence sites had this soil type. The next most suitable soil type was 71107 (which was the main soil type for eight of the presence sites) but the suitability dropped from 0.72 for 64303 to 0.55 for this soil type. One of the presence sites had the soil type 84102 which had very low suitability values. This is another example of how the soil type seemed to be the variable that experts were not as familiar with in terms of the values they assigned and the occurrence of the species on the soil type. It may be that the soil types were too broad and not discriminating enough for a species such as *G. illyricus*, which is believed to occur on a specific soil type (Stokes, 1987) and more detail may be required. The other factor that led to lower overall habitat suitability values was habitat type, as a lot of the presence sites were composed of a large proportion of dry heath, which had a highest suitability value of 0.58 in the ‘suitable habitat situation’ CPT.

Despite the relatively low values for the presence sites, the values for the absence sites were generally lower, showing a difference in average and the range of values for presence sites and absence sites, although the ranges overlapped considerably. Nonetheless, the model achieved an AUC value of 0.820 (see Table 38 below), demonstrating good discrimination. It may be difficult to create a model with outstanding discrimination because *G. illyricus* may not occur in apparently suitable habitats just because it is very rare, and this would likely become more apparent if more sites were used to test the model. It is also likely that *G. illyricus* occurs in several sites where the habitat is less suitable (such as sites with vigorous bracken density and deep bracken litter) because those sites have become unsuitable and the plant is just ‘hanging on’.

ROC plots (see Appendix 29) and corresponding AUC values were also generated for each of the models and the results, along with those for other measures of model discrimination ability, are shown in Tables 38 and 39 below. The measures used were the same as those used in Chapter 3 for evaluation of the GLM and GAM models (see Table 21, section 3.2.4 (Chapter 3) and Appendix 8.2 for further explanation).

Species	**AUC value**
C. nobile – 1	0.878
C. nobile – 2	0.846
G. constrictum	0.886
G. illyricus	0.820
H. semele – 1	0.965
H. semele – 2	1.00
N. sylvestris	0.963
P. argus – 1	0.999
P. argus – 2	0.914
P. globulifera – 1	0.965
P. globulifera – 2	0.933
P. punctata – all	0.831
P. punctata – dry	0.879
P. punctata – moist	0.684

Table 38. AUC values calculated from ROC plots for BBNs.

Species	***C. nobile - 1***	***C. nobile - 2***	***G. constrictum***	***G. illyricus***	***H. semele - 1***	***N. sylvestris***	***P. argus - 1***	***P. argus - 2***	***P. globulifera - 1***	***P. globulifera -2***	***P. punctata - all***	***P. punctata - dry***	***P. punctata - moist***
P_opt													
P_crit	0.490	0.213	0.083	0.025	0.333	0.255	0.550	0.413	0.103	0.115	0.560	0.750	0.513
Sensitivity	0.600	0.650	1.000	1.000	1.000	0.800	0.950	0.900	1.000	1.000	0.750	0.650	0.800
Specificity	0.900	0.950	0.800	0.600	0.900	1.000	0.950	0.650	0.900	0.800	0.850	0.950	0.600
False +ve	0.100	0.050	0.200	0.400	0.100	0.000	0.050	0.350	0.100	0.200	0.150	0.050	0.400
False -ve	0.400	0.350	0.000	0.000	0.000	0.200	0.950	0.100	0.000	0.000	0.250	0.350	0.200
CCR	0.750	0.800	0.900	0.800	0.950	0.900	0.950	0.775	0.950	0.900	0.800	0.800	0.700
Kappa	0.500	0.600	0.800	0.600	0.900	0.800	0.900	0.550	0.900	0.800	0.600	0.600	0.400
P_fair													
P_crit	0.285	0.130	0.238	0.125	0.365	0.175	0.533	0.448	0.318	0.348	0.555	0.625	0.520
Sensitivity	0.750	0.700	0.800	0.700	0.950	0.900	0.950	0.850	0.900	0.850	0.750	0.750	0.650
Specificity	0.750	0.750	0.800	0.700	0.900	0.900	0.950	0.750	0.900	0.800	0.800	0.800	0.650
False +ve	0.250	0.250	0.200	0.300	0.100	0.100	0.050	0.250	0.100	0.200	0.200	0.200	0.350
False -ve	0.250	0.300	0.200	0.300	0.050	0.100	0.050	0.150	0.100	0.150	0.250	0.250	0.350
CCR	0.750	0.725	0.800	0.700	0.930	0.900	0.950	0.800	0.900	0.825	0.775	0.775	0.650
Kappa	0.500	0.450	0.600	0.400	0.850	0.800	0.900	0.600	0.800	0.650	0.550	0.550	0.300
P_Kappa													
P_crit	0.490	0.210	0.083	0.025	0.333	0.218	0.548	0.413	0.103	0.115	0.560	0.750	0.513
Sensitivity	0.600	0.650	1.000	1.000	1.000	0.800	0.950	0.900	1.000	1.000	0.750	0.650	0.800
Specificity	0.900	0.950	0.800	0.600	0.900	0.950	0.950	0.650	0.900	0.800	0.850	0.950	0.600
False +ve	0.100	0.050	0.200	0.400	0.100	0.050	0.050	0.350	0.100	0.200	0.150	0.050	0.400
False -ve	0.400	0.350	0.000	0.000	0.000	0.200	0.050	0.100	0.000	0.000	0.250	0.350	0.200
CCR	0.750	0.800	0.900	0.800	0.950	0.875	0.950	0.775	0.950	0.900	0.800	0.800	0.700
Kappa	0.500	0.600	0.800	0.800	0.900	0.750	0.900	0.550	0.900	0.800	0.600	0.600	0.400
P=0.5													
P_crit	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500
Sensitivity	0.600	0.300	0.250	0.100	0.650	0.500	0.950	0.750	0.650	0.650	0.850	0.950	0.800
Specificity	1.000	1.000	0.950	1.000	0.950	1.000	0.950	0.900	0.950	0.950	0.650	0.550	0.550
False +ve	0.000	0.000	0.050	0.000	0.050	0.000	0.050	0.100	0.050	0.050	0.350	0.450	0.450
False -ve	0.400	0.700	0.750	0.900	0.350	0.500	0.050	0.250	0.350	0.350	0.150	0.050	0.200
CCR	0.800	0.650	0.600	0.550	0.800	0.750	0.950	0.825	0.800	0.800	0.750	0.750	0.875
Kappa	0.600	0.300	0.200	0.100	0.600	0.500	0.900	0.550	0.600	0.600	0.500	0.500	0.350

Table 39. Threshold-dependent evaluation criteria (as described in Appendix 8.2) with different cut-off values calculated on testing data in ROC_AUC software for BBN models. P_opt is the cut-off value which maximises the proportion of correct classifications. P_fair is the cut-off value at which sensitivity and specificity are the same. P_Kappa is the cut-off value at which Kappa is maximised. P=0.5 is a commonly used cut-off threshold. P_crit is the applied cut-off value and CCR is the correct classification rate.

All of the AUC values, excluding the ‘moist’ *P. punctata* model, were greater than 0.8, which indicates excellent discrimination (Hosmer and Lemeshow, 2000; see section 3.3.3, Chapter 3). The models for *H. semele*, *N. sylvestris*, *P. argus* and *P. globulifera* all had AUC values greater than 0.9, demonstrating outstanding discrimination (Hosmer and Lemeshow, 2000). This performance was also shown by the high sensitivity but low false positive values for these models across all of the thresholds in Table 39 (and by the shape of the ROC curves (see Apendix 29), which rise steeply at the origin and hug the top left hand corner). However this was lower (except for the *P. argus* (1) model, which had P_crit values close to 0.5 for the other cut-offs) for the arbitrary 0.5 threshold, indicating that this threshold was less suitable. For the *H. semele* model 2 (which had the highest AUC value of 1), there was clearly no overlap of the ranges of values between the presence sites and the absence sites (see Table 37 above) and the AUC value confirms this discrimination ability of the model. Although the *H. semele* (2) model provided greater discrimination, the model 1 version was slightly better because it is more accurate in assigning ‘grass suitability’.

The other highest AUC value was for the *P. argus* (1) model, which had the highest average overall habitat suitability value for the presence sites as well as a fairly low average for the absence sites. However, the models can still show good discriminatory ability without having a high average or range of values for the presence sites, as long as the values for the absence sites are lower (as for the *N. sylvestris* model). An issue with the model testing was that with a relatively small testing set of data, each site will have a relatively large influence on the average values. A more comprehensive, stratified sample of testing sites, particularly absence sites, may alter the apparent discriminatory ability of the models, which could potentially be exaggerated by the choice of sites.

Of the two models which achieved lower AUC values, one was the *P. punctata* ‘moist’ model, which was identified above as not performing as well as the other *P. punctata* models, and the other was the *G. illyricus* model (although this was considerably higher than the value for the *P. punctata* model). Also, as discussed above, this was due to several low values for the presence sites, although the average and range for the absence sites were lower. In terms of the threshold-dependent measures (Table 39), the fact that it was not possible to get an overall habitat suitability value of greater than 63.06 for the *G. illyricus* model meant that the P=0.5 cut-off was not appropriate, and this was shown by the very low sensitivity value using this threshold. However, using the other cut-offs (which gave very low threshold values) allowed for a high sensitivity rate (particularly for P_opt and P_Kappa), whilst still having a fairly low false positive rate. The P_crit values for the *G. constrictum* model were also very low (0.083, 0.238), but the model generally achieved high sensitivity and specificity rates whilst still maintaining low false positive and negative rates across the P_opt, P_fair and P_Kappa cut-offs, suggesting a good discriminatory ability.

The *P. punctata* models P_crit values for P_opt, P_fair and P_Kappa were amongst the highest across all the species, and particularly high for the 'dry' *P. punctata* model. These values reflected the fact that the values for the absence sites were quite high and that many sites are likely to be suitable for *P. punctata*, making it difficult for a model to discriminate between more and less suitable sites.

The lowest correct classification rates (0.750, 0.725) were generally for the *C. nobile* models (apart from the *P. punctata* (moist) model), although these values were still fairly high. The *C. nobile* (1) model also achieved the second lowest Kappa value (0.500; a moderate score (see Appendix 8.2)) at the P_opt cut-off. These values reflected the fairly high absence sites average, with a not particularly high presence site average, again a result of *C. nobile* occurring in a range of habitat situations, some of which appear less than optimal. If is difficult to apply a threshold value at which a site should be considered 'suitable' or 'unsuitable' as there is frequently an overlap in suitability values obtained for presence and absence sites. Further, the most suitable cut-off varied a great deal between different models. However, the 0.5 threshold was the least suitable in most cases.

The appropriate threshold will also depend on the implications of different errors (as discussed in section 3.2.4, Chapter 3). For this application false negatives would generally be more of a problem, although too many false positives could also waste time. However, for species such as *G. illyricus* which will not occur in many seemingly suitable sites just because it is rare, the false positive rate will be unavoidably high. Nonetheless, an important aspect of the models is to identify potentially suitable habitat, which means that there will inevitably be suitable locations which are unoccupied due to other factors, such as local extinctions (Pulliam, 2000; see section 1.2.3, Chapter 1).

All of the models worked well, as indicated by the high AUC values, and could be used with confidence (except for the *P. punctata* (moist) model). The *P. punctata* (dry) model should also be used as opposed to the other versions. In general the models where the variables were directly instantiated, rather than the variables influencing them instantiated, worked better. Although the *H. semele* (2) achieved a higher AUC value, with perfect discrimination, it provided a more conservative estimate of habitat suitability based on the likelihood of suitable grass species being present than the *H. semele* (1) model, which achieved a higher presence site average (although also a higher absence site average). Overall, the best models were for the two butterfly species, as well as for *P. globulifera* (1) and *N. sylvestris*.

4.3.11.3. Sensitivity analysis

Several of the models were run with different nodes instantiated. However, as can be seen from the results in section 4.3.11.2, the models performed better when important variables were instantiated directly, rather than the variables influencing them instantiated. However, using variables such as ‘habitat type’ to predict the likelihood of, for example, ‘suitable grass species’ for *H. semele*, did generally return values of a similar order, suggesting that they could be used to give an indication of suitable habitat based just on the GIS data without visiting every site. Running the *P. punctata* model with CPT values from experts with differing opinions identified a better model for that species, as discussed above.

The models were not run with different versions of nodes or their states or with different structures (Marcot *et al.*, 2006) as they all behaved realistically in terms of their outputs. Different models and states would likely produce different results but it was felt that the final models provided a good representation of current knowledge.

4.4. Discussion

The results of the testing of the models with fieldwork data show that the BBNs performed very well. Very few other studies have so far used BBNs in this way. However, the performance of the models in this study was generally better than achieved by Marcot (2006) for predicting suitability of sites in an old-growth forest (in the United States) for a rare fungus species. Although the true presence (sensitivity) rate was very high (1.0, using a cut-off of 0.5 for presence/absence) for the model developed by Marcot (2006) and higher than the sensitivity values for any of the models from this study (highest sensitivity value of 0.950 using the 0.5 cut-off (see Table 39, section 4.3.11.2), although this is not necessarily the most appropriate cut-off), the model developed by Marcot (2006) incurred a very high error rate in predicting true absence, with a true absence (specificity) rate of only 0.21. This was much lower than any of the specificity values achieved by the models in this study, where the lowest value was 0.550 for a cut-off of 0.5. The overall correct classification rate was 0.76 for the Marcot (2006) model, which was subsequently slightly improved to 0.78, with updating of the model. This was lower than the values for half of the models in this study and higher than the other half. However, the data set used to test the BBN models in this study was fairly limited in size and was not a random sample, so the performance may have been inflated. However, the data set used by Marcot (2006) was also small. Marcot (2006) also provided a ROC plot of the results, which approached the upper left corner of the graph, suggesting overall a fairly accurate predictive model, but an AUC value was not provided.

Smith *et al*. (2007) also did not provide an AUC value (or ROC plot) for their BBN model to predict suitable habitat for the endangered Julia Creek dunnart in an area of north-west Queensland, Australia. However, they did provide a Kappa value of 0.402 for their model, which was lower than the majority of the Kappa values from the models developed in this study (see Table 39, section 4.3.11.2), with the *P. argus* (1) model obtaining a Kappa score of 0.900 (using a threshold of 0.5). However, the overall accuracy value of 0.89 provided by Smith *et al*. (2007) was higher than all of the correct classification values for the models in this study except for the *P. argus* (1) model, although Smith *et al*. (2007) used three suitability categories (of low medium and high) as opposed to just suitable/unsuitable. Unfortunately comparable evaluation measures were not provided by Rowland *et al*. (2003) for their BBN model of wolverine habitat capability.

Poorer discrimination between presence and absence sites for the models resulted from either low predicted overall habitat suitability for presence sites or high predicted overall habitat suitability for absence sites. It is inevitable that this will occur as species may frequently be present in seemingly sub-optimal habitat as they can occur across a range of tolerances. However, the particularly sub-optimal sites would unlikely to be able to support the species in the long-term as the species is not occurring in favoured conditions and may just be ‘hanging on’ at a site where conditions have deteriorated. For example, *G. illyricus* may be recorded as sites with dense bracken because the bracken has become more vigorous over time. Related to this, it may be that although the dominant density of bracken at a site is vigorous (which would result in a lower habitat suitability from the model), *G. illyricus* may be surviving well on the edge of a dense bracken stand (although this may not be sustainable). In the case of more mobile species, such as butterflies, it may be that the species was recorded as it was just ‘passing through’ a site. Similarly, *N. sylvestris* can be found many metres away from the edge of a woodland in hot weather, but is unlikely to persist at such temporary sites.

There is a possibility that at some of the absence sites that appeared very suitable but did not contain the species, the species was overlooked. For example, *G. illyricus* can be very difficult to spot if it is not in flower. It can also be difficult to conclude with certainty that mobile species are absent from a site. For example, it could be that the butterfly species’ may not have been flying at the particular time that a site was visited. This is why it is necessary to return to a site several times in order to be certain of the absence of a species. However, this is very time consuming and was not feasible for this project, but measures such as visiting sites in good weather to maximise the chance of seeing the butterflies were taken to minimise the chance of recording false absences.

Species may also be absent from apparently very suitable sites because of factors such as local extinctions and dispersal limitation (Pulliam, 2000) or just because they are very rare, such as in the case of *G. illyricus*. However, identification of potentially suitable habitat is very useful, even if the species does not currently occur there because the species could potentially move or expand into such areas, particularly with future environmental change. In particular, areas with false positives located adjacent to known locations, i.e. true positives, could be considered good potential habitat (Marcot, 2006).

Species may absent from ‘suitable’ sites or present in ‘unsuitable’ sites because the model does not capture the habitat requirements of the species sufficiently. This may be due to a lack of knowledge or understanding of the species or a bias in expert knowledge. Further, as noted by Newton *et al.* (2007), absence of evidence is not the same as evidence of absence; a zero probability may simply reflect lack of appropriate evidence rather than absence of a probabilistic relationship. There may be currently unknown factors that are important for the species, or there may be variables that appear important but are actually related to or are influenced by other factors. For example, *P. punctata* is frequently found in acidic habitats such as heaths, but this is unlikely to be directly related to the occurrence of the fungus but more a function of the open nature of the habitat (most of the open (non-wooded) habitat in the New Forest is heathland) and where the ponies, and consequently their dung, tend to occur. However, as long as the models work sufficiently well, then they do not need to be complete.

An advantage of the BBNs is that as more information becomes available, it can be incorporated into the models. Something in particular that would be useful is a more detailed spatial soil map, which would mean that other soil variables such as soil acidity and fertility could be incorporated into the plant models and easily established.

A greater range of sites (in terms of sites with all the different states for the variables ranging from most to least suitable) to test the models would help to improve confidence in the discrimination ability of the models. This would also help to confirm, for example, whether the ‘dry’ *P. punctata* model was actually a better model. It is likely that slightly different versions of the models (in terms of structure, nodes, states and CPT values) could work as well as or better than the current models, as there is not a single best way of constructing each model. However, the most important thing is that the models perform well when tested and are usable.

An aspect of modelling habitat suitability to consider is the time frame. The period of time over which a habitat will remain suitable will differ between species. For example, *P. argus* requires pioneer heathland, meaning that a habitat may only remain suitable for a few years. Assessments of potentially suitable sites may therefore be ‘out-of-date’ relatively quickly, although in some cases it may be possible to work out future suitability from the initial assessment (such as by knowing how long it takes for heathland to progress through the different growth stages).

In general, the BBNs were an excellent approach for the purpose of this study and as reported by others (e.g. Uusitalo *et al.*, 2005; McNay *et al.*, 2006; Smith *et al.*, 2007) they can be very useful, particularly in capturing expert knowledge. Despite their relatively rare use for habitat suitability modelling to date (Newton, 2009a), the use of ROC curves and confusion matrices also proved a useful way of evaluating the models. The BBNs were relatively easy to construct and the experts found them easy to understand. The visual nature of the models helped to make the processes clear and facilitate discussion. Although population of the CPTs with probabilities is often cited as a difficult aspect of developing BBNs (Newton, 2009a), none of the experts struggled with this (although that is not to say that the values provided are necessarily accurate). A problem with using probabilities based on expert opinions is that the values provided by the experts may differ (even from the same expert on different occassions; see discussion below).

Although there are some widely acknowledged issues of using expert knowledge (e.g. Morgan and Henrion, 1990; Meyer and Booker, 1991; see section 4.1.3), the current results highlight the value of incorporating this source of information in habitat suitability models. However, some of the potential drawbacks should still be noted and an awareness of them should be borne in mind when using the models (Marcot *et al.*, 2001; Newton *et al.*, 2007). One factor is that experts have not carried out a full systematic survey of their species in the New Forest, as this would be unfeasible, and they may tend to only visit particular areas. Consequently, they may not know, or be familiar with, the full range of conditions in which the species occurs, particularly at the more extreme ends of its tolerances, and they may not know the likelihood of the species occurring in those different conditions.

Consistency of opinions among experts may be interpreted as a measure of reliability, whereas differences may reflect honest, valid differences of opinion (Burgman, 2005). Different experts may provide widely ranging estimates and Morgan and Henrion (1990) suggest that the analyst should try to understand why different experts reach different conclusions, through questions such as: have some experts ignored evidence that other experts consider very important; are motivational biases operating; are some (or all) of the experts just not very ‘expert’; and are the questions posed simply impossible for human experts to answer?

Differences of opinion (such as in the ordering of suitability for different states), as discussed in section 4.3.10, may be the result of experts differing in how frequently they have encountered the species in different habitat situations. For example, it may just be that different experts had encountered *P. punctata* more frequently in drier or more moist areas just because that is where they tend to spend more of their time. In this sense their knowledge is not ‘complete’ and they make not be able to accurately provide probability values for all of the combinations of states in the CPTs. Or some of those values may be biased towards their ‘coverage’ of the species. However, experts commented if they were unsure about any values and different versions of the models were run where particular differences arise. Further, testing data helped to verify the most suitable models.

Related to these issues, expert judgements (particularly about which conditions are most suitable) may be influenced by availability bias (see Appendix 22), with certain sites or situations being more readily called to mind, such as where they have personally encountered the species more frequently. Related to this, is that when experts convey information about where they tend to see the species, those conditions may not necessarily be optimal (although the experts are probably aware of this). It may also be possible that different experts gave different meanings to the different states, although this should have been minimised by the provision of definitions for each of the states.

Experts can also sometimes be ‘over-confident’ in the values they provide (known as overconfidence bias; see Appendix 22), believing that they have a greater knowledge than they actually have (e.g. the range of sites in which they have observed the species). It is also likely that if the experts were asked to fill in the CPTs again they would provide slightly different values (although they would probably follow the same pattern or order). Although this could be checked by asking the expert again, the large number of probabilities to assess, and the time involved means that really it is unfeasible to assess them more than once (Renooij, 2001) and the process may undermine the experts. The elicitation and averaging of probabilities from more than one expert helps to reduce the impact of this and the overall pattern of the model values is unlikely to be much different.

Another cause of differences between experts could be because the experts are less familiar with a certain variable. For example, differences of opinion occurred for both *P. globulifera* and *G. illyricus* over which soil type was more suitable and this was a variable that they found slightly more difficult to assign probabilities to. This could be because, particularly in casual observations of species, an observer may not be as aware of a variable that is less visible.

There were some CPTs that did not necessarily require knowledge of the species, but the likelihood of occurrence of different states in different habitats or conditions. For example, the probability of ‘suitable grassland’ in the *C. nobile* BBN, based on the occurrence of different habitat types, and the probability of ‘suitable grass species’ for the *H. semele* model. These sorts of questions were generally the more difficult ones for the species experts to answer as they were not directly related to the species, although experts are likely to have an idea of how often they observe such situations. This is one of the reasons why it is better to instantiate variables (such as presence of suitable grassland) directly where possible.

The decision on how to categorise (discretise) variables for the different states was the most difficult aspect of developing the BBN models. This information was generally not available in the literature, so this was where expert knowledge was particularly important, although the experts did not necessarily know the best cut-off point because this is often difficult to know. However, the groupings produced usable models that had realistic outputs that made sense, so although slightly different groupings may produce different results the groupings chosen appeared to be satisfactory. One other possible drawback of developing the BBNs is the time taken in conducting expert interviews. However, these provide a rich source of information and facilitate discussion of the species, which makes them very worthwhile. Further, as noted above, expert knowledge was required to establish suitable states.

From developing the BBN models, it became clear where knowledge was lacking. For example, there was relatively little information in the literature about *G. constrictum* and this was the species with which the experts were least familiar. There were also gaps in knowledge about the ecology of *P. punctata*, although a model could still be developed despite this, based on observable known factors. However, with a better understanding of the species, it may be possible to improve the model. Although there may be different versions of the models possible, and some aspects could be improved with better knowledge, the most important aspect is the development of a parsimonious, yet realistic model that does not necessarily represent the actual system, but more realistically represents current knowledge about that system (Borsuk *et al*., 2004).

A criticism of the models could be the use of an arbitrary fixed site of 100 x 100 metres. However, this was used to coincide with the models developed for Chapters 2 and 3, and the BBNs could easily be applied to whatever size (and boundary) of site required (such as in terms of management). This may be more appropriate for species such as *P. argus* (for which an indication of the surrounding habitat is included in the model) and for *N. sylvestris*, for which being present at a site which is part of, or near to, a larger woodland would make it more suitable. In addition, where the habitat suitability values for some of the *H. semele* presence sites were greatly reduced if there were no tussocks present in the site, it may have been that tussocks were present in adjoining areas outside of the survey site, which *H. semele* could use. Therefore the use of a less arbitrary site boundary would be more likely to capture this. However, it may still have been that this factor was not as important for *H. semele* as previously thought. Further, for species such as *P. globulifera*, it could just be the presence of a relatively small wetland feature that is important for the species at a site, so it may be more appropriate to focus on this feature and the immediate vicinity, rather than a whole one hectare site.

It is very difficult to capture the complex nature of species in a model and it is important to recognise that models are not reality, but rather are an interpretation of reality which reflect assumptions and current understanding of wildlife habitat relationships (Raphael *et al.*, 2001). As noted by Smith *et al*. (2007), the BBN modelling approach is not intended to replace empirical field research or species distribution and population models, but it provides a means for structuring knowledge of species-habitat relationships so that 'rapid appraisal' of habitat suitability can be conducted.

Another important consideration is that BBNs are explicitly tools for modelling belief, so evidence that is consistent with a given hypothesis (e.g. the likelihood of a particular state, given the influence of other factors) has a high likelihood (Newton *et al.*, 2007). Therefore 'suitability' does not strictly refer to a predicted state, but represents a strong belief based on accumulated evidence and, as with any other modelling approach, this will depend on the reliability, accuracy and precision of the information employed (Newton *et al.*, 2007). It should also be noted that the BBN models portray the relative suitability of habitat, not the actual habitat suitability (Raphael *et al.*, 2001) so should be only be used to rank sites in terms of their suitability. Importantly, the models should be used as a guide, rather than to dictate decisions (Marcot, 2006; Marcot *et al.*, 2006) as part of a larger process of management, research and monitoring (McCann *et al.*, 2006).

4.5. Conclusion

The BBNs proved to be an effective approach for capturing expert knowledge and developing models of potential habitat suitability that were shown to perform well when tested using field data. Results were similar, and in some cases better, than those achieved for the limited number of comparable studies. Although there are some issues in using expert knowledge, the benefits of using such a rich resource outweigh these. Therefore, although the application of BBNs is a relatively new approach to habitat suitability modelling, it appears to be a promising one.

CHAPTER 5

Chapter 5. The potential impact of climate change on selected species of conservation concern and their habitats in the New Forest

5.1. Introduction

Anthropogenic climate change is now recognised to be a global environmental concern, and a major potential cause of biodiversity loss (Rosenzweig *et al*., 2007). Ecosystems are sensitive not only to changes in climate and atmospheric trace gas concentrations (such as increased levels of carbon dioxide) but also to other anthropogenic changes such as land use, nitrogen deposition, pollution and invasive species, and also to natural disturbance regimes such as wildfire and insect outbreaks (Fischlin *et al*., 2007), which may be exacerbated by climate change. The Climate Change 2007 report (IPCC, 2007) suggests that the resilience of many ecosystems is likely to be exceeded this century by an unprecedented combination of climate change, associated disturbances (e.g. flooding, drought, wildfire, insects, ocean acidification) and other global change drivers (e.g. landuse change, pollution, fragmentation of natural systems, overexploitation of resources).

Research on climate change impacts have demonstrated that many ecosystems, habitats and species are already being affected by climate change and have the potential to be affected in the future (Berry *et al.*, 2003). Walmsley (2007) suggests that, given the strong evidence of species responding to climate changes to date, it is highly likely that the future impacts of climate change on biodiversity will be substantial, particularly as climate is a key factor determining species distribution and ecology. Both species and their habitats or communities will potentially be affected by future climate change, which could greatly affect the availability of suitable habitat and therefore the potential distributions of species. There is therefore a need to assess the potential impacts of climate change on biodiversity at a range of scales, so that appropriate nature conservation policy responses can be developed (Walmsley, 2007).

Studies that have modelled the effect of climate variables on species distributions, such as using climate envelopes, tend to have done so at relatively large scales, such as whole countries or continents (e.g. Zaniewski *et al.*, 2002; Huntley *et al.*, 2004).

Such models have also tended to be based purely on climatic variables such as temperature and precipitation and not on changes in other variables that might be affected by climate change. Willis and Bhagwat (2009) report that caution may be required in interpreting results from such coarse-scale models, not least because their coarse spatial scales may fail to capture important factors affecting the response of species to climate change.

With respect to the New Forest, climate change has been identified as a major potential issue (Trotter, 2007; New Forest National Park Authority, 2008). However, information is lacking on how climate change impacts might manifest at the scale of an individual protected area, such as the New Forest National Park. Potentially, the statistical models developed in Chapters 2 and 3 could be used to examine potential impacts of climate change. However, the results obtained did not identify climate-related variables as key determinants of species distributions at this scale. For example, only two (*G. constrictum* and *P. punctata*) of the Biomapper species models identified a climate variable as important (see section 2.3.2.2.1.2) and even those did not have the highest marginality or specialisation values.

The final GLM models (see section 3.3.1.4) for *G. constrictum* and *P. punctata* also included a single climate variable (related to the climate variable identified by Biomapper for those species), as did the *H. semele* model. However, none of the other species' models included climate variables, and in those that did, it may have just been an artefact of the data. Further, the only changes that could be made to simulate the impacts of climate change would be to the habitat type layers (and climate layers), whereas the BBN models could incorporate a wider range of (finer-scale) variables (such as ground moisture), not available as GIS layers, but which are more likely to capture the changes at this scale. Further, the BBNs had been shown in Chapter 4 to work well (with all models achieving AUC scores greater than 0.8) in predicting habitat suitability for the study species under current conditions.

In order to examine the potential impacts of climate change on biodiversity in the New Forest, it was decided first to review the available literature on the selected species and habitats. This was based on the assumption that the current understanding of the ecology of species and their associated habitats might provide insights into how they might be affected by climate change. Secondly, of the modelling approaches tested in this research, the BBNs were determined to be the most effective in terms of incorporating the potential impacts of climate change on habitat suitability for predictions at the scale of the New Forest. These models were therefore further explored in this chapter, based on the information obtained from the literature review, to examine the potential impacts of climate change on the distributions of the selected species.

Objectives

- To examine the potential impacts of environmental change on the study species and associated habitats in the New Forest, through a review of relevant literature.
- To explore the potential impacts of environmental change on the selected species using the Bayesian Belief Network (BBN) models (from Chapter 5).

First, this chapter provides a brief introduction to the issue of climate change, specifically in relation to southern England. Potential impacts on biodiversity in the New Forest are then considered. Results of a literature review are then presented, focusing on selected habitats and species in the New Forest, with a view to identifying the potential impacts of climate change. Lastly, results from an exploration of the BBN models, incorporating the potential impacts of climate change on the selected species, are presented.

5.2. Introduction to climate change

The IPCC (2007) refer to climate change as:

> "A change in the state of the climate that can be identified (e.g. using statistical tests) by changes in the mean and/or the variability of its properties, and that persists for an extended period, typically decades or longer. It refers to any change in climate over time, whether due to natural variability or as a result of human activity."

Warming of the climate system is unequivocal, as is now evident from observations of increases in global average air and ocean temperatures, widespread melting of snow and ice and rising global average sea level (IPCC, 2007). Global average temperatures have risen by nearly 0.8 °C since the late nineteenth century, and are rising at about 0.2 °C per decade over the past 25 years (Jenkins *et al.*, 2008). These changes are very unlikely to be due to natural variability in these systems (Rosenzweig *et al*., 2007) and it is very likely that man-made greenhouse gas emissions caused most of the observed temperature rise since the mid twentieth century (Jenkins *et al.*, 2008). Observational evidence from all continents and most oceans shows that many natural systems are being affected by recent climate changes, particularly regional temperature increases (Rosenzweig *et al*., 2007). From a global meta-analysis, Parmesan and Yohe (2003) report that species (or functional/biogeographic groups) show highly significant, non-random patterns of change in accord with observed climate warming in the twentieth century, indicating a very high confidence in a global climate change fingerprint.

Both climate and non-climate drivers (such as urbanisation and pollution) can influence systems directly as well as indirectly through their effects on climate variables such as soil-moisture regimes (Rosenzweig *et al*., 2007). Socio-economic processes, including land-use change (e.g., forestry to agriculture; agriculture to urban area) and land-cover modification (e.g., ecosystem degradation or restoration), as well as population growth and economic development, also affect multiple systems (Rosenzweig *et al*., 2007). Further, these drivers can operate either independently or in association with one another, with complex feedbacks and interactions occuring on all scales from local to global (Rosenzweig *et al*., 2007). The extent impacts of climate change will not be the same everywhere and these complexities make it difficult to predict the full implications.

5.2.1. Climate change in the UK and south-east England

Hulme *et al*. (2002) suggest that, in general, climate change in the UK is likely to include increased temperatures in all seasons, although this warming may be more rapid in summer than in winter, increasing the seasonal temperature contrast. Winters will become wetter (with heavy winter precipitation becoming more frequent) and summers become drier, continuing the trend observed over the last century. Greater changes in temperature and precipitation are expected to be experienced in the southeast than in the northwest of the UK. The UK climate may become sunnier than at present in summer, with cloud cover decreasing and solar radiation increasing during this season, most notably over southern parts of the UK, and relative humidity may generally decrease throughout the year. The effects on wind speeds are very uncertain, but it is likely that in winter (when most severe winds occur) more depressions across the UK could lead to stronger winds in southern and central Britain. Changes in soil moisture depend on changes in precipitation, temperature, wind speed and radiation and it is expected that in summer, the whole of the UK will experience a decrease in average soil moisture, with the southeast of England experiencing at least a 20% reduction (Hulme *et al.*, 2002). The temperature of UK coastal waters will also increase, although not as rapidly as over land and sea levels are expected to rise (Hulme *et al.*, 2002).

There have already been significant changes in the UK climate. Some of the recent changes in the south-east region, in which the New Forest lies, are described by Jenkins *et al.* (2008). The annual daily mean temperature from 1914 – 2006 has increased by 0.89 °C (based on a linear trend) and from 1961 – 2006, the change is more pronounced, with an increase of 1.62 °C. In addition, the annual number of days of air frost from 1961 – 2006 has decreased by 23.4 days. The percentage change in total annual precipitation in the south-east region from 1961 – 2006 has increased by 5.4%. However, in summer (June, July, August) it has decreased by 13.1%, but in winter (December, January, February) it has increased by 23.3%. Severe windstorms around the UK have also become more frequent in the past few decades (Jenkins *et al.*, 2008).

The most current projections of future UK climate are the UKCIP09 scenarios (UK Climate Impacts Programme, 2009a) based on the Hadley Centre Regional Climate Model (HadRM2), although they are still only projections. UKCIP09 provides outputs for Low, Medium and High emissions scenarios (see UK Climate Impact Programme (2009b) for explanation) for three different time-slices (2020s, 2050s and 2080s). Table 40 below shows the projected changes in temperature and precipitation for the south-east of England.

	2020s			2050s			2080s		
	Low	Medium	High	Low	Medium	High	Low	Medium	High
Winter mean temperature (^oC)	+1.3	+1.3	+1.4	+2	+2.5	+2.5	+2.6	+3	+3.7
Summer mean temperature (^oC)	+1.6	+1.6	+1.5	+2.5	+2.7	+3.1	+3	+3.9	+4.9
Summer mean daily max temperature (^oC)	+2.2	+2.1	+2.0	+3.5	+3.7	+4.3	+4.1	+5.3	+5.7
Annual mean precipitation (%)	+1	0	0	0	0	0	+2	+1	+1
Winter mean precipitation (%)	+7	+6	+7	+13	+16	+18	+18	+22	+30
Summer mean precipitation (%)	-6	-7	-3	-13	-18	-18	-14	-22	-26

Table 40. Changes in temperature and precipitation for Low, Medium and High scenarios for the three time-slices (2020s, 2050s and 2080s) for the South-East from the UKCIP09 output (UK Climate Impacts Programme, 2009a). The values in the table are the central estimates of changes in mean temperature (change in ^oC) and mean precipitation (percentage change).

Both summer and winter temperatures are projected to increase, even by the 2020s under the low emissions scenario, where the mean temperatures are expected to increase by over 1^oC and the maximum summer temperature by 2.2^oC. By the 2050s this increase range from over 2^oC increases for the means, up to 3.1^oC for the High emission scenario for mean summer temperature and up to 4.3^oC for summer maximum temperature. By the 2080s, for the High emissions scenario, this figure could be as much as 5.7^oC.

Although annual mean precipitation changes by 0% for most of the scenarios, there is a large difference in the seasonality of the precipitation. Winter precipitation will increase (potentially by 13 – 16% by the 2050s and up to a 30% increase for the 2080s High emissions scenario), but summer precipitation will decrease by about the same amount (potentially by 13 – 18% by the 2050s and decreasing by up to 26% for the 2080s High emissions scenario). Even by the 2020s, for the Low emissions scenario there could be a 7% increase in winter precipitation and a 6% decrease in summer precipitation.

It is important to note that projections of climate change and its impacts beyond about 2050 are strongly scenario- and model-dependent, with various sources of uncertainty (IPCC, 2007). Impacts research is therefore hampered by uncertainties surrounding regional projections of climate change, particularly precipitation (IPCC, 2007). However, it is still clear that there are going to be substantial increases in both summer and winter temperatures, as well as increased winter precipitation but decreased summer precipitation.

5.2.2. Climate change in the New Forest

The most recent management plan for the New Forest National Park, The New Forest National Park Plan Consultation Draft (New Forest National Park Authority, 2008) acknowledges climate change as a major pressure for the New Forest, with the following statement:

> "Climate change will alter the character of large parts of the National Park over the next 20-50 years and beyond, modifying landscapes, habitats, biodiversity and the way the Park is perceived and used".
>
> "Clear predictions relevant locally include:
>
> - a substantial rise in sea-levels affecting coastal habitats and recreation
> - increased summer droughts causing the decline of important wetland and woodland habitats
> - warmer winters allowing the spread of plant and animal diseases
> - changes to traditional land management."

The plan stresses that the main focus for the National Park must be on adaptation: finding appropriate ways to adjust to climate change now, based on the best available evidence, with the emphasis on conserving the special qualities of the area. This approach includes creating a robust and interconnected mosaic of habitats, to improve resilience of the New Forest landscapes, reduce local extinctions of species and enable migration of wildlife within and beyond the Park (New Forest National Park Authority, 2008). The Forestry Commission, who manage almost half of the New Forest National Park, also consider the potential impacts of climate change in their Management Plan for the Crown Lands (Forestry Commission, 2007).

As there is a limited understanding of the potential impacts of climate change on the New Forest, further research is clearly needed (New Forest National Park Authority, 2008). However, it is possible to suggest some likely impacts. The New Forest

National Park Authority (2007a) and Trotter (2007) suggest that the forecasted warmer and wetter winters (with increased rainfall intensity and less frost/ice/snow cover) and hotter, drier and longer summers are expected to bring both threats and possible opportunities to the New Forest. Some of the threats to biodiversity they suggest include:

- Loss of species (local extinction) due to changes in habitats leading to adverse conditions such as unsuitable breeding sites and lack of food sources;
- Risk of expansion of invasive species (e.g. bracken);
- Loss of acid grassland to heather and heath;
- Increased incidence of heathland, grassland and woodland fires in hot dry summers;
- Loss of vulnerable habitats, particularly wet heath and mires;
- Increased tree stress (summer drought, ozone and pollution effects) and loss, leading to changes in natural structure and species composition of woodlands;
- Reduction of flow rates in streams and rivers in summer and increase in winter with temperature, erosion and ecological issues;
- Increased soil erosion and pollutant leaching;
- Changes in agricultural practices in response to climate change may affect free-roaming grazing animals dependent on the survival of pastoralism and commoning;

The Forestry Commission (2007) also anticipate that climate change may exacerbate conditions favourable to the spread of disease or result in increased incidences of new pests and diseases in the New Forest. They further note that future climate change will have consequences for the plantations of the New Forest, as forest crops being planted today will potentially not be ready for harvesting for the next 70 to 200 years, and the issues of summer drought and wet winter ground conditions that already impact operations could potentially become much worse.

Climate change will, and already is, having a number of potential impacts upon management, and particularly heathland management, in the New Forest, especially in relation to seasonal changes which affect activities such as burning, where sensitive thresholds apply (Forestry Commission, 2007). For example, the Forestry

Commission (2007) describe how spring seems to have been coming earlier to the New Forest (reflected by an average temperature rise of 0.55 °C in March between the climate averages based on data from 1961 to 1990 and 1971 to 2000) and that this temperature rise is resulting in an earlier nesting season and earlier emergence of reptiles from hibernation. This has already been having potentially significant implications for the window of opportunity in which to undertake the annual burn programme. For example in 2007 the window of opportunity was reduced to two weeks primarily due to the wet winter weather conditions and mild spring (Forestry Commission, 2007).

As well as the more direct impacts of climate change on the habitats and species of the New Forest, there are also likely to be additional indirect impacts, such as higher visitor numbers as a result of longer and warmer summers, further increasing the pressures on the New Forest (New Forest Committee, 2003). There are an estimated 13.5 million day visits annually to the New Forest (New Forest National Park Authority, 2006b) and the range of recreational activities, such as walking and dog walking, cycling, horse riding and camping (Forestry Commission, 2004), can potentially cause disturbance to habitats and species (Gallagher *et al*., 2007).

However, it is also expected that the change in climate may provide opportunities, such as the distribution of some flora and fauna (particularly those with a pronounced southern distribution) possibly becoming more widespread and the range of some habitats possibly expanding, with development of new community types (New Forest National Park Authority, 2007a; Trotter, 2007). In addition, longer growing seasons could result from the warmer temperatures throughout the year and higher carbon dioxide concentrations could potentially increase growth rates and productivity (Trotter, 2007). The importance of ensuring that existing habitats are managed in better condition so that they are more resilient to changing climate is highlighted (Trotter, 2007).

One way of trying to understand how future climate changes may impact on the New Forest is to examine the impacts of past climate. Tubbs (2001) reports some of the effects of extreme weather events on the New Forest in the past, in particular during the decade of 1974-84 when the New Forest witnessed an uninterrupted sequence of hot, dry summers. These included numerous detrimental effects such as drying up of ponds and ephemeral wetlands and changes in the grasslands, as well as the death of many beech trees and population explosions of several moth and beetle pests. However, details were also provided by Tubbs (2001) of a number of “enriching events and trends” which coincided with the increased frequency of drier, warmer summers, including:

- An increasing incidence of good autumn seed crops on trees and shrubs – beech among them – after relatively frost and gale-free springs.
- An increased frequency of ‘good’ summers for many invertebrates, for example the Vespulid wasps, some migratory butterflies, moths and dragonflies, and several species of crickets. However, there was also what appears to be a population explosion of the tick (*Oxodes ricinus*), carrier of the causative agent of Lyme disease, the spriochaete (*Borrelia burgdoferi*).
- Several species of birds characteristic of heathland and on the edge of their range in Britain, notably Dartford warblers (*Sylvia undata*), woodlarks (*Lullula arborea*) and nightjars (*Caprimulgus europaeus*), undergoing substantial and apparently sustained increases in breeding populations. None of these changes was confined to the New Forest, but they do, perhaps, underline that it is not a static system, irrespective of the effects of human land use and management.

5.3. Results of literature review

A literature review was carried out to identify the potential impacts of climate change on the main New Forest habitats used by the study species: woodland, heathland, terrestrial wetland and grassland. Bracken (*Pteridium aquilinum*), although an individual species, was also included as it is an important habitat for one of the study species, *Gladiolus illyricus*. A search for relevant literature was carried out using tools such as ISI Web of Knowledge (www.isiknowledge.com) and Google (www.google.com). The term ‘climate change’ was used in combination with the names of the habitats and also in combination with the species names and with the names of the species groups (plants, insects, butterflies, fungi). Where possible, information specific to the New Forest or UK was used, but literature searches were not limited to these areas as there was frequently not enough relevant information available. Information about the effects of climate change on the study species were not explicitly elicited from the experts used to develop the BBN models in Chapter 4, but relevant comments from discussions with the experts were included.

5.3.1. Potential effects of climate change on habitats of the New Forest

As species are likely to respond individualistically to climate change (Huntley, 1991), it is difficult to predict the effects of climate change on habitats *per se* (Hossell and Rowe, 2006). Hossell and Rowe (2006) suggest that the initial effects of climate change may be subtle, affecting the composition of existing communities through alterations in the balance of inter-species competitive relationships. However, there are some general themes that arise in the literature, and some of the potential impacts on the main terrestrial habitats of the New Forest in the relatively short term (next few decades) are displayed in Table 41 below, summarising the more detailed information in Appendix 30.

Impact	**Description**	**References**
WOODLAND		
Changes in growth and phenology	A lengthened growing season and raised carbon dioxide levels are likely to result in an overall increase in tree growth, although it is not clear whether this would be sustained in the long-term. Advance in spring events and delay in autumn events. Implications for the composition and species dominance within forest habitats in the future, with knock-on impacts for other dependent species.	Broadmeadow (2000), Nisbet (2002) Menzel and Fabian (1999) DEFRA (2001)
Changes in species distributions and composition	Increases in temperatures are likely to change the boundaries of many tree species distributions in the UK, leading to a change in the community composition of current woodlands.	DEFRA (2001)
Die-back/changes in community composition and habitat structure	Increased occurrences of the death of trees may be caused by several factors related to climate change, including soil moisture stress as a result of drought, increased fire risk, increases in woodland pests and pathogens and increases in the frequency and severity of storms. Climate change and the resulting impacts in woodlands are expected to lead to changes in community composition and habitat structure.	Diack (1999), Mountford *et al*. (1999), DEFRA (2001), Tubbs (2001), Lonsdale and Gibbs (2002), Nisbet (2002), Broadmeadow and Ray (2005), Hossell and Rowe (2006) Diack (1999), DEFRA (2001)
HEATHLAND		
Changes in composition (wet heath replaced by dry heath)	A decrease in summer precipitation will cause many of the wet heaths to dry up and likely to revert to dry heath or be replaced by an expansion of the region's acid grasslands. The increased drying and its effects on the water table will cause the boundaries of the wet heath communities and their composition to change.	Cook and Harrison (2001), Hossell *et al*. (2005)

Competition with bracken	The competitive dynamics between bracken and heather may change as a result of climate change. See bracken section below.	
Increased risk of accidental fires	Dry heaths may be particularly at risk from reduced summer precipitation (and potential increase in visitor usage) leading to an increase in accidental fires. Such fires can be detrimental to the ecology of these heaths, removing vegetation, releasing nutrients into the system, and exacerbating erosion.	DEFRA (2001), Hossell and Rowe (2006)
TERRESTRIAL WETLANDS		
Changes in flow/ increased drying out	Changes in total precipitation, extreme rainfall events, and seasonality (such as increased autumn/winter precipitation and decreased summer precipitation) will affect the amount, timing and variability of flow in freshwater ecosystems. The 'flashiness' of flows is likely to increase, with the potential for more frequent winter flooding and summer drought. It is likely that flash flooding in winter may increase as the greater intensity of rainfall increases runoff into rivers and if flood events are severe enough these may scour river beds, washing away much aquatic life. Summer drought is likely to dry out bogs and result in widespread desiccation of many wetland habitats in southern and central England, including ponds, streams and ephemeral wetlands. Drought may exacerbate pollution problems in wetland habitats by concentrating pollutants in periods of low flow.	Carpenter *et al*. (1992) DEFRA (2001) Diack (1999) Cook and Harrison (2001), Tubbs (2001) Carpenter *et al*. (1992), Hossell *et al*. (2000)

	Changes in the timing, intensity and amount of rainfall could also affect water quality by increasing the leaching of fertilisers, herbicides and pesticides from surrounding land, as well as increasing levels of sedimentation. Combined with low flows, which may also concentrate nutrients, eutrophication is more likely to occur. An increase in temperature may produce an increase in pH in some water bodies as a result of increased decomposition of organic soils and weathering of rock.	Diack (1999), Hossell *et al*. (2000), DEFRA (2001) Eyre *et al*. (1993), Hossell *et al*. (2000), DEFRA (2001)
Changes in species composition/invasive species	The density, growth patterns and phenology of aquatic invertebrate communities are likely to be affected by temperature increases and there may also be a shift in the distribution of fauna and flora of these habitats especially in response to a switch from permanent to temporary standing water bodies. Non-native aquatic macrophytes (such as New Zealand Pigmyweed, Crassula helmsii) may spread more rapidly if winters become warmer and frost events are less frequent.	Eyre *et al*. (1993), Hossell *et al*. (2000), DEFRA (2001) Diack (1999)
Erosion/loss of peat (in bogs)	On disturbed bogs increased rates of decomposition, promoted by higher temperatures and drying and cracking in summer drought, on bare or partly vegetated peat surfaces may lead to increased erosion and loss of peat, exacerbated by increased winter rainfall intensity. In addition, if species such as *Sphagnum compactum* and *S. tenellum* are lost from former bog habitats as the summers become drier, this could increase the accessibility and use of some parts (for example, in the New Forest) for recreation, which	Hossell *et al*. (2000), DEFRA (2001) Hossell and Rowe (2006)

	in turn may increase erosion on drying peat soils.	
GRASSLAND		
Loss of grasslands	In terms of land cover in southern England, including Hampshire, grassland shows a large reduction in extent under models based on future climate change scenarios, although such models are based purely on suitable climate space. However, acid grassland may encroach into heaths as they dry out.	Hossell *et al*. (2005) Cook and Harrison (2001), Hossell *et al*. (2005)
Changes in productivity	Elevated carbon dioxide concentrations stimulate photosynthesis, leading to increased plant productivity and modified water and nutrient cycles in temperate grasslands. However, the response of grassland species to predicted increases in carbon dioxide and climate change are complex at a local level, and these factors, such as grazing pressure, water availability and increased carbon dioxide may offset each other, leading to no overall change in species productivity.	Soussana and Luscher (2007) DEFRA (2001)
Change in species composition/community structure	The diversity and botanical composition of temperate grasslands is likely to be affected by the rise in atmospheric carbon dioxide concentration, possibly through a decline in the relative abundance of grasses compared to forbs and legumes. Increased temperature and rainfall in winter may favour the increased growth of more competitive grasses, to the detriment of less-fast growing species. During droughts in the New Forest during 1974 – 1984, grass production was poor but the grassland flora increased in diversity	Teyssonneyre *et al*. (2002), Soussana and Luscher (2007) Diack (1999) Tubbs (2001)

	with the invasion of small annual plants which were able to become established with the suppression of more vigorous grasses. Although it is not entirely clear what the combined impact of different factors, such as drought, temperature, increased carbon dioxide concentration, competition between species, will be on grassland, it seems likely that there will be a shift in community composition. Changes in grazing in grassland habitats may have as great an influence as climate change and an increase in habitat degradation may occur if patterns or levels of grazing are not adjusted to take account of the changing precipitation patterns under climate change.	DEFRA (2001) Hossell *et al*. (2000)
BRACKEN		
Potential increase or decrease	Despite the advantages of warmer temperatures and fewer frosts, in the south-east of England bracken biomass is restricted by low water availability and that increased drought in the south may adversely affect bracken growth. However, once established bracken can be very tolerant of dry conditions, so may be tolerant of forecast drier summers for a while. The expansion of invasive species, such as bracken, may be a problem in some areas of the New Forest, particularly as it is very difficult to eradicate and expensive to control.	Pakeman and Marrs (1996), DEFRA (2001) Marrs and Watt (2006) Forestry Commission (2002b), Marrs and Watt (2006), New Forest National Park Authority (2007a), Trotter (2007)
Competition with heather	The effect of climate change on bracken may also be influenced	Whitehead *et al*. (1997), Anderson and

	by competition with heather (*Calluna vulgaris*), where they occur in combination, although the interacting effects of temperature, water availability, carbon dioxide levels and nutrient levels make it difficult to predict the precise outcome.	Hetherington (1999), Gordon *et al*. (1999), Werkman and Callaghan (2002)

Table 41. Summary of potential effects of climate change on the habitats (of the New Forest), based on the results of a literature review.

5.3.2. Potential effects of climate change on selected species of the New Forest

Recent analyses of long-term data sets from around the world and in the UK indicate that some species are already responding to climate and atmospheric change (Hughes, 2000; Hickling *et al.*, 2006) and some of the general effects of climate change on species are discussed in Appendix 31. In order to assess the potential impact on the selected study species in the New Forest, the variables identified in the development of the BBN models from Chapter 4 were used as a basis. Tables 42 to 49 below show the input variables used in the BBN models from Chapter 4, with a summary of the potential impacts of climate change on those variables, based on a literature review. Note that references are not necessarily in direct reference to the impacts of climate change on the particular species, but on the variables. A full discussion of this literature can be found in Appendix 34, which also includes a discussion of other variables which were not included in the BBN models as they were not directly related to habitat suitability, but which will influence the response of the species to climate change (such as species' dispersal ability).

Chamaemelum nobile

Variable/node	**Impact of climate change**	**References**
Habitat type	The main habitat types that are important for *C. nobile* (heathland and grassland) may change slightly in composition as they dry out. Apart from the potential loss of wet heathland, the overall impact on *C. nobile* is unlikely to be large as it occurs in a range of heathland and grassland habitats.	Cook and Harrison (2001), Hossell *et al*. (2005)
Suitable grassland	Acid grassland is likely to remain in the New Forest and may even increase in extent, although there may be increased encroachment of heather and invasive species. Grasslands may dry out and there may be changes in species and community composition.	Cook and Harrison (2001), DEFRA (2001), Tubbs (2001), Hossell *et al*. (2005)
Suitable grazing/mowing	Livestock grazing regimes may be adjusted to take account of changing climate. A substantial decrease in grazing could be detrimental to *C. nobile*. Mowing not affected.	Diack (1999), Hossell *et al*. (2000), New Forest National Park Authority (2007a), Trotter (2007)
Bare ground	This will largely depend on future grazing regimes. Increased trampling from increased visitor numbers may increase the amount of bare ground in some areas.	Gallagher *et al*. (2007)
Ground vegetation height	This will mainly depend on continuation of suitable grazing, particularly if there is increased growth of more vigorous plants which out-compete *C. nobile*.	Diack (1999), Hossell *et al*. (2000)
Tree canopy cover	Not affected in terms of *C. nobile* habitat.	
Soil type	Not affected.	
Soil/ground moisture	Summers are expected to be drier and the risk would be if ground conditions become too dry for too long for *C. nobile* to tolerate. In winter, the ground is likely to be wetter which should not be a problem for *C. nobile* as it favours seasonally wet grassland.	DEFRA (2001), UKCIP09 (UK Climate Impacts Programme, 2009a)
Winter flooding	Likely to increase with increased winter precipitation, which should be favourable for *C. nobile*.	DEFRA (2001), UKCIP09 (UK Climate Impacts Programme, 2009a)
Application of	Likely to be less of an issue in the New Forest habitats.	Diack (1999), Hossell *et al*. (2000),

herbicide		DEFRA (2001)

Table 42. Impacts of climate change on each of the input nodes for the *C.nobile* BBN. The variable highlighted in grey was not instantiated in running the BBN, but the impact of climate change on that variable would have an effect on the variables to which it is linked.

Galium constrictum

Variable/node	**Impact of climate change**	**References**
Tree canopy cover	Not affected in terms of *G. constrictum* habitat.	
Bare ground	This will largely depend on future grazing regimes. If grazing were to decrease and there was an increase in growth of more vigorous plants *G. constrictum* may be out-competed at some sites.	Diack (1999), Hossell *et al*. (2000), New Forest National Park Authority (2007a), Trotter (2007)
Wetland habitat	Drying out of wetland habitats in the summer could have a negative effect on *G. constrictum*.	Hossell *et al*. (2000), Cook and Harrison (2001), Hossell *et al*. (2005), Tubbs (2001)
Ground moisture	Summers are expected to be drier and the risk would be if ground conditions become too dry for too long for *G.constrictum* to tolerate. In winter, the ground is likely to be wetter which should not be a problem for *G.constrictum* and may be beneficial.	Hossell *et al*. (2000), DEFRA (2001), UKCIP09 (UK Climate Impacts Programme, 2009a)
Habitat type	Some of the wetter habitats may change to drier ones, which could have detrimental effects for *G. constrictum*. Encroachment of scrub or invasive species (such as bracken or rhododendron) may alter some habitats.	Hossell *et al*. (2000), DEFRA (2001)
Soil type	Not affected.	

Table 43. Impacts of climate change on each of the input nodes for the *G. constrictum* BBN.

Gladiolus illyricus

Variable/node	**Impact of climate change**	**References**
Habitat type	Bracken and heathland/grassland habitats are the most important habitats for *G. illyricus*. Heathland composition may change, but the drier heaths are likely to be less affected. Bracken habitat is discussed below.	Cook and Harrison (2001), Hossell *et al*. (2005)
Distance to woodland	Unlikely to change in terms of having an impact on *G. illyricus*.	DEFRA (2001)
Tree canopy cover	Not affected in terms of *G.illyricus* habitat, unless woodland encroaches onto sites which are close to woodland edges.	
Bracken density	Not clear, but likely to stay the same or increase at least in the short-term, which should not be a problem for *G. illyricus*, unless density greatly increases and the lack of frost damage create a more vigorous canopy, smothering *G. illyricus*. In the longer-term bracken may decline, which could pose a serious threat to *G. illyricus* (assuming grazing levels are still high). Bracken density will also depend on how it is managed, i.e. control/cutting of the bracken.	Stokes (1987), Pakeman and Marrs (1996), Stokes (2000), Marrs and Watt (2006), New Forest National Park Authority (2007a), Trotter (2007)
Bracken litter depth	Decomposition rates of bracken litter may be affected by future effects of climate change and nitrogen deposition. However, bracken litter depth will largely depend on bracken density (see above).	Anderson and Hetherington (1999)
Soil type	Not affected.	
Ground moisture	Summers are expected to be drier and the risk would be if ground conditions become too dry for too long for *G.illyricus* to tolerate. In winter, the ground is likely to be wetter with more flooding in some areas which may make some sites too wet and the *G.illyricus* corms may rot.	Stokes (1987), Hossell *et al*. (2000), DEFRA (2001), UKCIP09 (UK Climate Impacts Programme, 2009a)

Table 44. Impacts of climate change on each of the input nodes for the *G. illyricus* BBN.

Pilularia globulifera

Variable/node	**Impact of climate change**	**References**
Water acidity	Possible increase in pH in some water bodies (such as due to pollution or increased decomposition of organic soils), but level of increase and tolerance of *P. globulifera* not clear.	Eyre *et al*. (1993), Hossell *et al*. (2000), DEFRA (2001), UK Biodiversity Action Plan (2008)
Speed of water flow	*P. globulifera* mainly occurs in non-flowing water bodies, such as ponds, in the New Forest, which would not be affected. However, where it does occur in flowing water bodies such as streams and rivers, flow is likely to increase in winter, which could have detrimental effects (such as removing *P. globulifera*). Summer flow is likely to decrease, which should not be a problem unless it causes these wetland features to dry up for too long.	Carpenter *et al*. (1992), Diack (1999), DEFRA (2001)
Water depth	Likely to decrease in summer and increase in winter, but should not be too much of a problem unless the wetland features dry up for too long.	Carpenter *et al*. (1992), DEFRA (2001)
Water depth fluctuation	There is likely to be greater fluctuation in water depths, but this should not be too much of a problem for *P. globulifera* unless wetland features dry up for too long.	Carpenter *et al*. (1992), DEFRA (2001)
Disturbance/ grazing	Livestock grazing regimes may be adjusted to take account of changing climate but a significant decrease in open grazing by livestock could be detrimental by allowing more vigorous species to outcompete *P. globulifera*.	Diack (1999), Hossell *et al*. (2000), New Forest National Park Authority (2007a), Trotter (2007)
Bare/open substrate	This will depend on water depth fluctuation and disturbance/grazing (see above). If water levels are low for too long, *P. globulifera* may be vulnerable to competition from terrestrial plants or more tolerant aquatic plants. Invasive species (such as *Crassula helmsii*) may become more of a problem and out-compete *P. globulifera*.	Brock and van Vierssen (1992), Diack (1999), Scott *et al*. (1999)
Bank/shore substrate	Generally unlikely to change.	
Bank/shore angle	Unaffected, apart from possible erosion from increased winter precipitation and flash floods, which may create less suitable steeper banks.	Diack (1999)

Presence of wetland habitat	Expected drying up of many wetland habitats would have a detrimental effect. Although wetland features often occur within drier habitats, these will also be more prone to drying out for longer.	Hossell *et al*. (2000)
Tree canopy cover	Not affected in terms of *P. globulifera* habitat.	
Ground moisture	Summers are expected to be drier and the risk would be if ground conditions and wetland habitats become too dry for too long for *P.globulifera* to tolerate. In winter, the ground is likely to be wetter which should not be a problem for *P.globulifera* and may be beneficial.	Hossell *et al*. (2000), DEFRA (2001), UKCIP09 (UK Climate Impacts Programme, 2009a)
Soil type	Not affected.	

Table 45. Impacts of climate change on each of the input nodes for the *P.globulifera* BBN.

Hipparchia semele

Variable/node	**Impact of climate change**	**References**
Habitat type	There is likely to be an increase in dry heath and acid grassland over wet heath which is likely to be beneficial for *H. semele*.	Cook and Harrison (2001), Hossell *et al*. (2005)
Presence of suitable grass species	The main grass species used by *H. semele* are likely to be tolerant of drought and some may increase (see discussion below).	Hubbard and Hubbard (1984), Webb (1986), Hossell *et al*. (2005)
Tussocks	The main grass species used by *H. semele* are tussock-forming, so as long as they are not over-grazed they should persist.	
Time since burning	Frequency of accidental fires may increase and changes may need to be made in burning regimes. *H. semele* is unlikely to be greatly affected as it is able to recolonise recently burnt heathland fairly quickly.	DEFRA (2001), Hossell and Rowe (2006)
Suitable grazing (or disturbance)	Any potential negative changes in grazing regimes as a result of future environmental change could have a significant impact on habitat structure.	New Forest National Park Authority (2007a), Trotter (2007)
Pioneer heath	Maintenance of pioneer heath will be largely dependent on suitable management.	
Bare ground	Bare ground may increase as a result of more frequent fires and parching, which may be beneficial for *H. semele* but overall there may not be a large impact. Increased visitor numbers and associated trampling may lead to increases in bare ground in some areas, but such trampling may also cause greater erosion and damage to larval or pupal sites, particularly as *H. semele* lays eggs near bare ground.	Alcamo *et al*. (2007), Gallagher *et al*. (2007)
Tree canopy cover	Not affected in terms of most of *H. semele* habitat, unless woodland encroaches onto sites which are close to woodland edges.	
Ground moisture	The ground moisture is likely to be drier in summer (during the *H. semele* flight period), which *H. semele* prefers. In winter, the ground is likely to be wetter with more flooding in some areas which may make cause damage and mortality at larval sites.	Hossell *et al*. (2000), DEFRA (2001), UKCIP09 (UK Climate Impacts Programme, 2009a)

Table 46. Impacts of climate change on each of the input nodes for the *H. semele* BBN. Variables highlighted in grey are those which were not instantiated in running the BBN, but the impact of climate change on those variables would have an effect on the variables to which they are linked.

Plebejus argus

Variable/node	**Impact of climate change**	**References**
Habitat type	The boundaries of the different heathland types may change, with an expected drying up of wet heath. There may also be possible encroachment of acid grassland into heathland. This change to drier habitats is likely to have a negative impact on *P. argus*.	Cook and Harrison (2001), Hossell *et al*. (2005)
Suitable grazing/ management (burning)	Any potential negative changes in grazing regimes as a result of future environmental change could have a significant impact. Frequency of accidental fires may increase which may also be detrimental, with regeneration of pioneer heath potentially taking longer.	DEFRA (2001), Hossell and Rowe (2006), New Forest National Park Authority (2007a), Trotter (2007)
Suitable ericaceous species in pioneer phase	*E. tetralix* may be negatively affected by drought, but the other main foodplant, *C. vulgaris* is likely to be less affected. *E. cinerea* should also be largely unaffected. Maintenance of these species in the pioneer stage will be largely dependent on suitable management.	Webb (1986), Hossell *et al*. (2005)
Bare ground	Bare ground may increase as a result of more frequent fires (and trampling by visitors), although this is unlikely to have an overall beneficial (or negative) effect.	Alcamo *et al*. (2007), Gallagher *et al*. (2007)
Heathland patch size	Overall, the heaths are likely to become drier, which although *P. argus* is able to utilise both dry and wet heath, it may become a problem if too much of the heath becomes too dry, reducing the patch size of suitable habitat. A varied age structure (i.e. a larger patch size) is likely to be even more important, as the use of different microhabitats within sites can be used to overcome some of the negative impacts of drought.	Dennis and Bardell (1996), Cook and Harrison (2001), Hossell *et al*. (2005)

Table 47. Impacts of climate change on each of the input nodes for the *P.argus* BBN.

Nemobius sylvestris

Variable/node	Impact of climate change	References
Woodland type	Unlikely to change to have a significant impact on *N. sylvestris*.	DEFRA (2001)
Woodland cover	Unlikely to change to have a significant impact on *N. sylvestris*.	DEFRA (2001)
Leaf litter depth	Unlikely to change, although warmer winters may possibly mean that *N. sylvestris* does not require such a deep leaf litter.	
Leaf litter cover	Unlikely to change to have a significant effect on *N. sylvestris*.	
Ground vegetation cover	Opening up of woodland as a result of storm and drought damage may encourage more ground vegetation as more light reaches the ground. However, ground vegetation may also be affected by drought. It would also partly depend on grazing of ground vegetation in woodlands. As this is not too crucial a factor for *N. sylvestris* it is unlikely to have much effect.	Diack (1999), DEFRA (2001)
Ground moisture	Summers are expected to be drier and some woodlands may become drier, which may be beneficial for *N. sylvestris* by making more sites suitable. In winter, the ground is likely to be wetter with more flooding in some areas which may make some sites too wet and cause flooding of overwintering *N. sylvestris*.	Diack (1999); Hossell *et al*. (2000), DEFRA (2001), UKCIP09 (UK Climate Impacts Programme, 2009a)
Amount of woodland edge	Drought and storm damage may open up some woodlands as trees die, creating more edge, which could be beneficial for *N. sylvestris*.	Mountford *et al*. (1999), Hossell *et al*. (2000), Tubbs (2001)

Table 48. Impacts of climate change on each of the input nodes for the *N. sylvestris* BBN.

Poronia punctata

Variable/node	**Impact of climate change**	**References**
Habitat type	Any changes are unlikely to have much of an overall impact on *P. punctata*.	Cook and Harrison (2001), Hossell *et al*. (2005)
Soil type	Unlikely to be affected.	
Tree canopy cover	Some woodlands may become more open as a result of death of trees from drought and storms, which may increase the likelihood of *P. punctata* developing on dung deposited in these woodland sites.	Diack (1999), DEFRA (2001)
Ground vegetation height	Unlikely to change very much, unless grazing regimes change significantly.	Diack (1999), Hossell *et al*. (2000)
Ground moisture	Summers are expected to be drier, which is likely to dry out many sites, but *P. punctata* may not be unduly affected. In winter, when *P. punctata* is more likely to be fruiting, the ground is likely to be wetter with more flooding in some areas which may make some sites too wet so a decline in winter abundance may be possible.	Hossell *et al*. (2000), DEFRA (2001), UKCIP09 (UK Climate Impacts Programme, 2009a)
Presence of (dung from) ponies feeding on natural vegetation	This is the most important factor in *P. punctata* occurrence, so if significant reductions in the numbers of grazing ponies, as a result of changes in grazing regimes, occurred, this could have severe consequences for *P. punctata*.	New Forest National Park Authority (2007a), Trotter (2007)

Table 49. Impacts of climate change on each of the input nodes for the *P. punctata* BBN.

5.4. Results of BBN models

The BBN models from Chapter 4 were run with the field data (the same 20 presence sites and 20 absence sites for each species), but the states were changed, where appropriate, to indicate how they might be affected by future climate change (within approximately the next 40 years). This was based on the information obtained from the literature review summarised in Tables 42 to 49 in section 5.3.2 (and section 5.3.1). The changes in states made for each of the variables can be found in Appendix 35. Although this was based on a lot of assumptions and interpretation, it did not have the objective of producing precise predictions, but rather an indication of the potential impact of climate change on the suitability of habitats in the New Forest for the selected species. The full results are shown in Appendix 36 and a summary of the results is provided in Table 50 below.

Where for some of the species slightly different versions of the BBN models were run in Chapter 4 (see sections 4.2.11.3, 4.3.11.2 and 4.3.11.3 for explanation), only the models with the child nodes instantiated where there was the option of instantiating nodes at different levels (which, apart from for *H. semele*, had the higher AUC values (see section 4.3.11.2, Chapter 4), although both *H. semele* models had extremely high AUC values (0.965 and 1.00)) were run incorporating the changes in states. The best of the *P. punctata* models (the 'dry' model) was run for this species. The *P. punctata* model was run under the assumption that ground moisture would become drier (overall), as used for the other species. However, it was also run under the assumption that ground conditions would become wetter, as during the winter when *P. punctata* is more likely to fruit, the ground conditions are likely to be wetter, as increased rainfall is expected. Using the *P. argus* – 1 model (with the 'bare ground' and 'suitable ericaceous species in pioneer phase' variables instantiated) meant that no changes were made to the states (so the results are not shown in Table 50 below). Therefore, as the *P. argus* – 2 model (without these two variables instantiated, but their parent nodes, 'habitat type' and 'suitable management' instantiated) also performed very well (with an AUC score of 0.914; see section 4.3.11.2, Chapter 4), it was decided to run this model with the changed states.

Species	**Presence sites**			**Absence sites**		
	Average	**Change**	**Range**	**Average**	**Change**	**Range**
C. nobile – 1	61.26	+ 1.64	24.60 – 92.60	27.33	+ 9.81	0.75 – 72.84
G. constrictum	17.29	- 24.26	3.80 – 40.38	5.08	- 3.47	0 – 25.50
G. illyricus	29.17	+ 3.37	3.92 – 60.33	10.43	+ 0.81	0 – 42.00
H. semele – 1	70.21	+ 8.62	36.20 – 100	23.60	+ 12.53	0 – 100
N. sylvestris	57.41	+ 3.82	2.28 – 87.42	6.87	+ 2.49	0 – 43.04
P. argus – 2	65.18	- 5.99	26.48 – 92.24	21.19	- 0.25	0 – 76.64
P. globulifera – 1	16.94	- 40.30	2.85 – 29.21	1.78	- 3.64	0 – 15.35
P. punctata – dry – drier assumption	83.46	+ 6.04	39.84 – 98.81	64.04	+ 19.16	25.89 – 87.27
P. punctata – dry – wetter assumption	53.35	- 24.07	14.83 – 70.72	23.86	- 21.02	1.64 – 58.24

Table 50. Average (mean) and range of values, determined by the BBN models of probability of habitat being suitable for focus species for known presence sites and known absences sites from fieldwork data, with states changed to examine the effect of climate change. Values presented are percentages, representing likelihood of the habitat being suitable. Change is the change (increase or decrease) in the average value from the Chapter 4 BBN models.

Table 50 above shows that the average habitat suitability value for the *C. nobile* BBN model increased for both the presence and absence sites. This increase was greater for the absence sites, with the values for all of the absences sites increasing (see Table A-59, Appendix 36). The increases were largely the result of changing the ‘winter flooding’ state from ‘no to ‘yes’, although it should be borne in mind that although the assumption was made that all sites will flood in the winter, this may not be the case, particularly for sites on slopes for example. However, it appears that the predicted wetter winters may largely benefit this species, with more sites becoming suitable. Although this will depend on whether *C. nobile* can tolerate the drier summers as well as other factors, such as whether it suffers from competition from other species (see Appendix 34.1.4).

A slight increase in the average habitat suitability values for both the presence and absence sites is also shown in Table 50 for *G. illyricus*. All of the presence and absence sites increased (or stayed the same) in suitability, although the increase tended to be fairly small. The increase in suitability was due to the change of wet heath and mire (which had CPT values of either 0 or almost 0) to dry heath (HL1 and HL3), both of which were more suitable (see Appendix 27). Similarly, the change of states from 'moist/damp' or 'wet' to one state drier caused an increase in suitability. However, this did not take account of the fact the ground moisture may become wetter in the winter, which may be a significant factor if it causes *G. illyricus* corms to rot. Another important factor would be a change in bracken growth. No changes in states were made to 'bracken density' as it was not clear whether bracken density would increase (due to milder winters and less damage by frosts) or decrease (due to water availability). Further, bracken density would also depend on whether bracken was managed through cutting. An additional factor is that, although there may potentially be an increase in suitable habitat, *G. illyricus* is limited in its dispersal ability so would be unlikely to colonise such habitat.

The results from Table 50 show that *G. constrictum* may fare considerably worse. The average value decreases for both the presence and absence sites, particularly the presence sites and Table A-60 in Appendix 36 shows that the habitat suitability value for all of the sites decreased (except for the values which were already 0). This was in most part the result of the reduction of the presence of 'wetland habitat' to 50%. The presence of wetland habitat is a crucial aspect of habitat suitability for *G. constrictum*, but it is wetland habitats that are likely to be most affected by climate change. The tolerance of *G. constrictum* to long periods of drying during the summer months will be crucial, but it is not clear how tolerant *G. constrictum* may be. Invasive species such as New Zealand pigmyweed (*Crassula helmsii*) may pose an additional threat (see Appendix 34.1.2).

An even greater decrease in average habitat suitability values for the *P. globulifera* presence and absence sites is shown in Table 50. Again the habitat suitability values decreased (or stayed the same for the 0 values) for all of the presence and absence sites (see Table A-65, Appendix 36). As for *G. constrictum*, this was largely due to the reduction of the presence of 'wetland habitat' to 50%. Similarly, *P. globulifera* will also be vulnerable to invasive species such as *C. helmsii*. However, it has been suggested that *P. globulifera* may be able to persist or remain dormant in mud during periods of exceptional drought (Page, 1982; Jermy, 1994; see Appendix 34.1.4).

Further, the fact that it is a rapid opportunistic coloniser (Jermy, 1994; Botanical Society of the British Isles, 2008b), may be beneficial in being able to take advantage of suitable conditions quickly when they arise (see Appendix 34.1.4). However, other factors, such as maintenance of grazing by large herbivores, will also be important for this species.

In contrast, the average habitat suitability values for the presence and absence sites for *N. sylvestris* both increased, as did all of the individual values (or stayed the same). Only the ground moisture variable was changed in this BBN model and the increase to drier ground conditions was beneficial for *N. sylvestris*. Although there may be some changes to the other woodland variables, which were not included in the state changes, these would be expected to be beneficial if they did have an effect. Also, being ectothermic, it has been suggested that many insects will do well as temperatures warm, by feeding better, living longer and reproducing faster (Cook and Harrison, 2001; also see Appendix 34.3). Therefore, it is likely that this species will fare well with future climate change.

Another species which may fare well is *H. semele*. Both the presence site and absence site average values increased and all of the individual values increase or stayed the same. This was due to the change from wetter ground moisture conditions to drier ones, which *H. semele* favours. This means that more sites may become suitable for *H. semele* if wet heath is replaced by drier heath and acid grassland, as indicated by the increases in several of the absence site values. The fact that *H. semele* is able to utilise more than one hostplant is an advantage and the fact that it's hostplants are tolerant of drier sites. Again, being ectotherms, higher temperatures are expected to have beneficial impacts for butterflies, such as reducing the time required to raise body temperatures to flight activity thresholds and therefore increasing flight-dependent activities: mate-location, egg-laying, nectaring, predator-evasion and dispersal (Dennis and Shreeve, 1991; Dennis, 1993; see Appendix 34.2).

Using the *P. argus* – 1 BBN model resulted in no change to any of the states as 'habitat type' was the only variable which was expected to be significantly affected by climate change and this variable was not instantiated for the *P. argus* – 1 model. However, when the *P. argus* – 2 model was run, with 'habitat type' and 'suitable management' instantiated, the average value for the presence sites decreased slightly and the average for the absence sites decreased by an even smaller amount. All of the individual habitat suitability values either decreased or stayed the same (see Table A-64, Appendix 36) and this decrease was due to the loss of wet heath which was the most suitable habitat type for *P. argus* (see Appendix 27). One of the main foodplants of *P. argus*, cross-leaved heath (*Erica tetralix*), which is associated with wet heaths, is likely to decrease or even be lost in the New Forest. However, *P. argus* is able to utilise more than one foodplant, and the other species with which it is associated, heather (*Calluna vulgaris*) and possibly bell heather (*E. cinerea*) are expected to fare much better (see Appendix 34.2). Other factors, such as an increased risk of accidental fires may pose a threat to *P. argus*, but overall, it appears that although there may be some loss of certain types of habitat for this species, it may be able to persist in the New Forest in a changed climate.

Two different scenarios were run for *P. punctata* as although in general ground moisture is likely to become drier, during the main fruiting period for this species, ground moisture is likely to be wetter due to increased winter rainfall. Under the assumption that ground conditions become drier the average habitat suitability for both presence and absence sites increased. This was because the *P. punctata* 'dry' model assumed that habitat suitability increases as sites become drier. However, under the assumption that ground conditions will become wetter, which may be more appropriate for this species, the average for both the presence and absence sites decreased. Although it is not entirely clear how climate change may affect *P. punctata* in the New Forest. It seems less likely that *P. punctata* may fruit in the summer months if the summers are going to be drier and hotter. It may also be that for part of the winter ground conditions may be too wet at some sites for this species. Therefore, the period in which conditions are favourable for *P. punctata* may become smaller, but as long as it can persist over the summer months (and it is not clear exactly how this happens) and during other less favourable conditions, then there should still be a substantial number of sites which still provide suitable conditions. For example, although all the sites (apart from one absence site, 12) decreased in suitability, the presence site values (and some of the absence site values) were still quite high, with the majority of values greater than 50.

The results from Table 50 for all of the species show that where there was a decrease in habitat suitability for the presence sites (for *G. constrictum*, *P. argus* (2), *P. globulifera* (1) and *P. punctata* (dry – with 'wetter' assumption)) there was also a corresponding decrease in habitat suitability for the absence sites. Therefore, a decrease in habitat suitability or loss of suitable sites would not be made up by an increase in the number of (absence) sites becoming suitable. Although it should be noted that the range of sites was limited, and a wider range of sites (and conditions) may provide a different indication across the whole of the New Forest. For the species where there was an increase in habitat suitability at the presence sites there was also an increase in habitat suitability at the absence sites (although the increase may not be by enough for the sites to become suitable). Therefore, based on these limited results, it appears that habitat suitability will generally increase for a species or generally decrease.

5.5. Discussion

Although it is very difficult to predict the likely consequences of climate change for species due to the complex nature of interacting factors, the BBN models provide predictions to give an indication of the likely effect of climate change on the study species. There is an assumption that the BBN models developed to predict current habitat suitability are accurate. Although there was some variation in performance, testing of the BBN models from Chapter 4 indicated the models were able to discriminate between suitable and unsuitable habitat sufficiently well (AUC values all greater than 0.8).

The results of the BBN models incorporating the effects of climate change show that some species appear more vulnerable than others. For the two wetland species, *G. constrictum* and *P. globulifera,* the drying up of their habitat is likely to make them particularly vulnerable, but the actual impact of these changes will depend on how well they are able to tolerate such effects and associated factors, such as invasive species. The results show that *C. nobile* is likely to fare better, and *G. illyricus* may not be negatively affected, depending on factors such as changes in bracken density. Again, the ability of these species to tolerate summer drought will be crucial.

Of the two butterfly species, *H. semele* is likely to fare better than *P. argus*, but the fact that both species can utilise more than one foodplant is an advantage. *N. sylvestris* in particular may benefit the most from the effects of climate change and may be able to expand its range further northwards. It is expected that insects, being ectotherms will generally gain from the warmer temperatures. There is less certainty about the impact of climate change for *P. punctata*, although it seems likely that it may be able to persist. A clearer understanding of the likely effects for this species is hampered by a limited knowledge of its ecology. *G. constrictum* was another species for which there was less information, which makes it more difficult to predict the potential implications of climate change for these species.

The models were not able to provide an exact prediction of the potential impacts of climate change on the study species, and were based on numerous assumptions (such as a complete loss of wet heath and mire) and simplifications (such as not including the effect of other factors such as species' dispersal ability and drought tolerance (discussed in Appendix 34)). However, they were still able to provide an easily interpretable indication of the likely impacts and different scenarios could be run. Such quantitative information about the impacts of climate change on individual species at this scale has been very limited, despite the fact that such information would be most useful for management. It should be noted that the findings were based on a limited number of testing sites, which were not randomly selected. Therefore, the results (particularly for the absence sites) may show a slightly different pattern of changes in habitat suitability if the models were applied to a much wider range of sites across the New Forest.

An assessment of the effects of environmental change on individual species is important, but their ability to persist will largely depend on the effects on their habitats (and the species comprising those habitats) and whether they still exist in a favourable condition. Some of the habitats of the New Forest could be severely altered, with drying up of wetland habitats and shifts in dominant species and community composition. However, habitats are actually assemblages of species and the species within those habitats will largely respond individualistically to environmental change, such that new species assemblages may occur in the future

(DEFRA, 2001). It is therefore difficult to predict the effects of complex relationships between the species of a community (DEFRA, 2001).

Walther (2004) also suggests that the complexity of ecological interactions renders it difficult to extend the responses of individuals or populations to that of communities or ecosystems. This is because although the components of the considered system may have undergone considerable modifications, ecosystem properties may show less sensitivity to climate change than would be predicted from extrapolations of environmental responses of single species, due to compensatory effects among species or limitations by factors other than climatic parameters. A further issue is that there are still gaps in the knowledge of the ecological consequences of climate change. For instance, there is considerable uncertainty as to the effects of global change on soil biodiversity, as well as to the integration of aboveground and belowground processes under climate change scenarios (Walther, 2004).

It is therefore difficult to produce precise predictions as it is very difficult to determine, and incorporate into a model, all of the interacting factors which may influence species distributions. However, models can be used to indicate the likely direction and degree of the habitat or species response to climate change, but it should be borne in mind that they do not include all possible feedbacks and there are uncertainties in our understanding of species-climate relationships and in predicting the future climate (Hossell *et al.*, 2003). Walmsley and Harley (2007) note that it is likely that the response of many species and habitats to changed conditions will produce surprises, with unexpected winners and losers. Therefore, Hossell *et al.* (2003) suggest that monitoring changes in species and habitats is necessary to verify the direction, rate and extent of changes predicted by such models as well as providing new data, to improve the models themselves.

There is high confidence that the ability of many ecosystems to adapt naturally will be exceeded this century and this is hampered by the multiple barriers and constraints to effective adaptation that exist in human systems (IPCC, 2007). The responses could range from a complete loss of suitable ‘climate space’ (the potential limit, as defined by climate, of a species’ distribution), through to changes in regional distribution patterns, and extension into new areas (Hossell *et al.*, 2003).

Populations of many species are already perilously small, and further loss of habitat and stress associated with severe climate change may push many to extinction (Fischlin *et al.*, 2007). Walmsley and Harley (2007) suggest that many species will need to disperse to survive, but they will find it difficult across fragmented habitats. Therefore, the conservation of habitats and their maintenance in appropriate condition is essential for species trying to migrate (Berry *et al.*, 2003). It will also be important to reduce and manage other stresses on species and ecosystems, such as habitat fragmentation and eutrophication (Fischlin *et al.*, 2007).

Protected areas are expected to continue to be an important nature conservation structure under climate change, as areas of refuge as well as providing areas to aid dispersal (Hossell *et al.*, 2003). However, Hossell *et al.* (2003) suggest that the role of protected areas may need to change in the future to adapt to the impacts of climate change on species and habitats. Moreover changes may need to be made to both existing site management objectives and conservation targets. For example, management objectives will need to consider the potential to facilitate and/or manage the colonisation of non-native species, which may spread from continental Europe as climate changes (Hossell *et al.*, 2003). Decisions will also be needed in relation to cases where some species and habitats will be lost or severely restricted in terms of ‘when to let go’, i.e., when to change conservation priorities away from a particular location, habitat or species and when to adjust management regimes or to reassess the designation of sites (Hossell *et al.*, 2000). There is little point in trying indefinitely to conserve a habitat or species that is no longer sustainable, but the key is to know when to change the priority (Hossell *et al.*, 2000).

Hossell *et al*. (2000) suggest that the strict boundaries of protected sites are likely to be problematic as species distributions change in response to climate change, and buffers around sites will be important. Although this study has focused on the area of the National Park, conservation of New Forest species will also need to also be active beyond this boundary and consider surrounding habitat. For example, Berry *et al*. (2007a) report on opportunities for habitat re-creation of heathland, with potentially suitable sites around the edge of the New Forest.

Monitoring of species and habitats and an understanding of what aspects of habitats are important for species will be key. Unfortunately for some species, very little is known and environmental change projections are associated with a large degree of uncertainty. A limited understanding of the full impacts of climate change is an important issue and, as noted in the New Forest National Park Plan Consultation Draft (New Forest National Park Authority, 2008), more research is needed on the impacts of climate change in the New Forest. However, in order to understand the potential impacts of climate change on species and their habitats, it is important to know their requirements. The importance of capturing expert knowledge should not be underestimated as it can often be the most simple and effective approach to tackling complex problems (Walmsley and Harley, 2007).

As Root *et al*. (2003) note, anticipation of changes improves the capacity to manage, by acting proactively rather than reactively. However, climatic change impact studies are among the most complicated environmental assessments that scientists have ever faced and the risks and chances that policy-makers are facing are enormous (Varis and Kuikka, 1999).

5.6. Conclusion

There is no doubt that there are going to be significant impacts on the habitats, and species of which they contain, in the New Forest and elsewhere as a result of environmental change. The BBN models provided an effective way of predicting the potential impacts of climate change on the selected species at the scale of the New Forest National Park. Some species will fare better than others and some may benefit, such as *N. sylvestris*. A key problem in predicting the potential consequences of environmental change is a lack of understanding of the ecology of many species (such as *G. constrictum* and *P. punctata*) and the complexity of the possible effects on complex systems.

CHAPTER 6

Chapter 6. Synthesis

6.1. Summary of results

The results of the models showed that the project objective (see section 1.3.1, Chapter 1) of defining the potential distributions of selected species of conservation concern in the New Forest by carrying out habitat suitability modelling was met. Models were developed that could contribute towards improving monitoring and management of biodiversity in the New Forest where it is recognised that ‘data sets and analysis of the importance of many of the important species across the National Park are still far from adequate’ (Chatters, 2006). The evaluation results of the GLM (and GAM) models showed that all of the models achieved at least an acceptable level of discrimination (AUC values greater than 0.7) for modelling potential likelihood of occurrence of the species across the New Forest study area based on relatively ‘coarse’ environmental variables. In comparison, the BBN models were also shown to be effective (all AUC values greater than 0.8; excellent discrimination) for predicting potential habitat suitability based on finer-scale variables. A comparison of different approaches for modelling potential habitat suitability (objective 2) showed that the presence-absence models (GLMs and GAMs) were superior to the presence-only method (ENFA).

This work also showed that BBN modelling could provide a valid alternative (or accompanying) method for predicting habitat suitability based on finer-scale variables and the use of expert knowledge. Importantly, this method also provided an extremely valuable way of examining the potential impacts of climate change on the selected species and their habitats in the New Forest (objective 3). The modelling of climate change on individual species in this way at a small scale was unique, but an approach that would be particularly valuable for management of species and their habitats at this scale. The New Forest is a relatively small protected area and the application of habitat suitability modelling at the scale of an individual protected area of this size has so far been rare and has not been carried out in this way in the UK before.

The logistic regression (GLM and GAM) models (Chapter 3) and Bayesian Belief Network (BBN) models (Chapter 4) generally worked well, with measures of model performance within (and often higher) the range of values returned from other studies. The AUC values were above the 0.7 threshold for acceptable discrimination, with many showing excellent discrimination (AUC values greater than 0.8). The ENFA (Biomapper) approach (Chapter 2) did not appear to be as effective as the logistic regression and BBN approaches, with evaluation measures generally lower than those reported by other studies, except for the *G. illyricus* and *P. punctata* models. The ENFA approach also may not discriminate unsuitable habitat as well.

One of the difficulties in comparing model performance between models developed using presence-only data and those developed using presence-pseudo-absence data is the use of different evaluation measures, as although models should ideally be evaluated in the same manner, the best evaluation measures, such as AUC, require absences, so are prohibitive to presence-only modelling (Zaniewski *et al.*, 2002). However, overall, the regression models predicted potential habitat with higher accuracy than ENFA, as found by others (e.g. Zaniewski *et al.*, 2002; Brotons *et al.*, 2004; Olivier and Wotherspoon, 2006).

This suggests that, despite the lack of 'true' absence data, the presence-absence GLM (and GAM) models provided a more effective approach for predicting potential distributions. The fact that the use of pseudo-absences allowed for development of effective models is an advantage and means that the more discriminatory presence-absence models can be applied. However, it should be noted that these models were based on pseudo-absences, so the evaluation measures only provided an indication of how well the models predicted on those pseudo-absences and not how well they would predict true absences (Chefaoui and Lobo, 2008). Further, a different set of data may produce slightly different results.

Despite their superior performance, there are several issues in using a presence-absence modelling approach, such as the generation of pseudo-absences. Nonetheless, the straightforward random-generation approach used in this study appears to have been sufficient. Another issue is that of variable selection, particularly when there are a large number of potential variables. Although stepwise selection methods have been criticised by many (see Appendix 13.2), they proved to be an effective approach in this application (and in others, for example Matern *et al*. (2007)), identifying more significant variables than those identified through expert variable selection. Although the selected variables should preferably be examined ecologically, as spurious variables may be selected as a result of artefacts in the data, the models still performed well without the removal of such variables. It should also be noted that stepwise procedures are sensitive to the input order and other variables present (see section 3.2.2.3, Chapter 3, and Appendix 13.2), which means that different models could be selected. The fact that these straightforward approaches for pseudo-absence generation and variable selection proved effective is an advantage in developing a large number of models quickly.

Another important issue in the regression models is spatial autocorrelation (SAC), as the data used are frequently spatially autocorrelated. However, although most of the model residuals did show SAC, the use of the GLMMs, which relaxed the assumptions of the GLMs, allowing for random effects, did not select different variables or significance levels. Although other approaches for accounting for SAC may have produced different results, the most important thing is that the models produced accurate results in the evaluation. Further, Barry and Elith (2006) suggest that there is a danger that the inclusion, for example, of geographical space as a covariate, may be fitted to compensate for a missing predictor with strong spatial pattern and could potentially introduce more model error than it removes. Nonetheless, as the model results are not going to be projected outside of the study area used to develop the models, this does not appear to be too much of an issue in this case.

Biomapper (ENFA) has certain advantages on the regression methods; in particular that it does not require absence data. It also does not require variable selection in the sense that the regression methods do, instead weighting the environmental variables (Hirzel *et al.*, 2002). However, if there are a large number of variables compared to the number of presence data, some variables do have to be removed, as carried out in this application, which makes the process more time consuming. However, the permitted ratio of the maximum number of environmental variables to the number of species data for Biomapper was greater than that permitted for the GLM and GAM models. Some of the other issues that arise for regression, in particular SAC, are not such a problem in Biomapper because it is fundamentally a descriptive analysis and does not rely on any underlying hypothesis for the data (Basille *et al.*, 2008). However, the fact that the Biomapper models did not perform as well as the logistic regression models means that the logistic regression models should be favoured in this application. Nonetheless, Olivier and Wotherspoon (2006) suggest that ENFA may offer a simpler alternative to presence-absences models to define approximate habitat suitability.

A potential disadvantage of both ENFA and GLM models is that they only allow for linear relationships, when non-linear relationships may be more appropriate. GAM allows for non-linear relationships, although a problem can be that, particularly with small sample sizes, GAM models may over-fit the data (Gibson *et al.*, 2007). Although this is not so much of a problem if the models are not going to be used outside of the area used to develop them, the models may be unduly influenced by a few spurious values that do not fit the general trend, something which may be a particular issue when using pseudo-absences (Olivier and Wotherspoon, 2006). Despite this, using GAM, as opposed to GLM for *C. nobile* significantly improved the predictive ability for this species, although not for *P. punctata*. However, as noted above, it may just be more difficult to predict suitable habitat/potential distribution for this species as it is not as specialised (see discussion later). Although not used in this application (for reasons as discussed in section 3.4, Chapter 3), the use of transformations of the data (such as quadratic transformations) may improve the performance of GLMs.

The BBN models worked in a slightly different way and were not subject to any of the issues described above. Further, they were not limited by data availability in terms of requiring a certain number of unique locations for a species to develop the models (apart from for testing them), or spatial data. However, they did require input of expert knowledge to determine, in particular, the different states for the variables. This data is more time-consuming to collect, but provides a way of formally capturing expert knowledge, which may otherwise go unrecorded. The process of constructing the models also helped to develop a greater understanding of the species. Expert knowledge is particularly important for poorly known species.

A drawback to the BBN models was that they were not able to predict potentially suitable habitat across the whole of the study area as they were not based on spatial GIS data. This was the main difference between the BBN models and the ENFA and GLM/GAM coarser-scale models. A key advantage of the BBN models was that the potential impacts of climate change on the habitat suitability variables could be examined. This resulted in being able to generate easily interpretable and quantitative indications of the likely effect of climate change on the study species.

The different modelling approaches can therefore used in different ways, with the coarser scale models used to predict broadly suitable habitat across the whole of the study area and narrow down the choice of potentially suitable sites and the BBN models used to determine how suitable the habitat really is at a finer site-based scale, incorporating variables that are not available in a spatial GIS layer. However, this does require surveying on the ground, which is more time consuming.

Nonetheless, the benefit of the additional detail provided by the BBN models for predicting finer-scale habitat suitability can be seen from a comparison of the values obtained from the different models (see Table A-68 and discussion in Appendix 37), for one of the species (*P. punctata*) for which performance of the BBN model was high but the performance of the GLM model was one of the weaker ones. The fact that the BBN models were parameterised based on variables that were known to be important for the species, whereas the models based on the GIS data were dependent on the data available in that format (which was much more limited), means that they should provide more accurate predictions.

In terms of using a two-scale modelling approach, this appears to be a promising one, using the results of the regression models to predict potential habitat at a coarse scale, to narrow down the potentially most suitable sites and then visiting these sites to determine habitat suitability in more detail using the BBN models. Although surveying is more time consuming, the states used for the variables and the variables themselves, were selected partly so that they would be easy to survey and would not require any particular expertise. These would mean that they could be carried out by a wider range of people at any time of year. This would be particularly beneficial for cryptic species or those which are difficult to survey. Although this may not be the case for some of the study species, the fact that successful models were developed for a range of species from several functional groups (plants, insects, fungi) shows that this approach could potentially be applied to a much wider range of species (see discussion below).

Other studies using a similar multiple scale approach (e.g. Ben Wu and Smeins, 2000; Rowland *et al.*, 2003; Marcot, 2006; Brambilla *et al.*, 2009) have also reported on the effectiveness of this method. However, although the use of multi-scale modelling approaches is not new, the approach of using habitat suitability models at these two different scales, using both GIS and on-ground (much finer-scale) variables has so far been rare (Brambilla *et al.*, 2009). Further the suggestion of using both GLMs and BBNs does not appear to have been made before.

For some very rare species, for which there are very few records, it may not be possible to develop effective GLM models. In this case it may be better to use ENFA, or some of the newer methods that may require fewer records, such as Maxent (Elith *et al.*, 2006; Wisz *et al.*, 2008), or community models (see below). Although well-established methods were chosen for this study, other methods may have produced equally effective, or better (or worse) models and, particularly in the case of species with smaller sample size, it may be worth investigating their performance. For some species, it may also be possible (or necessary) to just use the BBN models (which do not require a minimum sample size, although they do require known sites to test the models), particularly if some of the GIS variables were incorporated. Where possible, the habitat type and soil type variables were built into the models for the study species so that a partial suitability value could be obtained just from those layers simply by querying the GIS layers, without visiting the sites. Therefore, it may be possible to reduce the number of predicted potentially suitable sites (to visit) for some species by a substantial amount, as carried out using the coarse-scale GLM models. This would save time in running the more complex coarse-scale models.

The process of developing the models also helped to identify what is really important in terms of habitat requirements for those species, something which is important for identifying other potentially suitable sites (particularly neighbouring sites) and management for those species and to obtain a better understanding of the implications of future environmental change for those species.

6.1.1. Comparison of selected variables and performance between different species

The variables selected by Biomapper (ENFA) and the GLM (and GAM) models for each species were generally similar, with between 1 and 3 identical variables selected. Several related variables were also selected, such as PA.HL1.HL3 for the Biomapper *H. semele* model and HL1.HL3.cover for the GLM model, or edge density of individual habitat types for the *C. nobile* Biomapper model but the merged habitat types (HL1.HL3) for the GLM model. Other examples included average annual temperature highlighted as an important variable in the *G. constrictum* Biomapper model, whereas minimum February temperature was selected by the GLM model, and Euclidean distance to dry heath for the *P. argus* Biomapper model and Euclidean distance to wet heath for the GLM model (and several of these Euclidean distance variables had to be removed for the GLMs as they were all correlated). A few different variables were selected by the different modelling methods, but this was likely to be due to the inclusion or not of pseudo-absences.

In terms of differences between model performances for different species, the BBN models all showed excellent performance (particularly for *H. semele*, *P. argus* and *P. globulifera*). The best performing Biomapper models were for *G. illyricus* and *P. punctata*, and to a lesser extent for *H. semele* and *P. globulifera*, whereas model performance was poorer for the other four species. In comparison, the *P. punctata* GLM and GAM models were the poorest performing presence-absence models, as was the *C. nobile* GLM model (which was also one of the poorer performing Biomapper models).

Although reported as a factor by others (e.g. Pearce and Ferrier, 2000b), sample size did not appear to be a factor affecting model performance for the Biomapper or regression models. *G. illyricus* and *P. punctata* had the fewest records, but the best performing Biomapper models and, although performance of the *P. puncata* was the worst for the regression models, *C. nobile* (the other species with the worst GLM performance), was one of the species with the most records. However, there was not a large range in sample size for the species and the influence of species specialism (as described below) may have had a greater influence. Nonetheless, the fact that in general effective models were achieved with relatively small sample sizes is promising.

The reason for the relative difference in performance of the Biomapper and the GLM/GAM models for *P. punctata* could be that although the potential presence/suitable locations for *P. punctata* may be predicted well by the ENFA model, the absence locations may not. This appeared to be the case for the GLM model because although the model achieved some of the higher sensitivity (true positive) values, it also achieved a relatively low specificity (true negative) and high false positive rate (see Table 33, section 3.3.3, Chapter 3). This suggests that the model was predicting large areas of suitable habitat and not discriminating unsuitable habitat. This was also likely to be occurring in the Biomapper model, as *P. punctata* can occur in a relatively wide range of habitats compared to some of the other species, so it is difficult to distinguish between suitable and unsuitable habitat. This was shown to some extent by the results of the *P. punctata* BBN model (which achieved one of the lower AUC values), with the values for the absence sites generally quite high, as there are unlikely to be many sites that are completely unsuitable. Further, *P. punctata* may frequently be absent from suitable habitat because it is more dependent on the conditions of the dung than the actual habitat (although the habitat does influence this). Therefore accurately modelling a species like this may be more difficult.

For *C. nobile*, (one of the comparatively less specialist species) it also appeared that the models were not as effective at predicting suitable or unsuitable habitat, with the GLM for this model achieving some of the lower sensitivity and specificity values and the highest false negative rate. However, these values were improved for the GAM model. For this application, a high false negative rate is more of a problem than a high false positive rate (although it should not be too high, so as not to waste too much time), because it is important not to miss potentially suitable habitat, and this was accounted for when applying the unequal weighting for the threshold-dependent evaluation measure. However, habitat suitability was better predicted at a finer scale using the BBN model, where more specific variables were taken into account.

Several other authors have noted that generalist species are often modelled with less accuracy than more specialist species (e.g. Hepinstall *et al.*, 2002; Brotons *et al.*, 2004; Elith *et al.*, 2006; Hernandez *et al.*, 2006; Tsoar *et al.*, 2007; Wisz *et al.*, 2008; see section 1.2.5, Chapter 1). Species mobility has also been reported to affect model discrimination ability, with poorer performance for more mobile species (Pearce *et al.*, 2001). This is also related to generalism, as more mobile species may be expected to be recorded in a wider range of habitats, with some of those not necessarily as suitable, but where they are just moving between patches of more suitable habitat. The most mobile of the study species was *H. semele*, but the models for this species did not perform any worse than for the non-mobile species, such as the plant species. In fact the BBN models for this species were among the best, along with the other butterfly species, *P. argus*. This may be because although *H. semele* is a fairly mobile butterfly, it is mobile across its main habitat type of dry heath/acid grassland rather than across a range of different habitat types and is still not particularly mobile when compared to species such as birds, for example.

6.2. Critical evaluation of methods

Further to the issues noted throughout the discussion above, there are some issues that relate more generally to the models. As well as there being many different modelling approaches which are likely to produce different results, there are also likely to be several different possible versions of equally effective or better models for each species using the same approach, depending on the variables and data used, the method of variable selection and so on. A different model may fit the data nearly as well, so there is a level of uncertainty associated with estimates of parameters (Whittingham *et al.*, 2006). However, this is the case with any model and as noted by Crawley (2007), it is important to remember the following truths about models: all models are wrong but some models are better [and more useful] than others and the correct model can never be known with certainty. The complexity of natural systems as well as imperfect knowledge and the fact that no model could (or should) be a perfect description of reality mean that the predictions will be imperfect (McCann *et al.*, 2006). Scott *et al.* (2002a) note that the complexity and inherent variation in species and their responses to physical and biological factors at multiple scales, as well as the dynamic nature of environments and species ranges can make predicting species occurrence with high levels of precision and accuracy difficult. Nonetheless, models can still contribute a great deal to knowledge, such as by invoking further field research leading to new insights, models revisions and more accurate predictions of the potential effects of management decisions (McCann *et al.*, 2006).

The models will only be as good as the data used to develop them, which is frequently limited and/or collected on an *ad hoc* basis. Further, models based on coarser variables will only be able to make coarse predictions of potentially suitable habitat. However, the most important thing is that the field data used to test the models shows that they are able to provide discriminatory predictions of habitat suitability (based on several evaluation measures), which is an advancement in predicting species distributions. It is extremely important that models are tested with field data to help gauge their degree of accuracy and identify ways in which the models could be improved (McCann *et al.*, 2006).

Obviously, because models are generalisations and simplifications of complex situations they will never be able to predict every single situation in which a species occurs as sometimes they do not occur in their ‘typical’ habitat. Species do not always occur in optimal habitat (as shown by the low values for some of the presence sites for the BBN models) and may be absent from seemingly suitable habitat (particularly if they are very rare, such as *G. illyricus*) or present in more unsuitable habitat (although they may not necessarily be able to persist for long), which may also be partly due to them not being in equilibrium with current conditions (see section 1.2.3, Chapter 1). Therefore, the models just predict where species are most likely to occur, based on current knowledge and known associations and a high predicted suitability does not in any way guarantee presence of a species. Similarly, when attempting to predict the potential impact of climate change, the models are based on current understanding, involving numerous assumptions, which may not be accurate.

Even when using the finer scale BBN models, there will still be other factors that have not been included (such as biotic interactions) that may contribute to whether a site is suitable or not. However, they may not need to include every factor to still be useful and 'over-prediction' of suitability can be advantageous, for example, by suggesting possible areas with high potential for (re)colonisation (Lütolf *et al.*, 2006). The models also make an assumption that presence of 'suitable' habitat/land-cover means that the habitat is of suitable quality and is available and accessible (Berry *et al.*, 2005a).

If too many factors are incorporated into the models (assuming there is sufficient knowledge or data to support this), they may become over-complicated. There is therefore a balance between complexity and prediction and simplicity/ease of use. Further, when less detail is required in a predictive model the resultant prediction is likely to be less uncertain (Reckhow, 1999).

6.2.1. Uncertainty and error in the models

A key issue of the models is the uncertainty and error inherent in their development at various stages, which is something that users should be aware of. In a broad sense, uncertainty refers to a lack of sureness or confidence about something and can be thought of as being synonymous with error, where error includes not only mistakes and faults, but also the statistical concept of variation (Elith *et al.*, 2002). Regan *et al*. (2002) classify uncertainty into two main categories: epistemic uncertainty (uncertainty associated with the knowledge of the state of a system) and linguistic uncertainty (uncertainty in language), both of which can be present in modelled predictions (Elith *et al.*, 2002).

Epistemic uncertainty can be classified into several main types as suggested by Regan *et al*. (2002). The first of these, measurement error, results from imperfection in measuring equipment and observational techniques (including operator error and instrument error). In this study this includes the errors in the species and environmental data as well as in the data processing. Systematic error occurs as a result of bias in the measuring equipment or the sampling procedure and the uncertainty in the species data is partly systematic error, for example recording bias and spatial autocorrelation (Elith *et al.*, 2002). Predictor variables can also be systematically biased (Elith *et al.*, 2002). A further type of epistemic uncertainty is model uncertainty, which occurs as a result of (simplified) representations of physical and biological systems. Another type is when subjective judgement is used to interpret data, and this is particularly applicable for the BBN models. Finally, natural variation exists in systems that change (with respect to time, space or other variables) in ways that are difficult to predict and although it is not a source of epistemic uncertainty *per se*; it is often regarded as a source of uncertainty because the true value of the quantity of interest is usually extraordinarily difficult to measure or predict across the full range of temporal and spatial values (or other related variables) (Regan *et al.*, 2002).

Linguistic uncertainty can also be classified into several main types (Regan *et al.*, 2002). The first is vagueness, which arises because natural and scientific language permits borderline cases. Related to this is ambiguity, which is uncertainty arising from the fact that a word can have more than one meaning and it is not clear which meaning is intended, and context dependence, which is uncertainty arising from a failure to specify the context in which a proposition is to be understood (Regan *et al.*, 2002). These types of uncertainty are particularly relevant for the BBN models (see discussion below). Two other types of linguistic uncertainty are suggested by Regan *et al*. (2002); underspecificity (which occurs when there is unwanted generality: the statement in question does not provide the desired degree of specificity) and indeterminacy of theoretical terms (the potential for future ambiguity), but these are not as relevant in this application.

The main sources of error and uncertainty in the development of the models for this study are outlined below.

Errors in the species data

The presence records were collected in an *ad hoc* way and although it is assumed, in carrying out the analyses, that the records are accurate there may be several issues, such as mis-identification of species (particularly those which are more difficult to identify, such as *G. constrictum*, which is very similar to *G. palustre*). However, many of the people who record the species have a great deal of experience in identifying those species. A further source of error in the species records is imprecise, or wrong, recording of the locations. This may become less of a problem with the increase in hand-held GPS ownership, but even these are prone to inaccuracy, particularly if there is poor satellite reception. However, the error is only likely to be a few metres, which in the majority of cases should not have a large impact, and cannot be rectified. Further, although they used a broader scale, Graham *et al*. (2008) report that species distribution modelling approaches in general are fairly robust to locational error and usable models can be built even when occurrence data are imprecise.

There is also an issue of (spatial) bias in the recording of species data, meaning that the data are unlikely to be fully representative of the range of environments in which the species occurs in the study area. It also means that the modelled relationships may be dominated by the patterns at sampled sites rather than the patterns across the entire study area (Barry and Elith, 2006). A likely cause of SAC in these models is from observer and sampling bias (Dormann *et al.*, 2007b). It is well known that the sort of *ad hoc* records usually available for species distribution modelling tend to be biased towards more accessible locations, such as closer to urban areas (Araújo and Guisan, 2004). The urban landuse/habitat type variables were selected for several of the GLM models, which may reflect recording bias, or reflect a genuine association, such as *C. nobile* frequently occurring on roadsides. Butterfly distribution maps, for example, have been found to reflect not just the distribution of individual species but also recorder effort (Dennis *et al.*, 1999). Records for species occurrences may also reflect a bias towards areas where recorders know they are more likely to, or expect to, find the species. Some species may also go unrecorded if they are more difficult to find. Species may also be biased towards locations where they are easier to find. For example, *P. punctata* may genuinely tend to occur in shorter ground vegetation, or it may be that it is easier to spot in shorter vegetation (S. Skeates, personal communication, February 25, 2009).

Biases in some of the records obtained as part of this study, may also have arisen as a result of a concentration of time spent near certain habitats, whilst recording other species. This can lead to artefacts or biases in the data trends. For example, it may have been the case that the relationship between *N. sylvestris* and lower Euclidean distance values to large ponds (ED.AQ6) was a result of opportunist recording of this species whilst visiting sites for wetland species.

A small sample size is also often an issue in species distribution modelling. Barry and Elith (2006) suggest that small sample size is an error in the sense that the sample may provide an insufficient basis for modelling. The minimum number of records required for a method depends partly on the complexity of the pattern being modelled and a smaller sample size will also limit the number of candidate environmental variables that can be supported numerically (Barry and Elith, 2006). The sample size limited the number of species that could be modelled, although model performance did not appear to be associated with the differences in sample sizes (but there was only a limited range of sample sizes between the species). It also meant that the number of variables had to be reduced, for example in the ENFA models, before running a final model, which meant that model development was more time-consuming. An alternative could be to reduce the number of candidate variables to a certain extent, using expert knowledge, although this may result in a failure to identify previously unknown associations (Steyerberg *et al.*, 1999) and may be more time-consuming in itself.

A lack of absence records is also a data error in the sense that it limits the creation of models that accurately discriminate between suitable and unsuitable habitats; in particular, it can lead to inaccurate identification of the attributes of unsuitable sites (Barry and Elith, 2006). Although pseudo-absences can be used, the absence of true zeros in models can result in different response shapes (Barry and Elith, 2006). This is an important consideration for the GLM and GAM models developed in this study, as pseudo-absences may have been generated at locations that are actually suitable for the species. Although the performance of the models was generally high, it is important to remember that the models were trained and tested using those pseudo-absences. It is likely that true absence data would improve model performance, but is very time consuming to collect.

A limited number of ‘true’ absence sites were recorded for testing the BBN models. However, error may be contained within this data as most sites were visited only once, which is generally suitable for identifying whether plant species are present or not (assuming they are relatively easy to spot), but may not be sufficient for the butterfly species in particular, which are more mobile and may not emerge if weather conditions are less suitable. However, where possible, sites were visited in favourable conditions, at the best time of year for surveying the species, so it was assumed that the classification of present or absent was correct.

Errors in the environmental data

A frequent issue in species distribution modelling is that at least some predictor variables are missing from most models (i.e. limited covariates), as even for mechanisms that are relatively well understood, directly relevant quantitative data that can be used for modelling are usually unavailable (Barry and Elith, 2006). This can increase spatial autocorrelation in the errors of the analysis (Barry and Elith, 2006) and spatial autocorrelation was present in the residuals for most of the models, so could have increased the systematic error. Even with all covariates known prediction is not perfect because of demographic variation (Tyre *et al.*, 2001; Barry and Elith, 2006).

The environmental variable data themselves also contained different sources of measurement error. There may be inaccuracies in some of the spatial layers, for example in deriving the habitat types from aerial photographs to create the habitat type GIS layer. However, an extensive amount of ground-truthing has been carried out to help improve the accuracy of the habitat layer so it was assumed that this layer was reasonably accurate. Another potential source of error, was that the habitat layer was based on a snapshot in time, and boundaries of habitats may change over time. This was also the case for the records of species occurrences, which were recorded over a period of several years (limiting the records only to very recent ones would have severely reduced the number of available records). Therefore, within that time the habitats may have changed, such as in terms of their quality (although it is not expected that the overall habitat type would change a great deal) and species occurrences may have changed, so that a species may no longer occur at some of the sites at which it was previously recorded. A potential effect of this mis-match may have been shown in some of the models. For example, a preference of *P. punctata* for sites with higher edge density values for dense scrub was shown in the ENFA models (section 2.3.2.2.1.2, Chapter 2, but this could reflect recordings of *P. punctata* at locations where scrub was present at the time that the habitat layer was created but has since been cleared, as part of management practices (see section 2.4, Chapter 2). Although ideally the habitat and other spatial environmental layers would originate from the same time as all of the presence records, this would be unfeasible.

During fieldwork, observed percentage cover of each of the habitat types was compared to the percentage cover of each of the habitat types predicted by the habitat type layer. Although it can be difficult to accurately visually estimate the percentage cover of each habitat type on the ground (and this may also be made more difficult if the GPS accuracy is poor and it is more difficult to determine the boundaries of the site) assessment was made at the sites by recording agreement or not with the predicted values from the habitat layer for those sites.

For 58% of the sites used to test the BBN models, there was full agreement with the habitat layer. Although this seems quite low, of the 42% of sites for which there was disagreement between predicted and observed percentage cover of the habitat types, 44% were due to the grassland habitat type being undefined (unknown grassland type predicted by the habitat layer), 28% were due to the absence of scrub when it was predicted to be present by the habitat layer (so had likely been cleared) and 28% were due to other reasons. These other reasons were: clearance of coniferous woodland, wrong woodland type predicted, wrong heathland type predicted (in particular several sites where it appeared that heather cover had become more dense, so that it was HL1 (dry heath, with more than 25% cover of heather) instead of HL3 (dry heath/acid grassland mosaic with less than 25% cover of heather)), or where previously dry heath/acid grassland mosaic (HL3) was now continuous bracken (GL8), or the reverse. Therefore, none of the differences were very significant. However, it may be that there were a number of sites, for *C. nobile* in particular, where the habitat layer indicated unknown grassland rather than, say acid grassland, so the relationship with this variable may not have been as strong when variables were selected for the models. This was a limitation of the available data but would only have have affected a small proportion of the sites for the species.

There is also another issue that arises with the habitat layer in that it is very difficult to draw boundaries around different habitat types (and soil types), as many of the habitats merge into one another gradually, rather than change from one to another abruptly as depicted by polygon borders or raster pixels. These errors are compounded as they are propagated through to the layers derived from the initial habitat layer (edge density, patch area, patch compactness and Euclidean distance). This has implications for the fragmentation measures in particular as it becomes difficult to define a 'patch' for example. Further, the imposed classifications may not be appropriate for the modelled species. This was partly considered by combining some of the habitat type layers for the modelling.

The habitat layer also did not include some smaller habitat features, such as small ponds. Additionally, finer-scale features such as wet ditches for example (which are important for species such as *G. constrictum*), were not captured by the habitat layer, which was part of the reason for developing the BBN models. Although measurement error could also have been introduced in recording the input variables for these models (Marcot, 2006). However, detailed information (the accompanying sheets in Appendix 26) was available to inform accurate recording for the BBN models to minimise this.

Including a greater number of classes in the GIS layers is difficult technically (as classification becomes more difficult to determine from aerial photographs) and in general, maps with greater numbers of cover-type classes have less accuracy than those with fewer classes, as the chance for misclassification increases with increasing numbers of cover types (Glenn and Ripple, 2004). Biases are often also linked to the spatial scale of the data (Barry and Elith, 2006). For example, as the grain at which data are recorded becomes coarser, units that exist at a finer grain may be subsumed into more prevalent ones, leading to a bias against unusual classes, such as rare vegetation classes (Barry and Elith, 2006). Aggregation of pixels may also result in loss of accurate cover values of each habitat type (see discussion below). There are also errors in digital elevation models (DEM), and although the average error may be small, local error values can be large, which can be propagated through to calculations of slope and aspect, derived from the DEM (Holmes *et al.*, 2000).

Although there are several errors inherent in the environmental data, Barry and Elith (2006) suggest that there is some justification in just ignoring them as in cases where the prediction sites have the same errors as those used for model building (as is the case in this application), the model will already reflect the errors and the predictions will be consistent with the data.

Errors in data processing

There are several aspects of the data preparation and processing that may lead to errors. One of these is the error resulting from the conversion of GIS layers into different formats, for example from vector to raster format (such as for the habitat and soil data), which may cause loss of information or accuracy. Related to this is the transformation of all of the spatial layers to the same resolution to match the resolution of the species data, which may have a similar effect, particularly when 10m x 10m cells were aggregated to 100 m x 100 m resolution and the pixel values averaged. However, this was required and the 100 m x 100 m resolution was felt to provide an appropriate level of detail for the application. As different procedures and transformations are carried out on the spatial layers any error present in the initial layers is propagated through the layers and compounded, potentially leading to even larger errors (Cablk *et al.*, 2002). Finally, there is also an issue of human error, such as naming of files so that the correct data are used for each species, and the correct data extracted. Checking of data was carried out to minimise this.

Model uncertainty and errors

Model uncertainty arises because models are used to represent biological or physical processes which requires simplifying assumptions and abstraction of ecological processes (Elith *et al.*, 2002) and this is largely unavoidable. Model errors may result due to mis-specification of the model. For example, if the 'true' relationship is not contained in the model, then over- and underestimation will typically result (Barry and Elith, 2006).

Only linear terms were included in the GLM (and ENFA) models and this could be improved upon, as suggested in section 3.4 (Chapter 3) by running univariate GAM models first to check the shape of the response curve and then either running a GLM with linear or transformed variables, or running a GAM model if required. However, this is more time consuming, particularly when there are a large number of potential candidate variables available. Further, if limited data are available, or pseudo-absences have been used, GAM models may over-fit the data and be overly influenced by spurious data (Olivier and Wotherspoon, 2006; Gibson *et al.*, 2007; see above). However, Barry and Elith (2006) suggest that if sufficient data are available, flexible regression-based techniques capable of fitting non-linear relationships (such as GAM) should be used, as these allow for the complicated response surfaces that are frequently observed in distributional data.

As discussed above, there are also likely to be several different versions of the models that would work as well as, or better than the 'final' models, and there is no single 'correct' model. Factors such as the use of different variables, different implementation and methods (such as in stepwise selection) and different modelling approaches may all produce different results, so there is some uncertainty in whether the model produced is optimal. Several factors can also influence which variables are selected by stepwise selection, such as the number of candidate variables, which means there is a significant amount of uncertainty regarding selection of the 'final' model (Derksen and Keselman, 1992; Whittingham *et al.*, 2006). Different species data would likewise lead to different influences in the model and different random splits could also produce slightly different models. One way of improving this would be to run the model with different random splits of the data, and take an average (Fielding and Bell, 1997; Harrell, 2001). However, this would be more time consuming and may not provide much benefit as the different models may not differ significantly. The most important thing is to develop a model that is shown by evaluation data to provide high predictive capability, so although the 'final' models may not be the 'best' models, the results showed that they worked.

Subjective judgement

The use of expert opinion in development of BBN models (as well as personal judgements on how to construct the influence diagrams) means that these models in particular were prone to errors of subjective judgement, a form of epistemic uncertainty. There are several sources of bias inherent in expert judgement (as discussed in section 4.1.3.4 and 4.3.4 (Chapter 4) and Appendix 22), but measures were taken to reduce this. There was a large amount of uncertainty and subjective judgement in particular in assigning different states to the climate change BBNs (Chapter 6) as this had to be based on numerous assumptions. However, it is accepted that the outputs were only to provide an indication of the potential impact and they may not necessarily be correct. There is a large element of uncertainty in the discretisation of variables, as this is frequently subjective and different results could be obtained by small shifts in the boundaries of the classifications. Likewise, the elicited probability values for the CPTs were estimates, based on expert judgement (which is prone to biases, as noted above), and it is not possible to test whether they are 'correct' or conform to 'reality', since the events for which the probabilities are assessed are often unobservable or would require a large amount of data (Renooij, 2001). Sensitivity analysis can be used as tool for exploring how different uncertainties influence the model output, and help identify the scope for reducing uncertainty due to epistemic sources (O'Hagan and Oakley, 2004). For example, the running of some of the BBN models with different values for the CPTs. For the *P. punctata* model in particular, this helped to identify which values may more closely reflect the true relationship. The use of field data to test the models was also very important in this respect. However, the values do not need to be exactly 'correct', just within the true order and the overall habitat suitability values should not be used as absolute values, but to indicate relative suitability (as for the coarse-scale models).

The nature of the BBN models means they are also more prone to linguistic uncertainty, such as vagueness. For example, some of the states for the variables may be particularly subject to this, although this (and another type of linguistic uncertainty, ambiguity) was minimised as much as possible by providing accompanying descriptions of each of the terms. However, the ground moisture variable states, for example, may be more prone to this problem, as different experts may have different perceptions about each of the categories. There may also be some subjective judgment in choosing which state is the most appropriate when recording the variables in the field (which is also a source of measurement error; see above).

Natural variation

Species data exhibit natural variation and it is unlikely that this variation would ever be fully characterised (Elith *et al.*, 2002). Therefore errors arise just as a result of this natural variation and are unavoidable. Models that are used to predict habitat suitability are usually incomplete because they do not incorporate all of the underlying mechanisms for variation (Elith *et al.*, 2002) and natural variation is inherent in species' habitat selection. For example, habitat selection can change with conditions on scales of days and hours for some species, such as butterflies (Dennis and Sparks, 2006) although this would not be applicable for non-mobile species, such as plants. However, changes in occurrence among years, such as due to fluctuating resources (Boyce *et al.*, 2002) or changes in habitat (quality) may occur for a wider range of species. *P. argus*, for example, requires heathland in its early growth stages, so although heathland may be present and suitable for a period of time at a site, as it matures, the site may no longer be suitable and the species will no longer be found there.

Overall error

Evaluation of the models provides details on the omission and commission errors. While these errors are inevitable as species will frequently be absent from suitable habitat and present in apparently unsuitable habitat (as discussed above and in section 1.2.3, Chapter 1) (a rate which may be increased by the use of pseudo-absences), clearly providing these values means that users are aware of their weaknesses. The rates of omission and commission errors are dependent on the suitable/unsuitable (or presence-absence) threshold applied, but by displaying the results from using a range of thresholds (as carried out for this study) and also weighting the errors, the differences can be seen. The omission and commission errors give an overall assessment of the (compound) error in the models (Guisan *et al.*, 2006b). Fewer false negative predictions could be seen as less error because predictions of unsuitability at sites where a species' presence has been observed are clear errors, whereas predictions of suitability at sites where no presence has been observed could be attributed to other factors that limit the actual distributions (such as biotic factors) or to insufficient sampling (Pearson *et al.*, 2004).

Despite the many sources of error discussed above the models generally worked very well, suggesting that this error did not have a serious impact. This may in part be because the prediction sites had the same errors as those used for model building (Barry and Elith, 2006). However, it should be noted that although the evaluation of the models provides an indication of their performance models, in the case of the BBN models in particular, because the limited number of testing sites were not selected at random, the results may be biased.

Environmental change uncertainty

The changes in states used to run the BBNs considering the impact of environmental change were only estimations of potential scenarios, based on available literature, which is unable to provide detailed predictions, particularly in terms of interactions between different factors. These models were therefore based on numerous assumptions and simplifications, so this should be understood when interpreting their results.

As noted in section 5.2.1 (Chapter 5), projections of climate change and its impacts beyond about 2050 are strongly scenario- and model-dependent, with various sources of uncertainty (IPCC, 2007). Impacts research is therefore hampered by uncertainties surrounding regional projections of climate change, particularly precipitation (IPCC, 2007). UKCIP projections (see section 5.2.1, Chapter 5) were determined from climate models which are themselves associated with uncertainties, such as natural variability in the climate system (Harrison *et al.*, 2007). They are also based on different emissions scenarios which result from uncertainty about future trends and behaviours, such as population growth, socio-economic development and technological progress, and how these might influence global emissions of greenhouse gases (Hulme *et al.*, 2002). However, there can be confidence in the overall general trends predicted.

Fischlin *et al*. (2007) suggest that the key uncertainties that limit an ability to project climate change impacts on ecosystems include: inadequate representation of the interactive coupling between ecosystems and the climate system and of the multiple interacting drivers of global change; major biotic feedbacks to the climate system (such as through trace gases from soils in all ecosystems); the net result of changing disturbance regimes (especially through fire, insects and land-use change) on biotic feedbacks to the atmosphere, ecosystem structure, function, biodiversity and ecosystem services; the magnitude of the carbon dioxide fertilisation effect in the terrestrial biosphere and its components over time; the synergistic role of invasive alien species in both biodiversity and ecosystem functioning; and the impacts of interactions between climate change and changes in human use and management of ecosystems as well as other drivers of global environmental change.

This great complexity of natural systems and limited understanding of how different factors may interact suggests that there are fundamental limits to the prediction of potential future species' distributions (Berry *et al.,* 2005b). However predictions can be useful as first approximations of the potential magnitude and broad pattern of future impacts as opposed to being accurate simulations of future species' distributions, providing the limitations are taken into account (Berry *et al.,* 2005b).

6.3. Knowledge contribution

This project was aimed at improving knowledge of, and investigating ways of predicting, distributions of species in the New Forest National Park. As many species are under-recorded, largely due to limited time and resources, effective models of habitat suitability would help to identify potentially suitable sites to focus survey efforts and reduce the time taken to do this. The investigations showed that effective predictive models were possible for projections of the potential distributions of the study species at the scale of the New Forest, an individual protected area. Previously, such an application had been relatively rare. It also showed that relatively straightforward and simple models could be developed (using methods such as random pseudo-absence generation and stepwise variable selection) and particularly that a presence-absence method (GLMs or GAMs) performed better than a presence-only method (ENFA) and could be used based on presence-only data and pseudo-absences. Further, the use of habitat suitability models at the two different scales implemented in this project has so far been rare (Brambilla *et al.,* 2009) and whilst there appears not to be any other studies using both logistic regression (or ENFA) as well as BBNs, the results suggest that this combination of models can be a useful approach.

Although a relatively new approach to modelling habitat suitability, the BBN models proved to be an effective way of capturing the habitat requirements of species and predicting potential habitat suitability at the site scale, demonstrating that this is an effective approach. Although the BBN models are more time consuming to use as they require ground surveys, they may be particularly beneficial for cryptic species or those which are difficult to survey, as well as identifying potentially suitable habitat for (re)colonisation (Lütolf *et al.,* 2006). It was also shown that the BBN approach provides an excellent way of formally capturing expert knowledge about a species and its habitat requirements. This is particularly important for less well known species, where published literature is scarce. The capturing of knowledge in this way also provided this information in a format which was easily interpretable to others.

The nature of the BBN models also allowed the estimation of the potential impacts of climate change on habitat suitability for the selected species habitats, a rare example demonstrating how the effects of climate change might be manifested at the scale of a relatively small protected area. Further, the ability of the models to provide a quantitative indication of such changes at a fine scale, based on models that have been shown to work in predicting current habitat suitability was really important, and an approach that has not been used before. Such models can be easily updated as a greater understanding and the models can also be used as an indication for management.

It is hoped that the success of these models will encourage further interest in developing them for a wider range of species. Having a very visible model of the important factors influencing habitat suitability for a species means that it is easier to apply what is known about the potential impacts of climate change to the different variables in the model. By focussing on what aspects of ‘habitat’ are really important for a species it is hoped that this will help to improve their management and conservation now and in the future. It is important that such changes are anticipated so that appropriate management adaptation can be taken (Hannah *et al.*, 2002). However, it should be noted that the models only predict potential distributions so should only be used as a guide and are not intended to replace current methods but instead add to the ‘toolbox’ of available methods.

None of the species had been modelled before in this way nor had species distribution modelling been carried out in the New Forest. Further, the application of predictive models to rare and endangered species has been relatively uncommon compared to common species (Engler *et al.*, 2004; Guisan *et al.*, 2006a; Matern *et al.*, 2007) and this study shows that successful models can be developed for such species. Application of species distribution modelling to individual protected areas, particularly of the relatively small size of the New Forest has also been limited, as had such modelling in the UK. The results also showed that it is possible to develop effective models at what is a relatively small scale in terms of species distribution modelling, which is frequently carried out at a much larger scales. Similarly, the results also showed that it is possible to model the potential impacts of climate change on species at this scale. Therefore, it is hoped that this work will set an example of how these methods can be used to help improve monitoring of species and understanding of their habitat requirements and aid in their conservation and management in protected areas such as the New Forest and encourage their use.

6.4. Further work

As it has been shown that effective models can be developed for the study species, it would be beneficial to develop models for a wider range of species. Although efforts should initially be aimed at rare or vulnerable species, more common species should not be overlooked as they too may decline and may not necessarily be able to cope any better with environmental change (Gaston and Fuller, 2007). Further, rare species may be dependent on common species and if declines in the more common species were seen this could have significant impacts on the rare species. Predictive models are particularly valuable for cryptic and less well-known species, such as invertebrates, bryophytes and fungi (Berg *et al.*, 2004).

It may also be beneficial to test the performance of some of the newer methods for species distribution modelling (such as some of the methods described in Elith *et al*. (2006)), particularly methods which may be more capable of effective prediction based on very small sample sizes, as this would increase the number of species that could be modelled using the coarse-scale models. However, it may be that successful GLMs could be developed with a smaller sample size, particularly if the sample data was of good quality and the species was fairly specialist, and this would be beneficial to test. Further, although presence-absence models were found to work better than the presence-only method (ENFA), other presence-only methods may perform better than ENFA (Tsoar *et al.*, 2007). Simple models can be used to help to stratify field sampling and identify new occurrence sites and add to the number of records, which can then be used to run potentially better models. Where possible, absence of a species should also be recorded.

Where there are very few records for species and there is limited knowledge on those species, it would be particularly beneficial to develop BBN models. This approach would also be the most useful for assessing the potential impacts of climate change for species. Experts could be encouraged to help develop models for the species for which they have expert knowledge so that such models are available for a larger proportion of the New Forest species. Taking this idea further, a database or index of species and their habitat requirements would mean that this information was readily accessible for those who may wish to use it to aid surveying or management or just to increase their understanding of the species. If new knowledge or understanding of a species becomes available then the models should be updated, and more experts can contribute to the models, in particular filling in the CPTs. It may also be beneficial to include the input from a greater number of experts for some of the current BBN models to be more confident that the values are accurate. Similarly, testing the BBN models with data from a greater range of sites (particularly absence sites) would also be beneficial, to determine whether the models still perform as well.

As well as running the BBN models with changes in the states to assess the impact of climate change, the BBN models could also be expanded further to take account of the potential influences of climate change and management on each of the variables influencing habitat suitability for the species. These could be incorporated using decision or utility nodes (see e.g. Newton *et al.*, 2007; Newton, 2009a). It may also be possible to incorporate some of the other variables discussed in the literature review of the impacts of climate change on species (Appendix 34), but which were not taken account of by the habitat variables in the BBN models. The habitat preferences of species may change as a result of climate change, such as the use of cooler micro-habitats (such as north-facing slopes) (see Appendix 34), so the models would need to be updated accordingly. A limited understanding of the full impacts of climate change is an important issue and, as noted in the New Forest National Park Plan Consultation Draft (New Forest National Park Authority, 2008), more research is needed on the impacts of climate change in the New Forest. The application of the impacts of climate change at the scale of individual species (or communities) in the New Forest is an effective step in that direction.

The use of expert knowledge may be the only way of developing models for some rare species for which relatively little is known and it is important to capture the knowledge of experts, particularly if there are very few experts with that knowledge, so that it is not lost, and can be used by others. BBNs can be constructed by species experts relatively quickly and then be used by experts or non-experts to look for suitable sites for the species in the field. This approach may be particularly useful for cryptic species, or those not present (or easily visible) for very long during the year. However, as the BBN models are not based on the spatial GIS data (although they could be), they cannot be used to predict potential habitat suitability across the whole of the New Forest, as for the models from Chapters 2 and 3. However, the models from Chapters 2 and 3 could be used to determine broadly suitable habitat (as long as these models were reliable) and reduce the number of potentially suitable sites to visit. The BBN models may then be used to determine the suitability in more detail by visiting those sites (in a similar approach to that used by Marcot (2006)). However, certain variables (habitat type and soil type) for some of the BBN models were available as GIS layers and could be used to give an indication of the potential suitability of sites by instantiating part of the model. Therefore, the models can be used to prioritise areas for surveys (Smith *et al.*, 2007).

Another option for poorly known species, or to reduce the number of required models, would be to develop models for key species, such as ‘umbrella’ species or ‘indicator’ species. Lambeck (1997) suggests that a multi-species umbrella approach can be used as a surrogate of vulnerability of all species in an ecosystem, where the requirements of selected ‘focal species’ are presumed to encapsulate those of an array of additional species, with each ‘focal species’ used to define the characteristics of different landscape attributes that must be represented in the landscape if that landscape is to meet the requirements of the species that occur there to persist (Lambeck, 1997). Similarly, certain types of species may make good monitoring or indicator species, such as those which are area-limited, dispersal-limited, resource-limited, process-limited, or keystone species (Noss, 1999). Species possessing these sorts of traits have also been found to frequently be those which are more vulnerable to environmental change and extinction (as discussed in Appendices 31 and 33).

Another possibility for modelling species with small sample sizes, and similar to the approach described above is to develop models for groups of species with similar habitat requirements that are frequently found together. Community models (see e.g. Ferrier *et al.*, 2002a) can use the data for more common species to help support the modelling of associated species with fewer records (Guisan *et al.*, 2006a) and can be used to model either the collective requirements of an assemblage or community, or to make predictions for individual species from a community model in which information from a wider set of species is used to construct a context in which individual species distributions are then described (Barry and Elith, 2006). In a review of methods for modelling species distributions, Elith *et al*. (2006) found that community methods performed well and suggest that they deserve further scrutiny, particularly where data for the species in question is sparse.

A related approach would be to develop BBNs for groups of species, particularly if their requirements are very similar (for example as suggested by Marcot (2006)), or to link models for associated species, such as a plant and its pollinator. This could lead to the modelling of communities (such as through the use of object-oriented Bayesian belief networks, which are networks that contain instance nodes, which represent an instance of another network, creating a hierarchical model; see Newton (2009a)), rather than viewing species in isolation, although this could become very complex, requiring good knowledge of the interactions between species. However, when assessing the potential impacts of environmental change it may be particularly helpful to develop models in this way, although it should be borne in mind that community interactions may change with future environmental change (see Chapter 5). It may also be useful to develop models for specific habitats and the factors influencing their condition, which could be linked in. This is something that would be particularly relevant in the context of environmental change. The use of decision nodes could also be used to indicate the impact of different management strategies for example and aid decision-making (Borsuk *et al.*, 2004; Newton, 2009a).

Due to the unique nature of the New Forest, the models developed as part of this study should not be used to project potential distributions of the species outside of this study area, as the species may have specific associations with the habitats in the New Forest, which are uniquely shaped by a long history of grazing and management for example. Several of the species are known to have slightly different habitat requirements in different parts of the country (or the world). For example, *P. argus* favours south-facing slopes in cooler northern parts of the country, whereas it has no aspect preference in the south (Joy, 1995).

Elith *et al*. (2006) suggest caution in making predictions for sites distant from the geographic domain from which the modelling data were drawn, due to predictions liable to be confounded by unrecognised environmental factors or by large-scale geographic variation in disturbance regimes. Randin *et al*. (2006) also found weak transferability of GLM and GAM models for a range of plant species, but suggest that the capacity for transferability is highly species, region and modelling technique specific. GLM models may be more robust when transferred than GAM models, as overfitting in GAMs can reduce transferability (Randin *et al.*, 2006). Possible factors suggested by Randin *et al*. (2006) for differences in transferability may be environmental (differences between geographic regions) and biotic (intrinsic to each individual species being modelled and to the regional species pool with which it is interacting). The presence-only (ENFA) models should also not be extrapolated to different locations as they are based on the comparison between the locations where the species has been observed and the available habitat (Hirzel, 2008). However, the approaches themselves could be used in areas outside the New Forest, particularly, as this research has shown that they can be used at a relatively small scale.

The identification of important features of habitat suitability from the models (particularly the BBNs) could be used as a starting point for determining potential habitat suitability of sites outside the New Forest. Further, the New Forest National Park should not be seen in isolation as, particularly in the face of environmental change, sites outside of the New Forest may become increasingly important for New Forest populations as species distributions potentially change (Hossell *et al.,* 2000; Berry *et al.*, 2007a).

6.5. Conclusion

The overwhelming complexity of the natural world presents a fundamental problem in modelling natural systems (O'Hanley, 2005b). However, it has been possible to develop effective predictive habitat suitability models for selected species in the

New Forest National Park and provide an indication of the potential impact of climate change on the selected species. However, it is important that such models are seen as approximations and guides rather than accurate predictions. Nonetheless, such models can help to improve surveying and knowledge of the current distributions of species and help to determine the potential impacts of climate change.

REFERENCES

References

Alcamo, J., Moreno, J. M., Nováky, B., Bindi, M., Coroboy, R., Devoy, R. J. N., Giannakopoulos, C., Martin, E., Olesen, J. E. & Shvidenko, A. (2007). Europe. IN: Parry, M. L., Canziani, O. F., Palutikof, J. P., Van Der Linden, P. J. & Hanson, C. E. (Eds.). *Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge: Cambridge University Press, 541-580.

Anderson, J. M. & Hetherington, S. L. (1999). Temperature, nitrogen availability and mixture effects on the decomposition of heather [*Calluna vulgaris* (L.) Hull] and bracken [*Pteridium aquilinum* (L.) Kuhn] litters. *Functional Ecology,* 13**,** 116-124.

Anderson, R. P., Lew, D. & Peterson, A. T. (2003). Evaluating predictive models of species' distributions: criteria for selecting optimal models. *Ecological Modelling,* 162(3)**,** 211-232.

Araújo, M. B. & Guisan, A. (2004). Five (or so) challenges for species distribution modelling. *Journal of Biogeography,* 33(10)**,** 1677-1688.

Araújo, M. B., Pearson, R. G., Thuiller, W. & Erhard, M. (2005). Validation of species-climate impact models under climate change. *Global Change Biology,* 11(9)**,** 1504-1513.

Asher, J., Warren, M., Fox, R., Harding, P., Jeffcoate, G. & Jeffcoate, S. (2001). *Millennium atlas of butterflies in Britain and Ireland.* Oxford: Oxford University Press.

Austin, M. (2007). Species distribution models and ecological theory: A critical assessment and some possible new approaches. *Ecological Modelling,* 200(1-2)**,** 1-19.

Austin, M. P. (2002a). Case studies of the use of environmental gradients in vegetation and fauna modelling: theory and practice in Australia and New Zealand. IN: Scott, J. M., Heglund, P. J., Morrison, M. L., Haufler, J. B., Raphael, M. G., Wall, W. A. & Samson, F. B. (Eds.). *Predicting species occurrences: issues of accuracy and scale.* Washington: Island Press, 73-82.

Austin, M. P. (2002b). Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling,* 157(2-3)**,** 101-118.

Austin, M. P. & Meyers, J. A. (1996). Current approaches to modelling the environmental niche of eucalypts: Implication for management of forest biodiversity. *Forest Ecology and Management,* 85(1-3)**,** 95-106.

Bailey, K. E. J., Baker, R. R., Brakefield, P. M., Bretherton, R. F., Emmet, A. M., Lear, N. W., Melling, T., Pollard, E., Porter, K. J., Pratt, C. R., Shreeve, T. G., Warren, M. S., Willmott, K. J. & Wilson, A. (1989). Nymphalidae. IN: Emmet, A. M. & Heath, J. (Eds.). *The Moths and Butterflies of Great Britain and Ireland.* Colchester: Harley Books, 180-285.

Bailey, S. (2007). Increasing connectivity in fragmented landscapes: An investigation of evidence for biodiversity gain in woodlands. *Forest Ecology and Management,* 238(1-3)**,** 7-23.

Bale, J. S., Masters, G. J., Hodkinson, I. D., Awmack, C., Bezemer, T. M., Brown, V. K., Butterfield, J., Buse, A., Coulson, J. C., Farrar, J., Good, J. E. G., Harrington, R., Hartley, S., Jones, T. H., Lindroth, R. L., Press, M. C., Symrnioudis, I., Watt, A. D. & Whittaker, J. B. (2002). Herbivory in global climate change research: direct effects of rising temperature on insect herbivores. *Global Change Biology,* 8(1)**,** 1-16.

Balmford, A., Crane, P., Dobson, A., Green, R. E. & Mace, G. M. (2005). The 2010 challenge: data availability, information needs and extraterrestrial insights. *Philosophical Transactions of the Royal Society B-Biological Sciences,* 360(1454)**,** 221-228.

Barry, S. & Elith, J. (2006). Error and uncertainty in habitat models. *Journal of Applied Ecology,* 43(3)**,** 413-423.

Basille, M., Calenge, C., Marboutin, E., Andersen, R. & Gaillard, J. M. (2008). Assessing habitat selection using multivariate statistics: Some refinements of the ecological-niche factor analysis. *Ecological Modelling,* 211(1-2)**,** 233-240.

Bates, D., Maechler, M. & Dai, B. (2008). *lme4: Linear mixed-effects models using S4 classes (R package)* (0.999375-28). [computer program]. Available from: http://lme4.r-forge.r-project.org/.

BBC News (2009). *Second brood for rare butterfly*. London: BBC. Available from: http://news.bbc.co.uk/1/hi/sci/tech/8208428.stm [Accessed 19 August 2009].

Ben Wu, X. & Smeins, F. E. (2000). Multiple-scale habitat modeling approach for rare plant conservation. *Landscape and Urban Planning,* 51(1)**,** 11-28.

Berg, A., Gardenfors, U. & von Proschwitz, T. (2004). Logistic regression models for predicting occurrence of terrestrial molluscs in southern Sweden - importance of environmental data quality and model complexity. *Ecography,* 27(1)**,** 83-93.

Berry, P. M., Dawson, T. P., Harrison, P. A., Pearson, R. & Butt, N. (2003). The sensitivity and vulnerability of terrestrial habitats and species in Britain and Ireland to climate change. *Journal for Nature Conservation,* 11(1)**,** 15-23.

Berry, P. M., Harrison, P. A., Dawson, T. P. & Walmsley, C. A. (Eds.) (2005a). *Modelling Natural Resource Responses to Climate Change (MONARCH): A Local Approach,* Oxford: UKCIP Technical Report.

Berry, P. M., Masters, G. J., Harrison, P. A., Hossell, J. E., Crick, H. Q. P., Freeman, S. N., Ellis, N. E., Orr, H. G., Walmsley, C. A. & Watts, O. (2005b). *Conclusions and future research*. Oxford: UKCIP Techincal Report.

Berry, P. M., O'Hanley, J. R. & Thomson, C. L. (2007a). *Spatial planning for chalk grassland and heathland habitats in Hampshire. Planning for biodiversity in a changing climate - BRANCH project final report*. UK: Natural England.

Berry, P. M., Thomson, C. L., Masters, G. J. & O'Hanley, J. R. (2007b). *MONARCH 3 - Technical report. Application of the MONARCH methodology to selected BAP species in Britain and Ireland*. Oxford: UK Climate Impacts Programme.

Beyer, H. L. (2006). *Hawth's Analysis Tools* (3.26). [computer program]. Available from: http://www.spatialecology.com/.

Bivand, R. (2009). *spdep: Spatial dependence: weighting schemes, statistics and models (R package)* (0.4-34). [computer program]. Available from: http://cran.r-project.org/web/packages/spdep/index.html.

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H. & White, J. S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution,* 24(3)**,** 127-135.

Borsuk, M. E., Stow, C. A. & Reckhow, K. H. (2004). A Bayesian network of eutrophication models for synthesis, prediction, and uncertainty analysis. *Ecological Modelling,* 173(2-3)**,** 219-239.

Botanical Society of the British Isles (2008a). *Gladiolus illyricus*. Available from: http://www.bsbi.org.uk/html/gladiolus_illyricus.html [Accessed 10 November 2008].

Botanical Society of the British Isles (2008b). *Pilularia globulifera*. Available from: http://www.bsbi.org.uk/html/pilularia_globulifera.html [Accessed 5 November 2008].

Boyce, M. S., Vernier, P. R., Nielsen, S. E. & Schmiegelow, F. K. A. (2002). Evaluating resource selection functions. *Ecological Modelling,* 157(2-3)**,** 281-300.

Brambilla, M., Casale, F., Bergero, V., Crovetto, G. M., Falco, R., Negri, I., Siccardi, P. & Bogliani, G. (2009). GIS-models work well, but are not enough: Habitat preferences of *Lanius collurio* at multiple levels and conservation implications. *Biological Conservation,* 142(10)**,** 2033-2042.

BRANCH Partnership (2007). *Planning for biodiversity in a changing climate - BRANCH project final report*. UK: Natural England.

Braunisch, V. & Suchant, R. (2005). A model for evaluating the 'habitat potential' of a landscape for capercaillie *Tetrao urogallus*: a tool for conservation planning. *Wildlife Biology,* 13(Suppl. 1)**,** 21-33.

Bretherton, R. F., Davies, M., Dempster, J. P., Emmet, A. M., Heath, J., Melling, T., Morton, A. C., Oates, M., Porter, K. J., Read, M. J., Thomas, J. A. & Willmott, K. J. (1989). Lycaenidae. IN: Emmet, A. M. & Heath, J. (Eds.). *The Moths and Butterflies of Great Britain and Ireland.* Colchester: Harley Books, 118-179.

Brewis, A., Bowman, P. & Rose, F. (1996). The vascular flora. IN: Brewis, A., Bowman, P. & Rose, F. (Eds.). *The Flora of Hampshire.* Colchester: Harley Books, 91 - 305.

Brito, J. C., Crespo, E. G. & Paulo, O. S. (1999). Modelling wildlife distributions: Logistic multiple regression vs overlap analysis. *Ecography,* 22(3)**,** 251-260.

Broadmeadow, M. (2000). *Climate change - implications for forestry in Britain. Forestry Commission information note 31*. Edinburgh: Forestry Commission.

Broadmeadow, M. & Ray, D. (2005). *Climate change and British woodland. Information note.* Edinburgh: Forestry Commission.

Brock, P. (2009). Grayling *Hipparchia semele*. IN: Barker, L., Norriss, T., Green, D., Bailey, K., Harper, A., Jones, J., Wall, M. & Welch, P. (Eds.). *Hampshire and Isle of Wight butterfly and moth report, 2008.* Wareham, Dorset: Hampshire and Isle of Wight Branch of Butterfly Conservation, 85-86.

Brock, T. C. M. & van Vierssen, W. (1992). Climatic change and hydrophyte-dominated communities in inland wetland ecosystems. *Wetlands Ecology and Management,* 2(1/2)**,** 37-49.

Broennimann, O., Thuiller, W., Hughes, G., Midgley, G. F., Alkemade, J. M. R. & Guisan, A. (2006). Do geographic distribution, niche property and life form explain plants' vulnerability to global change? *Global Change Biology,* 12(6)**,** 1079-1093.

Brotons, L., Herrando, S. & Pla, M. (2007). Updating bird species distribution at large spatial scales: applications of habitat modelling to data from long-term monitoring programs. *Diversity and Distributions,* 13(3)**,** 276-288.

Brotons, L., Thuiller, W., Araujo, M. B. & Hirzel, A. H. (2004). Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. *Ecography,* 27(4)**,** 437-448.

Brouwers, N. C. (2008). *Analysis of the ecological principles underpinning forest landscape restoration: A case study of Wood cricket (Nemobius sylvestris) on the Isle of Wight (UK)*. Thesis (Doctor of Philosophy). Bournemouth University.

Brown, V. K. (1978). Variation in voltinism and diapause intensity in *Nemobius sylvestris* (Bosc.) (Orthoptera: Gryllidae). *Journal of Natural History,* 12**,** 461-472.

Buchanan, A. (2007). *The taxonomic status of Gladiolus illyricus (Iridaceae) in Britain*. Thesis (Master of Science). Imperial College London.

Buckland, S. M., Grime, J. P., Hodgson, J. G. & Thompson, K. (1997). A comparison of plant responses to the extreme drought of 1995 in northern England. *Journal of Ecology,* 85(6)**,** 875-882.

Burgman, M. A. (2005). *Risks and decisions for conservation and environmental management.* Cambridge: Cambridge University Press.

Butterfly Conservation (2007). *Annual review of priority butterflies in Hampshire and Isle of Wight 2007*. Wareham, Dorset: Butterfly Conservation.

Butterfly Conservation (2009). *A-Z of butterflies: Grayling*. Available from: http://www.butterfly-conservation.org/Butterfly/32/Butterfly.html?ButterflyId=24 [Accessed 19 July 2009].

Cablk, M., White, D. & Kiester, A. R. (2002). Assessment of spatial autocorrelation in empirical models in ecology. IN: Scott, J. M., Heglund, P. J., Morrison, M. L., Haufler, J. B., Raphael, M. G., Wall, W. A. & Samson, F. B. (Eds.). *Predicting species occurrences: issues of accuracy and scale.* Washington: Island Press, 429-440.

Callaghan, T. V., Carlsson, B. A., Jonsdottir, I. S., Svensson, B. M. & Jonasson, S. (1992). Clonal plants and environmental change: introduction to the proceedings and summary. *Oikos*, 63(3)**,** 341-347.

Caporn, S. J. M., Brooks, A. L., Press, M. C. & Lee, J. A. (1999). Effects of long-term exposure to elevated CO_2 and increased nutrient supply on bracken (*Pteridium aquilinum*). *Functional Ecology,* 13**,** 107-115.

Carl, G. & Kühn, I. (2007). Analyzing spatial autocorrelation in species distributions using Gaussian and logit models. *Ecological Modelling,* 207(2-4)**,** 159-170.

Carpenter, S. R., Fisher, S. G., Grimm, N. B. & Kitchell, J. F. (1992). Global change and freshwater ecosystems. *Annual Review of Ecology and Systematics,* 23**,** 119-139.

Cayuela, L., Golicher, D. J., Newton, A. C., Kolb, M., de Alburquerque, F. S., Arets, E. J. M. M., Alkemade, J. R. M. & Pérez, A. M. (2009). Species distribution modelling in the tropics: problems, potentialities and the role of biological data for effective species conservation. *Tropical Conservation Science,* In press.

Chatters, C. (2006). The New Forest - National Park status for a medieval survivor. *British Wildlife,* 18(2)**,** 110-119.

Chefaoui, R. M., Hortal, J. & Lobo, J. M. (2005). Potential distribution modelling, niche characterization and conservation status assessment using GIS tools: a case study of Iberian *Copris* species. *Biological Conservation,* 122(2)**,** 327-338.

Chefaoui, R. M. & Lobo, J. M. (2007). Assessing the conservation status of an Iberian moth using pseudo-absences. *Journal of Wildlife Management,* 71(8)**,** 2507-2516.

Chefaoui, R. M. & Lobo, J. M. (2008). Assessing the effects of pseudo-absences on predictive distribution model performance. *Ecological Modelling,* 210(4)**,** 478-486.

Clapham, A. R., Tutin, T. G. & Moore, D. M. (1990). *Flora of the British Isles.* 3rd ed. Cambridge: Cambridge University Press.

Clark Labs (2006). *IDRISI Andes* (15.01). [computer program]. Worcester, MA, USA: Available from: http://www.clarklabs.org/.

Clarke, E. J. & Barton, B. A. (2000). Entropy and MDL discretization of continuous variables for Bayesian belief networks. *International Journal of Intelligent Systems,* 15(1)**,** 61-92.

Clemen, R. T. & Winkler, R. L. (1999). Combining probability distributions from experts in risk analysis. *Risk Analysis,* 19(2)**,** 187-203.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement,* 20(1)**,** 37-46.

Cook, C. & Harrison, P. A. (2001). *Climate change and nature conservation in Britain and Ireland. MONARCH - Modelling Natural Resource Responses to Climate Change: Summary report.* Oxford: UK Climate Impacts Programme.

Cowley, M. J. R., Wilson, R. J., Leon-Cortes, J. L., Gutierrez, D., Bulman, C. R. & Thomas, C. D. (2000). Habitat-based statistical models for predicting the spatial distribution of butterflies and day-flying moths in a fragmented landscape. *Journal of Applied Ecology,* 37**,** 60-72.

Cox, J. (1999). *The biodiversity of animal dung*. Eastleigh: Hampshire and Isle of Wight Wildlife Trust.

Cox, J. H. S. & Pickess, B. P. (1999). Observations concerning the ecology of Nail Fungus *Poronia punctata,* recently discovered in Dorset. *Dorset Proceedings,* 121**,** 129-132.

Cox, J. H. S., Pickess, B. P. & Peters, A. (2005). Nail Fungus *Poronia punctata* in Dorset, 1999 to 2005: population changes and ecological observations. *Dorset Proceedings,* 127**,** 95-99.

Crawley, M. J. (2007). *The R book.* Chichester: Wiley.

Davies, Z. G., Wilson, R. J., Coles, S. & Thomas, C. D. (2006). Changing habitat associations of a thermally constrained species, the silver-spotted skipper butterfly, in response to climate warming. *Journal of Animal Ecology,* 75(1)**,** 247-256.

Davis, M. B. (1989). Lags in vegetation response to greenhouse warming. *Climatic Change,* 15(1-2)**,** 75-82.

Dawson, T. P., Berry, P. M. & Kampa, E. (2003). Climate change impacts on freshwater wetland habitats. *Journal for Nature Conservation,* 11**,** 25-30.

DEFRA (2001). *Literature review of the implications of climate change for species, habitats and the wider UK countryside*. London: Department for Environment, Food and Rural Affairs.

deGroot, R. S., Ketner, P. & Ovaa, A. H. (1995). Selection and use of bio-indicators to assess the possible effects of climate change in Europe. *Journal of Biogeography,* 22(4-5)**,** 935-943.

Dennis, R. L. H. (1993). *Butterflies and climate change.* Manchester: Manchester University Press.

Dennis, R. L. H. & Bardell, P. (1996). The impact of extreme weather events on local populations of *Hipparchia semele* (L.) (Nymphalidae) and *Plebejus argus* (L.) (Lycaenidae): hindsight, inference and lost opportunities. *Entomologist's Gazette,* 47**,** 211-225.

Dennis, R. L. H. & Shreeve, T. G. (1991). Climatic change and the British butterfly fauna: opportunities and contraints. *Biological Conservation,* 55(1)**,** 1-16.

Dennis, R. L. H. & Sparks, T. H. (2006). When is a habitat not a habitat? Dramatic resource use changes under differing weather conditions for the butterfly *Plebejus argus*. *Biological Conservation,* 129(3)**,** 291-301.

Dennis, R. L. H., Sparks, T. H. & Hardy, P. B. (1999). Bias in butterfly distribution maps: the effects of sampling effort. *Journal of Insect Conservation,* 3(1)**,** 33-42.

Derksen, S. & Keselman, H. J. (1992). Backward, forward and stepwise automated subset-selection algorithms: Frquency of obtaining authentic and noise variables. *British Journal of Mathematical & Statistical Psychology,* 45**,** 265-282.

Dettmers, R., Buehler, D. A. & Bartlett, J. B. (2002). A test and comparison of wildlife-habitat modelling techniques for predicting bird occurrence at a regional scale. IN: Scott, J. M., Heglund, P. J., Morrison, M. L., Haufler, J. B., Raphael, M. G., Wall, W. A. & Samson, F. B. (Eds.). *Predicting species occurrences: issues of accuracy and scale.* Washington: Island Press, 607-615.

Diack, I. (1999). Ecology. IN: Wade, S., Hossell, J., Hough, M. & Fenn, C. (Eds.). *Rising to the challenge - The impacts of climate change in the South East: Technical report.* Epsom: WS Atkins, 44-49.

Dickson, G. (1997). Fungi are not plants - practical problems and conservation. *British Wildlife,* 9(1)**,** 17-21.

Dickson, G. C. & Leonard, A. (1996). *Fungi of the New Forest: A mycota.* Stourbridge: British Mycological Society.

Diniz-Filho, J. A. F., Bini, L. M. & Hawkins, B. A. (2003). Spatial autocorrelation and red herrings in geographical ecology. *Global Ecology and Biogeography,* 12(1)**,** 53-64.

Dormann, C. F. (2007). Effects of incorporating spatial autocorrelation into the analysis of species distribution data. *Global Ecology and Biogeography,* 16(2)**,** 129-138.

Dormann, C. F., McPherson, J. M., Araujo, M. B., Bivand, R., Bolliger, J., Carl, G., Davies, R. G., Hirzel, A., Jetz, W., Kissling, W. D., Kuhn, I., Ohlemuller, R., Peres-Neto, P. R., Reineking, B., Schroder, B., Schurr, F. M. & Wilson, R. (2007a). E5171. Electronic appendix for: Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography,* 30(5)**,** 609-628.

Dormann, C. F., McPherson, J. M., Araujo, M. B., Bivand, R., Bolliger, J., Carl, G., Davies, R. G., Hirzel, A., Jetz, W., Kissling, W. D., Kuhn, I., Ohlemuller, R., Peres-Neto, P. R., Reineking, B., Schroder, B., Schurr, F. M. & Wilson, R. (2007b). Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography,* 30(5)**,** 609-628.

Dreisig, H. (1995). Thermoregulation and flight activity in territorial male graylings, *Hipparchia semele* (Satyridae), and large skippers, *Ochlodes venata* (Hesperiidae). *Oecologia,* 101(2)**,** 169-176.

Eastman, J. R. (2006). *IDRISI Andes guide to GIS and image processing (manual version 15.00).* Worcester, MA: Clark Labs, Clark University.

Edwards, P. J. & Hollis, S. (1982). The distribution of excreta on New Forest grassland used by cattle, ponies and deer. *Journal of Applied Ecology,* 19(3)**,** 953-964.

Elith, J. & Burgman, M. (2002). Predictions and their validation: rare plants in the Central Highlands, Victoria, Australia. IN: Scott, J. M., Heglund, P. J., Morrison, M. L., Haufler, J. B., Raphael, M. G., Wall, W. A. & Samson, F. B. (Eds.). *Predicting species occurrences: issues of accuracy and scale.* Washington: Island Press, 303-313.

Elith, J. & Burgman, M. A. (2003). Habitat models for population viability analysis. IN: Brigham, C. A. & Schwartz, M. W. (Eds.). *Population viability in plants.* Heidelberg: Springer-Verlag, 203-235.

Elith, J., Burgman, M. A. & Regan, H. M. (2002). Mapping epistemic uncertainties and vague concepts in predictions of species distribution. *Ecological Modelling,* 157(2-3)**,** 313-329.

Elith, J., Graham, C. H., Anderson, R. P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R. J., Huettmann, F., Leathwick, J. R., Lehmann, A., Li, J., Lohmann, L. G., Loiselle, B. A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J. M., Peterson, A. T., Phillips, S. J., Richardson, K., Scachetti-Pereira, R., Schapire, R. E., Soberon, J., Williams, S., Wisz, M. S. & Zimmermann, N. E. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography,* 29(2)**,** 129-151.

Emmett, B. A., Beier, C., Estiarte, M., Tietema, A., Kristensen, H. L., Williams, D., Penuelas, J., Schmidt, I. & Sowerby, A. (2004). The response of soil processes to climate change: Results from manipulation studies of shrublands across an environmental gradient. *Ecosystems,* 7(6)**,** 625-637.

Engler, R., Guisan, A. & Rechsteiner, L. (2004). An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. *Journal of Applied Ecology,* 41(2)**,** 263-274.

English Nature (1996). *Reasons for designating the New Forest SSSI.* Available from: http://www.english-nature.org.uk/citation/citation_photo/1003036.pdf [Accessed 6 October 2006].

ESRI (2005). *ArcView* (Version 9.1). [computer program]. Available from: http://www.esri.com/.

Evans, H., Straw, N. & Watt, A. (2002). Climate change: implications for insect pests. IN: Broadmeadow, M. (Ed.) *Climate change: impacts on UK forests. Forestry Commission bulletin 125*. Edinburgh: Forestry Commission, 99-118.

Eveling, D. W., Wilson, R. N., Gillespie, E. S. & Bataille, A. (1990). Environmental effects on sporocarp counts over fourteen years in a forest area. *Mycological Research,* 94(7)**,** 998-1002.

Ewald, N. C. (2008). *The impact of climate change on temporary pond macroinvertebrate communities*. Thesis (Doctor of Philosophy). University of Sussex.

Ewers, R. M. & Didham, R. K. (2006). Confounding factors in the detection of species responses to habitat fragmentation. *Biological Reviews,* 81(1)**,** 117-142.

Eyre, M. D., Foster, G. N. & Young, A. G. (1993). Relationships between water-beetle distributions and climatic variables - a possible index for monitoring global climatic change. *Archiv Fur Hydrobiologie,* 127(4)**,** 437-450.

Eyre, M. D., Rushton, S. P., Luff, M. L. & Telfer, M. G. (2004). Predicting the distribution of ground beetle species (Coleoptera, Carabidae) in Britain using land cover variables. *Journal of Environmental Management,* 72(3)**,** 163-174.

Eyre, M. D., Rushton, S. P., Luff, M. L. & Telfer, M. G. (2005). Investigating the relationships between the distribution of British ground beetle species (Coleoptera, Carabidae) and temperature, precipitation and altitude. *Journal of Biogeography,* 32(6)**,** 973-983.

Fei, S. L., Schibig, J. & Vance, M. (2007). Spatial habitat modeling of American chestnut at Mammoth Cave National Park. *Forest Ecology and Management,* 252(1-3)**,** 201-207.

Ferrier, S., Drielsma, M., Manion, G. & Watson, G. (2002a). Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. II. Community-level modelling. *Biodiversity and Conservation,* 11(12)**,** 2309-2338.

Ferrier, S. & Watson, G. (1997). *An evaluation of the effectiveness of environmental surrogates and modelling techniques in predicting the distribution of biological diversity. Consultancy report to the Biodiversity Convention and*

Strategy Section of the Biodiversity Group, Environment Australia. Armidale: NSW National Parks and Wildlife Service.

Ferrier, S., Watson, G., Pearce, J. & Drielsma, M. (2002b). Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. I. Species-level modelling. *Biodiversity and Conservation,* 11(12)**,** 2275-2307.

Fielding, A. H. (2002). What are the appropriate characteristics of an accuracy measure? IN: Scott, J. M., Heglund, P. J., Morrison, M. L., Haufler, J. B., Raphael, M. G., Wall, W. A. & Samson, F. B. (Eds.). *Predicting species occurrences: issues of accuracy and scale.* Washington: Island Press, 271-280.

Fielding, A. H. & Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation,* 24(1)**,** 38-49.

Fischer, M. & Stocklin, J. (1997). Local extinctions of plants in remnants of extensively used calcareous grasslands 1950-1985. *Conservation Biology,* 11(3)**,** 727-737.

Fischlin, A., Midgley, G. F., Price, J. T., Leemans, R., Gopal, B., Turley, C., Rounsevell, M. D. A., Dube, O. P., Tarazona, J. & Vellchko, A. A. (2007). Ecosystems, their properties, goods and services. IN: Parry, M. L., Canziani, O. F., Palutikof, J. P., Van Der Linden, P. J. & Hanson, C. E. (Eds.). *Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge, UK: Cambridge University Press, 211-272.

Fitter, A. H. & Fitter, R. S. R. (2002). Rapid changes in flowering time in British plants. *Science,* 296(5573)**,** 1689-1691.

Ford, H., Peat, H. & Fitter, A. (2008a). *The ecological characteristics for Chamaemelum nobile*. York: University of York. Available from: http://www.ecoflora.co.uk/search_ecochars.php?plant_no=1690590010 [Accessed 5 November 2008].

Ford, H., Peat, H. & Fitter, A. (2008b). *The ecological characteristics for Gladiolus illyricus*. York: University of York. Available from:

http://www.ecoflora.co.uk/search_ecochars.php?plant_no=1880100010 [Accessed 5 November 2008].

Ford, H., Peat, H. & Fitter, A. (2009). *The ecological characteristics for Galium constrictum*. York: University of York. Available from: http://www.ecoflora.co.uk/search_ecochars.php?plant_no=1440050150 [Accessed 20 July 2009].

Forestry Commission (2002a). *New Forest Fact File: Ecology and management: The New Forest habitats*. Lyndhurst, Hampshire: Forestry Commission. Available from: http://www.forestry.gov.uk/forestry/INFD-62LHYL [Accessed 18 August 2009].

Forestry Commission (2002b). *New Forest Fact File: Open forest management: Bracken control in the New Forest*. Lyndhurst, Hampshire: Forestry Commission. Available from: http://www.forestry.gov.uk/forestry/INFD-62LHYL [Accessed 18 August 2009].

Forestry Commission (2004). *New Forest Fact File: Tourism and recreation: Conflicts in the New Forest*. Lyndhurst, Hampshire: Forestry Commission. Available from: http://www.forestry.gov.uk/forestry/INFD-62LHYL [Accessed 18 August 2009].

Forestry Commission (2006a). *New Forest - Conservation*. Available from: http://www.forestry.gov.uk/forestry/infd-6a5kw3 [Accessed 6 October 2006].

Forestry Commission (2006b). *New Forest - Fungi*. Available from: http://www.forestry.gov.uk/forestry/infd-6e3gaz [Accessed 6 October 2006].

Forestry Commission (2007). *Management Plan. Part B: The Crown Lands. Draft: November 2007*. Lyndhurst: Forestry Commission.

Fox, J. (2002). *An R and S-Plus companion to applied regression.* Thousand Oaks, California: Sage.

Fox, R., Warren, M. S., Asher, J., Brereton, T. M. & Roy, D. B. (2007). *The state of Britain's butterflies 2007.* Wareham, Dorset: Butterfly Conservation and the Centre for Ecology and Hydrology.

Franklin, J. (1998). Predicting the distribution of shrub species in southern California from climate and terrain-derived variables. *Journal of Vegetation Science,* 9(5)**,** 733-748.

Gabbutt, P. D. (1959). The bionomics of the wood cricket, *Nemobius sylvestris* (Orthoptera: Gryllidae). *The Journal of Animal Ecology,* 28(1)**,** 15-42.

Gallagher, K., Graham, M. & Colas, S. (2007). *PROGRESS (Promotion and Guidance for Recreation on Ecologically Sensitive Areas) Project Handbook* Lyndhurst: PROGRESS project.

Gange, A. C., Gange, E. G., Sparks, T. H. & Boddy, L. (2007). Rapid and recent changes in fungal fruiting patterns. *Science,* 316(5821)**,** 71-71.

García-Barros, E. (1988). Delayed ovarian maturation in the butterfly *Hipparchia semele* as a possible response to summer drought. *Ecological Entomology,* 13(4)**,** 391-398.

Garthwaite, P. H., Kadane, J. B. & O'Hagan, A. (2005). Statistical methods for eliciting probability distributions. *Journal of the American Statistical Association,* 100(470)**,** 680-700.

Gaston, K. J. & Fuller, R. A. (2007). Biodiversity and extinction: losing the common and the widespread. *Progress in Physical Geography,* 31(2)**,** 213-225.

Gibson, L., Barrett, B. & Burbidge, A. (2007). Dealing with uncertain absences in habitat modelling: a case study of a rare ground-dwelling parrot. *Diversity and Distributions,* 13(6)**,** 704-713.

Gibson, L. A., Wilson, B. A., Cahill, D. M. & Hill, J. (2004). Spatial prediction of rufous bristlebird habitat in a coastal heathland: a GIS-based approach. *Journal of Applied Ecology,* 41(2)**,** 213-223.

Glenn, E. M. & Ripple, W. J. (2004). On using digital maps to assess wildlife habitat. *Wildlife Society Bulletin,* 32(3)**,** 852-860.

Gordon, C., Woodin, S. J., Mullins, C. E. & Alexander, I. J. (1999). Effects of environmental change, including drought, on water use by competing *Calluna vulgaris* (heather) and *Pteridium aquilinum* (bracken). *Functional Ecology,* 13**,** 96-106.

Graham, C. H., Elith, J., Hijmans, R. J., Guisan, A., Peterson, A. T. & Loiselle, B. A. (2008). The influence of spatial errors in species occurrence data used in distribution models. *Journal of Applied Ecology,* 45(1)**,** 239-247.

Grant, M. J. & Edwards, M. E. (2005). Conserving idealized landscapes: past history, public perception and future management in the New Forest (UK). *Vegetation History and Archaeobotany,* 17(5)**,** 551-562.

Green, D. G. (2000). *The status of Lepidoptera in the New Forest*. Wareham: Butterfly Conservation.

Gu, Y. Q., McNicol, J. W., Peiris, D. R., Crawford, J. W., Marshall, B. & Jefferies, R. A. (1996). A belief network-based system for predicting future crop production. *Ai Applications,* 10(1)**,** 13-24.

Guisan, A. (2002). Semiquantitative response models for predicting the spatial distribution of plant species. IN: Scott, J. M., Heglund, P. J., Morrison, M. L., Haufler, J. B., Raphael, M. G., Wall, W. A. & Samson, F. B. (Eds.). *Predicting species occurrences: issues of accuracy and scale.* Washington: Island Press, 315-326.

Guisan, A., Broennimann, O., Engler, R., Vust, M., Yoccoz, N. G., Lehmann, A. & Zimmermann, N. E. (2006a). Using niche-based models to improve the sampling of rare species. *Conservation Biology,* 20(2)**,** 501-511.

Guisan, A., Edwards, T. C. & Hastie, T. (2002). Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling,* 157(2-3)**,** 89-100.

Guisan, A. & Hofer, U. (2003). Predicting reptile distributions at the mesoscale: relation to climate and topography. *Journal of Biogeography,* 30(8)**,** 1233-1243.

Guisan, A., Lehmann, A., Ferrier, S., Austin, M., Overton, J. M. C., Aspinall, R. & Hastie, T. (2006b). Making better biogeographical predictions of species' distributions. *Journal of Applied Ecology,* 43(3)**,** 386-392.

Guisan, A., Theurillat, J. P. & Kienast, F. (1998). Predicting the potential distribution of plant species in an Alpine environment. *Journal of Vegetation Science,* 9(1)**,** 65-74.

Guisan, A. & Thuiller, W. (2005). Predicting species distribution: offering more than simple habitat models. *Ecology Letters,* 8(9)**,** 993-1009.

Guisan, A. & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling,* 135(2-3)**,** 147-186.

Hannah, L., Midgley, G. F. & Millar, D. (2002). Climate change-integrated conservation strategies. *Global Ecology and Biogeography,* 11(6)**,** 485-495.

Harrell, F. E. (2001). *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis.* New York: Springer-Verlag.

Harrell, F. E., Lee, K. L. & Mark, D. B. (1996). Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistic in Medicine,* 15(4)**,** 361-387.

Harrison, P. A., Berry, P. M. & Dawson, T. P. (Eds.) (2001). *Climate change and nature conservation in Britain and Ireland: Modelling natural resource responses to climate change (the MONARCH project),* Oxford: UKCIP Techincal Report.

Harrison, P. A., Thomson, C. L., O'Hanley, J. R. & Berry, P. M. (2007). *Quantifying uncertainties in the simulation of potential suitable climate space for species due to projections of future climate* Oxford: UKCIP Technical Report.

Hastie, T. (2009a). *gam: Generalized Additive Models* (R package version 1.0). [computer program]. Available from: http://CRAN.R-project.org/package=gam.

Hastie, T. (2009b). *Package 'gam' reference manual*. Available from: http://cran.r-project.org/web/packages/gam/gam.pdf [Accessed 12 June 2009].

Hegland, S. J., Nielsen, A., Lazaro, A., Bjerknes, A. L. & Totland, O. (2009). How does climate warming affect plant-pollinator interactions? *Ecology Letters,* 12(2)**,** 184-195.

Heijmans, M., Mauquoy, D., van Geel, B. & Berendse, F. (2008). Long-term effects of climate change on vegetation and carbon dynamics in peat bogs. *Journal of Vegetation Science,* 19(3)**,** 307-320.

Henle, K., Davies, K. F., Kleyer, M., Margules, C. & Settele, J. (2004). Predictors of species sensitivity to fragmentation. *Biodiversity and Conservation,* 13(1)**,** 207-251.

Hepinstall, J. A., Krohn, W. B. & Sader, S. A. (2002). Effects of niche width on the performance and agreement of avian habitat models. IN: Scott, J. M., Heglund, P. J., Morrison, M. L., Haufler, J. B., Raphael, M. G., Wall, W. A. & Samson, F. B. (Eds.). *Predicting species occurrences: issues of accuracy and scale.* Washington: Island Press, 593-606.

Hernandez, P. A., Graham, C. H., Master, L. L. & Albert, D. L. (2006). The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography,* 29(5)**,** 773-785.

Hickling, R., Roy, D. B., Hill, J. K., Fox, R. & Thomas, C. D. (2006). The distributions of a wide range of taxonomic groups are expanding polewards. *Global Change Biology,* 12(3)**,** 450-455.

Hickling, R., Roy, D. B., Hill, J. K. & Thomas, C. D. (2005). A northward shift of range margins in British Odonata. *Global Change Biology,* 11(3)**,** 502-506.

Hill, J. K., Thomas, C. D. & Huntley, B. (1999). Climate and habitat availability determine 20th century changes in a butterfly's range margin. *Proceedings of the Royal Society of London Series B-Biological Sciences,* 266(1425)**,** 1197-1206.

Hill, M. O., Preston, C. D. & Roy, D. B. (2004). *PLANTATT: Attributes of British and Irish plants: status, size, life history, geography and habitats.* Huntington: Centre for Ecology & Hydrology.

Hirzel, A. (2004). *Biomapper 3 user's manual*. Lausanne: Laboratory for Conservation Biology, University of Lausanne.

Hirzel, A. (2008). *Biomapper - Frequently Asked Questions*. Lausanne: Laboratory for Conservation Biology, University of Lausanne.

Hirzel, A. H. & Arlettaz, R. (2003). Environmental-envelope based habitat-suitability models. IN: Manly, B. F. J. (Ed.) *1st Conference on Resource Selection by Animals, Laramie, Wyoming, USA*. Omnipress, Laramie, USA.

Hirzel, A. H., Hausser, J., Chessel, D. & Perrin, N. (2002). Ecological-niche factor analysis: How to compute habitat-suitability maps without absence data? *Ecology,* 83(7)**,** 2027-2036.

Hirzel, A. H., Hausser, J. & Perrin, N. (2006a). *Biomapper 3.2.* (version 3.2.2, build 308). [computer program]. Lausanne: Laboratory for Conservation Biology. Available from: http://www.unil.ch/biomapper.

Hirzel, A. H., Helfer, V. & Metral, F. (2001). Assessing habitat-suitability models with a virtual species. *Ecological Modelling,* 145(2-3)**,** 111-121.

Hirzel, A. H., Le Lay, G., Helfer, V., Randin, C. & Guisan, A. (2006b). Evaluating the ability of habitat suitability models to predict species presences. *Ecological Modelling,* 199(2)**,** 142-152.

Hirzel, A. H., Posse, B., Oggier, P. A., Crettenand, Y., Glenz, C. & Arlettaz, R. (2004). Ecological requirements of reintroduced species and the implications for release policy: the case of the bearded vulture. *Journal of Applied Ecology,* 41(6)**,** 1103-1116.

Holmes, K. W., Chadwick, O. A. & Kyriakidis, P. C. (2000). Error in a USGS 30-meter digital elevation model and its impact on terrain modeling. *Journal of Hydrology*, 233(1-4)**,** 154-173.

Hopkins, J. (2007). British wildlife and climate change 1. evidence of change. *British Wildlife,* 18(3)**,** 153-159.

Hora, S. C. & VonWinterfeldt, D. (1997). Nuclear waste and future societies: A look into the deep future. *Technological Forecasting and Social Change,* 56(2)**,** 155-170.

Hosmer, D. W. & Lemeshow, S. (2000). *Applied logistic regression.* 2nd ed. New York ; Chichester: Wiley.

Hossell, J. & Rowe, K. (2006). *Headline indicators on the impact of climate change on south eastern protected landscapes - final report*. Wolverhampton: ADAS.

Hossell, J. E., Austin, G. E., Berry, P. M., Butt, N., Crick, H. Q. P., Freeman, S., Harrison, P. A., Masters, G. J., Morrison, A., Rehfisch, M. M., Scholefield, P., Ward, N. & Wilde, I. (2005). *Impacts for the Hampshire case study area*. Oxford: UK Climate Impacts Programme.

Hossell, J. E., Briggs, B. & Hepburn, I. R. (2000). *Climate change and UK nature conservation: A review of the impact of climate change on UK species and habitat conservation policy*. London: Department of the Environment, Transport and the Regions.

Hossell, J. E., Ellis, N. E., Harley, M. J. & Hepburn, I. R. (2003). Climate change and nature conservation: Implications for policy and practice in Britain and Ireland. *Journal for Nature Conservation,* 11(1)**,** 67-73.

Hubbard, C. E. & Hubbard, J. C. E. (1984). *Grasses: a guide to their structure, identification, uses, and distribution in the British Isles.* 3rd ed. Harmondsworth: Penguin.

Hughes, C. L., Hill, J. K. & Dytham, C. (2003). Evolutionary trade-offs between reproduction and dispersal in populations at expanding range boundaries. *Proceedings of the Royal Society of London Series B-Biological Sciences,* 270**,** S147-S150.

Hughes, L. (2000). Biological consequences of global warming: is the signal already apparent? *Trends in Ecology & Evolution,* 15(2)**,** 56-61.

Hugin Expert A/S (2007). *Hugin Researcher 6.9* [computer program]. Aalborg, Denmark: Available from: http://www.hugin.com.

Hulme, M., Jenkins, G. J., Lu, X., Turnpenny, J. R., Mitchell, T. D., Jones, R. G., Lowe, J., Murphy, J. M., Hassell, D., Boorman, P., McDonald, R. & Hill, S. (2002). *Climate Change Scenarios for the United Kingdom: The UKCIP02 Scientific Report*. Norwich, UK: Tyndall Centre for Climate Change Research, School of Environmental Sciences, University of East Anglia.

Huntley, B. (1991). How plants respond to climate change: Migration rates, individualism and the consequences for plant communities. *Annals of Botany,* 67**,** 15-22.

Huntley, B., Green, R. E., Collingham, Y. C., Hill, J. K., Willis, S. G., Bartlein, P. J., Cramer, W., Hagemeijer, W. J. M. & Thomas, C. J. (2004). The performance of models relating species geographical distributions to climate is independent of trophic level. *Ecology Letters*, 7(5)**,** 417-426.

Huston, M. A. (2002). Introductory essay: critical issues for improving predictions. IN: Scott, J. M., Heglund, P. J., Morrison, M. L., Haufler, J. B., Raphael, M. G., Wall, W. A. & Samson, F. B. (Eds.). *Predicting species occurrences: issues of accuracy and scale.* Washington: Island Press, 7-21.

Hutchinson, G. E. (1957). Concluding Remarks. *Cold Spring Harbor Symposium on Quantitative Biology,* 22**,** 415-427.

Insightful Corporation (2001). *S-Plus 6 for Windows guide to statistics, volume 1.* Seattle, Washington: Insightful Corporation.

IPCC (2007). *Climate change 2007: synthesis report. Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Geneva: IPCC.

IUCN - International Union for Conservation of Nature and Natural Resources. Species Survival Commission (2001). *IUCN red list categories and criteria: version 3.1.* Gland, Cambridge: IUCN.

James, F. C. & McCulloch, C. E. (1990). Multivariate analysis in ecology and systematics: Panacea or Pandora's box? *Annual Review of Ecology and Systematics,* 21**,** 129-166.

Jenkins, G. J., Perry, M. C. & Prior, M. J. O. (2008). *The climate of the United Kingdom and recent trends*. Exeter: Met Office Hadley Centre.

Jensen, F. V. & Nielsen, T. D. (2007). *Bayesian networks and decision graphs.* 2nd ed. New York: Springer.

Jermy, A. C. (1994). *Pilularia globulifera* L., Pillwort. IN: Stewart, A., Pearman, D. A. & Preston, C. D. (Eds.). *Scarce plants in Britain.* Peterborough: JNCC, 311-312.

Jimenez-Valverde, A. & Lobo, J. M. (2006). The ghost of unbalanced species distribution data in geographical model predictions. *Diversity and Distributions,* 12(5)**,** 521-524.

Jimenez-Valverde, A. & Lobo, J. M. (2007). Threshold criteria for conversion of probability of species presence to either-or presence-absence. *Acta Oecologica-International Journal of Ecology,* 31(3)**,** 361-369.

Johnson, C. J. & Gillingham, M. P. (2005). An evaluation of mapped species distribution models used for conservation planning. *Environmental Conservation,* 32(2)**,** 117-128.

Joint Nature Conservation Committee (2004). *Common Standards Monitoring guidance for lowland heathland*. Peterborough: JNCC.

Joint Nature Conservation Committee (2005). *Common Standards Monitoring guidance for standing waters*. Peterborough: JNCC.

Jordano, D., Rodriguez, J., Thomas, C. D. & Haeger, J. F. (1992). The distribution and density of a lycaenid butterfly in relation to *Lasius* ants. *Oecologia,* 91(3)**,** 439-446.

Jordano, D. & Thomas, C. D. (1992). Specificity of an ant-lycaenid interaction. *Oecologia,* 91(3)**,** 431-438.

Joy, J. (1995). *Heathland management for the Silver-studded Blue butterfly*. Peterborough: English Nature.

Karlsson, B. & Wiklund, C. (2005). Butterfly life history and temperature adaptations; dry open habitats select for increased fecundity and longevity. *Journal of Animal Ecology,* 74(1)**,** 99-104.

Kauserud, H., Stige, L. C., Vik, J. O., Økland, R. H., Høiland, K. & Stenseth, N. C. (2008). Mushroom fruiting and climate change. *Proceedings of the National Academy of Sciences,* 105(10)**,** 3811-3814.

Killick, H. J. (2002). *Chamaemelum nobile* Chamomile. IN: Preston, C. D., Pearman, D. A. & Dines, T. D. (Eds.). *New Atlas of the British and Irish*

Flora: An Atlas of the Vascular Plants of Britain, Ireland, The Isle of Man and the Channel Islands. Oxford: Oxford University Press, 648.

Klute, D. S., Lovallo, M. J. & Tzilkowski, W. M. (2002). Autologistic regression modelling of American Woodcock habitat use with spatially dependent data. IN: Scott, J. M., Heglund, P. J., Morrison, M. L., Haufler, J. B., Raphael, M. G., Wall, W. A. & Samson, F. B. (Eds.). *Predicting species occurrences: issues of accuracy and scale.* Washington: Island Press, 335-343.

Kotiaho, J. S., Kaitala, V., Komonen, A. & Paivinen, J. (2005). Predicting the risk of extinction from shared ecological characteristics. *Proceedings of the National Academy of Sciences of the United States of America,* 102(6)**,** 1963-1967.

Lachat, T. & Butler, R. (2009). Identifying conservation and restoration priorities for saproxylic and old-growth forest apecies: a case study in Switzerland. *Environmental Management,* 44(1)**,** 105-118.

Lambeck, R. J. (1997). Focal species: A multi-species umbrella for nature conservation. *Conservation Biology,* 11(4)**,** 849-856.

Landis, J. R. & Koch, G. G. (1977). Measurement of observer agreement for categorical data. *Biometrics,* 33(1)**,** 159-174.

Lauritzen, S. L. & Spiegelhalter, D. J. (1988). Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society Series B-Methodological,* 50(2)**,** 157-224.

Lavergne, S., Molina, J. & Debussche, M. (2006). Fingerprints of environmental change on the rare Mediterranean flora: a 115-year study. *Global Change Biology,* 12(8)**,** 1466-1478.

Legendre, P. (1993). Spatial autocorrelation: trouble or new paradigm? *Ecology,* 74(6)**,** 1659-1673.

Lehmann, A., Overton, J. M. & Leathwick, J. R. (2002). GRASP: generalized regression analysis and spatial prediction. *Ecological Modelling,* 157(2-3)**,** 189-207.

Lehmkuhl, J. F., Kie, J. G., Bender, L. C., Servheen, G. & Nyberg, H. (2001). Evaluating the effects of ecosystem management alternatives on elk, mule deer, and white-tailed deer in the interior Columbia River basin, USA. *Forest Ecology and Management,* 153(1-3)**,** 89-104.

Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J. & Aulagnier, S. (1996). Application of neural networks to modelling nonlinear relationships in ecology. *Ecological Modelling,* 90(1)**,** 39-52.

Lennon, J. J. (2000). Red-shifts and red herrings in geographical ecology. *Ecography,* 23(1)**,** 101-113.

Lichstein, J. W., Simons, T. R., Shriner, S. A. & Franzreb, K. E. (2002). Spatial autocorrelation and autoregressive models in ecology. *Ecological Monographs,* 72(3)**,** 445-463.

Lindenmayer, D. B. & Franklin, J. F. (2002). *Conserving forest biodiversity: a comprehensive multiscaled approach* Washington: Island Press.

Liu, C. R., Berry, P. M., Dawson, T. P. & Pearson, R. G. (2005). Selecting thresholds of occurrence in the prediction of species distributions. *Ecography,* 28(3)**,** 385-393.

Lobo, J. M., Jimenez-Valverde, A. & Real, R. (2008). AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography,* 17(2)**,** 145-151.

Lonsdale, D. & Gibbs, J. (2002). Effects of climate change on fungal diseases of trees. IN: Broadmeadow, M. (Ed.) *Climate change: impacts on UK forests. Forestry Commission bulletin 125.* Edinburgh: Forestry Commission, 83-98.

López-López, P., Garcia-Ripollés, C., Soutullo, A., Cadahia, L. & Urios, V. (2007). Identifying potentially suitable nesting habitat for golden eagles applied to 'important bird areas' design. *Animal Conservation,* 10(2)**,** 208-218.

Lütolf, M., Kienast, F. & Guisan, A. (2006). The ghost of past species occurrence: improving species distribution models for presence-only data. *Journal of Applied Ecology,* 43(4)**,** 802-815.

Mackey, B. G. & Lindenmayer, D. B. (2001). Towards a hierarchical framework for modelling the spatial distribution of animals. *Journal of Biogeography,* 28(9)**,** 1147-1166.

Malcolm, G. M., López-Gutiérrez, J. C., Koide, R. T. & Eissenstat, D. M. (2008). Acclimation to temperature and temperature sensitivity of metabolism by ectomycorrhizal fungi. *Global Change Biology,* 14(5)**,** 1169-1180.

Manel, S., Dias, J. M. & Ormerod, S. J. (1999). Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with a Himalayan river bird. *Ecological Modelling,* 120(2-3)**,** 337-347.

Marcot, B. G. (2006). Characterizing species at risk I: Modeling rare species under the Northwest Forest Plan. *Ecology and Society,* 11(2).

Marcot, B. G., Holthausen, R. S., Raphael, M. G., Rowland, M. M. & Wisdom, M. J. (2001). Using Bayesian belief networks to evaluate fish and wildlife population viability under land management alternatives from an environmental impact statement. *Forest Ecology and Management,* 153(1-3)**,** 29-42.

Marcot, B. G., Steventon, J. D., Sutherland, G. D. & McCann, R. K. (2006). Guidelines for developing and updating Bayesian belief networks applied to ecological modeling and conservation. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere,* 36(12)**,** 3063-3074.

Marrs, R. H. & Watt, A. S. (2006). Biological flora of the British Isles: *Pteridium aquilinum* (L.) Kuhn. *Journal of Ecology,* 94(6)**,** 1272-1321.

Marshall, J. A. & Haes, E. C. M. (1988). *Grasshoppers and allied insects of Great Britain and Ireland.* Colchester, Essex: Harley Books.

Martin, T. G., Kuhnert, P. M., Mengersen, K. & Possingham, H. P. (2005). The power of expert opinion in ecological models using Bayesian methods: Impact of grazing on birds. *Ecological Applications,* 15(1)**,** 266-280.

Martinez, I., Carreno, F., Escudero, A. & Rubio, A. (2006). Are threatened lichen species well-protected in Spain? Effectiveness of a protected areas network. *Biological Conservation,* 133(4)**,** 500-511.

Masters, G. J., Brown, V. K., Clarke, I. P., Whittaker, J. B. & Hollier, J. A. (1998). Direct and indirect effects of climate change on insect herbivores: Auchenorrhyncha (Homoptera). *Ecological Entomology,* 23(1)**,** 45-52.

Matern, A., Drees, C., Kleinwachter, M. & Assmann, T. (2007). Habitat modelling for the conservation of the rare ground beetle species Carabus variolosus (Coleoptera, Carabidae) in the riparian zones of headwaters. *Biological Conservation,* 136(4)**,** 618-627.

McCann, R. K., Marcot, B. G. & Ellis, R. (2006). Bayesian belief networks: applications in ecology and natural resource management. *Canadian Journal of Forest Research,* 36(12)**,** 3053-3062.

McCarty, J. P. (2001). Ecological consequences of recent climate change. *Conservation Biology,* 15(2)**,** 320-331.

McNay, R. S., Marcot, B. G., Brumovsky, V. & Ellis, R. (2006). A Bayesian approach to evaluating habitat for woodland caribou in north-central British Columbia. *Canadian Journal of Forest Research,* 36(12)**,** 3117-3133.

McPherson, J. M. & Jetz, W. (2007). Effects of species' ecology on the accuracy of distribution models. *Ecography,* 30(1)**,** 135-151.

McPherson, J. M., Jetz, W. & Rogers, D. J. (2004). The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact? *Journal of Applied Ecology,* 41(5)**,** 811-823.

Meek, W. R. (2002). *Galium constrictum* Slender Marsh-bedstraw. IN: Preston, C. D., Pearman, D. A. & Dines, T. D. (Eds.). *New Atlas of the British and Irish Flora: An Atlas of the Vascular Plants of Britain, Ireland, The Isle of Man and the Channel Islands.* Oxford: Oxford University Press, 592.

Menzel, A. & Fabian, P. (1999). *Climate change and the phenology of European trees and shrubs*. Sandy, Bedfordshire: RSPB.

Met Office (2009). *Everton 1971-2000 averages*. Available from: http://www.metoffice.gov.uk/climate/uk/averages/19712000/sites/everton.html [Accessed 26 February 2009].

Meyer, M. A. & Booker, J. M. (1991). *Eliciting and analyzing expert judgement: a practical guide.* London: Academic Press.

Meynard, C. N. & Quinn, J. F. (2007). Predicting species distributions: a critical comparison of the most common statistical models using artificial species. *Journal of Biogeography,* 34(8)**,** 1455-1469.

Millennium Ecosystem Assessment (2005). *Ecosystems and human well-being: biodiversity synthesis*. Washington, DC: World Resources Institute.

Milsom, T. P., Langton, S. D., Parkin, W. K., Peel, S., Bishop, J. D., Hart, J. D. & Moore, N. P. (2000). Habitat models of bird species' distribution: an aid to the management of coastal grazing marshes. *Journal of Applied Ecology,* 37(5)**,** 706-727.

Moisen, G. G., Freeman, E. A., Blackard, J. A., Frescino, T. S., Zimmermann, N. E. & Edwards, T. C. (2006). Predicting tree species presence and basal area in Utah: A comparison of stochastic gradient boosting, generalized additive models, and tree-based methods. *Ecological Modelling,* 199(2)**,** 176-187.

Moisen, G. G. & Frescino, T. S. (2002). Comparing five modelling techniques for predicting forest characteristics. *Ecological Modelling,* 157(2-3)**,** 209-225.

Morgan, M. G. & Henrion, M. (1990). *Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis.* Cambridge: Cambridge University Press.

Morgan, M. G., Pitelka, L. F. & Shevliakova, E. (2001). Elicitation of expert judgments of climate change impacts on forest ecosystems. *Climatic Change,* 49(3)**,** 279-307.

Mountford, E. P., Peterken, G. E., Edwards, P. J. & Manners, J. G. (1999). Long-term change in growth, mortality and regeneration of trees in Denny Wood, an old-growth wood-pasture in the New Forest (UK). *Perspectives in Plant Ecology, Evolution and Systematics,* 2(2)**,** 223-272.

National Biodiversity Network (2009). *NBN Gateway*. Nottingham: National Biodiversity Network. Available from: http://data.nbn.org.uk/ [Accessed 13 August 2009].

Natural England (2009). *Condition of SSSI units in the New Forest - compiled 01 Sep 2009*. Peterborough: Natural England.

New Forest Committee (2003). *Strategy for the New Forest (adopted as the New Forest National Park interim management plan)*. Lyndhurst: New Forest Committee.

New Forest National Park Authority (2006a). *History*. Available from: http://www.newforestnpa.gov.uk/index/learningabout/le-generalinformation/le-history.htm [Accessed 13 December 2006].

New Forest National Park Authority (2006b). *New Forest National Park pocket guide*. Lymington, UK: New Forest National Park Authority.

New Forest National Park Authority (2007a). *Conservation in the New Forest National Park: 2 - Climate change*. Lymington, UK: New Forest National Park Authority.

New Forest National Park Authority (2007b). *Static map*. Available from: http://www.newforestnpa.gov.uk/staticmap_august07e.pdf [Accessed 5 March 2008].

New Forest National Park Authority (2008). *New Forest National Park plan consulation draft: National Park management plan and local development* *framework core strategy and development policies*. Lymington, UK: New Forest National Park Authority.

Newton-Cross, G., White, P. C. L. & Harris, S. (2007). Modelling the distribution of badgers *Meles meles*: comparing predictions from field-based and remotely derived habitat data. *Mammal Review,* 37(1)**,** 54-70.

Newton, A. C. (2007). *Forest ecology and conservation. A handbook of techniques.* Oxford: Oxford University Press.

Newton, A. C. (2009a). Bayesian belief networks in environmental modelling: a review of recent progress. IN: Findley, P. N. (Ed.) *Environmental modelling: new research.* Hauppauge, New York: Nova Science Publishers, 13-50.

Newton, A. C. (2009b). Fungi. IN: Newton, A. C. (Ed.) *Biodiversity in the New Forest.* Newbury: Pisces Publications.

Newton, A. C. (2009c). Use of a Bayesian network for Red Listing under uncertainty. *Environmental Modelling and Software,* In press.

Newton, A. C., Stewart, G. B., Diaz, A., Golicher, D. & Pullin, A. S. (2007). Bayesian belief networks as a tool for evidence-based conservation management. *Journal for Nature Conservation,* 15(2)**,** 144-160.

Nicholson, I. A. & Patterson, I. S. (1976). The ecological implications of bracken control to plant/animal systems. IN: Perring, F. H. & Gardiner, B. G. (Eds.). *The biology of bracken.* London: Academic Press, 269-285.

Nisbet, T. (2002). Implications of climate change: soil and water. IN: Broadmeadow, M. (Ed.) *Climate change: impacts on UK forests. Forestry Commission bulletin 125.* Edinburgh: Forestry Commission, 53-68.

Noss, R. F. (1999). Assessing and monitoring forest biodiversity: A suggested framework and indicators. *Forest Ecology and Management,* 115(2-3)**,** 135-146.

Nyberg, J. B., Marcot, B. G. & Sulyma, R. (2006). Using Bayesian belief networks in adaptive management. *Canadian Journal of Forest Research,* 36(12)**,** 3104-3116.

O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E. & Rakow, T. (2006). *Uncertain judgements: eliciting experts' probabilities.* Chichester: Wiley.

O'Hagan, A. & Oakley, J. E. (2004). Probability is perfect, but we can't elicit it perfectly. *Reliability Engineering & System Safety,* 85(1-3)**,** 239-248.

O'Hanley, J. R. (2005a). *Automation of the SPECIES model*. Oxford: UKCIP Technical Report.

O'Hanley, J. R. (2005b). *Automation of the SPECIES model. MONARCH 3 - Technical report*. Oxford: UK Climate Impacts Programme.

Oates, M., Taverner, J. & Green, D. (2000). *The butterflies of Hampshire.* Newbury: Pisces Publications.

Olivier, F. & Wotherspoon, S. J. (2006). Modelling habitat selection using presence-only data: Case study of a colonial hollow nesting bird, the snow petrel. *Ecological Modelling,* 195(3-4)**,** 187-204.

Orians, G. H. & Groom, M. J. (2006). Global biodiversity: patterns and processes IN: Groom, M. J., Meffe, G. K. & Carroll, C. R. (Eds.). *Principles of conservation biology* 3rd ed., 27-62.

Page, C. N. (1982). *The ferns of Britain and Ireland.* Cambridge: Cambridge University Press.

Pakeman, R. J. & Marrs, R. H. (1996). Modelling the effects of climate change on the growth of bracken (*Pteridium aquilinum*) in Britain. *Journal of Applied Ecology,* 33(3)**,** 561-575.

Parmesan, C. (2006). Ecological and evolutionary responses to recent climate change. *Annual Review of Ecology, Evolution, and Systematics,* 37**,** 637-669.

Parmesan, C., Root, T. L. & Willig, M. R. (2000). Impacts of extreme weather and climate on terrestrial biota. *Bulletin of the American Meteorological Society,* 81(3)**,** 443-450.

Parmesan, C., Ryrholm, N., Stefanescu, C., Hill, J. K., Thomas, C. D., Descimon, H., Huntley, B., Kaila, L., Kullberg, J., Tammaru, T., Tennent, W. J., Thomas, J. A. & Warren, M. (1999). Poleward shifts in geographical ranges of butterfly species associated with regional warming. *Nature,* 399(6736)**,** 579-583.

Parmesan, C. & Yohe, G. (2003). A globally coherent fingerprint of climate change impacts across natural systems. *Nature,* 421(6918)**,** 37-42.

Pearce, J. & Ferrier, S. (2000a). Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling,* 133(3)**,** 225-245.

Pearce, J. & Ferrier, S. (2000b). An evaluation of alternative algorithms for fitting species distribution models using logistic regression. *Ecological Modelling,* 128(2-3)**,** 127-147.

Pearce, J., Ferrier, S. & Scotts, D. (2001). An evaluation of the predictive performance of distributional models for flora and fauna in north-east New South Wales. *Journal of Environmental Management,* 62(2)**,** 171-184.

Pearce, J. L., Vernier, L. A., Ferrier, S. & McKenney, D. W. (2002). Measuring prediction uncertainty in models of species distribution. IN: Scott, J. M., Heglund, P. J., Morrison, M. L., Haufler, J. B., Raphael, M. G., Wall, W. A. & Samson, F. B. (Eds.). *Predicting species occurrences: issues of accuracy and scale.* Washington: Island Press, 383-390.

Pearson, R. G. & Dawson, T. P. (2003). Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology and Biogeography,* 12(5)**,** 361-371.

Pearson, R. G., Thuiller, W., Araujo, M. B., Martinez-Meyer, E., Brotons, L., McClean, C., Miles, L., Segurado, P., Dawson, T. P. & Lees, D. C. (2004). Model-based uncertainty in species range prediction. *Journal of Biogeography,* 33(10)**,** 1704-1711.

Peñuelas, J., Gordon, C., Llorens, L., Nielsen, T., Tietema, A., Beier, C., Bruna, P., Emmett, B., Estiarte, M. & Gorissen, A. (2004). Nonintrusive field experiments show different plant responses to warming and drought among sites, seasons, and species in a north-south European gradient. *Ecosystems,* 7(6)**,** 598-612.

Peterson, A. T. & Cohoon, K. P. (1999). Sensitivity of distributional prediction algorithms to geographic data completeness. *Ecological Modelling,* 117(1)**,** 159-164.

Peterson, A. T., Stockwell, D. R. B. & Kluza, D. A. (2002). Distributional prediction based on ecological niche modelling of primary occurrence data. IN: Scott, J. M., Heglund, P. J., Morrison, M. L., Haufler, J. B., Raphael, M. G., Wall, W. A. & Samson, F. B. (Eds.). *Predicting species occurrences: issues of accuracy and scale.* Washington: Island Press, 617-623.

Phillips, S. J., Anderson, R. P. & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling,* 190(3-4)**,** 231-259.

Phillips, S. J., Dudik, M. & Schapire, R. E. (2004). A maximum entropy approach to species distribution modelling. *Proceedings of the 21st International Conference on Machine Learning, Banff, Canada*.

Pimm, S. L., Ayres, M., Balmford, A., Branch, G., Brandon, K., Brooks, T., Bustamante, R., Costanza, R., Cowling, R., Curran, L. M., Dobson, A., Farber, S., da Fonseca, G. A. B., Gascon, C., Kitching, R., McNeely, J., Lovejoy, T., Mittermeier, R. A., Myers, N., Patz, J. A., Raffle, B., Rapport, D., Raven, P., Roberts, C., Rodriguez, J. P., Rylands, A. B., Tucker, C., Safina, C., Samper, C., Stiassny, M. L. J., Supriatna, J., Hall, D. H. & Wilcove, D. (2001). Environment - Can we defy nature's end? *Science,* 293(5538)**,** 2207-2208.

Pimm, S. L., Russell, G. J., Gittleman, J. L. & Brooks, T. M. (1995). The future of biodiversity. *Science,* 269(5222)**,** 347-350.

Plantlife (2001). *Managing your land for Chamomile*. London: Plantlife International.

Plantlife (2005). *Under pressure: Climate change and the UK's wild plants*. Salisbury: Plantlife International.

Plantlife (2006). *Pillwort (Pilularia globulifera)*. Salisbury: Plantlife International.

Plantlife (2007a). *Chamaemelum nobile Chamomile*. Available from: http://www.plantlife.org.uk/uk/plantlife-saving-species-under-our-care-chamaemelum-nobile.htm [Accessed 21 August 2007].

Plantlife (2007b). *Pilularia globulifera Pillwort*. Available from: http://www.plantlife.org.uk/uk/plantlife-saving-species-under-our-care-pilularia-globulifera.htm [Accessed 27 June 2007].

Plantlife International (2001). *Managing your land for Chamomile*. London: Plantlife International.

Plantlife International (2006). *Pillwort (Pilularia globulifera)*. Salisbury: Plantlife International.

Podchong, S., Schmidt-Vogt, D. & Hond, K. (2009). An improved approach for identifying suitable habitat of Sambar Deer (*Cervus unicolor* Kerr) using ecological niche analysis and environmental categorization: Case study at Phu-Khieo Wildlife Sanctuary, Thailand. *Ecological Modelling,* 220(17)**,** 2103-2114.

Poel, L. W. (1951). Soil aeration in relation to *Pteridium aquilinum* (L.) Kuhn. *Journal of Ecology,* 39(1)**,** 182-191.

Poland, J. (2004). *Survey of nail fungus Poronia punctata in the New Forest during 2003-4*. A report produced for the Hampshire and Isle of Wight Wildlife Trust.

Pollard, E. (1988). Temperature, rainfall and butterfly numbers. *Journal of Applied Ecology,* 25(3)**,** 819-828.

Pollino, C. A., Woodberry, O., Nicholson, A., Korb, K. & Hart, B. T. (2007). Parameterisation and evaluation of a Bayesian network for use in an ecological risk assessment. *Environmental Modelling & Software,* 22(8)**,** 1140-1152.

Porley, R. & McDonnell, A. (1997). *Rare and scarce vascular plants and Bryophytes in Natural Areas*. Sheffield: English Nature, 267.

Pratt, R. M., Putman, R. J., Ekins, J. R. & Edwards, P. J. (1986). Use of habitat by free-ranging cattle and ponies in the New Forest, southern England. *Journal of Applied Ecology,* 23(2)**,** 539-557.

Preston, C. D. (2002). *Pilularia globulifera* Pillwort. IN: Preston, C. D., Pearman, D. A. & Dines, T. D. (Eds.). *New Atlas of the British and Irish Flora: An Atlas of the Vascular Plants of Britain, Ireland, The Isle of Man and the Channel Islands.* Oxford: Oxford University Press, 63.

Preston, C. D. & Croft, J. M. (1997). *Aquatic plants in Britain and Ireland.* Colchester: Harley.

Proess, R. & Baden, R. (2000). Survey of the orthopteran species *Barbitistes serricauda* (Fabricus, 1798), *Leptophyes punctatissima* (Bosc, 1972), *Meconema thalassinum* (De Geer, 1773) and *Nemobius sylvestris* (Bosc, 1972) in Luxembourg (Insecta, Saltatoria). *Bulletin de la Societe des Naturalistes Luxembourgeois***,** 159-170.

Puddu, G., Maiorano, L., Falcucci, A., Corsi, F. & Boitani, L. (2009). Spatial-explicit assessment of current and future conservation options for the endangered Corsican Red Deer (*Cervus elaphus corsicanus*) in Sardinia. *Biodiversity and Conservation,* 18(8)**,** 2001-2016.

Pulliam, H. R. (2000). On the relationship between niche and distribution. *Ecology Letters,* 3(4)**,** 349-361.

Pullin, A. S. (2002). *Conservation biology* Cambridge: Cambridge University Press.

Purvis, A., Gittleman, J. L., Cowlishaw, G. & Mace, G. M. (2000). Predicting extinction risk in declining species. *Proceedings of the Royal Society of London Series B-Biological Sciences,* 267(1456)**,** 1947-1952.

Qi, D. W., Hu, Y. B., Gu, X. D., Li, M. & Wei, F. W. (2009). Ecological niche modeling of the sympatric giant and red pandas on a mountain-range scale. *Biodiversity and Conservation,* 18(8)**,** 2127-2141.

R Development Core Team (2009). *R: A language and environment for statistical computing* [computer program]. Vienna: R Foundation for Statistical Computing. Available from: http://www.R-project.org.

Rahel, F. J., Bierwagen, B. & Taniguchi, Y. (2008). Managing aquatic species of conservation concern in the face of climate change and invasive species. *Conservation Biology,* 22(3)**,** 551-561.

Rahel, F. J. & Olden, J. D. (2008). Assessing the effects of climate change on aquatic invasive species. *Conservation Biology,* 22(3)**,** 521-533.

Randin, C. F., Dirnbock, T., Dullinger, S., Zimmermann, N. E., Zappa, M. & Guisan, A. (2006). Are niche-based species distribution models transferable in space? *Journal of Biogeography,* 33(10)**,** 1689-1703.

Raphael, M. G., Wisdom, M. J., Rowland, M. M., Holthausen, R. S., Wales, B. C., Marcot, B. G. & Rich, T. D. (2001). Status and trends of habitats of terrestrial vertebrates in relation to land management in the interior Columbia river basin. *Forest Ecology and Management,* 153(1-3)**,** 63-88.

Ravenscroft, N. O. M. & Warren, M. S. (1996). *Species action plan: the Silver-studded Blue Plebejus argus*. Wareham: Butterfly Conservation.

Real, R., Barbosa, A. M. & Vargas, J. M. (2006). Obtaining environmental favourability functions from logistic regression. *Environmental and Ecological Statistics,* 13(2)**,** 237-245.

Reckhow, K. H. (1999). Water quality prediction and probability network models. *Canadian Journal of Fisheries and Aquatic Sciences,* 56(7)**,** 1150-1158.

Regan, H. M., Colyvan, M. & Burgman, M. A. (2002). A taxonomy and treatment of uncertainty for ecology and conservation biology. *Ecological Applications,* 12(2)**,** 618-628.

Reid, D. A. (1986). A collection of *Poronia punctata* from Surrey. *Bulletin of the British Mycological Society,* 20(1)**,** 58-59.

Renooij, S. (2001). Probability elicitation for belief networks: issues to consider. *Knowledge Engineering Review,* 16(3)**,** 255-269.

Renooij, S. & Witteman, C. (1999). Talking probabilities: communicating probabilistic information with words and numbers. *International Journal of Approximate Reasoning,* 22(3)**,** 169-194.

Reutter, B. A., Helfer, V., Hirzel, A. H. & Vogel, P. (2003). Modelling habitat-suitability using museum collections: an example with three sympatric *Apodemus* species from the Alps. *Journal of Biogeography,* 30(4)**,** 581-590.

Richards, T. J. (1952). *Nemobius sylvestris* in S.E. Devon. *The Entomologist,* 85**,** 83-87, 108-111, 136-141, 161-166.

Rodríguez, J. P., Brotons, L., Bustamante, J. & Seoane, J. (2007). The application of predictive modelling of species distribution to biodiversity conservation. *Diversity and Distributions,* 13(3)**,** 243-251.

Rogers, J. D. (1979). The Xylariaceae: systematic, biological and evolutionary aspects. *Mycologia,* 71(1)**,** 1-42.

Root, T. L., Price, J. T., Hall, K. R., Schneider, S. H., Rosenzweig, C. & Pounds, J. A. (2003). Fingerprints of global warming on wild animals and plants. *Nature,* 421(6918)**,** 57-60.

Rose, F. (1996). The habitats and vegetation of present-day Hampshire. IN: Brewis, A., Bowman, P. & Rose, F. (Eds.). *The Flora of Hampshire.* Great Horkesley, Colchester: Harley Books, 14-62.

Rosenzweig, C., Casassa, G., Karoly, D. J., Imeson, A., Liu, C., Menzel, A., Rawlins, S., Root, T. L., Saguin, B. & Tryjanowski, P. (2007). Assessment of observed changes and responses in natural and managed systems. IN: Parry, M. L., Canziani, O. F., Palutikof, J. P., Van Der Linden, P. J. & Hanson, C. E. (Eds.). *Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge, UK: Cambridge University Press, 79-131.

Rowland, M. M., Wisdom, M. J., Johnson, D. H., Wales, B. C., Copeland, J. P. & Edelmann, F. B. (2003). Evaluation of landscape models for wolverines in the interior northwest, United States of America. *Journal of Mammalogy,* 84(1)**,** 92-105.

Roy, D. B., Rothery, P., Moss, D., Pollard, E. & Thomas, J. A. (2001). Butterfly numbers and weather: predicting historical trends in abundance and the future effects of climate change. *Journal of Animal Ecology,* 70(2)**,** 201-217.

Roy, D. B. & Sparks, T. H. (2000). Phenology of British butterflies and climate change. *Global Change Biology,* 6(4)**,** 407-416.

Rushton, S. P., Ormerod, S. J. & Kerby, G. (2004). New paradigms for modelling species distributions? *Journal of Applied Ecology,* 41(2)**,** 193-200.

Saetersdal, M. & Birks, H. J. B. (1997). A comparative ecological study of Norwegian mountain plants in relation to possible future climatic change. *Journal of Biogeography,* 24(2)**,** 127-152.

Santos, X., Brito, J. C., Sillero, N., Pleguezuelos, J. M., Llorente, G. A., Fahd, S. & Parellada, X. (2006). Inferring habitat-suitability areas with ecological modelling techniques and GIS: A contribution to assess the conservation status of Vipera latastei. *Biological Conservation,* 130(3)**,** 416-425.

Sattler, T., Bontadina, F., Hirzel, A. H. & Arlettaz, R. (2007). Ecological niche modelling of two cryptic bat species calls for a reassessment of their conservation status. *Journal of Applied Ecology,* 44(6)**,** 1188-1199.

Schneider, S. H., Semenov, S., Patwardhan, A., Burton, I., Magadza, C. H. D., Oppenhelmer, M., Pittock, A. B., Rahman, A., Smith, J. B., Suarez, A. & Yarmin, F. (2007). Assessing key vulnerabilities and the risk from climate change. IN: Parry, M. L., Canziani, O. F., Palutikof, J. P., Van Der Linden, P. J. & Hanson, C. E. (Eds.). *Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge, UK: Cambridge University Press, 779-810.

Schröder, B. (2006). *ROC_AUC* (Dec 2006). [computer program]. Potsdam: Universität Potsdam. Available from: http://brandenburg.geoecology.uni-potsdam.de/users/schroeder/download.html.

Schröder, B. (2008). *Software - ROC/AUC calculation*. Available from: http://brandenburg.geoecology.uni-potsdam.de/users/schroeder/download.html [Accessed 25 January 2008].

Schröder, B. & Richter, O. (1999/2000). Are habitat models transferable in space and time? *Zeitschrift fur Okologie and Naturschutz,* 8**,** 195-205.

Schultz, A., Klenke, R., Lutze, G., Voss, M., Wieland, R. & Wilkening, B. (2003). Habitat models to link situation evaluation and planning support in agricultural landscapes. IN: Bissonette, J. A. & Storch, I. (Eds.). *Lanscape ecology and resource management: linking theory with practice.* Washington: Island Press, 261-282.

Scott, J. M., Heglund, P. J., Morrison, M. L., Haufler, J. B., Raphael, M. G., Wall, W. A. & Samson, F. B. (2002a). Introduction. IN: Scott, J. M., Heglund, P. J., Morrison, M. L., Haufler, J. B., Raphael, M. G., Wall, W. A. & Samson, F. B. (Eds.). *Predicting species occurrences: issues of accuracy and scale.* Washington: Island Press, 1-5.

Scott, J. M., Heglund, P. J., Morrison, M. L., Haufler, J. B., Raphael, M. G., Wall, W. A. & Samson, F. B. (Eds.) (2002b). *Predicting species occurrences: issues of accuracy and scale,* Washington: Island Press.

Scott, M., Scott, S. & Sydes, C. (1999). A Scottish perspective on the conservation of Pillwort. *British Wildlife,* 10(5)**,** 297-302.

Segurado, P. & Araújo, M. B. (2004). An evaluation of methods for modelling species distributions. *Journal of Biogeography,* 31(10)**,** 1555-1568.

Segurado, P., Araújo, M. B. & Kunin, W. E. (2006). Consequences of spatial autocorrelation for niche-based models. *Journal of Applied Ecology,* 43(3)**,** 433-444.

Sell, P. D. & Murrell, G. (1996). *Flora of Great Britain and Ireland. Volume 5: Butomaceae - Orchidaceae.* Cambridge: Cambridge University Press.

Seoane, J., Justribó, J. H., García, F., Retamar, J., Rabadán, C. & Atienza, J. C. (2006). Habitat-suitability modelling to assess the effects of land-use changes on Dupont's lark *Chersophilus duponti*: A case study in the Layna Important Bird Area. *Biological Conservation,* 128(2)**,** 241-252.

Sérgio, C., Figueira, R., Draper, D., Menezes, R. & Sousa, A. J. (2007). Modelling bryophyte distribution based on ecological information for extent of occurrence assessment. *Biological Conservation,* 135(3)**,** 341-351.

Smith, C. S., Howes, A. L., Price, B. & McAlpine, C. A. (2007). Using a Bayesian belief network to predict suitable habitat of an endangered mammal - The Julia Creek dunnart (Sminthopsis douglasi). *Biological Conservation,* 139(3-4)**,** 333-347.

Smith, P. A. (1994). Autocorrelation in logistic regression modelling of species' distributions. *Global Ecology and Biogeography Letters,* 4(2)**,** 47-61.

Smith, S. & Gilbert, J. (2001). *National inventory of woodland and trees - England.* Edinburgh: Forestry Commission.

Soares, C. & Brito, J. C. (2007). Environmental correlates for species richness among amphibians and reptiles in a climate transition area. *Biodiversity and Conservation,* 16(4)**,** 1087-1102.

Soussana, J. F. & Luscher, A. (2007). Temperate grasslands and global atmospheric change: a review. *Grass and Forage Science,* 62(2)**,** 127-134.

Sparks, T., Crick, H., Woiwod, I. & Beebee, T. (1999). Climate change and phenology in the United Kingdom. IN: Green, R. E., Harley, M., Spalding, M. & Zöckler, C. (Eds.). *Impacts of climate change on wildlife.* Sandy, Bedfordshire: RSPB, 53-55.

Sparks, T. H., Dennis, R. L. H., Croxton, P. J. & Cade, M. (2007). Increased migration of Lepidoptera linked to climate change. *European Journal of Entomology,* 104(1)**,** 139-143.

Sparks, T. H., Jeffree, E. P. & Jeffree, C. E. (2000). An examination of the relationship between flowering times and temperature at the national scale using long-term phenological records from the UK. *International Journal of Biometeorology,* 44(2)**,** 82-87.

Sparks, T. H., Roy, D. B. & Dennis, R. L. H. (2005). The influence of temperature on migration of Lepidoptera into Britain. *Global Change Biology,* 11(3)**,** 507-514.

Stace, C. A. (1997). *New flora of the British Isles.* 2nd ed. Cambridge: Cambridge University Press.

Sterry, P. (2006). *Complete British wild flowers.* London: Collins.

Stewart, A. J. A. & Hutchings, M. J. (1996). Conservation of populations. IN: Spellerberg, I. F. (Ed.) *Conservation biology.* Harlow: Pearson Education Limited, 122-140.

Stewart, N. F., Fitzgerald, R., Lansdown, R. V. & Jermy, A. C. (2000). *The conservation of Pillwort Pilularia globulifera in England. Interim report no. 2.* Salisbury: Plantlife.

Steyerberg, E. W., Eijkemans, M. J. C. & Habbema, J. D. F. (1999). Stepwise selection in small data sets: A simulation study of bias in logistic regression analysis. *Journal of Clinical Epidemiology,* 52(10)**,** 935-942.

Stockwell, D. & Peters, D. (1999). The GARP modelling system: problems and solutions to automated spatial prediction. *International Journal of Geographical Information Science,* 13**,** 143-158.

Stockwell, D. R. B. (1997). Generic predictive systems: an empirical evaluation using the learning base system (LBS). *Expert Systems with Applications,* 12(3)**,** 301-310.

Stockwell, D. R. B. & Peterson, A. T. (2002a). Controlling bias in biodiversity data. IN: Scott, J. M., Heglund, P. J., Morrison, M. L., Haufler, J. B., Raphael, M. G., Wall, W. A. & Samson, F. B. (Eds.). *Predicting species occurrences: issues of accuracy and scale.* Washington: Island Press, 537-546.

Stockwell, D. R. B. & Peterson, A. T. (2002b). Effects of sample size on accuracy of species distribution models. *Ecological Modelling,* 148(1)**,** 1-13.

Stokes, J. (1987). *The ecology of the Wild Gladiolus (Gladiolus illyricus) in the New Forest, Hampshire*. Thesis (MSc Dissertation in Conservation). University College London.

Stokes, J. (2000). Plant of the season: Wild gladiolus, *Gladiolus illyricus*. *Plantlife,* Summer 2000. Salisbury: Plantlife, 13.

Strauss, B. & Biedermann, R. (2005). The use of habitat models in conservation of rare and endangered leafhopper species (Hemiptera, Auchenorrhyncha). *Journal of Insect Conservation,* 9(4)**,** 245-259.

Tabachnick, B. G. & Fidell, L. S. (2001). *Using multivariate statistics.* 4th ed. Boston: Allyn and Bacon.

Taylor, I. (2002). *Gladiolus illyricus* Wild Gladiolus. IN: Preston, C. D., Pearman, D. A. & Dines, T. D. (Eds.). *New Atlas of the British and Irish Flora: An Atlas of the Vascular Plants of Britain, Ireland, The Isle of Man and the Channel Islands.* Oxford: Oxford University Press, 832.

Teyssonneyre, F., Picon-cochard, C., Falcimagne, R. & Soussana, J. F. (2002). Effects of elevated CO_2 and cutting frequency on plant community structure in a temperate grassland. *Global Change Biology,* 8(10)**,** 1034-1046.

The National Trust (2009). *Duke of Burgundy butterfly defies odds with second brood*. Swindon: The National Trust. Available from:

http://www.nationaltrust.org.uk/main/w-global/w-news/w-latest_news/w-news-duke-of-burgundy-defies-odds.htm [Accessed 20 August 2009].

The UK Biodiversity Steering Group (1995) Biodiversity: The UK Steering Group Report - Volume II: Action Plans (Annex G - species action plans, habitat action plans and habitat statements). HMSO, London.

Thomaes, A., Kervyn, T. & Maes, D. (2008). Applying species distribution modelling for the conservation of the threatened saproxylic Stag Beetle (*Lucanus cervus*). *Biological Conservation,* 141(5)**,** 1400-1410.

Thomas, C. D. (2000). Dispersal and extinction in fragmented landscapes. *Proceedings of the Royal Society of London Series B-Biological Sciences,* 267(1439)**,** 139-145.

Thomas, C. D., Bodsworth, E. J., Wilson, R. J., Simmons, A. D., Davies, Z. G., Musche, M. & Conradt, L. (2001). Ecological and evolutionary processes at expanding range margins. *Nature,* 411(6837)**,** 577.

Thomas, C. D., Cameron, A., Green, R. E., Bakkenes, M., Beaumont, L. J., Collingham, Y. C., Erasmus, B. F. N., de Siqueira, M. F., Grainger, A., Hannah, L., Hughes, L., Huntley, B., van Jaarsveld, A. S., Midgley, G. F., Miles, L., Ortega-Huerta, M. A., Peterson, A. T., Phillips, O. L. & Williams, S. E. (2004). Extinction risk from climate change. *Nature,* 427(6970)**,** 145-148.

Thomas, C. D., Franco, A. M. A. & Hill, J. K. (2006). Range retractions and extinction in the face of climate warming. *Trends in Ecology & Evolution,* 21(8)**,** 415-416.

Thomas, C. D., Glen, S. W. T., Lewis, O. T., Hill, J. K. & Blakeley, D. S. (1999a). Population differentiation and conservation of endemic races: the butterfly, Plebejus argus. *Animal Conservation,* 2(1)**,** 15-21.

Thomas, C. D., Hill, J. K. & Lewis, O. T. (1998). Evolutionary consequences of habitat fragmentation in a localized butterfly. *Journal of Animal Ecology,* 67(3)**,** 485-497.

Thomas, J. A., Rose, R. J., Clarke, R. T., Thomas, C. D. & Webb, N. R. (1999b). Intraspecific variation in habitat availability among ectothermic animals near their climatic limits and their centres of range. *Functional Ecology,* 13**,** 55-64.

Thuiller, W. (2003). BIOMOD - optimizing predictions of species distributions and projecting potential future shifts under global change. *Global Change Biology,* 9(10)**,** 1353-1362.

Thuiller, W. (2004). Patterns and uncertainties of species' range shifts under climate change. *Global Change Biology,* 10(12)**,** 2020-2027.

Thuiller, W., Brotons, L., Araujo, M. B. & Lavorel, S. (2004). Effects of restricting environmental range of data to project current and future species distributions. *Ecography,* 27(2)**,** 165-172.

Thuiller, W., Lavorel, S. & Araujo, M. B. (2005). Niche properties and geographical extent as predictors of species sensitivity to climate change. *Global Ecology and Biogeography,* 14(4)**,** 347-357.

Tilman, D. & Lehman, C. (2000). Human-caused environmental change: Impacts on plant diversity and evolution. *National Academy of Sciences Colloquium on the Future of Evolution, Irvine, California*. 5433-5440.

Tobalske, C. (2002). Effects of spatial scale on the predictive ability of habitat models for the green woodpecker in Switzerland. IN: Scott, J. M., Heglund, P. J., Morrison, M. L., Haufler, J. B., Raphael, M. G., Wall, W. A. & Samson, F. B. (Eds.). *Predicting species occurrences: issues of accuracy and scale.* Washington: Island Press, 197-204.

Toberman, H., Freeman, C., Evans, C., Fenner, N. & Artz, R. R. E. (2008). Summer drought decreases soil fungal diversity and associated phenol oxidase activity in upland *Calluna* heathland soil. *FEMS Microbiology Ecology,* 66(2)**,** 426-436.

Toone, G. (2005). Debates on the nativity of Wild Gladiolus. *BSBI Recorder,* February 2005**,** 17-19.

Trotter, S. (2007). *Climate change update (New Forest National Park Authority Meeting - 18 January 2007).* Lymington, UK: New Forest National Park Authority.

Tsoar, A., Allouche, O., Steinitz, O., Rotem, D. & Kadmon, R. (2007). A comparative evaluation of presence-only methods for modelling species distribution. *Diversity and Distributions,* 13(4)**,** 397-405.

Tubbs, C. R. (2001). *The New Forest: history, ecology and conservation.* Lyndhurst: New Forest Ninth Centenary Trust.

Tyler, S. J. (1972) The behaviour and social organisation of the New Forest ponies. IN: Cullen, J. M., Beer, C. G. & Newark, N. J. (Eds.) *Animal behaviour monologues (Volume 5, Part 2).*

Tyre, A. J., Possingham, H. P. & Lindenmayer, D. B. (2001). Inferring process from pattern: can territory occupancy provide information about life history parameters? *Ecological Applications,* 11(6)**,** 1722-1738.

Uhmann, T. V., Kenkel, N. C. & Baydack, R. K. (2001). Development of a habitat suitability index model for Burrowing Owls in the eastern Canadian prairies. *Journal of Raptor Research,* 35(4)**,** 378-384.

UK Biodiversity Action Plan (2006). *Species Action Plan - New Forest Cicada (Cicadetta montana).* Available from: http://www.ukbap.org.uk/UKPlans.aspx?ID=216 [Accessed 13 November 2006].

UK Biodiversity Action Plan (2007). *Species Action Plan - Nail Fungus (Poronia punctata).* Available from: http://www.ukbap.org.uk/UKPlans.aspx?ID=528 [Accessed 8 October 2007].

UK Biodiversity Action Plan (2008). *Species Action Plan - Pillwort (Pilularia globulifera).* Available from: http://www.ukbap.org.uk/UKPlans.aspx?ID=518 [Accessed 6 March 2008].

UK Climate Impacts Programme (2009a). *Key findings: South East England*. Oxford: UK Climate Impacts Programme. Available from: http://ukclimateprojections.defra.gov.uk/content/view/2155/528/index.html [Accessed 30 June 2009].

UK Climate Impacts Programme (2009b). *Reference: emissions scenario*. Oxford: UK Climate Impacts Programme. Available from: http://ukclimateprojections.defra.gov.uk/content/view/551/9/ [Accessed 6 October 2009].

United Nations Environment Programme (2002). *Report on the Sixth Meeting of the Conference of the Parties to the Convention on Biological Diversity (UNEP/CBD/COP/6/20/Part 2) Strategic Plan Decision VI/26 (Convention on Biological Diversity, 2002).* Available from: http://www.biodiv.org/decisions/default.aspx?m=COP-06&id=7200&lg=0 [Accessed 11 January 2007].

Uusitalo, L., Kuikka, S. & Romakkaniemi, A. (2005). Estimation of Atlantic salmon smolt carrying capacity of rivers using expert knowledge. *Ices Journal of Marine Science,* 62(4)**,** 708-722.

van der Gaag, L. C., Renooij, S., Witteman, C. L. M., Aleman, B. M. P. & Taal, B. G. (1999). How to elicit many probabilities. IN: Laskey, K. B. P. H. (Ed.) *15th Conference on Uncertainty in Artificial Intelligence, Stockholm, Sweden*. 647-654.

Varis, O. & Kuikka, S. (1999). Learning Bayesian decision analysis by doing: lessons from environmental and natural resources management. *Ecological Modelling,* 119(2-3)**,** 177-195.

Vaughan, I. P. & Ormerod, S. J. (2005). The continuing challenges of testing species distribution models. *Journal of Applied Ecology,* 42(4)**,** 720-730.

Vaughan, P. (2008). Grayling *Hipparchia semele*. IN: Barker, L. & Norriss, T. (Eds.). *Hampshire and Isle of Wight butterfly and moth report, 2007.* Wareham, Dorset: Hampshire and Isle of Wight Branch of Butterfly Conservation, 80-81.

Venables, W. N. & Ripley, B. D. (2002). *Modern applied statistics with S.* 4th ed. New York: Springer.

von Winterfeldt, D. & Edwards, W. (1986). *Decision analysis and behavioral research.* Cambridge: Cambridge University Press.

Walker, K. J. & Preston, C. D. (2006). Ecological predictors of extinction risk in the Flora of Lowland England, UK. *Biodiversity and Conservation,* 15(6)**,** 1913-1942.

Walmsley, C. A. (2007). Introduction. IN: Walmsley, C. A., Smithers, R. J., Berry, P. M., Harley, M., Stevenson, M. J. & Catchpole, R. (Eds.). *MONARCH - Modelling Natural Resource Responses to Climate Change - a synthesis for biodiversity conservation.* Oxford: UK Climate Impacts Programme.

Walmsley, C. A. & Harley, M. (2007). Conclusions and the implications for policy and practice. IN: Walmsley, C. A., Smithers, R. J., Berry, P. M., Harley, M., Stevenson, M. J. & Catchpole, R. (Eds.). *MONARCH - Modelling Natural Resource Responses to Climate Change - a synthesis for biodiversity conservation.* Oxford: UKCIP.

Walther, G. R. (2004). Plants in a warmer world. *Perspectives in Plant Ecology Evolution and Systematics,* 6(3)**,** 169-185.

Walther, G. R., Post, E., Convey, P., Menzel, A., Parmesan, C., Beebee, T. J. C., Fromentin, J. M., Hoegh-Guldberg, O. & Bairlein, F. (2002). Ecological responses to recent climate change. *Nature,* 416(6879)**,** 389-395.

Warren, M. & Wigglesworth, T. (2008). *Silver-studded Blue factsheet*. Wareham: Butterfly Conservation.

Warren, M. S., Hill, J. K., Thomas, J. A., Asher, J., Fox, R., Huntley, B., Roy, D. B., Telfer, M. G., Jeffcoate, S., Harding, P., Jeffcoate, G., Willis, S. G., Greatorex-Davies, J. N., Moss, D. & Thomas, C. D. (2001). Rapid responses of British butterflies to opposing forces of climate and habitat change. *Nature,* 414(6859)**,** 65-69.

Webb, N. (1986). *Heathlands.* London: Collins.

Webster, J. (1999). Fungi of dung. IN: Cox, J. (Ed.) *The biodiversity of animal dung.* Eastleigh: Hampshire and Isle of Wight Wildlife Trust, 7-12.

Werkman, B. R. & Callaghan, T. V. (2002). Responses of bracken and heather to increased temperature and nitrogen addition, alone and in competition. *Basic and Applied Ecology,* 3(3)**,** 267-276.

Wessel, W. W., Tietema, A., Beier, C., Emmett, B. A., Penuelas, J. & Riis-Nielsen, T. (2004). A qualitative ecosystem assessment for different shrublands in western Europe under impact of climate change. *Ecosystems,* 7(6)**,** 662-671.

Whalley, A. J. S. & Dickson, G. C. (1986). *Poronia punctata*, a declining species? *Bulletin of the British Mycological Society,* 20(1)**,** 54-57.

Whitehead, S. J., Caporn, S. J. M. & Press, M. C. (1997). Effects of elevated CO_2, nitrogen and phosphorus on the growth and photosynthesis of two upland perennials: *Calluna vulgaris* and *Pteridium aquilinum*. *New Phytologist,* 135(2)**,** 201-211.

Whittaker, R. J., Araujo, M. B., Paul, J., Ladle, R. J., Watson, J. E. M. & Willis, K. J. (2005). Conservation biogeography: assessment and prospect. *Diversity and Distributions,* 11(1)**,** 3-23.

Whittingham, M. J., Stephens, P. A., Bradbury, R. B. & Freckleton, R. P. (2006). Why do we still use stepwise modelling in ecology and behaviour? *Journal of Animal Ecology,* 75(5)**,** 1182-1189.

Wicklow, D. T. & Hirschfield, B. J. (1979). Evidence of a competitive heirarchy among coprophilous fungal populations. *Canadian Journal of Microbiology,* 25(7)**,** 855-858.

Wiens, J. A. (2002). Predicting species occurrences: progress, problems and prospects. IN: Scott, J. M., Heglund, P. J., Morrison, M. L., Haufler, J. B., Raphael, M. G., Wall, W. A. & Samson, F. B. (Eds.). *Predicting species occurrences: issues of accuracy and scale.* Washington: Island Press, 739-749.

Willis, K. J. & Bhagwat, S. A. (2009). Biodiversity and climate change. *Science,* 326(5954)**,** 806-807.

Willott, S. J. & Hassall, M. (1998). Life-history responses of British grasshoppers (Orthoptera : Acrididae) to temperature change. *Functional Ecology*, 12(2)**,** 232-241.

Wilson, E. O. (2001). *The diversity of life.* New ed. London: Penguin Books.

Wilson, K. A. (2003). *Uncertainty and vulnerability in conservation planning*. Thesis (Doctor of Philosophy). University of Melbourne.

Wilson, K. A., Westphal, M. I., Possingham, H. P. & Elith, J. (2005). Sensitivity of conservation planning to different approaches to using predicted species distribution data. *Biological Conservation,* 122(1)**,** 99-112.

Windschitl, P. D. & Wells, G. L. (1996). Measuring psychological uncertainty: verbal versus numeric methods. *Journal of Experimental Psychology: Applied,* 2(4)**,** 343-364.

Winship, H. & Chatters, C. (1994). *Chamaemelum nobile* (L.) All., Chamomile. IN: Stewart, A., Pearman, D. A. & Preston, C. D. (Eds.). *Scarce plants in Britain.* Peterborough: JNCC, 110.

Winship, H. R. (1993). *The current distribution and conservation requirements of Chamomile (Chamaemelum nobile (L.) All.) in Britain*. Curdridge, UK: Unpublished report, obtained from Hampshire and Isle of Wight Wildlife Trust.

Winship, H. R. (1998). *A survey of some extant and extinct locations of Chamomile (Chamaemelum nobile) in southern England*. Salisbury: Plantlife International, 100.

Wisz, M. S., Hijmans, R. J., Li, J., Peterson, A. T., Graham, C. H., Guisan, A. & NCEAS Predicting Species Distributions Working Group (2008). Effects of sample size on the performance of species distribution models. *Diversity and Distributions,* 14(5)**,** 763-773.

Witteman, C. & Renooij, S. (2003). Evaluation of a verbal-numerical probability scale. *International Journal of Approximate Reasoning*, 33(2)**,** 117-131.

Wollan, A. K., Bakkestuen, V., Kauserud, H., Gulden, G. & Halvorsen, R. (2008). Modelling and predicting fungal distribution patterns using herbarium data. *Journal of Biogeography,* 35(12)**,** 2298-2310.

Wright, R. N. & Westerhoff, D. V. (2001). *New Forest Special Area of Conservation management plan 2001*. Lyndhurst: English Nature.

Yamada, K., Elith, J., McCarthy, M. & Zerger, A. (2003). Eliciting and integrating expert knowledge for wildlife habitat modelling. *Ecological Modelling*, 165(2-3)**,** 251-264.

Zaniewski, A. E., Lehmann, A. & Overton, J. M. C. (2002). Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. *Ecological Modelling*, 157(2-3)**,** 261-280.

APPENDICES

Appendix 1. Information about the New Forest National Park study site

1.1. Introduction to the New Forest

The New Forest, designated a national park in March 2005, is a unique semi-natural landscape covering an area of 571 square kilometres in south-west Hampshire, England (OS Grid Reference: 505159N, 014050W) (New Forest Committee, 2003). It lies immediately to the north of the Solent, between the major conurbations of Bournemouth to the west and Southampton to the east (Wright and Westerhoff, 2001). Figure A-1 below shows a map of the New Forest.

Figure A-1. Map of the New Forest National Park (from: New Forest National Park Authority, 2007b)

The New Forest encompasses a matrix of habitats, shaped by its use throughout history as a Royal hunting forest, designated by William the Conqueror in 1079. In addition, the rights of common (attached to New Forest properties) developed from medieval times have also been a huge influence. These common rights cover an area of over 38,000 hectares of the forest, known as the historic ‘Perambulation’, where commoners’ animals can roam freely (New Forest National Park Authority, 2006b). Many of these rights are no longer practised but the rights of pasture (for ponies and cattle) and pannage (the foraging of pigs) are still very important, with between 400 and 500 commoners still practising their rights (New Forest National Park Authority, 2006a). The exercise of common rights is an intimate part of the ecosystem of the New Forest, with the influence of these large herbivores able to be seen throughout the Forest (Tubbs, 2001).

Habitats of the New Forest include ancient pasture woodland, lowland heath, grassland, valley and seepage step mire, or fen, and numerous wetland environments, but according to Tubbs (2001) it is the unique mosaic of these habitats that create a biological wealth that is greater than the sum of its parts. Nowhere else do these habitats occur in combination and on so large a scale and many of these habitats are now rare in lowland western Europe (English Nature, 1996). This is in part due to the demise of grazing, as it may only be in the New Forest that a pastoral economy based on the exercise of common rights over unenclosed common land truly survives, albeit in a modified form (Tubbs, 2001). Grazing is important for many of the species which can be found in the New Forest, such as the rare marsh gentian (*Gentiana pneumonanthe*) (Tubbs, 2001).

1.2. Climate and biophysical characteristics of the New Forest

The New Forest has a mild temperate climate, with westerly winds bringing a succession of low and high-pressure systems leading to frequent variations of small amplitude in the weather, characterised as wet and mild (Wright and Westerhoff, 2001). The summers are warm, with the mean temperature of the warmest month between 19^{o}C and 22^{o}C, and the winters are mild, with a mean temperature for the coldest month between 0^{o}C and 8^{o}C and frequent frosts, but rare incidence of snowfall (Wright and Westerhoff, 2001). Rainfall occurs throughout the year, with an average total annual precipitation of 763.7 mm, with the driest month (July) having an average rainfall of 37.7mm (Met Office, 2009).

The New Forest lies in part of a broad and shallow chalk syncline, on Tertiary sands and clays with overlying gravels (Rose, 1996; Wright and Westerhoff, 2001). More recent material continues to be deposited in the form of river alluvium and peat (Wright and Westerhoff, 2001). The relative fertility of the New Forest soils follows a north-south gradient with the underlying rock strata, generally becoming less impoverished from north to south, with less leaching of nutrients and lower acidity on the finer clays towards the south (Wright and Westerhoff, 2001). The soils of the New Forest influence its ecology, with a greater diversity in the flora in the south of the Forest on the more nutrient-rich, less base-poor soils than on the more acid, nutrient poor sands and gravels of the north (Wright and Westerhoff, 2001). The soils have also impacted on the way the land has been used traditionally: the fact that much of the New Forest soils are generally infertile, acidic and often poorly drained means the Forest has not been extensively used for crop cultivation (New Forest Committee, 2003).

1.3. Biodiversity of the New Forest

The diverse habitats of the New Forest include a mosaic of heathlands, grasslands, wetlands and ancient woodland and support a rich and characteristic flora and fauna, many species (or elements) of which are rare and nationally or internationally important (English Nature, 1996; Tubbs, 2001; Wright and Westerhoff, 2001).

The heathlands comprise a series of plant communities, including the dry heath (and associated dry grasslands), which grades into the wetter humid heath (and associated valley mires, streams, ponds, temporary pools and wet grasslands) (English Nature, 1996; Tubbs, 2001; Wright and Westerhoff, 2001). With approximately 19,500 hectares of lowland heath, the largest area of this rare habitat remaining in the UK, the New Forest supports many species that are partly or wholly dependent on the area (New Forest National Park Authority, 2006b), for example the Dartford warbler (*Sylvia undata*), is listed in Annex 1 of the EU Directive on the Conservation of Wild Birds (English Nature, 1996).

Within the heathland mosaic, on pockets of richer soils, acid grassland can occur. The species that occur in these habitats can generally be grouped into either species that benefit from heavy grazing, which grow as a prostrate or dwarf form (such as rosette-forming species and small herbs), or less palatable species such as the wild gladiolus (*Gladiolus illyricus*), which is confined in Britain to the New Forest (English Nature, 1996; Tubbs, 2001).

The more neutral grasslands (or lawns) vary with factors such as soils, topography floodwater nutrient quality and grazing pressure and the species present reflect this (English Nature, 1996). There are also areas in the New Forest of unimproved grasslands or meadows comprising communities that are now rare or scarce in England, and again their species composition is modified by factors such as soil type, water content and grazing (English Nature, 1996).

The New Forest contains ninety separate valley mires; there are no more than twenty in the rest of the English lowlands, and only a handful around the European littoral from Denmark to Spain (Tubbs, 2001). This international importance is reflected in the New Forests' designation as a Ramsar site (a wetland of international importance). The richest mires, for example, contain more than 150 plant species, some of which are locally distributed and rare (English Nature, 1996). The term 'mire' is used to encompass the variation in vegetation, which is a combination of bog (typically rain-fed, mineral and nutrient poor and acidic) and fen (groundwater-fed, with a higher nutrient status and generally neutral or alkaline) communities (Wright and Westerhoff, 2001).

Of particular note is the slender cottongrass (*Eriophorum gracile*), confined in England to sites in the New Forest and one in Surrey, and listed under Schedule 8 of the Wildlife and Countryside Act, as well as being a Red Data Book species (English Nature, 1996). Many invertebrate species also depend on the New Forest wetlands. For example, the black bog ant (*Formica transkaucasica*), is known only from five New Forest mires and some others in Purbeck (Dorset, southern England), so the New Forest provides vital support for these populations (Tubbs, 2001).

The New Forest wetlands support all but one (the natterjack toad, *Bufo calamita*) of Britain's native amphibians, including the rare great crested newt (*Triturus cristatus*), a Schedule 5 of the Wildlife and Countryside Act protected species, found in the less acidic ponds of the Forest (English Nature, 1996; Tubbs, 2001). Some of the New Forest ponds dry out in the summer (known as ephemeral or temporary ponds), providing ideal conditions for specially adapted invertebrates. The fairy shrimp (*Chirocephalus diaphanous*) and tadpole shrimp (*Triops cancriformis*) are both nationally rare, Red Data Book and Schedule 5 protected species, with the tadpole shrimp only known to occur in one of these New Forest ponds in the whole of Britain (English Nature, 1996).

The rivers and streams of the New Forest support a range of specialist plants and animals now rare in lowland Britain because of habitat loss, and because of a requirement by some for grazing (which has declined), to provide the resulting muddy conditions (Tubbs, 2001). Indeed, the plant community associated with these streams, the result of a combination of nutrient-poor acid waters and outcrops of neutral-enriched soils, is restricted to the New Forest and one other location (Bodmin Moor, Cornwall, England) (English Nature, 1996).

The New Forest also contains areas of marine wetland habitat. The National Park extends to the coast of the Solent from Keyhaven to Southampton Water (approximately 26 miles), which incorporates saltmarsh, mudflats, lagoons, shingle spits and low cliff habitats (New Forest Committee, 2003; New Forest National Park Authority, 2006b). These areas are of national and international importance for nature conservation and attract large numbers of wildfowl and waders and some unusual plant and invertebrate species (New Forest Committee, 2003).

Perhaps the habitat that the New Forest is best known for is its woodlands. The unenclosed (pasture) woodlands extend to some 4,430 hectares (excluding riverine and bog woodland) and are dominated by oak (*Quercus robur*) and beech (*Fagus sylvatica*), depending largely on the soil type, with holly (*Ilex aquifolium*) as the dominant shrub layer species (English Nature, 1996; Wright and Westerhoff, 2001). The older trees (some of which are of early 17th century origin) support the richest known woodland lichen flora in lowland Europe, including the rare *Catillaria laureri* and *Paramelia minarium* species which are both Schedule 8 protected (English Nature, 1996).

These older trees also provide for an exceptionally species-rich deadwood fauna, with 75-90% of the 600-800 species inhabiting deadwood in Britain occurring in the

New Forest (Tubbs, 2001). One of these species is the stag beetle (*Lucanus cervus*), which is rare in Europe but remains common in the New Forest area (Tubbs, 2001). The ancient woodlands of the New Forest are also particularly rich in fungi, although fungi can also be found on the grassland and heathland areas as well. Of the 12,000 species of fungi present in the UK, 2,700 can be found in the New Forest, including many rare species (Forestry Commission, 2006b).

In addition the older trees provide ideal nesting and roosting sites for birds, such as the redstart (*Phoenicurus phoenicurus*), and bats, such as the very rare Schedule 5 protected Bechstein's bat (*Myotis bechsteini*) (English Nature, 1996). The New Forest is also home to the only representative of the *Cicadidae* in the UK; the New Forest Cicada (*Cicadetta montana*), which occurs at warm woodland edges and open scrub, and is a Schedule 5 protected species (UK Biodiversity Action Plan, 2006).

The riverine woodlands (approximately 212 hectares) comprise those floodplain woodland communities dominated by oak and ash or by alder which, where not damaged by over-deepening of drainage channels, flood seasonally as water levels rise along meandering natural flood channels (Wright and Westerhoff, 2001). Bog woodland, of which there is approximately 250 hectares in the New Forest, occurs on peat in which there is a significant component of bog species in the ground flora (Wright and Westerhoff, 2001).

The sylvicultural Inclosures can provide for some different species to the pasture woodland. Where grazing pressure is lower it can allow for a ground flora which includes, for example, important populations of three nationally rare plants; narrow-leaved lungwort (*Pulmonaria longifolia*), bastard balm (*Melittis melissophyllum*) and soft-leaved sedge (*Carex montana*) (Tubbs, 2001). Certain invertebrates, which use this flora as food plants such as the silver-washed fritillary butterfly (*Argynnis paphia*) can also be found (English Nature, 1996). However, many of the plantation Inclosures, particularly those dominated by conifers, have low light availabilities beneath the trees which results in a sparse ground flora (Tubbs, 2001). Nevertheless, the conifers, mainly Scots pine (*Pinus sylvestris*), have been found to provide important nest sites in the New Forest for several bird species, such as buzzards (*Buteo buteo*) and hobbys (*Falco subbuteo*) (Tubbs, 2001).

1.4. Conserving the New Forest and the organisations responsible

The uniqueness and international importance of the New Forest is described by Tubbs (2001) as due to its size, intactness and high biodiversity, the existence of habitats largely lost elsewhere, and a socio-economic system which is inextricably part of its ecology, which has also been mostly lost in north-west Europe. This uniqueness of the New Forest has resulted in many forms of protection and special designations, particularly because of its large areas of lowland heath, valley mire and ancient woodland pasture (Forestry Commission, 2006a).

In 1923 the Forestry Commission was made responsible for the management of almost half (27,000 hectares) of the total area of the National Park, mainly the Crown Lands, and is still responsible for the day-to-day management of the New Forest (New Forest National Park Authority, 2006a). Responsibility for the New Forest is also in combination with other bodies such as Natural England, the New Forest Committee, the Verderers, landowners, farmers and commoners. Tubbs (2001) suggests that the Verderers can be seen as the guardians of the commoners and common rights as well as of the landscape and natural habitats of the New Forest, a duty exercised through their power to veto development.

On 1st March 2005, the New Forest was designated a National Park, making it the first in the south-east and the first in England for nearly 50 years. National Park status is the highest level of countryside designation, ensuring the strongest permanent protection for the future (New Forest National Park Authority, 2006b). As a result of the National Park designation, the New Forest National Park Authority took on its full duties in April 2006 and is responsible for making decisions, setting policies and priorities and for making sure that resources are used properly in the Forest (New Forest National Park Authority, 2006b).

As well as being designated a National Park, the New Forest has been given protection from several other designations. The New Forest forms part of the South Hampshire Coast Area of Outstanding Natural Beauty (1967) and contains 37 Countryside Heritage Sites identified by Hampshire County Council (Forestry Commission, 2006a). The New Forest Heritage Area was identified in 1985 and special planning policies were adopted to protect it, for which the New Forest

Committee is responsible for promoting and coordinating (New Forest National Park Authority, 2006a). Other conservation designations include; Sites of Special Scientific Interest (SSSI), a Wetland of International Importance under the Ramsar Convention in 1993 and a Special Protection Area under the EC Wild Birds Directive (1979) in 1993 (Forestry Commission, 2006a).

The New Forest contains over 600 species of conservation concern as identified in the Special Area of Conservation Management Plan (Wright and Westerhoff, 2001). Many of these species are red listed (IUCN) or protected by Schedule 5 or 8 of the Wildlife and Countryside Act (1981). In addition it contains over 100 species for which Species Action Plans have been implemented, as part of the UK Biodiversity Action Plan (UK BAP) and many of the habitats of the New Forest are also protected, such as by Habitat Action Plans, including ancient semi-natural woodland, heathland, acid grassland and bog (Wright and Westerhoff, 2001).

The New Forest is very much a managed landscape, with intervention required to ensure the persistence of many of the habitats and species as they are at present. For example, if the New Forest was left to itself, in the total absence of grazing, cutting and burning, the drier habitats would quickly succeed to scrub and woodland and the wetlands would rapidly become dominated by *Molinia* tussocks and would progress towards closed carr habitats (Wright and Westerhoff, 2001). Not everyone agrees on how the New Forest should be managed and conflicts of interest occur. Obviously, management of the New Forest is a complex task and is made even more complicated by the partly unknown challenges for the future.

Appendix 2. Information about species selected for modelling

2.1. Wild chamomile (*Chamaemelum nobile* (*L.*) All.)

Figure A-2. Wild chamomile (*Chamaemelum nobile*) close up (left) and in poached ground (right).

Description:

Chamaemelum nobile (Family: Asteraceae) is a creeping, downy, greyish, native perennial herb that is pleasantly aromatic, with yellow disc florets and white ray florets in solitary heads (18-24mm, from June to August); and can reach a height of up to 25 cm (Sterry, 2006; see Figure A-2 above). *C. nobile* can be a long-lived plant in the right conditions, reproducing both by clonal spread, and in some populations, by seed (Winship and Chatters, 1994; Plantlife, 2001). Under heavy grazing pressures it adopts a semi-prostrate, often non-flowering form where the stems creep out parallel to the ground and thus avoid being nibbled (Winship and Chatters, 1994); in the New Forest the plants are generally of this prostrate form (Tubbs, 2001). *C. nobile* is very much a flagship species indicative of exceptional habitat and is associated with many rare and scarce species (Winship, 1998) and is classified as vulnerable in Great Britain, as well as being a UK Biodiversity Action Plan priority species (www.ukbap.org.uk).

***Habitat*:**

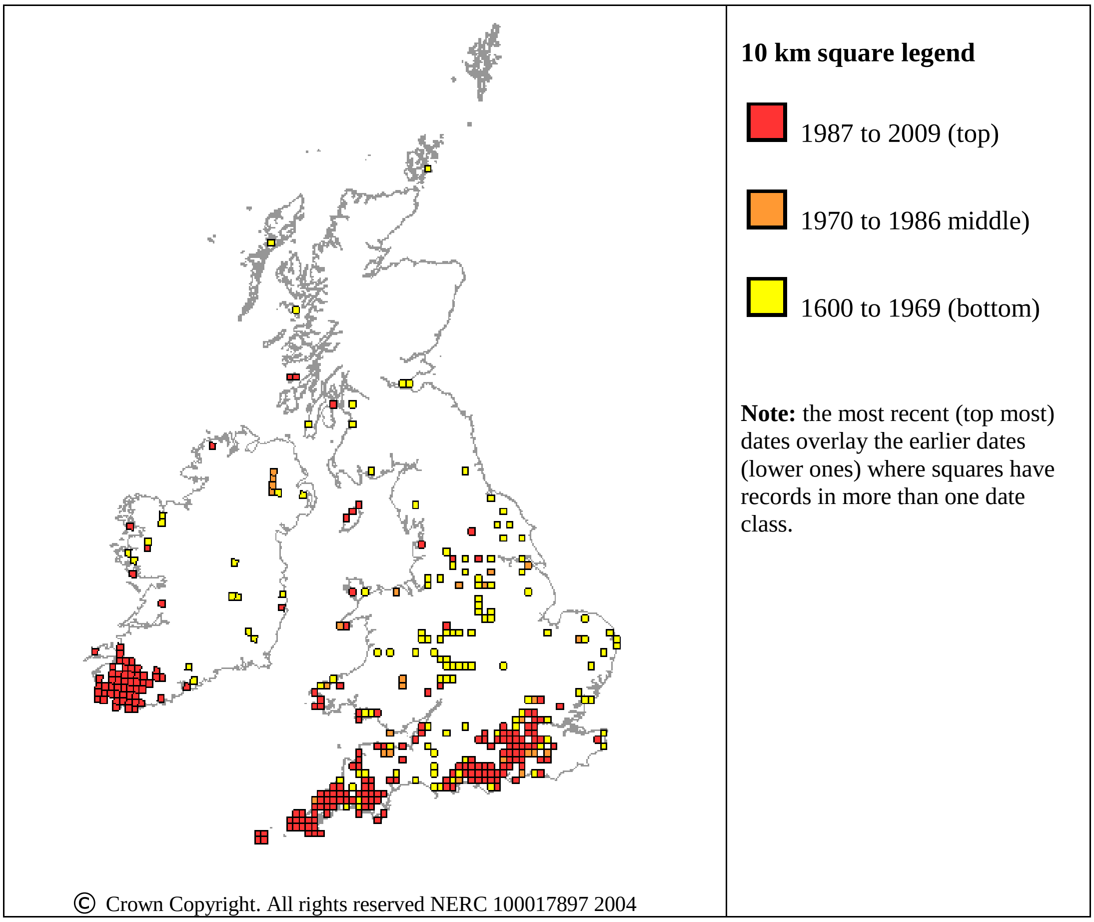
Five hundred years ago *C. nobile* was used for lawns, and it still prefers mown or grazed grassland in open places, such as sandy heaths, cricket pitches, open glades in light woodland (Plantlife, 2001; 2007a) or New Forest lawns (Tubbs, 2001). It is adapted to living in closely grazed grass, and so is tolerant both of stock-grazing and of some mowing regimes where the mowing, trampling or grazing discourages competitors (Plantlife, 2001). However, on sports pitches, although the mowing mimics grazing pressure, the lack of soil disturbance restricts opportunist *C. nobile* plants and produces a different, less diverse sward (Winship, 1993).

C. nobile likes open sunny situations, and thrived in the past on land where grazing animals kept down taller plants and created areas of open ground which it could colonise (Plantlife, 2001). It generally occurs on moderately acidic grassland (Killick, 2002) and can grow in fairly rich soils, but if artificial fertilisers are heavily applied to pastures, these encourage vigorous, nitrogen-hungry plants which may out-compete *C. nobile* (Plantlife, 2001). It is also killed by the herbicides which are often applied to lawns and pastures to control 'weeds'. *C. nobile* flourishes best in areas which are prone to winter flooding (Plantlife, 2001).

The main reasons for the decline of *C. nobile* are the loss of former grasslands and heaths, drainage of winter-wet areas in grasslands, conversion of pastures to arable land, decline of grazing on old pastures, heaths and commons, loss of unmade tracks and their associated flower-rich verges, and increased use of artificial fertilisers and herbicides (Plantlife, 2001).

Distribution:

Formerly *C. nobile* was fairly common in Britain, but northern populations have all but disappeared and it has become scarce even in the south, where it was once widespread (see Figure A-3 below). Strongholds remain in the Scilly Isles, the south-west of England and the New Forest (Plantlife, 2001). It is widespread in western Europe from Belgium, where it may be adventive, southwards to Algeria and the Azores (Winship and Chatters, 1994).



Note: the most recent (top most) dates overlay the earlier dates (lower ones) where squares have records in more than one date class.

Figure A-3. 10 km squares with records for *Chamaemelum nobile* (Wild chamomile) in Great Britain and Ireland (from National Biodiversity Network (2009)).

2.2. Slender marsh-bedstraw (*Galium constrictum* Chaub.)

Figure A-4. Slender Marsh-bedstraw (*Galium constrictum*). Close-up (left) and location in a roadside ditch (*G. constrictum* is at the bottom centre of the picture on the right).

***Description*:**

Galium constrictum (Family Rubiaceae) is a native, perennial herb (Meek, 2002), which is classified as rare in Great Britain. It is described as decumbent to ascending or scrambling, smooth or slightly scabrid perennial to 40 cm (Stace, 1997). It has white flowers, 2-3 mm across with four petals in few-flowered clusters (Sterry, 2006; see Figure A-4 above).

***Habitat*:**

G. constrictum is found in marshy places, ditches and pond-sides (Stace, 1997), particularly around the margins of ponds which dry out in summer, as well as on New Forest ‘lawns’ and in track ruts, and locally in marl-pits and ditches in water-meadows (Meek, 2002). It occurs mainly on acid soils (Hill *et al*., 2004).

Distribution:

G. constrictum is very local in south Hampshire, south Wiltshire, south Devon, and the Channel Islands (Stace, 1997; see Figure A-5 below). Meek (2002) suggests that the many new records for *G. constrictum* in recent years probably result from a genuine increase in the number of sites as well as greater recording effort and that the New Forest populations seem reasonably secure as long as current management practices continue. *G. constrictum* also occurs in southern and south western Europe from France, Spain and Portugal to southern Balkans and Crete (Clapham *et al.*, 1990).

Ⓒ Crown Copyright. All rights reserved NERC 100017897 2004

10 km square legend

🟥 1987 to 2009 (top)

🟧 1970 to 1986 middle)

🟨 1600 to 1969 (bottom)

Note: the most recent (top most) dates overlay the earlier dates (lower ones) where squares have records in more than one date class.

Figure A-5. 10 km squares with records for *Galium constrictum* (Slender marsh-bedstraw) in Great Britain and Ireland (from National Biodiversity Network (2009)).

2.3. Wild gladiolus (*Gladiolus illyricus* Koch)

Figure A-6. Wild gladiolus (*Gladiolus illyricus*). Close-up on left and situation in bracken habitat on right.

Description:

Gladiolus illyricus is a cormous perennial with bluish-green erect stems which are generally 25-50 cm (although can be up to 90cm) and reddish-purple flowers which are 3.5 – 4 cm in diameter (Sell and Murrell, 1996; see Figure A-6 above). *G. illyricus* reproduces primarily by offsets, as flowering and seed production appear to be limited (Stokes, 2000; Taylor, 2002). Each dying flowering plant contains up to 60 cormlets, which in the following season become new plants. These new plants wait at least four years before venturing to produce blooms themselves (Stokes, 2000).

Stokes (1987) reported that pollination was found to be mainly by the large skipper butterfly (*Ochlodes venata*) and occasionally other insects (gate keeper (*Pyronia tithonus*), and two bee species (*Bombus pascorum* and *Bombus lucorum*)), although many other species are likely to be involved. The only data available on this is from Stokes's (1987) study where the sample sizes do not permit highly confident statements (Toone, 2005). The means of dispersal are not yet known, apart from gradual site migration (Toone, 2005).

There is some debate about taxonomic status of Wild Gladiolus in Britain, as to whether it should be treated as a species or whether it is sufficiently distinct to be worthy of subspecific rank (Stokes, 1987; Toone, 2005; Botanical Society of the British Isles, 2008a). Hamilton (personal communication, cited in Stokes (1987)) is of the opinion that *G. illyricus* in Britain is sufficiently different from its European counterparts to warrant its designation as the separate subspecies ‘britanicus’. Attempts to clarify this matter, such as by Buchanan (2007), have been inconclusive (Botanical Society of the British Isles, 2008a). *G. illyricus* is currently listed under Schedule 8 of the Wildlife and Countryside Act 1981.

Habitat:

Stokes (1987) reports that of the 50 known colonies in the New Forest at the time of his study (the most comprehensive study of the ecology of *G. illyricus* to date), all but two of the known populations were in bracken dominated acid grass heath. Most of the colonies were found in bracken alongside a woodland edge, although colonies did exist up to one kilometre from the nearest woodland. There is a widely held assumption that *G. illyricus* is a relic species of ancient wood pasture, due to this association with woodland and because several species frequently used as ancient woodland indicators have been recorded as associates of *G. illyricus* (Stokes, 1987). However, Stokes (1987) reports that indicator of other habitats are also present.

As noted by Stokes (1987), grazing pressure in the acid grass heaths of the New Forest is controlled by the density and height of the gorse and bracken, as large herbivores do not graze in dense bracken stands during the summer, and this period during which grazing animals avoid the bracken stands is also the period during which *G. illyricus* flowers, resulting in fewer potentially viable seeds being lost to grazing (Stokes, 1987).

Stokes (1987) did not find any *G. illyricus* plants outside of bracken stands and suggests that since newly emerging juvenile Gladioli strongly resemble grass, it is unlikely that new plants could ever get established outside the bracken, where grazing is heavy, even if they are unpalatable. Whether there are in fact Gladioli in open areas of the Forest whose visible presence is suppressed only by grazing is unknown, but there is historical evidence that Gladiolus was found in grassy open tracts of one of the plantations (where grazing pressure would be lower).

The density of the bracken was also recognised as important by Stokes (1987) who reported that *G. illyricus* was not found in sparse bracken (with an open canopy and a ground flora essentially the same as in adjacent areas where the bracken is absent), but favoured areas of less vigorous bracken (with an incomplete canopy cover and a relatively well developed ground flora), where the flowering plants produced significantly larger numbers of seed pods, rather than vigorous bracken (with a complete canopy cover, and impoverished ground flora). The three bracken communities were those recognised by Nicholson (1976).

The more open less vigorous bracken sites appeared to be created by the action of frost, which depletes the density of bracken fronds, and the greater the level of grass under bracken (and the less bracken litter) the more likely it is that bracken will be killed by frost in subsequent years (Stokes, 1987). Stokes (1987) reported that the balance between this system is critical to *G. illyricus* because a dramatic decrease in the amount of bracken could lead to eradication of *G. illyricus* through overgrazing of the sward, whereas too much bracken litter could ‘smother’ *G. illyricus* out of existence and impede the penetration of the seeds and roots to the soil. The more open bracken sites, which are better for *G. illyricus* are also better for grass feeding butterflies like the large skipper to survive. There can therefore be higher pollination of *G. illyricus* in open areas, due to more insects and greater ease with which insects find plants (Stokes, 1987; 2000).

Despite the benefits that *G. illyricus* gains from growing within bracken, Stokes (1987) found that a build up of bracken litter is detrimental to its survival, with a significant decline in the number of *G. illyricus* plants with an increase in the depth of the bracken litter. Stokes (1987) also found an association of *G. illyricus* with a particular less acidic, fine sandy loam (brown earth soil) which was relatively high in its clay and silt fractions compared to other soils in the New Forest, but low in nutrients. Whether *G. illyricus* requires some specific nutrient or physical characteristic of the soil is unknown (Stokes, 1987), however Stokes (1987) suggests that it seems likely that *G. illyricus* is growing on soils that are: relatively unpodsolised; provide the necessary nutrients for survival without allowing competition to become too great; are damp enough to prohibit bracken from becoming dominant, without being so wet that the corms rot; are not too acidic; and are stoneless and hence do not prohibit the downwards movement of the developing corms.

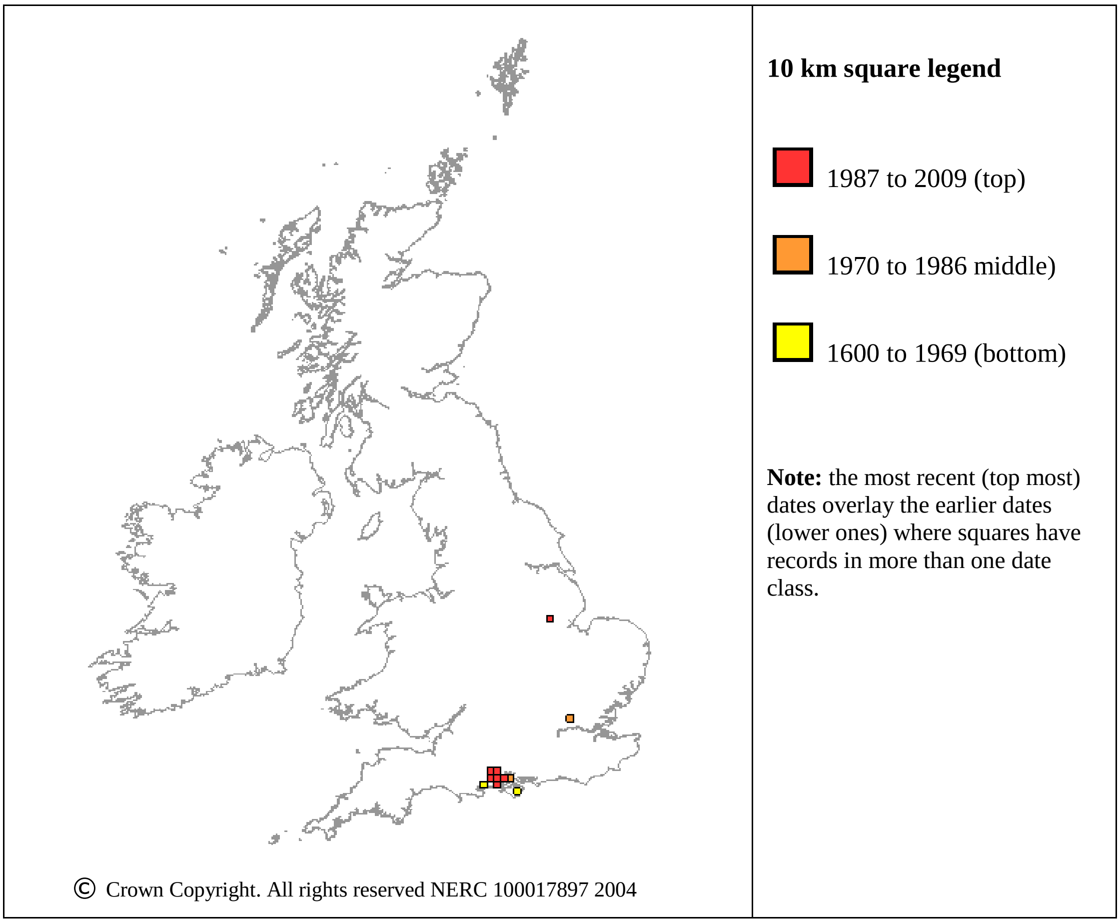
A further observation made by Stokes (1987) was that the aspect of the site did not appear to matter, but most of the sites were on a slight slope (up to 70 degrees), close to water (bogs, streams etc., from on site up to 100 metres, apart from one site), which suggests that the *G. illyricus* sites are flushing areas through which water from the upper slopes moves. Stokes (1987) suggests that the constant turnover of water and nutrients ensures that the soils do not become heavily podsolised, like much of the New Forest, while remaining damp throughout the year.

Distribution:

G. illyricus is, in Britain, largely restricted to the New Forest (see Figure A-7 below), although it was first discovered in Britain on the Isle of Wight in 1855 and then found a year later in the New Forest (Botanical Society of the British Isles, 2008a). The last sighting of *G. illyricus* in the Isle of Wight was in 1931 (Buchanan, 2007) and most recent accounts state that it is in decline in the New Forest, possibly as a result of changes in management, such as *Pteridium* control (Taylor, 2002), although there are more records, and more populations, now recorded than ever before (Botanical Society of the British Isles, 2008a). This could be because it is better recorded and because in the past botanists often used to keep localities secret in order to protect it from collectors (Botanical Society of the British Isles, 2008a). It is also a difficult plant to find because it often occurs under bracken, and by the time it flowers it can be completely hidden from view (Botanical Society of the British Isles, 2008a). There are several other recorded sites of *G. illyricus* outside of Hampshire but taxonomic uncertainties have led these records sometimes to be dismissed (Botanical Society of the British Isles, 2008a).

In Europe, *G. illyricus* is found in a much wider range of habitats than in Britain, including heaths, scrub, open woodland and coastal calcaereous cliffs (Stokes, 1987). This has lead to the suggestions that *G. illyricus* exists on the edge of its range in

England, in which case, its' restricted distribution to the warm heaths of southern Hampshire where the climate is mild and oceanic, is not surprising (Stokes, 1987). It occurs in south and west Europe (Sell and Murrell, 1996) and it (or its close relatives) is common around the Mediterranean basin as an agricultural weed and a casual of waste ground (Botanical Society of the British Isles, 2008a).



Note: the most recent (top most) dates overlay the earlier dates (lower ones) where squares have records in more than one date class.

© Crown Copyright. All rights reserved NERC 100017897 2004

Figure A-7. 10 km squares with records for *Gladiolus illyricus* (Wild gladiolus) in Great Britain and Ireland (from National Biodiversity Network (2009)).

2.4. Grayling (***Hipparchia semele*** subsp. *semele* (Linnaeus, 1758))

Figure A-8. Grayling (*Hipparchia semele*). The remarkable camouflage is demonstrated, particularly in the image on the right.

***Description*:**

Hipparchia semele (Lepidoptera: Nymphalidae) has cryptic colouring which provides it with excellent camouflage (see Figure A-8 above). The wings are kept closed when not in flight and the fore wings are usually tucked behind the hind wings, concealing the eyespots and making the butterfly appear smaller. In flight it is a distinctive, large butterfly (male wingspan is 51 – 56mm and female wingspan is 54 – 62mm) with a looping and gliding flight, during which the paler bands on the upper wings are visible (Bailey *et al.*, 1989; Asher *et al.*, 2001). *H. semele* is single brooded and lays eggs singly on plants growing in full sun, usually surrounded by bare ground. The larvae feed on the grass leaves at night and hibernate in tussocks during the winter and resume feeding in spring, becoming full-grown in June (Asher *et al.*, 2001). The flight period for *H. semele* is mid-July to mid-September (Oates *et al.*, 2000). In 2007 *H. semele* was named as a UKBAP priority species.

***Habitat*:**

H. semele is a butterfly of arid places, breeding in grass tussocks amongst bare ground, including bare pockets and along paths where vegetation is worn or heavily grazed (Oates *et al.*, 2000). On New Forest heaths, the species can be abundant where heathers are regenerating after burns, whereas in adjoining dense heather *H.*

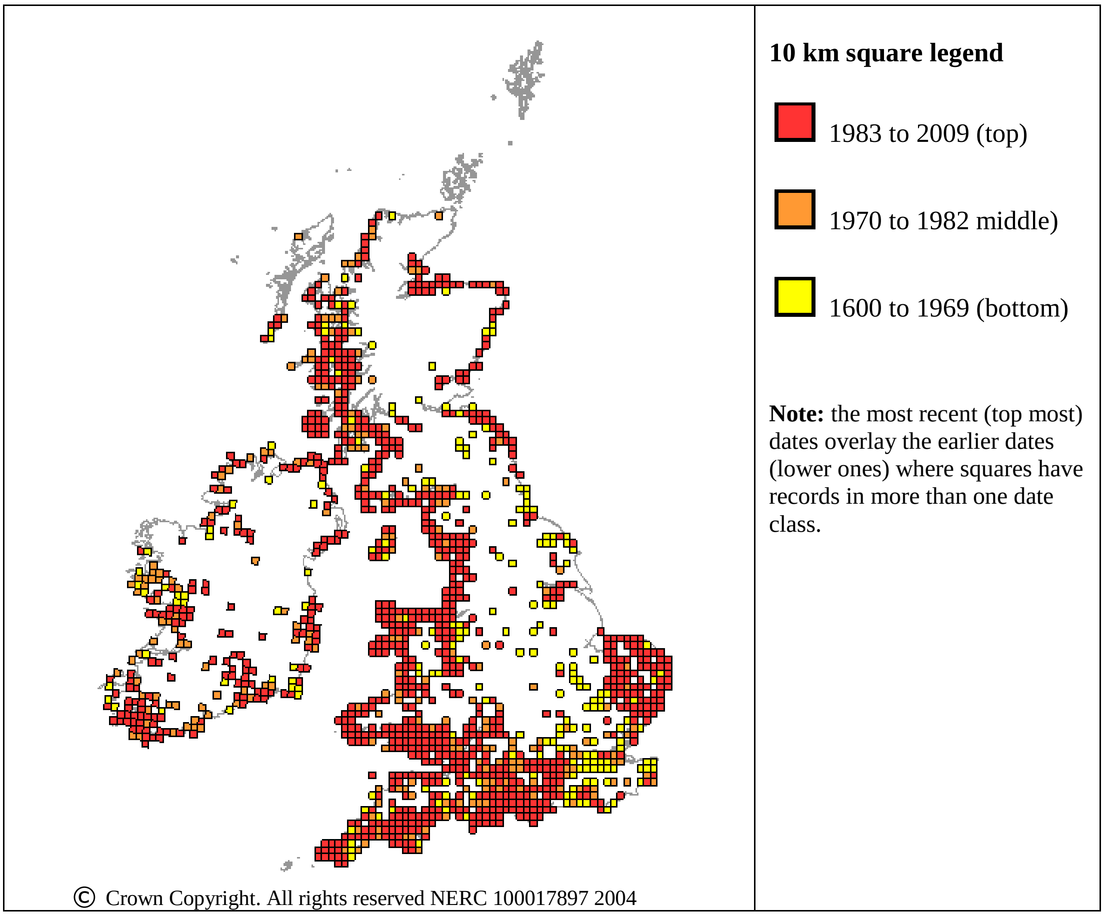
semele may be missing or present in sparse numbers. It is therefore very much a species of managed heathland and is absent, or in the process of dying out, from neglected heaths (Oates *et al.*, 2000). *H. semele* has such strong status within the New Forest that small colonies can be found along sunny rides with suitable vegetation in Inclosures, whilst individuals can be seen just about anywhere (Oates *et al.*, 2000).

In the New Forest, *H. semele* is strongly associated with bristle bent (*Agrostis curtisii*) but females tend to select isolated clumps for ovipositing, especially where two or three tussocks grow close together amongst bare ground (Oates *et al.*, 2000). Solid carpets of bristle bent, which are heavily grazed, are mostly ignored. Pathside tussocks of purple moor-grass (*Molinia caerulea*) and wavy hair-grass (*Deschampsia flexuosa*) are also used (Oates *et al.*, 2000). In hot weather, adults will readily take nectar from flowers, especially bramble and heathers; *Buddleia* is also visited (Oates *et al.*, 2000).

The New Forest is *H. semele*'s main stronghold in Hampshire and has not suffered any major change in status throughout the 20th century (Oates *et al.*, 2000). However, it has disappeared from many heaths outside the New Forest and has been drastically reduced on others, due primarily to afforestation, fragmentation by development, and neglect (especially in the form of the abandonment of grazing on common land). The butterfly's status is highly vulnerable now in the north and east of Hampshire (Oates *et al.*, 2000).

Distribution:

H. semele is a widespread coastal species in Britain and Ireland (see Figure A-9 below), although it is absent from much of the east coast between Norfolk and Northumberland, and sparse in the extreme north of Scotland. It is also present on many offshore islands as well as some inland areas, especially in Dorset, the New Forest, in Surrey and in the Breckland of Suffolk and Norfolk (Asher *et al.*, 2001). *H. semele* can be found throughout Europe (except for parts of south-east Europe (Butterfly Conservation, 2009)) as far north as 63 degrees north, and extending into western and northern Asia; however, it is declining in many European countries (Asher *et al.*, 2001).



Note: the most recent (top most) dates overlay the earlier dates (lower ones) where squares have records in more than one date class.

Figure A-9. 10 km squares with records for *Hipparchia semele* (Grayling) in Great Britain and Ireland (from National Biodiversity Network (2009)).

2.5. Wood cricket (*Nemobius sylvestris* (Bosc, 1792))

Figure A-10. Wood cricket (*Nemobius sylvestris*) (photo courtesy of N. Brouwers) and woodland edge location (right).

***Description*:**

Nemobius sylvestris (Orthoptera: Gryllidae) is a small dark brown cricket with lighter markings, with females 7-11 mm and males 7-9 mm in length (Marshall and Haes, 1988; see Figure A-10). It is a non-flying cricket and in both sexes the hindwings are totally absent and in the male the forewings cover about half of the abdomen but in the female the wings are shorter lateral lobes and an ovipositor is also present (Marshall and Haes, 1988). *N. sylvestris* has an unusual two year life-cycle (Marshall and Haes, 1988): their eggs hatch during June; nymphs reach their fifth or sixth instar by autumn and overwinter at this stage. In the following spring the nymphs complete their eight nymphal instars, becoming adult during June and July. Adults may survive until late November and a few will overwinter, though probably not living long into their third year. The distinctive purring of a group of males can be heard from about the end of June until November (Marshall and Haes, 1988).

Dead leaves and perhaps associated fungi are probably the main food for this omnivorous species (Marshall and Haes, 1988), although the litter is not edible immediately after leaf fall (Gabbutt, 1959). From studies in south-east Devon, Richards (1952) also notes that it seems likely that fungi, including the smaller moulds and mildews which batten on decaying oak leaves and twigs, form a substantial proportion of the natural diet of *N. sylvestris* and supplementary items would doubtless embody dead insects, fallen acorns and leaf-galls, and honeysuckle foliage.

Richards (1952) reports that *N. sylvestris* dislikes direct hot sunshine, which appears to be injurious to them. They like heat, provided there is shelter from direct sunlight (Richards, 1952). Richards (1952) also found that *N. sylvestris* does not appear to have many predators, and only observed Wolf Spiders (*Lycosidae, Pisauridae*) and Red Ants (*Myrmica rubra*) taking a few nymphs at study sites in south-east Devon. Richards (1952) suggests that the main enemies of *N. sylvestris* are probably adverse climatic conditions, such as wet summers and hard winters. *N. sylvestris* is classified as nationally scarce in Great Britain.

***Habitat*:**

N. sylvestris occurs in deep leaf-litter, mainly under oaks, holly and bracken (but also including beech and sweet chestnut (Brown, 1978)), in warm and sunny clearings or along sunny margins of deciduous woods, along the edges of woodland tracks and railways, and often in very large populations (Richards, 1952; Marshall and Haes, 1988). In some places, colonies occur in old stone walls and earthbanks, but only below scrub or on woodland margins. *N. sylvestris* is not known from any open or exposed localities, although, in warm summers, later instar individuals may wander some metres from woodland cover (Brown, 1978; Marshall and Haes, 1988). The eggs are laid singly in the soil beneath the surface litter layers (Brown, 1978).

In a study of *N. sylvestris* in the Isle of Wight (UK), Brouwers (2008) confirmed its preference for open wooded edges and found that the factors positively influencing *N. sylvestris* presence within woodland included the presence of a well-developed leaf litter layer, relatively low ground vegetation cover and height and relatively short distances between individual populations. Woodland patch size and age were also found to be significant.

Distribution:

In the UK, at the northern limit of its European distribution, *N. sylvestris* has a very restricted range with three separate centres of distribution: Hampshire, the Isle of

Wight and South Devon (see Figure A-11 below), although the largest population is in the New Forest where it is widespread and numerous (Marshall and Haes, 1988). *N. sylvestris* is also found in wooded areas in Jersey in the Channel Islands as well as a colony in Surrey which became established very quickly. However, its distribution is still surprising as it is apparently absent from many seemingly suitable sites between the two large mainland populations in Devon and Hampshire (Marshall and Haes, 1988). The origin of *N. sylvestris* in the UK is not clear but it has been suggested that it is native to the Isle of Wight but was accidentally introduced to the New Forest and South Devon on forestry trees of Continental origin about the beginning of the nineteenth century (Marshall and Haes, 1988). *N. sylvestris* occurs in southern and central Europe as far north as Holland and northern Germany, eastwards through Poland to Russia, as well as in North Africa, the Azores and the Canary Islands (Marshall and Haes, 1988).

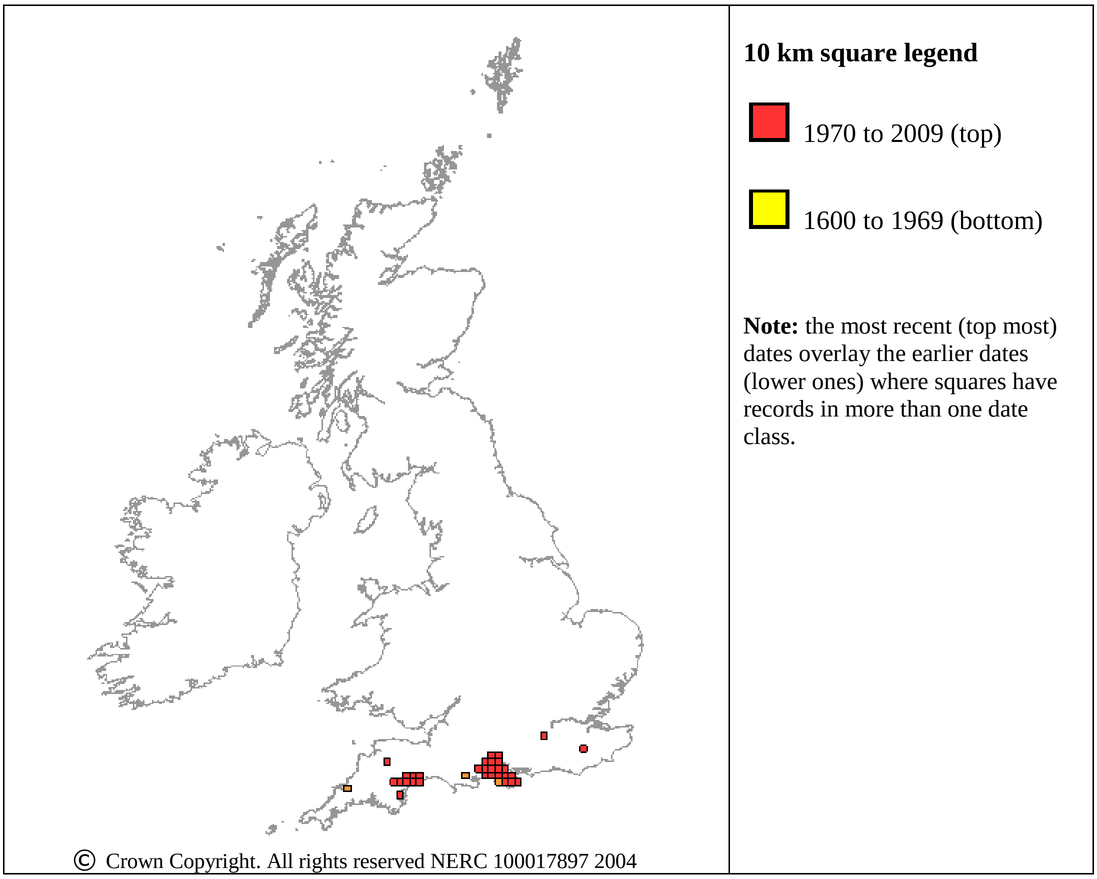


Figure A-11. 10 km squares with records for *Nemobius sylvestris* (Wood cricket) in Great Britain and Ireland (from National Biodiversity Network (2009)).

2.6. Silver-studded blue (*Plebejus argus*) (subsp. *argus*) (Linnaeus, 1758)

Figure A-12. Silver-studded blue butterfly (*Plebejus argus*). Photo on the left shows the underside of the female (top) and male (bottom). Photo on the right shows the upper side of a male.

***Description*:**

Plebejus argus (Lepidoptera: Lycaenidae) is a small butterfly (male wingspan is 26 – 32mm, female wingspan is 25 – 31mm) with the males possessing silvery-blue wings (Bretherton *et al.*, 1989; Asher *et al.*, 2001). The females are brown and far less conspicuous but, like the male, have distinct metallic spots on the hindwing (Asher *et al.*, 2001; see Figure A-12 above). *P. argus* has one generation per year (Joy, 1995) and the eggs are laid singly, close to the ground, where they pass the winter (Asher *et al*., 2001). On heathland the eggs are often laid on woody stems of the foodplants where there is sparse vegetation and patches of bare ground (Asher *et al*., 2001). The larvae hatch in spring and feed on the buds, flowers, young leaves, or growing tips of the foodplants. The larvae also have a close association with ants (most commonly *Lasius niger* and *L. alienus* on heathland), whereby the ants tend and protect them in return for sugar-rich liquids produced from special glands on the larvae’s bodies (Asher *et al*., 2001). The flight period of *P. argus* is mid-June to mid-August (Oates *et al*., 2000) and although most adults are reported to move less than 20m per day, with only a few travelling more than 50m (Asher *et al*., 2001), Oates *et al*. (2000) report that *P. argus* is reasonably mobile in the New Forest and so can readily colonise new habitat when it develops. *P. argus* is a UK Biodiversity Action Plan priority species, classified as nationally scarce in Great Britain and protected under Schedule 5 of the Wildlife and Countryside Act 1981 (with respect to sale only).

***Habitat*:**

In Hampshire, *P. argus* breeds exclusively on heathland, including the edges of wet heaths and pockets of heathland in coniferous plantations. It forms discrete colonies in areas sheltered by dykes, the local relief, or clumps of trees and bushes (Oates *et al*., 2000). The species requires managed heathland and colonies die out on neglected heaths as mature heathers are unsuitable. Therefore, grazing, rotational burning or periodic disturbance are essential (Oates *et al*., 2000). Its food plants have not been studied in Hampshire, but it appears that in the New Forest the butterfly uses heather (*Calluna vulgaris*) and cross-leaved heath (*Erica tetralix*) and possibly bell heather (*Erica cinerea*) (Oates *et al*., 2000). Some interesting occurrences have been noted, for example some specimens have been found miles from heaths, but this occurred in exceptionally hot weather and in the 1976 drought they were found in fair numbers on downs and in woods (Oates *et al*., 2000).

Joy (1995) summarises the requirements of *P. argus* as: flat or gently sloping land; a warm microclimate; heathers in the pioneer or early building stages; a continuous supply of pioneer heathland within one kilometre of a colony so that the species can move should conditions on the site deteriorate; a flight area containing varied age structures; and the correct species of ant – more work needs to be carried out on this in the Hampshire heaths, but pupae are attractive to ants of the genus *Lasius* with which larvae and pupae are strongly associated. *P. argus* has a broader niche in the south than in the northern heathland (Asher *et al*., 2001).

Distribution:

P. argus declined steadily in Britain during the twentieth century and became extinct in most of central and northern England, parts of Wales and on the North Downs in Surrey and Kent (Asher *et al.*, 2001; see Figure A-13 below). Its decline appears to have been due to two main factors; the destruction of heathland and the loss of early successional habitats (Asher *et al*., 2001). It is now widespread only on the heaths of southern England (where the New Forest is one of its main strongholds (Oates *et al*., 2000)), although there are large populations on the limestone in north Wales and certain sand dunes in Cornwall (Asher *et al*., 2001). Outside these areas it occurs in smaller colonies on the remnant heathlands of north Wales, the Suffolk Sandlings, the Norfolk Breckland, Shropshire, and along the coast of Cornwall and south Devon; it also occurs on Sark in the Channel Islands and has been reintroduced to a number of sites, including in north Wales and Suffolk (Asher *et al*., 2001). *P. argus* exists as a number of races in Britain, each of which is important for the conservation of genetic diversity (Asher *et al*., 2001).

In Britain *P. argus* is at the northern limit of its range (Joy, 1995) with a fairly southerly current distribution, except for some sites in Wales and Norfolk (see Figure A-13 below). It is widespread across temperate Europe and Asia to Japan and its range appears to be stable in much of Europe but declines have been recorded in some central and west European countries, and expansions in parts of south-east Europe and Russia (Asher *et al*., 2001).

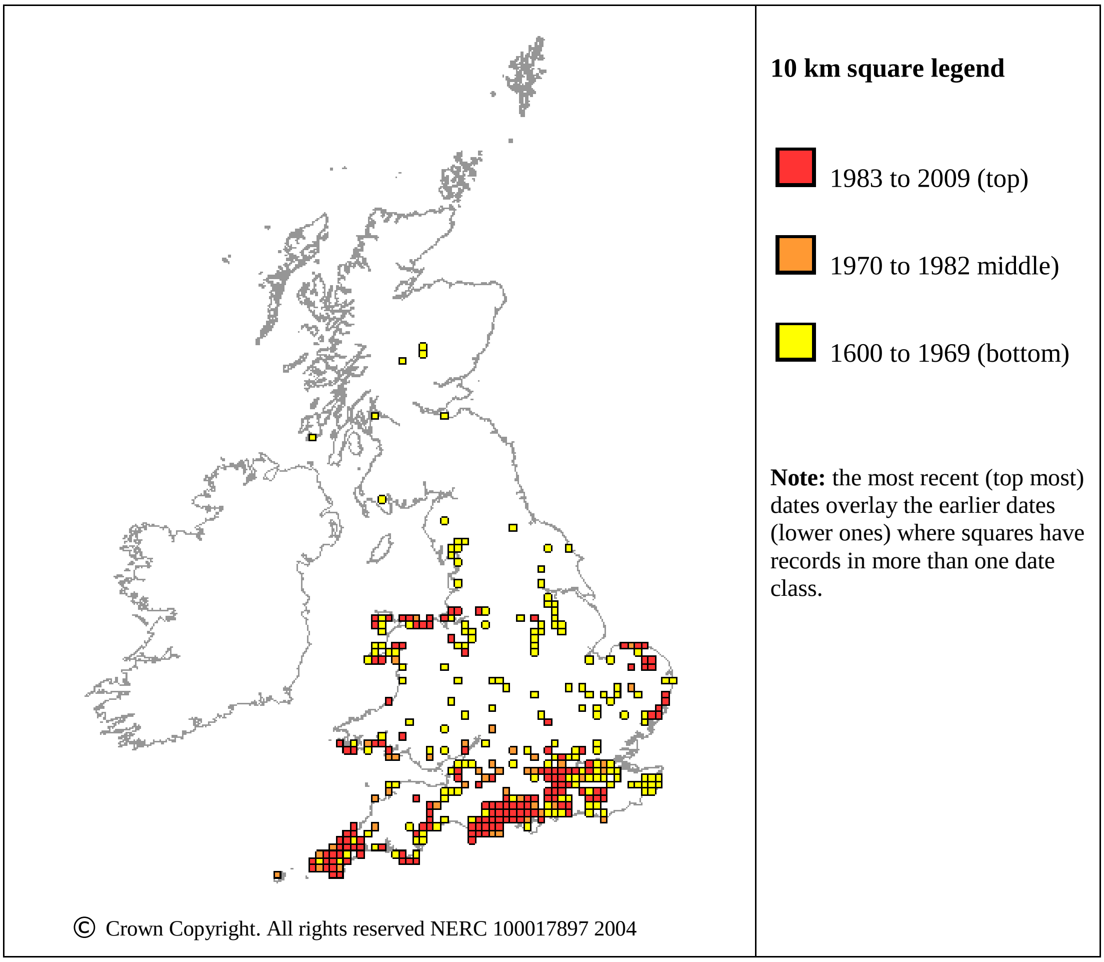


Figure A-13. 10 km squares with records for *Plebejus argus* (Silver-studded blue) in Great Britain and Ireland (from National Biodiversity Network (2009)).

2.7. Pillwort (*Pilularia globulifera* L.)

Figure A-14. Pillwort (*Pilularia globulifera*).

***Description*:**

Pilularia globulifera (Rubiaceae) is a native perennial, delicate grass-like rhizomatous aquatic fern, about 3 – 8 cm (but up to 15 cm) tall, with creeping stems carrying round pill-like spore bodies (sporocarps) during summer months (Stace, 1997; Preston, 2002; Plantlife, 2006; see Figure A-14 above). Populations can vary greatly from year to year, often responding to low water levels leaving exposed bare substrate (Botanical Society of the British Isles, 2008b). Sporocarp formation only takes place as water levels decline and expose the plant (Plantlife, 2006), with spores ripening from June to September (Clapham *et al.*, 1990). Spores released from the sporocarp in late summer can develop through the gametophyte phase to produce new sporophytes within seventeen days (Jermy, 1994), making *P. globulifera* an opportunistic coloniser (Botanical Society of the British Isles, 2008b). Page (1982) suggests that after colonies dies back during years of exceptional droughts, sporocarps may persist and remain dormant in mud, perhaps for many years, but further work is required to determine this. Jermy (1994) also notes that the sporocarps appear to have the potential of long-term storage in mud or silt, but there is no evidence of this happening. Sporocarps are probably carried from pool to pool on the feet of livestock and waterfowl (Plantlife, 2006).

Plants die back considerably in winter, but survive as sub-evergreen fragements to resume growth the following spring; once established from such fragments, plants are able to make substantial growth in one season (Page, 1982). Growth can vary widely from year to year, even in the same locality, depending on climatic conditions and water levels, and their reappearance can be sporadic (Page, 1982). *P. globulifera* is a UK Biodiversity Action Plan priority species, classified as near threatened in Great Britain and receives general protection under the Wildlife and Countryside Act, 1981.

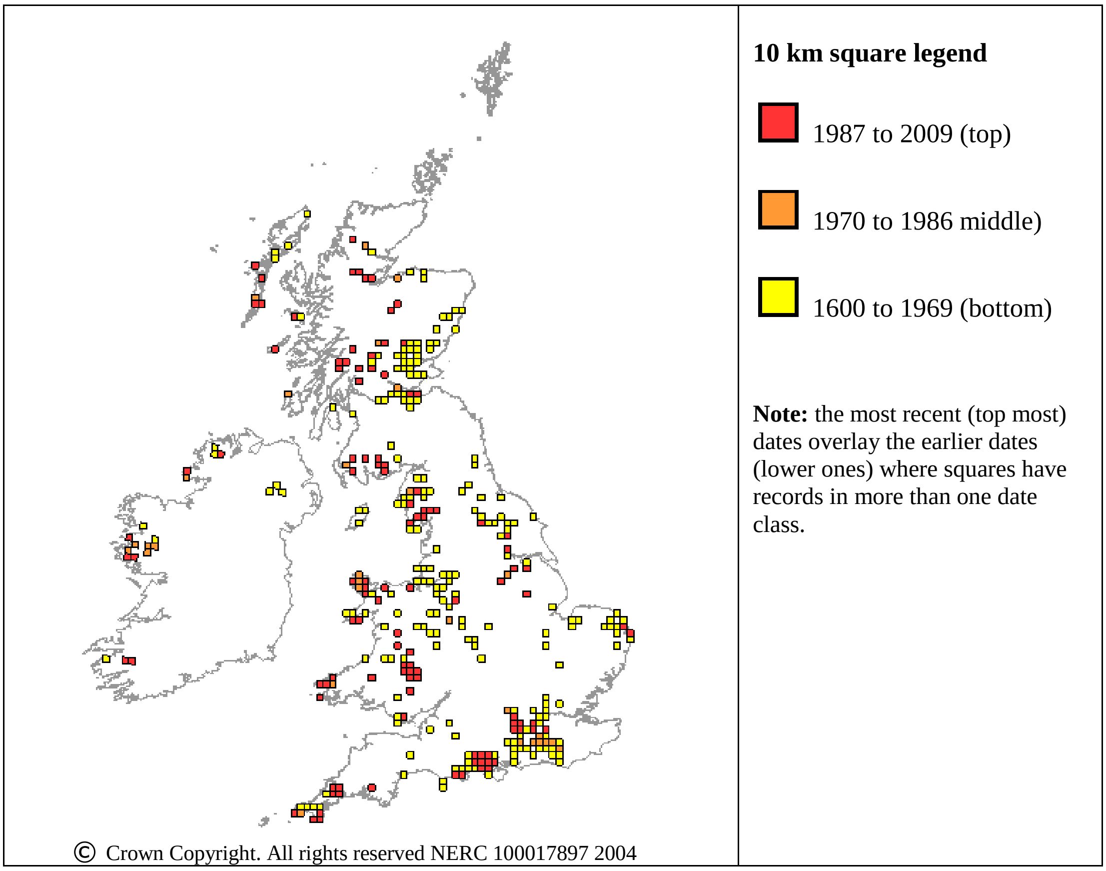
***Habitat*:**

P. globulifera is typical of the seasonally flooded margins of mildly acid (non-calcareous) lakes, reservoirs, ponds, pools or slow-flowing rivers on clays, sand and peaty substrates within heathland and upland grassland (Preston, 2002; Plantlife, 2006). It is an opportunist species, which can rapidly colonise open substrate exposed by falling water levels or newly created in disused clay and gravel pits (Preston and Croft, 1997). *P. globulifera* sometimes occurs on damp mine workings or as a submerged aquatic (Preston, 2002). It is not a competitive species, and at some of its sites suitable conditions are maintained by the trampling of cattle and horses (Preston and Croft, 1997). It requires areas where competition is reduced by fluctuating water levels or disturbance (Preston, 2002). Interestingly, in the north of Scotland it seems to occupy a different niche, with most of the records from the sides of rivers and lakes where, perhaps, the scouring effect of running water creates the open conditions it requires (Botanical Society of the British Isles, 2008b).

Continuing threats to its recovery are water pollution, particularly by fertilisers, which encourage the growth of coarse plants and eutrophication; the decline of cattle grazing and the resultant loss of trampling and disturbance; drainage; the ploughing of old pastures; and invasion by the vigorous non-native water plant New Zealand Pigmyweed (*Crassula helmsii*) (Preston, 2002; Plantlife, 2007b).

Distribution:

P. globulifera is found in scattered sites throughout the British Isles (Plantlife, 2006) (see Figure A-15 below). The UK populations are amongst the most important in Europe due to their relatively large size (Plantlife, 2006) and because *P. globulifera* is internationally threatened, as it is declining throughout its European range (Plantlife, 2007b). *P. globulifera* is endemic to western Europe, with lowland areas in Britain, France, northern Germany and southern Sweden containing the bulk of the populations and outlying populations extending to Portugal, Italy and the Czech Republic (Scott *et al*., 1999).



Note: the most recent (top most) dates overlay the earlier dates (lower ones) where squares have records in more than one date class.

Figure A-15. 10 km squares with records for *Pilularia globulifera* (Pillwort) in Great Britain and Ireland (from National Biodiversity Network (2009)).

2.8. Nail fungus (*Poronia punctata* **(L.) Fr.**)

Figure A-16. Nail fungus (*Poronia punctata*) on single dung boluse (left) and the boluse as part of a whole dung pile (right).

Description:

Poronia punctata (Xylariaceae family) is a saprotrophic ascomycete fungus which forms its fruiting bodies (perithecial stromata) on the dung of horses and ponies which have grazed on unimproved (unfertilised) acid-loving vegetation (Webster, 1999; Poland, 2004). The fruiting bodies, or stromata, of *P. punctata* are nail-like in shape with a dark cylindrical stalk attached to the dung (and usually buried within the dung) with a whitish, expanded, roughly circular upper disc, which is about 5 mm to 15 mm in diameter (Whalley and Dickson, 1986; Webster, 1999; Poland, 2004; see Figure A-16 above). The whitish, flattened, upper surface is dotted with small black ostioles opening into the perithecia (Poland, 2004) and the stalked stromata function to raise the perithecia above the surface of the dung and thus increase the efficiency of spore dispersal (Rogers, 1979; Whalley and Dickson, 1986). The perithecia contain asci which enclose the sexually produced ascospores and the ascospores are enclosed by a gelatinous coat which, following discharge from the asci into the air, attach (by their gelatinous coats) to surrounding vegetation and are ingested by herbivores (Webster, 1999; Poland, 2004). Ascospores ingested with the vegetation by grazing herbivores germinate to form a mycelium within the dung and new fruiting bodies develop from the mycelium (Webster, 1999).

Although *P. punctata* can occur at any time of year, it is most often observed from September to February (Poland, 2004) and is most frequently recorded from the New Forest in October (Dickson and Leonard, 1996). *P. punctata* is a UK Biodiversity Action Plan priority species and is a species of global conservation concern listed on the GB Red Data List as Endangered and in the IUCN Red Data List as Indeterminate (UK Biodiversity Action Plan, 2007).

Habitat:

In a survey carried out in the New Forest by Poland (2004), the majority (68%) of *P. punctata* records were from amongst NVC-classified H3 humid heath (*Ulex minor-Agrostis curtsii*), with 18% from M16 wet heath (*Erica tetralix-Sphagnum compactum*), 13% from H2c dry heath (*Calluna vulgaris-Ulex-minor*) and only 3% found on U3 acid grassland (*Agrostis curtisii*). Several species were found to almost always be present with *P. punctata* in the 2m by 2m sample areas, in particular *Molinia caerulea*, *Calluna vulgaris*, and *Erica tetralix*. However, the study was fairly limited in its coverage, so the results may not reflect a true trend.

Poland (2004) comments that *P. punctata* was rarely recorded on open acid grassland, which is supposedly a favoured habitat. Similar findings were reported from a study in Dorset (Cox and Pickess, 1999), which reported that all the *P. punctata* locations comprised acidic vegetation types, with 65% of occurrences in H2 (dry heath), 17% in H3 (humid heath), 10% U3 (Bristle Bent acidic grassland), 4% in M16 (wet heath) and 4% in recent gorse-clearance areas (acidic bare ground). Cox and Pickess (1999) also reported that many of the *P. punctata*-bearing dung deposits were found in open habitats, frequently on or near mown firebreaks and on former fire sites where the vegetation was short.

A survey of the Dorset occurrences of *P. punctata* from 1999 to 2005 (Cox *et al*., 2005) suggests that the population has more than doubled in that time. Cox *et al*. (2005) reported that *P. punctata* was most commonly found in H2 (dry heath) (48%), with 17% of occurrence in H3 (humid heath), 15% in U3, 12% in U1 (acidic grassland), with low occurrences in a few other acidic habitats. They also consider it likely that movements of ponies between suitable sites facilitates the spread of *P. punctata* and that changes in the movements of the ponies may be one of the causes of the year-to-year differences between the numbers of *Poronia*-bearing dung deposits occurring in the different NVC communities.

Whalley and Dickson (1986) report that in the New Forest *P. punctata* only occurs on horse droppings on open grass or heathland and not on dung dropped in woodland when animals take shelter. However, there are a few more recent records from woodland (A. Lucas, personal communication, March 5, 2009).

Poland (2004) suggests that *P. punctata* requires a level of drainage impedance as some moisture is required and it may not develop on arid sites. Whalley and Dickson (1986) also suggest that *P. punctata* apparently does not like drought (or really wet weather). However, the Coprophilous (i.e. living or growing on dung) Xylariaceae family of fungi, of which *P. punctata* is a member, are considered to have adaptations that allow them to inhabit dry sites (Rogers, 1979). The Xylariaceae produce and release ascospores over periods of weeks or months and the ascospores are usually released during a drying period following thorough wetting (Rogers, 1979). Rogers (1979) also notes that perithecia embedded in a stroma can undergo severe desiccation, then release ascospores after a short period of wetting.

From the survey, Poland (2004) suggests that *P. punctata* would also not generally favour waterlogged conditions and noted that on undulating heathland sites, *P. punctata* did not occur on dung deposited in the seasonally flooded short turf hollow dominated by *Molinia* and *Carex* sp. (primarily (*C. panacea*)). However, it did occur on slightly raised ground or heathy hummocks (a height differential of around 20 cm) that remained unflooded. *P. punctata* was also absent from dung deposited in M1 *Sphagnum auriculatum* bog pools, M21 *Narthecium ossifragum-Sphagnum papillosum* valley mires or M29 *Hypericum elodes-Potamogeton polygonifolius* soakways. These communities are constantly wet with a high water table, although Poland (2004) notes that they were less frequently transverse on transect routes. However, Poland (2004) suggests that the absence of *P. punctata* in these very wet communities, combined with the fact that it was rarely recorded on saturated dung, may indicate that, for that survey season at least, excessive moisture may inhibit growth of *P. punctata* by some means.

Vegetation height was also found to be an important factor in the presence of *P. punctata*. The majority of *P. punctata* was found within vegetation with a maximum height between 10-15 cm and was absent from vegetation above 30 cm (Poland, 2004). The sub-shrub canopy was always open and vegetation usually did not totally surround the dungpiles, leaving them exposed to varying aspects. Cox and Pickess (1999) also reported that *P. punctata* was mostly found on short vegetation (e.g. where recently burnt), although some occurred amongst mature *Calluna* up to 30 cm high. The mean vegetation height was 12.8 cm. Poland (2004) therefore suggests that a semi-open environment may be an important factor in the abundance of *P. punctata*. However, as suggested by Cox and Pickess (1999), it is probable that dung deposits with P. punctata would have been more likely to have been overlooked during the surveys if they occurred amongst taller vegetation.

Dung

Poland (2004) found that *P. punctata* occurs and apparently thrives on normal moist dung (relatively recent, i.e. several weeks old, containing some moisture and dark brown in colouration) and desiccated dung (characteristically bleached grey in colour and dried due to exposure to sunlight, several months old and does not appear to re-hydrate even after heavy rain). Only one saturated dungpile (retaining much water after rain or partially submerged on waterlogged ground) was found supporting *P. punctata*. However, *P. punctata* fruiting bodies do not occur on very fresh dung (Cox and Pickess, 1999).

Poland (2004) found that the structure of the dung pile was not as important. A total of 43% of *P. puntata* colonies were found on normal intact boluses (firm and perhaps several weeks old, with a normal amount of weathering). However, the rest of the records were roughly evenly distributed between disturbed dungpiles (in which the otherwise firm boluses have been broken up by minor trampling or other causes: 20%), collapsing dungpiles (those which were heavily weathered by rain/trampling etc, causing the boluses to sink and lose their original shape: 18%) and flat dungpiles (weathered boluses that had fully collapsed and completely lost their shape: 20%). However, Cox and Pickess (1999) reported that *P. punctata* was not observed on extensively disrupted boluses, possibly because if the dung has been extensively fragmented any fungal mycelium would be more likely to desiccate before stromata could be formed. Whalley and Dickson (1986) suggest that *P. punctata* is typically found on dung which is about one month old and still in lumps. Although Poland (2004) reports that other fungi were frequently recorded from dungpiles, he also notes that other fungi rarely shared dungpiles with *P. punctata* and were only once found to share an individual bolus.

In the New Forest *P. punctata* only occurs on horse droppings, but in Europe it has been reported on both horse and cattle dung (Whalley and Dickson, 1986). In the United States it is typically found on cattle dung and not on horse dung (Wicklow and Hirschfield, 1979). It has also been found to fruit on sterilised sheep, cow and rabbit faeces, although it is less prolific than on horse (Whalley and Dickson, 1986). Reid (1986) reports of an imperfectly localised herbarium record on cow dung (but does not provide a date) in the UK as well as a record in 1933 on rabbit pellets in Norfolk.

Grazing

In order to understand more about *P. punctata*, it is beneficial to know a bit about pony grazing. For example, the *P. punctata*-containing dung may be deposited a fair distance away from where the *P. punctata* spores were ingested. In hind-gut fermenting herbivores such as horses, food takes approximately 30 to 40 hours to pass through the gut (Cox, 1999). It is not clear from the literature how far ponies are likely to move within a day or in a 30 – 40 hour period, but each social group of New Forest ponies has a well-defined home range in which it confines all its routine activities throughout the year (Tubbs, 2001). These home ranges may vary in size between 82 hectares and 1020 hectares and the home ranges of different groups of ponies often coincide with or considerably overlap each other (Tyler, 1972). Within their home ranges, ponies follow daily patterns of movement which remain constant for long periods but show seasonal variations mainly associated with changes in food availability, the summer habit of shading and the greater need for shelter in the winter (Tubbs, 2001). In general, most groups concentrate daily on the grassland forming the focus of their home range (Tubbs, 2001).

In addition it appears that the distribution of excretia by free-ranging ponies (and cattle and fallow deer) is non-random, with distinct latrine areas where faeces and urinations are concentrated (Edwards and Hollis, 1982). From their study on three areas of reseeded grassland in the New Forest Edwards and Hollis (1982) found that at all times both cattle and ponies would avoid grazing in the immediate vicinity of their dung so that fresh dung soon became surrounded by a fringe 10-20 cm wide of tall herbage. This exclusion effect persisted for about 3 months by which time the dung had largely disappeared.

Pratt *et al*. (1986) found that New forest ponies use grassland and heathland the most in summer, but less in winter when there is an increased use of gorse-brake and woodland, which provides cover and shelter, although less food. In addition, use of habitat differs between day and night, particularly in summer when ponies tend to move from the open grazing communities (grassland, heathland) occupied in daylight towards the various cover communities (woodland, gorsebrake) at night. Pratt *et al*. (1986) also reported that New Forest ponies spent more than 75% of their time feeding throughout the year, but there were distinct seasonal trends with feeding activity peaks in April and May.

The use of different habitats and the diet of the ponies is important, as this will determine whether the *P. punctata* spores are ingested and the occurrence of *P. punctata* is largely determined by where the ponies are ingesting and excreting spores. Tyler (1972) found that large clumps of *Molinia caerulea* in the valley bogs and a sparser growth on moist heaths seemed to form the bulk of food throughout the summer. In addition, *Agrostis* species and *Festuca* species, the dominant species of grass on many woodland and streamside ‘lawns’ and on patches of ‘grass heath’ were grazed throughout the year.

Distribution:

P. punctata is possibly the rarest fungus in Europe (Poland, 2004). In the UK, it is virtually confined to the New Forest and Dorset heaths (see Figure A-17 below), and is otherwise found only in a few places in Europe, including Denmark, Sweden and Holland (Whalley and Dickson, 1986).

No extant persistent colonies of *P. punctata* exist outside the New Forest or Dorset heaths and it is also absent from potentially suitable locations such as Exmoor and

Dartmoor (Poland, 2004). Although occasional records have turned up at various UK locations in the past, such as at a site in Farnborough (Hampshire), where it occurred from 1997 to 2000 on the dung of Przewalski's horses (Poland, 2004), and a small population at Eelmoor Marsh near Aldershot (Hampshire) since 1998 and at Snelsmore Common in Berkshire in 2005, they do not appear to persist for long, suggesting that the habitat requirements of *P. punctata* are very specific (Webster, 1999). Webster (1999) suggests that the sporadic records of *P. punctata* from other localities in Britain are most probably from the dung of New Forest ponies, having ingested vegetation in the New Forest bearing spores of the fungus and transported to new sites with the fungus in their guts (G. Dickson, personal communication, cited in Webster (1999)). However, *P. punctata* is not restricted to the dung of New Forest ponies, as shown by its occurrence on the dung of Przewalski's horses.

P. punctata was first recorded in the New Forest in 1899 but there were very few records after that (Whalley and Dickson, 1986) until fairly recently. However, it could be that *P. punctata* has always been in the New forest and is widespread but no one has bothered to record it (Whalley and Dickson, 1986). Newton (2009b) questions whether, given its dependence on pony dung as a substrate, the fluctuations in pony numbers in the New Forest over the years are associated with the frequency of *P. punctata* occurrence. It may be the case, therefore, that *Poronia* has undergone a recent increase in abundance in the Forest, and may even still be increasing, as a result of an increase in pony numbers.

Whalley and Dickson (1986) report that *P. punctata* was widespread in Britain and Europe until the end of the nineteen century but has since declined. They therefore note that it is tempting to suggest that in those countries where *P. punctata* has declined this may be related to the reduction of horses grazing on natural vegetation, as most horses today have a diet supplemented with various types of fodder and additives and the New Forest is one of only a few areas where horses live in the 'wild' and forage for themselves.

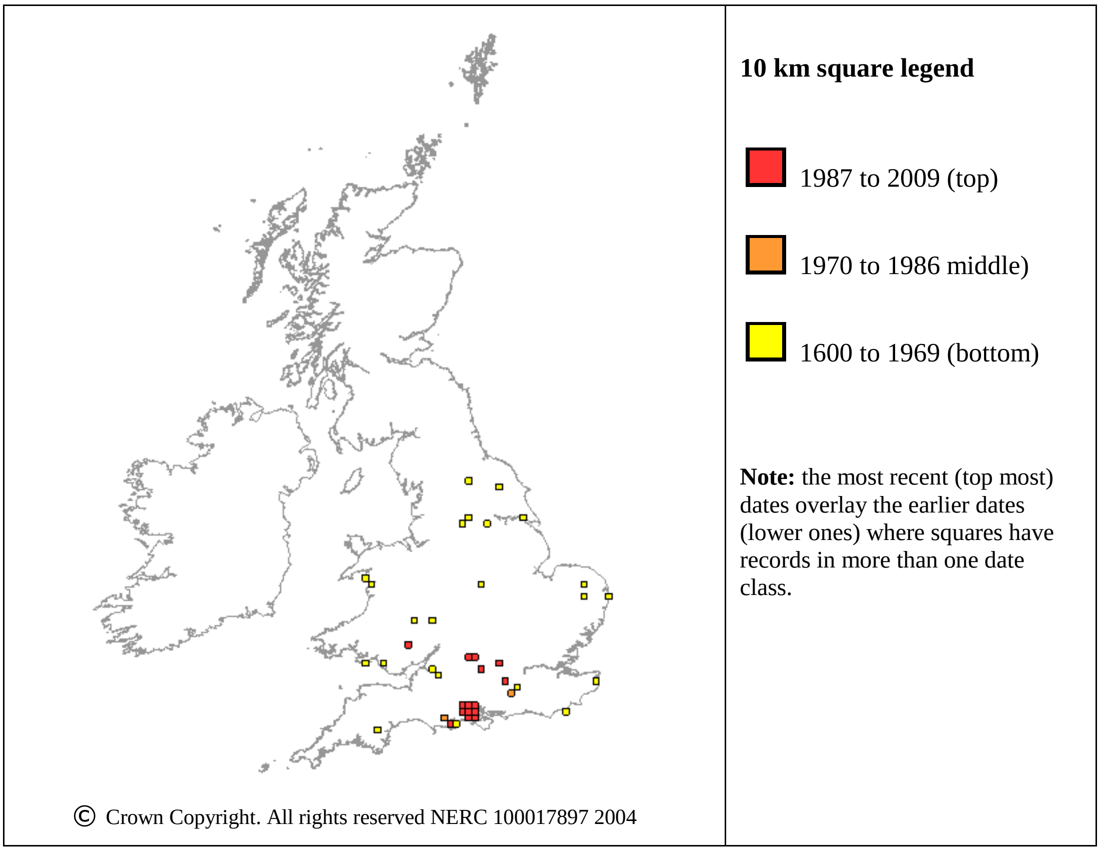


Figure A-17. 10 km squares with records for *Poronia punctata* (Nail fungus) in Great Britain and Ireland (from National Biodiversity Network (2009)).

Appendix 3. Information on species data (including providers)

I am extremely grateful for the help and time given by the data suppliers, as well as to the recorders, without whom it would not be possible to develop the models in this study, and to whom I hope the outputs will be useful.

Species occurrence data:

Species	**Total number of presence records**	**Number of training records (70%)**	**Number of testing records (30%)**	**Records provided by**
Wild Chamomile (*Chamaemelum nobile*)	194 (1995 onwards)	136	58	M. Rand (Botanical Society of the British Isles, VC11 recorder)
Slender marsh-bedstraw (*Galium constrictum*)	122 (1990 onwards)	85	37	M. Rand (Botanical Society of the British Isles, VC11 recorder)
Wild Gladiolus (*Gladiolus illyricus*)	94 (1990 onwards)	66	28	M. Rand (Botanical Society of the British Isles, VC11 recorder)
Pillwort (*Pilularia globulifera*)	141 (1990 onwards)	99	42	M. Rand (Botanical Society of the British Isles, VC11 recorder)
Grayling (*Hipparchia semele*)	186 (1995 onwards)	130	56	D. Green (Hampshire) and R. Fox (Wiltshire), both of Butterfly Conservation
Silver-studded Blue (*Plebejus argus*)	257 (1995 onwards)	180	77	D. Green (Hampshire) and R. Fox (Wiltshire), both of Butterfly Conservation
Wood cricket (*Nemobius sylvestris*)	147 (1995 onwards)	103	44	B. Pinchen (Hampshire recorder) and P. Budd (Southampton Natural History Society)

Nail fungus (*Poronia punctata*)	100	70	30	Hampshire Fungus Recording Group and J. Poland (records from survey and report by Poland (2004). Additional records collected by A. Newton (Bournemouth University).

Table A-1. Sources of species data and total number of records (the number of occurrence/presence records available to use in the analyses after pre-1990 (or pre-1995) and duplicated 6-figure grid reference records have been removed), and the number of presence records in the training and testing data set after being randomly split (see section 2.2.1.1). (This also applies to the number of pseudo-absence data in each category (for Chapter 3)).

Additional records were also obtained during fieldwork carried out as part of this project for the following species: *C. nobile, H. semele, N. sylvestris, P. globulifera* and *P. punctata*.

Appendix 4. Information on environmental data (including providers) and derived spatial layers

Table A-2 below shows the sources for the environmental data. All layers were converted to raster format for use in IDRISI Andes (Clark Labs, 2006).

Environmental data layer	**Source**	**Notes**
Habitat type	Hampshire Biodiversity Information Centre (HBIC) (A. Foy, March 2007)	Only the most relevant habitat type/landuse types were used. The habitat layer was also used to derive the following additional layers in IDRISI Andes: patch area, patch compactness and edge density (see below for more details).
Soil type	National Soil Resources Institute (NSRI), Cranfield University at Silsoe (May 2007)	Only the most abundant soil types that covered a substantial area were included in the analyses and these are listed in Appendix 6.
Digital Terrain Model (DTM; Elevation)	Relevant OS Land-Form PROFILE DTM 1:10000 tiles for the New Forest National Park were downloaded from the Edina Digimap website (http://edina.ac.uk/digimap/) on 14/06/2007.	The DTM layer was also used to derive Slope and Aspect layers using the SLOPE and ASPECT functions in IDRISI Andes (see below).
Climate data	Downloaded the appropriate 30 arc-seconds (~1 km) tiles for current conditions (~1950 – 2000) from the WORLDCLIM website (http://www.worldclim.org).	Data was obtained for the following variables: August (hottest month) average maximum temperature (Tmax8); February (coldest month) average minimum temperature (Tmin2); July (driest month) average precipitation (Prec7); December (wettest month) average precipitation (Prec12); Annual mean temperature (Bio1); Annual mean precipitation (Bio12).
New Forest National Park	Downloaded from the MAGIC website	This was to define the area of the National Park for all of the

boundary	(http://www.magic.gov.uk/ datadoc/metadata.asp? dataset=41&x=16&y=7) on 21/06/2007.	other GIS layers.

Table A-2. Sources of spatial GIS environmental data.

Derived spatial layers:

Aspect

The values of Aspect (in degrees) range from 0 to 359.9, where 0 is directly North-facing, 180 is directly South-facing, 359.9 is also North-facing and flat is given a value of –1. This means that the values of 0 and 359.9 indicate a very similar direction, but in any statistical analyses they would not be recognised as similar, so significant patterns in the data may be missed. Therefore, it was decided to classify the Aspect data into three categories: North-facing (0 to 89.9 and 270 to 359.9 classed as 1, everything else as 0), South-facing (90 to 269.9 and 270 to 359.9 classed as 1, everything else as 0), or Flat (-1 classed as 1, everything else as 0) using the RECLASS function in IDRISI Andes (Clark Labs, 2006).

Creation of edge density, patch compactness and patch area layers in IDRISI Andes

Edge density, patch compactness and patch area GIS layers were created in IDRISI for each of the habitat types to provide simple measures of habitat configuration and fragmentation. Patch size of heathland, for example, has been reported to be an important factor for *P. argus* habitat suitability (Joy, 1995). Such measures are likely to be more relevant at the scale (in terms of both resolution and extent) used in this study. The habitat layer was used as the earlier and later land cover image for the Land Change Modeller (LCM) function in IDRISI Andes (Clark Labs, 2006).

Edge density

This simple measure of fragmentation was calculated for each of the individual habitat types using the ‘edge density’ function in LCM, with the neighbourhood size selected as 3x3. Edge density is tabulated as the number of adjacent pairs of pixels within the neighbourhood that are different from each other relative to the maximum number of different pairs possible (Eastman, 2006). Each pixel is assigned a value between 0 and 1, where 1 indicates all the neighbouring cells are the same and a low value indicates a lot of edge of the particular habitat type (a value of 0 indicates that habitat type is not present). (Note that the edge density values were multiplied by 100, so that they were composed of a wider range of values (more continuous) for the models).

Patch compactness

This was calculated for each of the habitat types using the ‘Patch compactness’ function in LCM. Patch compactness groups adjacent pixels of similar landcover category into patches, calculates their compactness, and outputs an image where each pixel expresses the compactness of the patch to which it belongs (Eastman, 2006). Compactness is calculated as:

$$C = SQRT(Ap/Ac)$$

Where SQRT is the square root function, Ap is the area of the patch being calculated, and Ac is the area of the circle having the same perimeter of the patch being calculated (Eastman, 2006). A value between 0 and 1 is assigned, where 1 means that the patch is as compact as a circle. (Note that the patch compactness values were multiplied by 100, so that they were composed of a wider range of values (more continuous) for the models).

Patch area

It was discovered that the ‘Patch area’ function in LCM does not work properly, so the following work-around was used (as suggested by Clark Labs Technical Support, personal communication, October 15, 2007):

The GROUP function was used to group adjacent (including diagonal) cells of the same habitat type into ‘patches’. The AREA function was then used to calculate the area of each of these patches in hectares (square metres generated huge numbers). Each pixel then expressed the area of the patch to which it belonged.

Multiplication of layers:

The habitat type, soil type, aspect, patch compactness and edge density layers originally all had cell values ranging between 0 and 1. As this could cause problems with some of the analyses later on because the values do not show enough variation, these layers were multiplied by 100 to give a wider range of values (i.e. between 0 and 100), using the OVERLAY function in IDRISI Andes and multiplying each of the layers by an equivalent layer with pixels classified as 0 or 100.

Conversion of environmental layers to the same resolution:

So that the environmental layers were all at the same resolution as the species data, the environmental layers (at 10 m x 10 m pixel size) were aggregated to 100 m x 100 m pixel size using the CONTRACT function in IDRISI Andes (using a contraction factor of 10 in both X and Y). This meant that for the habitat types and soil types and aspect layers, each 100 m x 100 m pixel was assigned a value indicating the percentage of that e.g. habitat type in that pixel. For the DTM (elevation), slope, patch area, patch compactness edge density and Euclidean distance layers, the value of each pixel became an average of the values for the one hundred 10 m x 10 m pixels contained within it. The climate layers were originally at a resolution of 1 km, but these cells were split up to a 100 m resolution (using the EXPAND function in Idrisi), so that all 100 m pixels within the original 1 km pixel were assigned the value of that original 1 km pixel. All layers were masked by the National Park boundary layer, so that values outside of this boundary were -1 (a value that was not included in the range of values for any of the layers).

Appendix 5. Habitat types included in the HBIC habitat layer and brief description about how the habitat layer was derived

Information about the HBIC habitat layer

(Extracted from a document ‘The creation and maintenance of the habitats layer on the GIS – an explanatory note’ supplied by A. Foy of Hampshire Biodiversity Information Centre (HBIC) on 07/11/2007).

The system used by Hampshire County Council to describe land use types in Hampshire was based primarily on the English Nature Phase 1 types, with an additional 20 or so categories added to cover the urban environment in more detail. A total of 87 codes are now in use, many of which map to one or more of the Phase 1 codes. Mapping of all of Hampshire’s land use types onto the GIS using the categories described below was done by consultants in 1996/7 from interpretation of aerial photographs taken during 1995/6 at 1:20,000. Once this was complete an ecologist was employed during 1997/8 to go through all the detailed Phase 2 surveys of land in Hampshire to “ground truth” the aerial photographic interpretation. Approximately 5000 - 6000 ‘sites’ were ground truthed in this way. Since 1997 all new Phase 2 and Phase 1 ground surveys have been used to further update and refine the GIS layer. This amounts to approximately 1500 sites being surveyed between 1997 and 2002, of which around 50% cover ‘new’ sites. Work is ongoing.

Only the most appropriate habitat types for this application were used and only those habitat types with a significant presence in the New Forest. It is important to note that the habitat types do not indicate the quality of the habitat. Data were available from Natural England about the quality of SSSI (Sites of Special Scientific Interest) sites within the New Forest, but unfortunately these did not cover the whole of the National Park and the SSSI unit habitat classifications did not correspond to the habitat classifications of the HBIC habitat layer, so were not included.

Habitat classifications (with codes as used in the text)

Woodland

W1 - Broadleaved woodland including carr woodland with associated scrub of the New Forest and other heathlands.

W2 - Broadleaved plantation

W5 - Coniferous woodland

W6 - Coniferous plantation

W7 - Mixed woodland

W8 - Mixed plantation

Heathland

HL1 - Dry heath (includes humid heath), with >25% cover of *Calluna* – Dry acidic dwarf shrub heath (may include scattered bracken).

HL2 - Wet heath – Wet dwarf shrub heath (grading into valley mire). Applies to areas of wet dwarf shrub heath and set heath and grassland mosaic. (These mosaics are impossible to disaggregate to meaningful boundaries).

HL3 - Dry heath/acid grassland mosaic – Applies to areas where heather and ericaceous species occupy less than 25% cover and includes areas of burnt heather.

Grassland

GL3 - Improved grassland – Stable grassland which is not part of the arable rotation.

GL11 - Unimproved acidic grassland - often associated with heathland

GL12 - Unimproved neutral grassland

GL13 - Semi-improved neutral grassland

GL8 - Bracken (continuous cover) – Continuous bracken assumed to overly neutral/acid grassland; minimum patch size 1 hectare

Wetland

AQ1 - Acid fen/flush/valley mire in heathland situations

AQ5 - Ponds <0.5ha

AQ6 - Ponds >0.5ha

AQ7 - Running water including canals

Other

ST1 - Dense scrub

ST3 - Parkland/scattered trees over unknown grassland (<30% tree cover)

ED.UR.res – merged layers of residential area habitat/land-use types, including amenity grassland, sports fields, medium density residential (15-40% cover), low density residential (<15% cover).

ED.UR.nonres – merged layers of non-residential buildings and transport areas.

Merged habitat layers

Some of the habitat type layers were merged if they were similar as it was felt that stronger relationships may be obtained by using the layers in this way, as for some species the difference between the classifications (from a human use point of view) may not be relevant in terms of the species preferences. These were:

AQ5.AQ6 – Ponds of any size.

HL1.HL3 – Dry heath (including humid heath), with any percentage cover of heather and ericaceous species.

HL2.AQ1 – Wet heath or mire.

W1.W2 – Broadleaved woodland (woodland or plantation).

W5.W6 – Coniferous woodland (woodland or plantation).

W7.W8 – Mixed woodland (woodland or plantation).

Wood.edge – Woodland edge of any woodland type (broadleaved, coniferous or mixed).

Bld.edge – Broadleaved woodland (woodland or plantation) edge.

Con.edge – Coniferous woodland (woodland or plantation) edge.

Mix.edge – Mixed woodland (woodland or plantation) edge.

Appendix 6. Information on soil types

Soil type	**S64301**	**S64303**	**S71107**	**S84102**
Description	Deep sandy to clay.	Deep loam to clay.	Seasonally wet loam to clayey over shale.	Seasonally wet deep loam.
Dominant soils	Naturally very acid sandy over clayey and loamy over clayey soils locally with humose or peaty surface horizons, slowly permeable subsoils and slight seasonal waterlogging.	Naturally very acid coarse loamy over clayey soils with a bleached subsurface horizon, slowly permeable subsoils and slight seasonal waterlogging.	Slowly permeable seasonally waterlogged fine loamy over clayey and coarse loamy over clayey soils and similar more permeable soils with slight waterlogging.	Coarse and fine loamy permeable soils mainly over gravel variably affected by groundwater.
Associated soils	Some very acid well drained sandy soils, and some deep sandy soils, affected by groundwater with humose surface horizons.	Humose or peaty surface horizons locally. Some shallow and very flinty soils.	Some deep coarse loamy soils affected by groundwater.	-
Crop/Land use	Wet lowland heath habitats and coniferous woodland recreation; some agriculture and horticulture.	Lowland heath habitats and deciduous woodland; recreation; gravel extraction.	Cereals and grassland with dairying; deciduous and coniferous woodland and wet lowland heath habitats in the New Forest.	Permanent grassland, deciduous wood land and rough grazing in the South East Region; gravel extraction.

Table A-3. Descriptions of the soil types used in this study. Note that only the most abundant soil types were selected for use in the models. Information (and the soil data) supplied by the National Soil Resources Institute (NSRI), Cranfield University at Silsoe.

Appendix 7. Summary of environmental variables and range of values

Environmental layer	**Description**	**Units**	**Range**
Habitat type/cover (.cover)	Percentage of specified habitat type in cell.	%	0 – 100.00
Soil type (S.....)	Percentage of specified soil type in cell.	%	0 – 100.00
Patch area (PA.)	Patch area of the patch of specified habitat type to which the cell is part of.	hectares	0 - 2310.00
Patch compactness (PC.)	Average patch compactness of the patch of specified habitat type to which the cell is part of. (A value of 0 indicates that habitat type not present).	compactness value (multiplied by 100)	0 – 100.00
Edge density (Eden.)	Average edge density of the specified habitat type for that cell. (A value of 0 indicates that habitat type not present).	edge density value (multiplied by 100)	0 – 100.00
Euclidean distance to habitat type (ED.)	Euclidean distance to the specified habitat type. Euclidean distance to different woodland edges is also included, so that locations inside a large patch of woodland, do not just get assigned a value of 0 (which is the just the distance to woodland), but a value that indicates distance to a woodland edge. (This was mainly included for *Nemobius sylvestris*).	metres	0 – 10300.00
DTM/Elevation (DTM)	Average elevation in cell.	metres	0 - 152.54
Slope (Slope)	Average angle of slope in cell.	degrees	0 – 16.70
Aspect – South (Aspect.south)	Percentage of south-facing aspect in cell. (90° to 269.9°)	%	0 - 100
Aspect – North (Aspect.north)	Percentage of north-facing aspect in cell. (0 to 89.9° and 270° to 359.9°)	%	0 – 100
Aspect – Flat (Aspect.flat)	Percentage of flat terrain in cell (classed as -1 by Idrisi).	%	0 - 100
August	Average August (hottest month)	°C	20.80 –

maximum temperature (Tmax8)	mean temperature.		21.80
February minimum temperature (Tmin2)	Average February (coldest month) mean temperature (~1950-2000).	°C	0.60 – 1.90
Average annual temperature (Bio1)	Mean annual temperature (~1950-2000).	°C	9.60 – 10.70
July precipitation (Prec7)	Average July (driest month) precipitation (~1950-2000).	mm	43.00 – 46.00
December precipitation (Prec12)	Average December (wettest month) precipitation (~1950-2000).	mm	84.00 – 94.00
Average annual precipitation (Bio12)	Average annual precipitation (~1950-2000).	mm	753.00 – 816.00

Table A-4. Information about the environmental data layers used for modelling. Names in brackets in the environmental layer column are the abbreviations used for these variables in the models. Note that all environmental layers are in raster format where a raster cell = 100 m x 100 m. See Appendices 4 – 6 for further information on the environmental layers.

‘DTM/Elevation’, ‘Slope’, ‘Aspect.south’, ‘Aspect.north’ and ‘Aspect.flat’ are referred to as ‘terrain layers’.

‘August maximum temperature’, ‘February minimum temperature’, ‘Average annual temperature’, ‘July precipitation’, ‘December precipitation’ and ‘Average annual precipitation’ are referred to as ‘climate layers’.

Habitat cover, patch area, patch compactness, edge density and Euclidean distance layers include the merged layers.

The abbreviations used in discussion of the variables for the models in the main document are used in the following way. For example, edge density of wet heath would be Eden.HL2 (see Appendix 5 for abbreviations for habitat types).

Appendix 8. Evaluation of presence-only and presence-absence species distribution models

As discussed in section 1.2.6 (Chapter 1), species data can be split into a calibration (or training) set and an evaluation (or testing). This testing set can then be used to evaluate the predictive performance of the model in various different ways, depending on the modelling approach (presence-only or presence-absence)

8.1. Evaluation of presence-only models

There are numerous measures which can be applied to the evaluation of presence-absence models, such as Kappa or the Area Under the Curve (AUC) of a Receiver Operating Characteristic (ROC) plot. However, the lack of absence data makes it difficult to assess the predictive power of presence-only based models (Hirzel and Arlettaz, 2003). Guisan *et al*. (2006b) note that although there are now several methods available to fit models using presence-only data, problems remain in the evaluation of predictions from these models.

The development of techniques for evaluation of presence-only models is lagging behind that of presence/absence models (Hirzel *et al.*, 2006b). With respect to this, Hirzel *et al.* (2006b) used a presence-absence dataset of 114 plant species to test how common presence-absence indices (adjusted D^2, Kappa, MaxKappa, AUC) compared to presence-only evaluation indices (Absolute Validation Index (AVI), Contrast Validation Index (CVI) and the Boyce index), all of which were implemented in the Biomapper software. The AVI is the proportion of presence evaluation points falling above some fixed habitat suitability threshold (e.g. 50) and varies from 0 to 1 (Hirzel *et al.*, 2006b). It indicates how well the model discriminates high-suitability (i.e. likely presence) from low-suitability areas (likely absence) (Sattler et al., 2007). The CVI is the AVI minus the AVI of a model predicting presence everywhere (chance model), and varies from 0 to 0.5. Both the AVI and CVI suffer from having to use an arbitrary threshold (Hirzel *et al.*, 2006b).

The Boyce index (Boyce *et al.*, 2002) provides a more continuous assessment of model predictive power by partitioning the habitat suitability range into *b* classes (or bins), instead of only two (using one arbitrary threshold) (Hirzel *et al.*, 2006b). Then for each class it calculates the predicted frequency of evaluation points and the expected frequency of evaluation points (i.e. the frequency expected from a random distribution across the study area) and finally a predicted-to-expected ratio for each class (Hirzel *et al.*, 2006b). If the habitat model properly delineates the areas suitable for the species, a low suitability class should contain fewer evaluation presences than expected by chance. Boyce index values range from –1 to 1 with positive values indicating a model whose predictions are consistent with the presences distribution in the evaluation dataset, with values close to zero indicating that the model is not different from a chance model, and negative values indicating an incorrect model (Hirzel *et al.*, 2006b).

The main shortcoming of the Boyce index is its sensitivity to the number of suitability classes *b* and to their boundaries, so Hirzel *et al.* (2006b) derived a new threshold-independent evaluator based on a ‘moving window’ of width W (say W = 0.1) instead of fixed classes, to produce a smooth predicted-to-expected ratio curve, on which a ‘continuous Boyce index’ is computed. This continuous Boyce index is available for use within the Biomapper software (Hirzel *et al.*, 2006a).

Hirzel *et al.* (2006b) found all evaluators conveyed correlated information, meaning that presence-only evaluators can be trusted. However, the agreement between presence/absence and presence-only measures tended to be lower when the species prevalence was less than fifty presences. Hirzel *et al.* (2006b) provide an explanation for this, that a low number of presences prevent presence-only evaluators from assessing the overall quality of the model, whilst presence/absence evaluators can still rely on the fit between predicted and observed absences. Among the presence-only evaluators, the continuous Boyce index was found to be most accurate for characterising predictive capability.

In their comparison study of GAM and ENFA modelling approaches, Zaniewski *et al.* (2002) used cross-validation on a ROC statistic to evaluate their GAM models and a split-sample-sample approach (with the data randomly divided into a calibration set and a validation set) to evaluate the ENFA models by analysing the proportion of presences from the validation data set found in cells with a predicted suitability index (calculated from the calibration data set) greater than 50 [AVI]. This threshold value could be changed depending on the individual study. Zaniewski *et al.* (2002) recognised that while models ideally should be evaluated in the same manner, evaluation measures such as ROC require absences so are not suitable for presence-only modelling. Phillips *et al.* (2004) suggest that to use ROC curves with presence-only data, all grid cells with no occurrence localities must be interpreted as negative, even if they support good environmental conditions for the species. However, this could lead to erroneous results.

8.2. Evaluation of presence-absence models

Evaluating the predictive performance of models is a vital step in model development as it assists in determining the suitability of a model for specific applications and provides a basis for comparing different modelling techniques and competing models (Pearce and Ferrier, 2000a). The discrimination ability of logistic regression models is often quantified by calculating statistics from a 2 x 2 classification table (also known as a confusion or error matrix; see Figure A-18 below) of predictions and observations where a species is predicted to be present or absent at a site based on whether the predicted probability for the site is higher or lower than a specified threshold probability (Pearce *et al.*, 2002). It requires presence and absence (or pseudo-absence) data.

	Actual presence	Actual absence
Predicted presence	A	B
Predicted absence	C	D

Figure A-28. A confusion matrix (Fielding and Bell, 1997)

Several error or accuracy measures can be derived from a confusion matrix (Fielding and Bell, 1997) and include those shown in Table A-5 below.

Measure	**Calculation**
Correct classification rate	(A + D)/N
Sensitivity (true positive rate)	A/(A + C)
Specificity (true negative rate)	D/(B + D)
False positive rate	B/(B + D)
False negative rate	C/(A + C)
Kappa	[(A + D) – (((A + C)(A + B) + (B + D)(C + D))/N]/[N – (((A + C)(A + B) + (B + D)(C +D))/N)]

Table A-5. Error or accuracy measures that can be derived from the confusion matrix.

The Kappa statistic (Cohen, 1960) is the proportion of correctly classified units after accounting for the probability of chance agreement (Moisen *et al.*, 2006) and ranges form 0 to 1, with the following ranges of agreement between observed and predicted values: Kappa scores of 0.00 - 0.20 is slight, 0.21 - 0.40 is fair, 0.41 - 0.60 is moderate, 0.61 - 0.80 is substantial and 0.81 - 1 is almost perfect (Landis and Koch, 1977; Fielding and Bell, 1997).

False positives are sometimes referred to as commission errors and false negatives as omission errors; predicting the species to be absent when it is present or present when it is absent (Franklin, 1998). In most cases, an optimal threshold should be chosen that minimises omission and commission errors, particularly when a species is rare in the sample (i.e. low prevalence), because a model that predicted it to be absent everywhere would have high overall accuracy, measured by the correct classification rate and *vice versa* (Fielding and Bell, 1997; Franklin, 1998). Fielding and Bell (1997) also note that it is possible to obtain high overall accuracy when, for example, prevalence is low. For example, if prevalence was 5% it would be possible to achieve a 95% correct classification rate by labelling all cases as negative (Fielding and Bell, 1997).

Strauss and Biedermann (2005) note that false-positive predictions do not necessarily indicate a poor model fit, since sites recorded as non-use are not always unsuitable habitat (for example, see section 1.2.3). This is particularly true in declining populations, where many false positive predictions might result (Wilson *et al.*, 2005) as due to an increased extinction rate, suitable habitat might not be inhabited (Strauss and Biedermann, 2005). Therefore, habitat suitability maps may help to identify areas for the reintroduction of endangered or rare species by showing potentially suitable habitat (Strauss and Biedermann, 2005).

One problem with some of these evaluation measures is that unequal group sizes (prevalence) can influence the scores for many of the classifier methods (Fielding and Bell, 1997; see Appendix 13.1.2). Another important issue is that these measures are sensitive to the location of the threshold probability (i.e. the cutoff point for binary classification of presence or absence), which can often be an arbitrary choice, such as 0.5 (Fielding and Bell, 1997). This threshold can be adjusted and the effects examined, for example by selecting a threshold that maximises Kappa or minimises omission and commission errors. However, it may be the case that false negative errors are more serious than false positive errors, in which case the threshold can be adjusted to decrease the false negative rate at the expense of the increased false positive rate (Fielding and Bell, 1997). For example, if a model is to be used to define suitable protected areas or occurrence locations for a rare or endangered species, failure to correctly predict positive locations would be more costly (Fielding, 2002). Although it could be argued that the allocations of costs is subjective, a failure to explicitly apply costs implies equal costs, which is rarely justified (Fielding, 2002).

Jimenez-Valverde and Lobo (2006) suggest that the intuitively appealing 0.5 cut-off makes no sense, as each model has its own characteristics related to prevalence and fit. For example, in the case of rare species data, a 0.5 cut-off would convert presences to absences and would yield a false sensitivity value (true predicted presences) of zero in the most extreme case.

The performance of twelve different approaches to selecting a presence/absence threshold for probabilities/suitability's of species occurrence was examined by Liu *et al*. (2005). They found that the prevalence approach (taking the prevalence of model-building data as the threshold) and the average probability/suitability approach (taking the average predicted probability/suitability of the model building data as the threshold) were the most simple and effective. They also found that when the prevalence of model-building data was 0.5 (or 50%), there was little difference among the twelve approaches and recommend that, if possible, datasets with a prevalence of 0.5 should be used to build models since, in addition to other advantages (including that most optimisation criteria might be satisfied or nearly satisfied at the same time), it is easier to find the optimal threshold.

A more universal accuracy measure should describe the accuracy of the system, not just its performance in a given scenario, such as for a given threshold value (Pearce *et al.*, 2002). One such measure is the area under the receiver operating characteristic (ROC) curve (AUC). A ROC plot is obtained by plotting all sensitivity values (true positive fraction) on the *y* axis against their equivalent (1 minus specificity) values (false positive fraction) for all available thresholds (from 0 to 1) on the *x* axis (Fielding and Bell, 1997).

A good model will achieve a high true positive rate while the false positive rate is still relatively small; thus the ROC plot will rise steeply at the origin, then level off at a value near the maximum of 1 (Moisen *et al.*, 2006). The ROC plot for a poor model (whose predictive ability is the equivalent of random assignment) will lie near the diagonal, where the true positive rate equals the false positive rate for all thresholds (Moisen *et al.*, 2006). Therefore, the area under the curve (AUC) is a threshold-independent measure of overall model performance, with good models achieving an AUC near 1, while poor models achieve an AUC near 0.5 (equivalent to a random model) from a range of 0 to 1. Hosmer and Lemeshow (2000) provide a general rule of performance for AUC values (which can be seen in section 3.3.3, Chapter 3).

AUC is widely used in species distribution modelling to evaluate models (including large studies such as Berry *et al*. (2007b) and Elith *et al*. (2006) for example) and Austin (2007) states that current best practice for assessing model success for presence/absence data is AUC. However, Austin (2007) remarks that the procedure depends on the relationship between observed and predicted values; that is on predictive success, not on explanatory value. In addition, Austin (2007) also notes that models with the same or very similar AUC value may predict very different patterns of species distributions, and therefore reliance on AUC as a sufficient test of model success needs to be re-examined.

Although ROC has been widely used in species distribution modelling studies, several papers have recently emerged cautioning its use. In particular Lobo *et al.* (2008) outline several features of the AUC that may cause it to be a misleading measure of the performance of predictive distribution models. They suggest that the real value of AUC is that it provides a measure of the degree to which a species is restricted to a part of the variation range of the modelled predictors, so that presences can be told apart from absences. Lobo *et al*. (2008) conclude that accuracy measures proposed in the literature can be used to compare techniques for the same species at the same extent and that instead of using only the AUC, sensitivity and specificity should be also reported, so that the relative importance of commission and omission errors can be considered to assess the method performance. However, despite their reported problems with AUC, Lobo *et al*. (2008) are unable to recommend an alternative method to compare model performance among species.

Appendix 9. Variables used for final ENFA (Biomapper) for each species

Species	**C. no**	**G. co**	**G. il**	**H. se**	**N. sy**	**P. ar**	**P. gl**	**P. pu**
No. of presence training records	136	85	66	130	103	180	99	70
Maximum no. of variables	45	28	22	43	34	60	33	23
No. of selected variables	15	28	22	32	34	45	32	11
Selected variables:								
S64301	/		/		/	/	/	
S64303	/		/	/	/	/	/	/
S71107					/			
S84102	/	/			/		/	
DTM		/		/	/	/	/	
Slope	/	/	/	/	/		/	
Aspect.flat	/		/	/			/	
Aspect.north		/	/	/	/	/	/	
Aspect.south		/		/	/	/	/	
Bio1		/			/	/		
Bio12	/			/				
Prec7					/	/	/	/
Prec12			/	/	/			
Tmin2						/	/	
Tmax8						/		
AQ1.cover						/		
AQ5.cover		/	/					
AQ6.cover				/				/
AQ5.AQ6.cover						/	/	
GL8.cover			/			/		
HL1.cover				/		/	/	/
HL2.cover				/		/		
HL3.cover		/	/	/	/	/	/	
HL1.HL3.cover						/		
HL2.AQ1.cover		/				/	/	
W1.W2.cover					/			
W5.W6.cover					/			
W7.W8.cover					/			
ST1.cover				/		/		
Eden.AQ1		/		/		/		
Eden.GL8			/	/		/		
Eden.HL1	/	/		/		/	/	
Eden.HL2		/	/	/		/		
Eden.HL3	/	/		/	/	/	/	/
Eden.HL1.HL3			/					
Eden.HL2.AQ1	/					/	/	
Eden.W1.W2	/	/			/			
Eden.W5.W6					/			
Eden.W7.W8					/			
Eden.ST1	/			/		/	/	/
Eden.ST3		/						
PC.AQ1				/	/	/		
PC.GL3								
PC.GL8			/			/		

PC.HL1		/	/	/		/	/	
PC.HL2				/		/		
PC.HL3		/	/	/	/	/	/	
PC.HL2.AQ1		/					/	
PC.W1.W2		/			/			
PC.W5.W6					/			
PC.W7.W8					/			
PC.ST1			/	/		/	/	
PA.AQ1						/		
PA.AQ5.AQ6		/				/	/	
PA.GL8			/	/		/		/
PA.HL2				/		/		
PA.HL3		/					/	/
PA.HL1.HL3				/		/		
PA.HL2.AQ1							/	
PA.W1.W2					/			
PA.W5.W6					/			
PA.W7.W8					/			
PA.ST1						/		
ED.AQ1		/	/	/		/	/	
ED.AQ5	/						/	/
ED.AQ6	/					/		/
ED.GL8			/	/	/	/		
ED.GL11	/	/	/			/	/	
ED.HL1		/		/		/	/	/
ED.HL2		/	/	/	/	/	/	
ED.HL3	/	/		/	/	/	/	
ED.W1.W2					/			
ED.W5.W6					/			
ED.W7.W8					/			
ED.Bld.edge		/						
ED.con.edge		/	/			/	/	
ED.wood.edge					/			

Table A-6. Variables (including total number) used in the final ENFA for each species, and the suggested maximum number of variables (no more than one third of the number of presence training records (Hirzel, 2008)) for each species.

Appendix 10. Score matrices for final ENFA for each species

The score matrices (up to the number of retained factors for the habitat suitability computation (see section 2.2.2.3, Chapter 2 for more details)) for the final ENFA for each species are shown in Tables A-7 to A-14 below and were used to determine the most important ecological variables for each species. The first column of the matrix is the marginality factor. The other columns are the *V*-1 specialisation factors. (*V* is the number of variables). The rows are the EGV contributions to each factor (see section 2.2.2.2 for further explanation). The amount of specialisation accounted for by each factor (see Appendix 11 for details) is also provided in the tables.

The score matrix indicates how the factors are correlated with the variables. Providing that a factor explains enough information (the broken-stick criterion was used to select the significant factors; see section 2.2.2.3 for details), those variables that show the highest coefficient (absolute value) are the more important in explaining the species' distribution (Hirzel, 2008).

C. nobile

EGV/Factor	**Factor 1 (11%)**	**Factor 2 (18%)**	**Factor 3 (11%)**	**Factor 4 (10%)**	**Factor 5 (9%)**	**Factor 6 (7%)**
Edge density HL3	0.640	0.081	-0.035	0.076	0.189	-0.172
Edge density ST1	0.345	0.020	-0.053	-0.023	0.064	-0.019
% Soil 64303	0.305	-0.017	0.078	0.503	-0.353	0.029
Euclidean distance HL3	-0.293	0.719	-0.338	0.350	0.345	-0.196
Slope	-0.287	-0.643	0.191	0.271	-0.064	-0.282
Euclidean distance AQ6	-0.198	-0.003	0.164	0.556	0.129	0.287
Edge density HL1	0.195	-0.022	-0.088	-0.105	0.470	-0.143
% Soil 84102	0.174	-0.033	0.089	0.198	0.272	0.161
Annual temperature (Bio1)	0.158	0.014	0.523	0.294	-0.455	-0.474
% Soil 64301	0.141	-0.029	-0.012	0.169	-0.010	0.190
% Aspect - Flat	0.129	-0.119	0.109	0.048	0.061	0.389
Edge density W1_W2	0.115	0.033	-0.202	0.160	0.054	0.306
Edge density HL2_AQ1	0.104	0.097	0.108	-0.116	0.294	-0.206
Euclidean distance GL11	-0.100	0.183	0.669	-0.106	0.125	0.208
Euclidean distance AQ5	-0.086	-0.041	0.110	0.132	0.292	-0.362

Table A-7. Score matrix from ENFA for *C. nobile*. EGVs are sorted by decreasing absolute value of coefficients on the marginality factor (factor 1). Positive values on this factor mean the species prefers locations with higher values on the corresponding EGV than the mean location in the New Forest study area, whereas negative values on this factor mean the species prefers locations with lower values. Signs of coefficient have no meaning on the specialisation values (subsequent factors), where a higher absolute value means the range of the species is more restricted on that variable. The amount of specialisation accounted for by each factor is given in parentheses in each column heading. The number of factors included in the table is the number of factors retained for the habitat suitability map.

G. constrictum

EGV/Factor	**Factor 1 (21%)**	**Factor 2 (21%)**	**Factor 3 (11%)**	**Factor 4 (8%)**	**Factor 5 (6%)**	**Factor 6 (5%)**
% Soil 84102	0.366	0.032	-0.003	0.069	-0.069	-0.044
% AQ5	0.319	0.002	-0.002	0.013	0.001	0.014
Edge density HL3	0.301	-0.059	0.014	0.026	0.006	0.029
Patch area HL3	0.294	0.044	0.018	0.018	0.085	0.062
% HL3	0.271	-0.183	-0.058	-0.075	-0.257	-0.195
Slope	-0.234	0.180	-0.288	0.111	-0.341	0.346
Euclidean distance HL1	-0.228	-0.132	-0.029	0.157	-0.244	-0.149
Edge density HL1	0.219	0.062	-0.031	0.067	0.054	0.022
Annual temperature (Bio1)	0.215	0.362	0.305	0.365	0.350	0.256
Elevation	-0.214	0.531	0.329	0.353	0.285	-0.200
Euclidean distance GL11	-0.212	0.008	0.121	-0.047	0.391	-0.049
Euclidean distance HL3	-0.211	-0.402	0.403	0.201	-0.066	0.392
Euclidean distance AQ1	-0.193	-0.031	-0.253	0.071	-0.175	-0.200
% Aspect - North	-0.186	-0.043	0.053	-0.127	0.057	-0.207
Euclidean distance HL2	-0.157	0.113	0.095	-0.055	0.217	0.126
Patch compactness HL3	0.157	0.113	0.022	0.097	0.168	0.169
Edge density W1_W2	0.101	-0.098	-0.046	-0.074	0.085	-0.100
Edge density AQ1	0.094	-0.023	-0.012	-0.019	0.006	-0.015
Patch compactness W1_W2	0.092	-0.078	-0.042	0.01	-0.081	-0.050
% Aspect - South	0.090	-0.055	0.082	0.012	-0.138	-0.060
Patch compactness HL1	0.088	-0.195	-0.015	-0.019	-0.102	-0.014
Euclidean distance broadleaved (W1_W2) edge	-0.074	-0.200	-0.540	-0.413	0.217	0.214
Edge density HL2	0.066	0.067	-0.087	0.176	0.065	0.148
Patch compactness HL2_AQ1	0.049	-0.122	-0.199	-0.001	0	-0.271
Euclidean distance coniferous (W5_W6) edge	-0.025	-0.360	-0.024	0.188	0.019	-0.437
Patch area HL2_AQ1	0.016	0.111	0.218	-0.066	-0.113	0.182
Patch area ponds (AQ5_AQ6)	0.013	-0.145	0.240	-0.594	-0.213	0.046
Edge density ST3	-0.011	0.159	-0.009	-0.111	0.346	0.201

Table A-8. Score matrix from ENFA for *G. constrictum*. EGVs are sorted by decreasing absolute value of coefficients on the marginality factor (factor 1). Positive values on this factor mean the species prefers locations with higher values on the corresponding EGV than the mean location in the New Forest study area, whereas negative values on this factor mean the species prefers locations with lower values. Signs of coefficient have no meaning on the specialisation values (subsequent factors), where a higher absolute value means the range of the species is more restricted on that variable. The amount of specialisation accounted for by each factor is given in parentheses in each column heading. The number of factors included in the table is the number of factors retained for the habitat suitability map.

G. illyricus

EGV/Factor	Factor 1 (50%)	Factor 2 (23%)	Factor 3 (10%)	Factor 4 (5%)
Patch area GL8	0.540	-0.003	-0.006	-0.006
% GL8	0.444	0.012	0.012	0.016
Patch compactness GL8	0.349	-0.012	-0.011	-0.010
Edge density GL8	0.299	-0.042	-0.005	-0.026
Patch compactness HL3	0.196	0.030	0.005	-0.025
Edge density HL1_HL3	0.189	-0.022	0.004	-0.025
% AQ5	0.178	0.004	0.001	-0.001
Euclidean distance AQ1	-0.167	0.575	0.471	-0.666
Euclidean distance HL2	-0.164	0.051	-0.807	0.288
% HL3	0.147	-0.038	-0.026	0.038
Euclidean distance GL8	-0.143	-0.795	0.322	0.436
% Soil 64301	0.139	-0.003	-0.025	-0.008
Patch compactness ST1	0.132	0.005	-0.004	0.003
Euclidean distance GL11	-0.124	0.072	0.018	0.176
Edge density HL2	0.120	0.042	-0.027	-0.021
Slope	0.100	0.062	-0.009	0.005
% Soil 64303	0.084	-0.022	-0.029	-0.042
December precipitation (Prec12)	0.066	-0.051	0.011	0.130
Euclidean distance coniferous (W5_W6) edge	-0.063	-0.115	-0.134	-0.466
Patch compactness HL1	0.053	0.024	0.010	0.045
% Aspect - Flat	-0.046	0.012	0.003	0.074
% Aspect - North	0.024	0.048	0.018	0.064

Table A-9. Score matrix from ENFA for *G. illyricus*. EGVs are sorted by decreasing absolute value of coefficients on the marginality factor (factor 1). Positive values on this factor mean the species prefers locations with higher values on the corresponding EGV than the mean location in the New Forest study area, whereas negative values on this factor mean the species prefers locations with lower values. Signs of coefficient have no meaning on the specialisation values (subsequent factors), where a higher absolute value means the range of the species is more restricted on that variable. The amount of specialisation accounted for by each factor is given in parentheses in each column heading. The number of factors included in the table is the number of factors retained for the habitat suitability map.

H. semele

EGV/Factor	Factor 1 (16%)	Factor 2 (18%)	Factor 3 (13%)	Factor 4 (9%)	Factor 5 (7%)	Factor 6 (6%)	Factor 7 (5%)
Patch area HL1_HL3	0.414	0	0	0	0	0	0
% Soil 64303	0.334	0	0	0	0	0	0
Edge density HL1	0.316	0	0	0	0	0	0
% HL1	0.296	0	0	0	0	0	0
Euclidean distance HL1	-0.240	0	0	0	0	0	0
Euclidean distance HL2	-0.233	0	0	0	0	0	0
Euclidean distance AQ1	-0.222	0	0	0	0	0	0
Patch compactness HL1	0.216	0	0	0	0	0	0
Edge density ST1	0.213	0	0	0	0	0	0
Patch compactness ST1	0.197	0	0	0	0	0	0

Euclidean distance HL3	-0.173	0	0	0	0	0	0
Edge density HL3	0.170	0	0	0	0	0	0
% ST1	0.166	0	0	0	0	0	0
Edge density HL2	0.158	0	0	0	0	0	0
% HL3	0.146	0	0	0	0	0	0
Euclidean distance GL8	-0.134	0	0	0	0	0	0
Edge density GL8	0.121	0	0	0	0	0	0
Elevation	0.111	0	0	0	0	0	0
% AQ6	0.107	0	0	0	0	0	0
Patch compactness HL2	0.103	0	0	0	0	0	0
Patch compactness HL3	0.098	0	0	0	0	0	0
% HL2	0.092	0	0	0	0	0	0
Slope	0.079	0	0	0	0	0	0
Edge density AQ1	0.075	0	0	0	0	0	0
Patch compactness AQ1	0.052	0	0	0	0	0	0
Annual precipitation (Bio12)	0.044	0	0	0	0	0	0
Patch area HL2	0.035	0	0	0	0	0	0
% Aspect - North	-0.031	0.610	0.610	-0.610	-0.610	0.610	0.610
% Aspect - South	0.026	0.627	0.627	-0.627	-0.627	0.627	0.627
Patch area GL8	0.022	0	0	0	0	0	0
December precipitation (Prec12)	0.014	0	0	0	0	0	0
% Aspect - Flat	0.006	0.485	0.485	-0.485	-0.485	0.485	0.485

Table A-10. Score matrix from ENFA for *H. semele*. EGVs are sorted by decreasing absolute value of coefficients on the marginality factor (factor 1). Positive values on this factor mean the species prefers locations with higher values on the corresponding EGV than the mean location in the New Forest study area, whereas negative values on this factor mean the species prefers locations with lower values. Signs of coefficient have no meaning on the specialisation values (subsequent factors), where a higher absolute value means the range of the species is more restricted on that variable. The amount of specialisation accounted for by each factor is given in parentheses in each column heading. The number of factors included in the table is the number of factors retained for the habitat suitability map.

N. sylvestris

EGV/Factor	**Factor 1 (74%)**	**Factor 2 (10%)**	**Factor 3 (2%)**	**Factor 4 (2%)**
Patch area W1_W2	0.430	-0.039	-0.085	0.076
% W1_W2	0.373	-0.126	0.046	-0.212
Euclidean distance W7_W8	-0.290	0.190	-0.003	0.066
Edge density W7_W8	0.259	0.006	0.013	-0.042
Euclidean distance HL3	-0.211	-0.087	-0.080	-0.673
Edge density HL3	0.204	-0.003	0.059	0.005
Patch compactness W7_W8	0.203	0.011	-0.075	0.144
Euclidean distance W1_W2	-0.198	-0.229	0	-0.056
Euclidean distance W5_W6	-0.194	-0.122	-0.174	-0.259
% W7_W8	0.194	-0.074	0.062	-0.192
Edge density W1_W2	0.185	-0.095	0.003	0.078

Patch compactness W1_W2	0.182	0.006	-0.090	0.073
Euclidean distance wood edge	-0.180	-0.843	0.325	0.168
Euclidean distance GL8	-0.180	0.125	-0.510	0.494
Euclidean distance HL2	-0.169	0.088	0.329	-0.013
Patch area W5_W6	0.158	-0.007	-0.083	0.043
Edge density W5_W6	0.142	-0.053	-0.046	-0.043
% Soil 64301	0.138	-0.061	0	0.003
% W5_W6	0.110	-0.067	0.145	-0.140
Slope	0.085	-0.009	-0.027	-0.025
Patch compactness HL3	0.081	0.012	-0.060	0.021
% HL3	0.074	-0.071	-0.067	-0.180
Patch compactness AQ1	0.061	-0.046	-0.052	-0.028
December precipitation (Prec12)	-0.059	-0.077	-0.467	0.097
Patch area W7_W8	0.057	-0.005	-0.074	0.069
% Soil 71107	0.056	-0.108	0.069	-0.008
Patch compactness W5_W6	0.048	0.009	-0.139	0.032
July precipitation (Prec7)	0.044	-0.026	0.160	-0.002
% Soil 64303	0.043	-0.033	0.020	0.007
Elevation	0.043	-0.160	-0.100	-0.048
Annual temperature (Bio1)	0.039	-0.236	-0.360	0.006
% Aspect - North	0.039	-0.030	0.046	-0.014
% Soil 84102	0.032	-0.058	0.038	0.054
% Aspect - South	-0.024	-0.001	-0.010	-0.008

Table A-11. Score matrix from ENFA for *N. sylvestris*. EGVs are sorted by decreasing absolute value of coefficients on the marginality factor (factor 1). Positive values on this factor mean the species prefers locations with higher values on the corresponding EGV than the mean location in the New Forest study area, whereas negative values on this factor mean the species prefers locations with lower values. Signs of coefficient have no meaning on the specialisation values (subsequent factors), where a higher absolute value means the range of the species is more restricted on that variable. The amount of specialisation accounted for by each factor is given in parentheses in each column heading. The number of factors included in the table is the number of factors retained for the habitat suitability map.

P. argus

EGV/ Factor	**Factor 1 (17%)**	**Factor 2 (15%)**	**Factor 3 (9%)**	**Factor 4 (6%)**	**Factor 5 (5%)**	**Factor 6 (5%)**	**Factor 7 (4%)**	**Factor 8 (4%)**	**Factor 9 (3%)**
Edge density HL1	0.304	0.001	0.001	-0.001	-0.001	0	-0.002	0.003	0
% HL1_HL3	0.302	-0.620	0.568	-0.539	-0.595	-0.640	-0.206	-0.610	0.553
Patch area HL1_HL3	0.298	-0.004	0.004	0.002	0	0	-0.006	0.004	-0.003
% HL1	0.251	0.502	-0.469	0.437	0.487	0.524	0.151	0.504	-0.438
Patch compactness HL1	0.236	0.003	0.003	0.002	0	-0.001	0.014	-0.001	-0.009
Euclidean distance HL1	-0.232	0.060	-0.013	0.010	-0.028	0.006	-0.017	0.029	-0.059
Euclidean distance HL2	-0.223	0.064	-0.038	-0.062	0.015	-0.010	0.002	0.011	0.061
Euclidean distance	-0.215	-0.024	0.045	0.073	0.010	-0.003	-0.033	0.002	-0.017

AQ1									
% Soil 64301	0.184	-0.004	0	-0.002	0	0	0.004	0.004	0.003
Edge density HL2_AQ1	0.180	0.003	0.002	0	-0.001	0	-0.018	-0.002	0.004
Euclidean distance HL3	-0.179	-0.099	0.023	-0.032	0.016	0	0.039	-0.017	-0.037
% Soil 64303	0.174	0	0.001	-0.001	0	0	0.012	0.002	-0.002
% HL3	0.167	0.352	-0.327	0.319	0.344	0.371	0.113	0.355	-0.317
Edge density HL2	0.161	-0.006	-0.004	-0.003	0.002	-0.004	0.018	0.003	-0.002
Edge density HL3	0.157	-0.002	-0.001	-0.007	0	0.002	-0.006	0.001	-0.003
Patch compactness HL3	0.149	-0.002	0.002	-0.009	0.001	-0.001	0.007	-0.003	-0.001
Euclidean distance GL8	-0.148	-0.076	-0.012	-0.007	-0.002	0.005	0.016	-0.012	0.054
Edge density AQ1	0.138	-0.003	0.002	-0.002	0.001	0	0.023	-0.001	-0.003
Edge density ST1	0.137	-0.001	-0.002	0.001	0.003	0.003	-0.005	0.001	-0.002
Patch compactness HL2	0.128	-0.004	-0.004	-0.001	0.003	0	0.036	0.004	-0.001
Patch compactness ST1	0.126	-0.001	0.004	-0.003	0	-0.004	-0.014	0.003	-0.008
% HL2_AQ1	0.124	-0.330	0.434	-0.465	-0.394	-0.306	-0.651	-0.362	0.462
Euclidean distance GL11	-0.121	0.011	0.011	-0.022	-0.009	0.006	0.022	-0.003	0.012
Edge density GL8	0.105	-0.016	-0.007	-0.005	-0.002	0	-0.054	-0.002	-0.002
% GL8	0.104	0.031	0.017	0.010	0.001	-0.002	0.340	0.020	0.018
% ST1	0.104	-0.001	-0.002	0.007	-0.001	-0.001	0.016	0	0.012
% HL2	0.099	0.276	-0.338	0.373	0.311	0.250	0.443	0.284	-0.361
% AQ1	0.095	0.163	-0.206	0.219	0.188	0.148	0.264	0.175	-0.203
Patch compactness GL8	0.093	-0.023	-0.009	-0.004	0	0.001	-0.202	-0.008	-0.008
Patch compactness AQ1	0.092	-0.005	0.002	0.008	-0.002	-0.004	0.018	0.001	-0.011
Patch area AQ1	0.082	-0.001	-0.001	0.001	-0.001	-0.001	0.011	0	-0.004
Euclidean distance AQ6	-0.071	0.007	-0.006	0.028	0	-0.002	0.053	-0.011	0.002
% AQ5_AQ6	0.055	-0.002	0.009	-0.008	-0.010	-0.015	-0.004	-0.008	0.007
February minimum temperature (Tmin2)	0.055	0.037	-0.013	-0.003	-0.004	-0.002	0.106	-0.039	-0.047

Annual temperature (Bio1)	0.051	0	-0.001	0.004	0.016	-0.001	-0.073	0.027	0.001
Patch area HL2	0.050	-0.005	-0.002	-0.003	0	-0.001	0.027	0.003	0
August maximum temperature (Tmax8)	0.050	-0.015	0.002	0.008	-0.016	0.008	0.039	-0.004	0.014
Euclidean distance coniferous edge (W5_W6)	-0.049	-0.014	-0.011	0.026	-0.008	-0.002	-0.027	0.009	-0.008
% Aspect - North	-0.048	0.005	-0.003	0.010	0.001	0.002	0.026	-0.007	0.003
Patch area GL8	0.047	-0.011	-0.007	-0.008	-0.001	0.003	-0.204	-0.017	-0.014
Patch area ST1	0.041	0.001	0.001	-0.004	0.001	0.008	0	-0.003	0.002
Patch area AQ5_AQ6	0.037	0.002	-0.010	0.011	0.012	0.017	0.006	0.010	-0.007
Elevation	0.036	0.007	-0.012	0.009	0	0.001	0.023	-0.007	-0.019
July precipitation (Prec7)	0.035	0.011	-0.005	-0.001	-0.001	-0.001	0.029	-0.006	-0.010
% Aspect - South	0.031	-0.006	-0.005	0.002	0.001	-0.001	0.008	-0.001	0.007

Table A-12. Score matrix from ENFA for *P. argus*. EGVs are sorted by decreasing absolute value of coefficients on the marginality factor (factor 1). Positive values on this factor mean the species prefers locations with higher values on the corresponding EGV than the mean location in the New Forest study area, whereas negative values on this factor mean the species prefers locations with lower values. Signs of coefficient have no meaning on the specialisation values (subsequent factors), where a higher absolute value means the range of the species is more restricted on that variable. The amount of specialisation accounted for by each factor is given in parentheses in each column heading. The number of factors included in the table is the number of factors retained for the habitat suitability map.

P. globulifera

EGV/Factor	**Factor 1 (58%)**	**Factor 2 (8%)**	**Factor 3 (6%)**	**Factor 4 (5%)**	**Factor 5 (4%)**	**Factor 6 (3%)**
% AQ5_AQ6	0.358	0	0	0	0	0
Edge density HL2_AQ1	0.294	0	0	0	0	0
Edge density HL3	0.287	0	0	0	0	0
% HL2_AQ1	0.234	0	0	0	0	0
% HL3	0.233	0	0	0	0	0
Euclidean distance HL2	-0.232	0	0	0	0	0
Euclidean distance AQ1	-0.230	0	0	0	0	0
Patch area HL2_AQ1	0.219	0	0	0	0	0
Euclidean distance HL3	-0.212	0	0	0	0	0
Euclidean distance HL1	-0.208	0	0	0	0	0
Patch compactness HL2_AQ1	0.205	0	0	0	0	0
Edge density HL1	0.186	0	0	0	0	0
Patch area HL3	0.186	0	0	0	0	0
% Aspect - Flat	0.179	-0.485	0.485	-0.485	-0.485	-0.485

Patch compactness HL3	0.178	0	0	0	0	0
Slope	-0.158	0	0	0	0	0
Patch area AQ5_AQ6	0.150	0	0	0	0	0
Edge density ST1	0.128	0	0	0	0	0
Euclidean distance GL11	-0.128	0	0	0	0	0
% Aspect - North	-0.122	-0.610	0.610	-0.610	-0.610	-0.610
Patch compactness ST1	0.120	0	0	0	0	0
% Soil 64301	0.115	0	0	0	0	0
% Soil 84102	0.103	0	0	0	0	0
Euclidean distance AQ5	-0.092	0	0	0	0	0
Patch compactness HL1	0.090	0	0	0	0	0
% HL1	0.087	0	0	0	0	0
Euclidean distance coniferous edge (W5_W6)	-0.070	0	0	0	0	0
% Soil 64303	0.067	0	0	0	0	0
February minimum temperature (Tmin2)	0.065	0	0	0	0	0
% Aspect - South	-0.019	-0.627	0.627	-0.627	-0.627	-0.627
July precipitation (Prec7)	-0.037	0	0	0	0	0
Elevation	-0.053	0	0	0	0	0

Table A-13. Score matrix from ENFA for *P. globulifera*. EGVs are sorted by decreasing absolute value of coefficients on the marginality factor (factor 1). Positive values on this factor mean the species prefers locations with higher values on the corresponding EGV than the mean location in the New Forest study area, whereas negative values on this factor mean the species prefers locations with lower values. Signs of coefficient have no meaning on the specialisation values (subsequent factors), where a higher absolute value means the range of the species is more restricted on that variable. The amount of specialisation accounted for by each factor is given in parentheses in each column heading. The number of factors included in the table is the number of factors retained for the habitat suitability map.

P. punctata

EGV/Factor	**Factor 1 (100%)**	**Factor 2 (0%)**
% HL1	0.623	0.004
% Soil 64303	0.517	0.001
Euclidean distance HL1	-0.368	0.066
Edge density HL3	0.252	-0.001
July precipitation (Prec7)	0.246	0
Euclidean distance AQ6	-0.199	0.004
Edge density ST1	0.152	0.001
Euclidean distance AQ5	-0.129	-0.003
Patch area HL3	0.051	0
Patch area GL8	-0.049	0.046
% AQ6	-0.024	-0.997

Table A-14. Score matrix from ENFA for *P. punctata*. EGVs are sorted by decreasing absolute value of coefficients on the marginality factor (factor 1). Positive values on this factor mean the species prefers locations with higher values on the corresponding EGV than the mean location in the New Forest study area, whereas negative values on this factor mean the species prefers locations with lower values. Signs of coefficient have no meaning on the specialisation values (subsequent factors), where a higher absolute value means the range of the species is more restricted on that variable. The amount of specialisation accounted for by each factor is given in parentheses in each column heading. The number of factors included in the table is the number of factors retained for the habitat suitability map.

Appendix 11. Weightings for factors used in habitat suitability computation

The ecological niche factors convey two kinds of information: marginality and specialisation. The first factor always explains 100% of the marginality and some varying part of specialisation; the subsequent factors explain no marginality and the rest of the specialisation (Hirzel, 2008).

Explained specialisation only includes the explained specialisation and not the marginality. Mathematically, if $L1$, $L2$, ..., Lf are the eigenvalues of the f retained factors, and SL is the sum of all n eigenvalues ($= L1+L2+...+Lf+...+Ln$), then:

$$\text{Explained specialisation} = Se = (L1+L2+...+Lf)/SL$$

In the explained information index, the marginality gets the same weight as the specialisation and therefore:

$$\text{Explained information} = Ie = (Se+1)/2$$

Explained information (Ie) is therefore always greater than explained specialisation (Se) and can never be smaller than 0.5. The Ie value is used as a decision support value to choose how many factors to include in the habitat suitability analysis (Hirzel, 2008).

The weighting for each factor is calculated based on the explained marginality and specialisation. The first factor, which always explains 100% of the marginality and some varying part of specialisation (which is why it is given at least half of the weighting for HS computation (Hirzel, 2008)), is given a weighting of 1 for the marginality, plus the specialisation weighting (given by the eigenvalues, which give an indication of how much variance is explained by the factors, and are included in the ENFA output). For example, for *C. nobile* factor 1 explained 0.112 of the specialisation, so was given a weight of 1.112 (1 + 0.112). The subsequent factors explain no marginality and the rest of the specialisation (Hirzel, 2008) and are weighted according to the amount of specialisation that they explain (e.g. *C. nobile* factor 2 explained 0.184 of the specialisation). Tables A-15 to A-22 below show the weightings for each factor used in the habitat suitability map computation.

Factor	**Weight**
1	1.11
2	0.184
3	0.105
4	0.096
5	0.086
6	0.072

Table A-15. Weightings for each factor for *C. nobile* habitat suitability computation. (Explained information = 0.828; Explained specialisation = 0.655).

Factor	**Weight**
1	1.21
2	0.214
3	0.114
4	0.080
5	0.061
6	0.053

Table A-16. Weightings for each factor for *G. constrictum* habitat suitability computation. (Explained information = 0.868; Explained specialisation = 0.736).

Factor	**Weight**
1	1.50
2	0.232
3	0.103
4	0.053

Table A-17. Weightings for each factor for *G. illyricus* habitat suitability computation. (Explained information = 0.943; Explained specialisation = 0.887).

Factor	**Weight**
1	1.17
2	0.176
3	0.127
4	0.094
5	0.067
6	0.057
7	0.051

Table A-18. Weightings for each factor for *H. semele* habitat suitability computation. (Explained information = 0.868; Explained specialisation = 0.737).

Factor	**Weight**
1	1.74
2	0.097
3	0.023
4	0.022

Table A-19. Weightings for each factor for *N. sylvestris* habitat suitability computation. (Explained information = 0.941; Explained specialisation = 0.883).

Factor	**Weight**
1	1.17
2	0.153
3	0.088
4	0.063
5	0.050
6	0.043
7	0.043
8	0.036
9	0.034

Table A-20. Weightings for each factor for *P. argus* habitat suitability computation. (Explained information = 0.842; Explained specialisation = 0.685).

Factor	**Weight**
1	1.58
2	0.079
3	0.062
4	0.055
5	0.040
6	0.026

Table A-21. Weightings for each factor for *P. globulifera* habitat suitability computation. (Explained information = 0.922; Explained specialisation = 0.843).

Factor	**Weight**
1	2
2	1.32E-12

Table A-22. Weightings for each factor for *P. punctata* habitat suitability computation. (Explained information = 1; Explained specialisation = 1).

Appendix 12. Biomapper habitat suitability maps with maps of environmental variables with highest marginality value for each species

The Biomapper habitat suitability maps are shown below alongside maps of the variables with the highest marginality values for each species (see Tables 2 to 9 in section 2.3.2.2.1.2, Chapter 2).

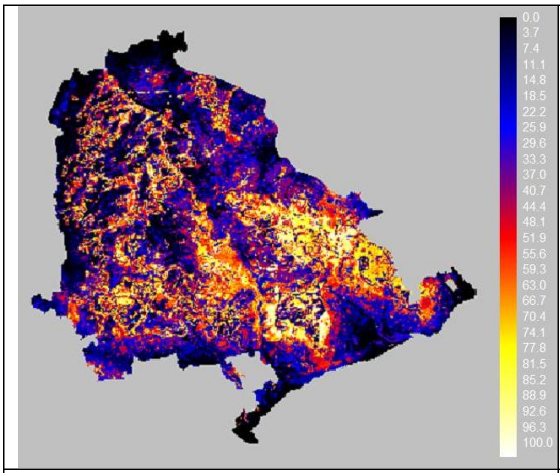


Figure A-19. Biomapper habitat suitability map for *C. nobile*.

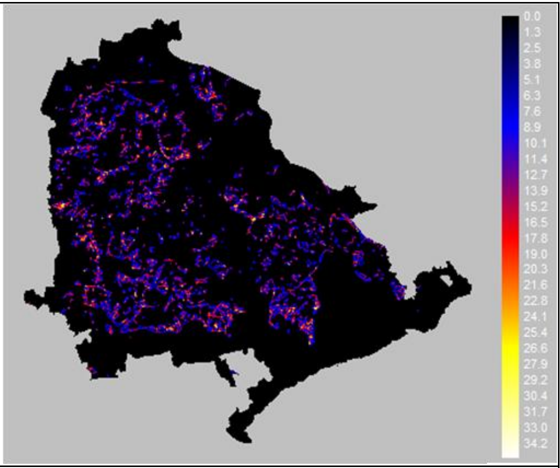


Figure A-20. Map of edge density of dry heath/acid grassland mosaic (HL3).

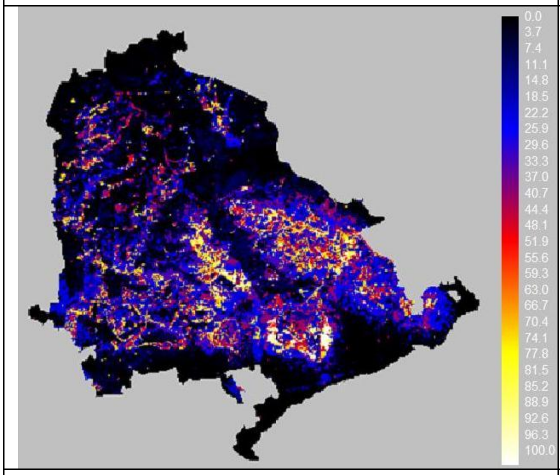


Figure A-21. Biomapper habitat suitability map for *G. constrictum*.

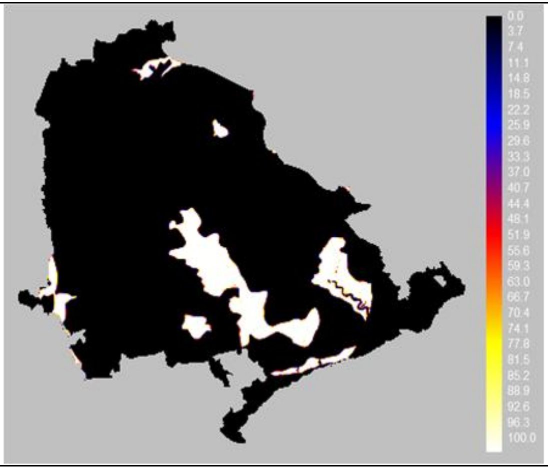


Figure A-22. Map of percentage cover of soil type 84102.

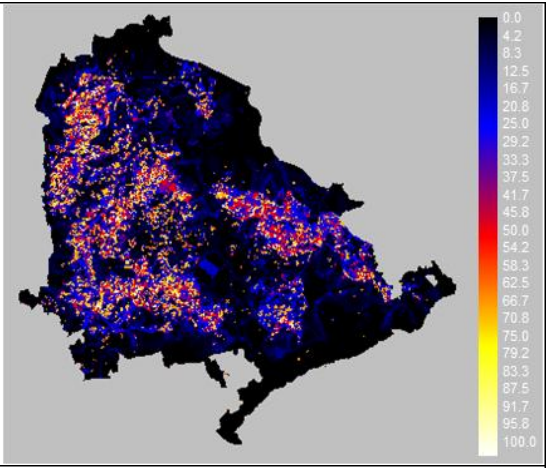


Figure A-23. Biomapper habitat suitability map for *G. illyricus*.

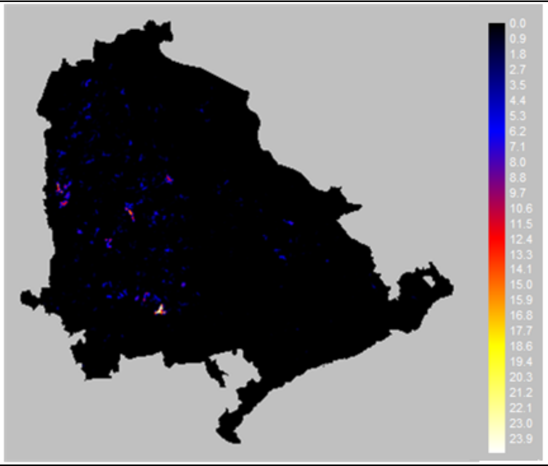


Figure A-24. Map of patch area of continuous bracken cover (GL8).

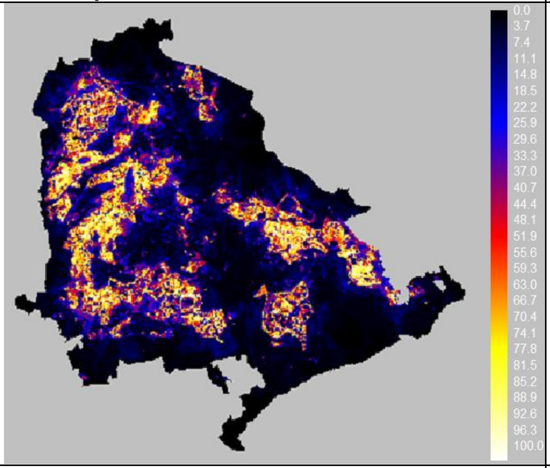


Figure A-25. Biomapper habitat suitability map for *H. semele*.

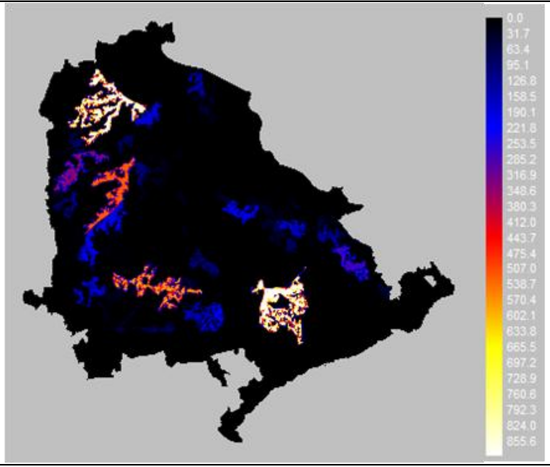


Figure A-26. Map of patch area of merged dry heath and dry heath acid grassland mosaic (HL1_HL3).

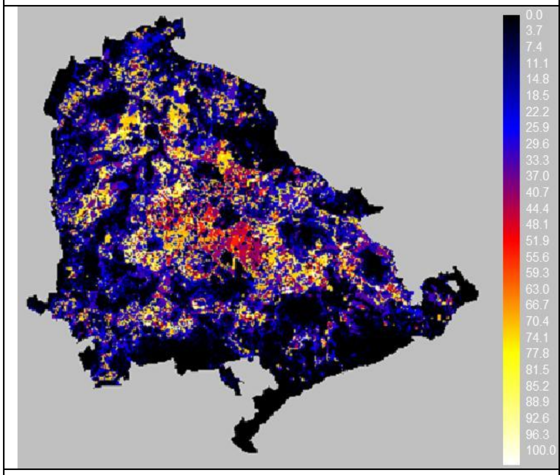


Figure A-27. Biomapper habitat suitability map for *N. sylvestris*.

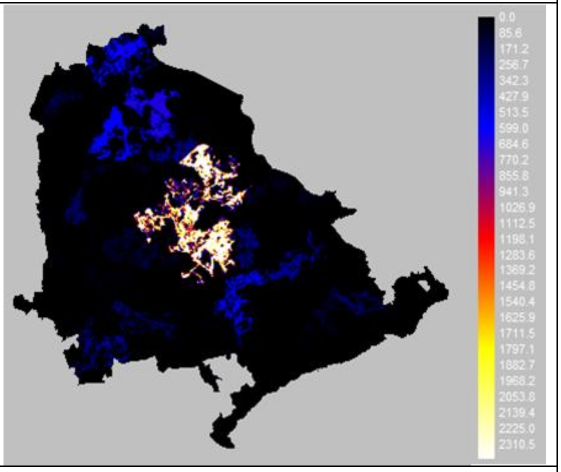


Figure A-28. Map of patch area of deciduous woodland (W1_W2).

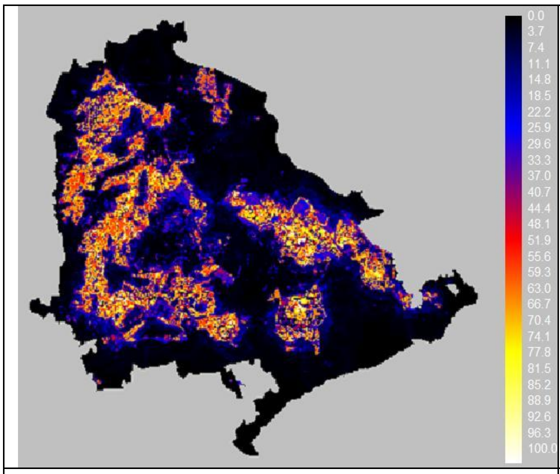


Figure A-29. Biomapper habitat suitability map for *P. argus*.

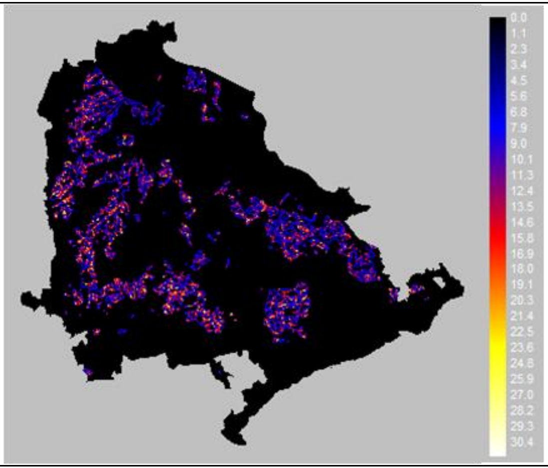


Figure A-30. Map of edge density of dry heath (HL1).

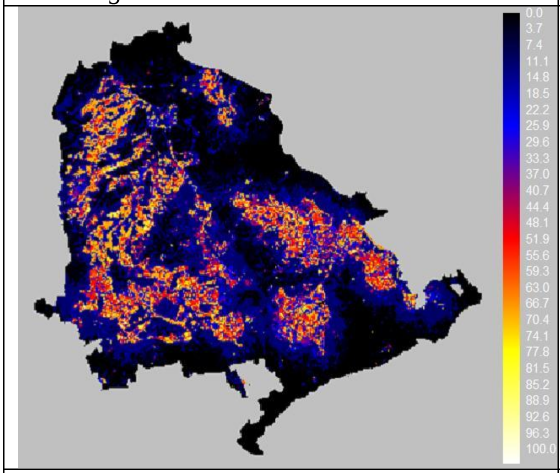


Figure A-31. Biomapper habitat suitability map for *P. globulifera*.

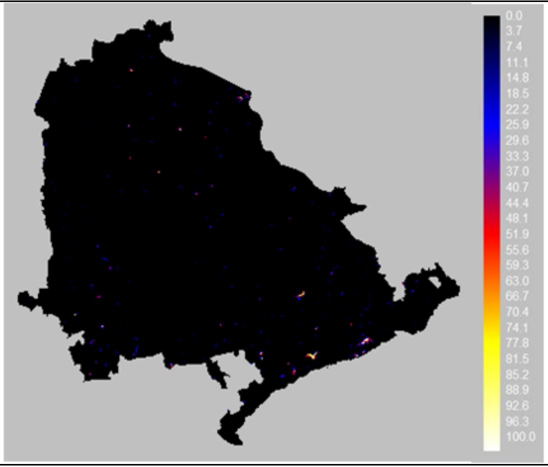


Figure A-32. Map of percentage cover of ponds (AQ5_AQ6).

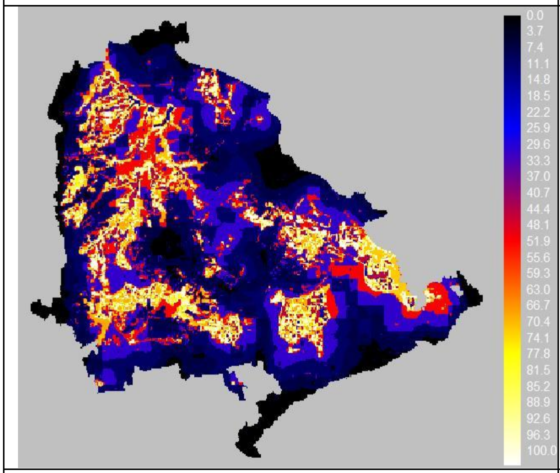


Figure A-33. Biomapper habitat suitability map for *P. punctata*.

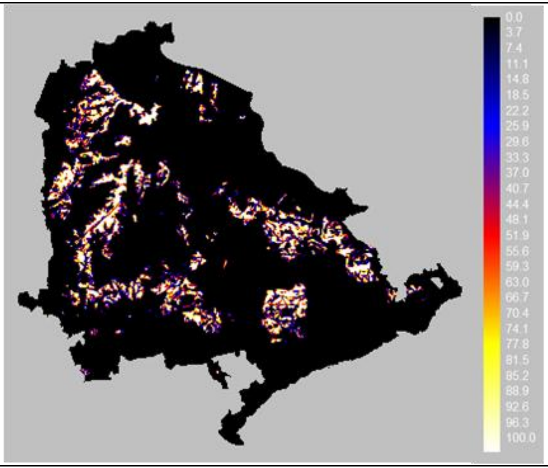


Figure A-34. Map of percentage cover of dry heath (HL1).

Appendix 13. Methodological issues relating to presence-absence species distribution models

13.1. Generation of pseudo-absences

GLMs and GAMs require presence and absence data. However, the majority of species data (particularly for rare and endangered species) is presence-only and good absence data is frequently unavailable. To enable the use of presence-absence methods, ‘pseudo-absence’ data can be generated for a study area. Various strategies have been proposed to generate pseudo-absence data to enhance the performance of species distribution modelling (Lütolf *et al.*, 2006). The easiest way to select pseudo-absences is to generate them totally at random over the study area, for example as used by Ferrier and Watson (1997). However, this runs a risk of generating an absence in an area that is actually favourable to the species (Engler *et al.*, 2004) and the model could be trained on false absence sites with favourable environmental conditions (O'Hanley, 2005a). This is particularly the case when working with rare species, for which data are often scarce, and choosing a wrong absence could significantly reduce the quality of a model (Engler *et al.*, 2004). An alternative is to draw pseudo-absence locations from specific areas known not to hold the species, based on the field notes of collectors or the recollection of experienced field workers (Stockwell and Peterson, 2002a), if this is available.

Guisan *et al*. (2006a) used a novel approach for generating pseudo-absences for a rare plant species based on the occurrences of eleven other rare species. As the study species was easily detectable and well-known, this method worked well because providers of rare species observations usually have a good knowledge of most other rare species. Thus, for each single observation of a rare species one can confidently assume the absence of all other possible rare species that usually share the same type of habitat. However, this sort of data may not be available for a large number of species and the technique would not be suitable for more cryptic species. A new strategy was used by Engler *et al.* (2004) who generated pseudo-absences based on a preliminary ecological niche factor analysis (ENFA) habitat suitability map which they then used in a GLM.

The generation of pseudo-absences beyond the feasible range of a species is known as creating ‘naughty noughts’ (Austin and Meyers, 1996) and if large numbers of these zero values are included in regression analysis they will reduce the explanatory power of the model (Austin and Meyers, 1996). The resulting models can provide over-optimistic evaluation of predictive ability from inspection of ROC curves (Cayuela *et al.*, 2009). In effect, the models are attempting to predict the occurrence of a species beyond the domain where it can exist (Austin and Meyers, 1996).

13.1.2. Proportion of presences and absences (prevalence)

Classification success using logistic regression is sensitive to the relative proportion of presences (the prevalence) and absences in the sample, independently of the fit of the model (Hosmer and Lemeshow, 2000; Real *et al.*, 2006). The logistic function is symmetric by definition, and its inflection point corresponds to a probability value of 0.5; the value is commonly used as a default threshold above which to assume that the model predicts species presence (Real *et al.*, 2006). However, when the proportions of presences and absences are not equal within the sample, the logistic regression output within the function’s domain is not symmetrical, and the probability values become biased toward the highest number of either presence or absences, whichever has the greater number of cases (Hosmer and Lemeshow, 2000; Jimenez-Valverde and Lobo, 2006; Real *et al.*, 2006). The scores for many of the measures used to evaluate models can also be influenced by unequal group sizes (Fielding and Bell, 1997; see Appendix 8.2).

One strategy proposed to avoid this problem is to use, where possible, an equal number of presences and absences (Liu *et al*., 2005), for example as used by Brito *et al.* (1999) and Engler *et al*. (2004). However, this may involve discarding valuable information, particularly for rare species with very few presence records (Jimenez-Valverde and Lobo, 2006; Real *et al.*, 2006). If good absence data are unavailable, then the number of pseudo-absences generated can be easily matched to the number of presence data, but if good absence data are available, then they should all be used. One method suggested for ensuring equal prevalence (0.5) between presences and absences is to down-weight the absences, for example as used by Guisan *et al*. (2006a), Gibson *et al*. (2007) and Ferrier *et al*. (2002b) for GLMs with unequal numbers of presences and absences.

Jimenez-Valverde and Lobo (2006) note that the effect of prevalence may also be mediated by selecting an appropriate cut-off value for converting decimal fraction probabilities to presence/absence and for evaluating the model correctly when using measures such as sensitivity, specificity or the Kappa statistic (see Appendix 8.2).

13.2. Environmental variable selection

One problem often encountered when dealing with a rare species, is that the number of presence sites is likely to be relatively small and this then limits the number of candidate variables that should be used in model development, as a high ratio of candidate models to the number of species observations can lead to overfitting of the model, i.e. inclusion of spurious variables (Harrell, 2001; Gibson *et al.*, 2007). Therefore, when there are a large number of environmental variables available for use in species modelling, or the species ecological requirements are not well known, a procedure known as stepwise variable selection is frequently used to select the appropriate environmental variables to be included in the regression analysis for species distribution modelling (Hosmer and Lemeshow, 2000; Pearce and Ferrier, 2000b). This may involve forward selection of the most significant candidate variable or backward elimination of the least significant predictor in the model or a combination of both. In forward selection, the model initially contains no variables, and variables are added sequentially until a final model is obtained, whereas in backward selection/elimination, all variables are included in the initial model, and these are then removed sequentially until a final model is produced (Pearce and Ferrier, 2000b). Both-directional stepwise selection is a variation on forward selection, in which each forward step is followed by a backward step to remove variables that are no longer significantly related to the response (Pearce and Ferrier, 2000b).

In forward selection, it is possible that a variable selected at an early stage may become unimportant at a later stage, as other variables enter the model. Similarly, in backward elimination, a variable deleted at an early stage could become important at a later stage, as other variables are eliminated from the model (Derksen and Keselman, 1992). Both-directional stepwise selection was aimed at overcoming these problems (Derksen and Keselman, 1992).

Any stepwise procedure for selection or deletion of variables from a model is based on a statistical algorithm that checks for the 'importance' of variables, and either includes or excludes them on the basis of a fixed decision rule (Hosmer and Lemeshow, 2000). The 'importance' of a variable is defined in terms of a measure of the statistical significance of the coefficient for the variable. A commonly used stopping rule (as used in R; see section 3.2.2.3) is based on Akaike's Information Criterion (AIC).

The AIC is an index of fit that takes into account the parsimony of the model by penalising for the number of parameters; it is defined as:

$$\text{AIC} = -2 \text{ x maximum log-likelihood} + 2(p + 1)$$

Where p is the number of parameters and 1 is added for the estimated variance (this could be called another parameter) (Crawley, 2007). Smaller AIC values are indicative of a better fit to the data (Fox, 2002). The more parameters there are in a model the better the fit; a perfect fit could be obtained if the model has a separate parameter for every data point, but this model would have absolutely no explanatory power (Crawley, 2007). There will always be a trade-off between the goodness of fit and the number of parameters required by parsimony, so AIC is useful because it explicitly penalises any superfluous parameters in the model, by adding $2(p + 1)$ to the deviance (Crawley, 2007). The AIC can be used to compare the fit of alternative models with different numbers of parameters, and is typically employed for model selection (Fox, 2002).

Stepwise variable selection procedures have been criticised by many, for example James and McCulloch (1990) who, although recognising that the procedure of screening variables may improve prediction, warn that it may also eliminate variables that are in fact important, and stepwise procedures are not intended to rank variables by their importance. The use of stepwise selection is also cautioned by Harrell (2001) who provides a list of problems with this method, including it yielding biased high R-squared values and *P*-values that are too small, as well as removing the need to think about the problem. Fox (2002) also suggests that researchers using these methods are prone to over-interpret the results: there are often many subsets of predictors of a given size that are nearly equally good.

A further issue is that, by chance, the occurrence of a species may be more strongly negatively correlated with an unsuitable habitat than it is positively correlated with a suitable habitat. This would mean that the unsuitable habitat variable would be selected and the suitable habitat variable may not, which is not as preferable for trying to describe species occurrence (R. Clarke, personal communication, February 19, 2008). A disadvantage of all stepwise methods is that they may fail to find optimal subsets of predictors (Fox, 2002).

It has also been found that several factors affect the frequency with which authentic and noise variables are selected by automated stepwise algorithms (Derksen and Keselman, 1992). Derksen and Keselman (1992) found that as the degree of correlation between predictor variables increased, fewer authentic variables gained entry into the final model (however, correlated variables can be removed prior to stepwise selection). In addition, for a fixed number of authentic variables, as the number of candidate predictor variables increased (from 12 to 24), the frequency with which noise variables entered the final models also increased. Finally, the size of the sample positively affected the number of authentic variables in the final model, although the effect (e.g. from a sample size of 30 to 90) was fairly small (Derksen and Keselman, 1992). Whittingham *et al*. (2006) also note that inconsistencies among model selection algorithms (effects of the direction, order of parameter entry or deletion, and the number of candidate parameters can all affect the selected model) and an inappropriate focus or reliance on a single best model, are issues in the stepwise multiple regression approach.

However, stepwise variable selection is widely used in ecology and species distribution modelling studies (Pearce and Ferrier, 2000b; Whittingham *et al.*, 2006; and see examples below) and it has the advantage, as noted by Tobalske (2002), who used both forward and backwards stepwise procedures, of providing an objective, repeatable approach to model building, as well as allowing for the rapid development of models (Pearce and Ferrier, 2000b). Additionally, Guisan and Zimmermann (2000) do not recommend that selection of predictors be made arbitrarily.

Although Steyerberg *et al.* (1999) highlight some alternatives to stepwise selection, they also have drawbacks and were proposed to construct a regression model with a limited number of variables (which may often not be the case). These include limiting the number of candidate variables, for example by critically reviewing the plausibility of the effect of the candidate variables using knowledge and findings from other studies. A further reduction may be achieved by clustering of variables in related groups. However, Steyerberg *et al.* (1999) note that a drawback of complete pre-specification of the variables for the multivariable model is that no new associations of variables with the outcome will be identified from the data. If current knowledge is used to select variables then it could be possible that no new associations will be identified, because only those variables that are already known to be important will be selected. In addition, variables that are not necessarily considered important may be related indirectly to a species' presence or absence.

Hosmer and Lemeshow (2000) suggest that, particularly when the outcome being studied is relatively new and the important variables and their associations with the outcome may not be well known, stepwise selection can provide a fast and effective means to screen a large number of variables, and to fit a number of logistic regression equations simultaneously. However, data-driven selection with stepwise methods means that true predictors may be excluded because of a lack of power (Steyerberg *et al.*, 1999). Hosmer and Lemeshow (2000) also suggest that stepwise selection provides a useful and effective data analysis tool, which is intuitively appealing in that it builds models in a sequential fashion and it allows for the examination of a collection of models which might not otherwise have been examined. However, they caution that the procedure identifies variables as candidates for a model solely on statistical grounds and therefore following stepwise selection all variables should be carefully scrutinised for biological plausibility. This also helps overcome one of the problems of stepwise selection suggested by Harrell (2001) above of lack of thinking about the problem.

Harrell (2001) suggests that the step-down (i.e. backwards) method is preferred to forward stepwise selection for several reasons, including it usually performing better than forward stepwise methods, and it making one examine a full model fit, which is the only fit providing accurate standard errors, error mean square, and *P*-values.

Forwards selection is also particularly sensitive to the order in which variables enter the model. For example, a variable that is not significant on its own, but is significant in combination with other variables, might be omitted from the model resulting in under-fitting (Pearce and Ferrier, 2000b). Schröder and Richter (1999/2000) found that backwards stepwise developed models always yielded higher prediction accuracy than the forward stepwise developed ones. Backwards stepwise selection was also used by Manel *et al.* (1999) and Matern *et al.* (2007) for example.

13.3. Spatial autocorrelation

An important and, until recently, often neglected (Klute *et al.*, 2002), issue in species distribution modelling is the impact of spatial autocorrelation (Legendre, 1993; Guisan and Zimmermann, 2000). Spatial autocorrelation (SAC) is loosely defined by Legendre (1993) as the property of random variables taking values, at pairs of locations a certain distance apart, that are more similar (positive autocorrelation) or less similar (negative autocorrelation) than expected for randomly associated pairs of observations.

SAC may occur due to external environmental factors (e.g. climate, soil type) which may lead to a similar occurrence probability in neighbouring sites, simply because the external factors show a specific autocorrelation pattern (Dormann, 2007). Historical factors (such as disturbance) can also play a part in spatial autocorrelation, as well as intrinsic (endogenous) factors, due to the biology of the species under consideration: dispersal, colonial breeding, home-range size, competition, host availability, predation or parasitisation risk, and other behavioural factors causing the spatial aggregation of populations and species in landscapes (Dormann, 2007). These causes of SAC are usually much more difficult to quantify, and data are often scarce. They also tend to occur at small spatial scales (e.g. less than 1 km) (Dormann, 2007). Dormann (2007) found no evidence that models for plants and different animal groups differ in their susceptibility to SAC.

As well as ecological reasons, SAC may also be caused by mapping bias from observer bias and differences in sampling schemes and sampling effort (Dormann *et al.*, 2007b). In addition, SAC in model residuals may also be caused by omitting an important autocorrelated environmental variable from the model (Lichstein *et al.*, 2002; Guisan and Thuiller, 2005; Dormann *et al.*, 2007b) or mis-specifying its functional relationship with the response (Legendre, 1993; Dormann *et al.*, 2007b). Dormann (2007) suggests that while this may be true, most of the time the ‘correct’ environmental variables will not be available for the analysis at the required spatial resolution (e.g. abundances of competitors) or at the necessary biological accuracy (e.g. habitat quality). So, for the time being, surrogate variables have to suffice for such factors and as a result, model residuals will probably display SAC. Additionally, moderate collinearity among environmental variables may lead models to exclude one or more variables which would be important in explaining the species’ spatial patterning (Dormann *et al.*, 2007b).

From an ecological point of view, SAC contains information, such as on dispersal, that one might not want to ‘correct for’ in the analysis (Dormann, 2007). In most cases, however, the presence of SAC is seen as posing a serious shortcoming for hypothesis testing and prediction and some previous analyses might be flawed because of ‘red herrings’ generated by SAC with both the estimated predictive power and the choice of variables being seriously biased (Lennon, 2000; Segurado *et al.*, 2006; Dormann *et al.*, 2007b). SAC violates the assumption of data independence (on which analyses such as regression analyses are based) and can result in incorrect inference when using classical statistical techniques (Klute *et al.*, 2002). Positive SAC results in overestimating the effects of ecological covariates in descriptive or predictive models and in declaring too often that computed test statistics are significant under the null hypothesis (type I errors) (Legendre, 1993; Klute *et al.*, 2002).

Sample size is a crucial parameter in the outcome of classical hypothesis testing as it determines the necessary degrees of freedom for pattern detection (Segurado *et al.*, 2006). However, samples from within the range of spatial autocorrelation around a data point will add little independent information (depending on the strength of autocorrelation), compared to more widely spaced samples (Dormann *et al.*, 2007b). Therefore they will overestimate the effective sample size (a form of pseudo-replication), and thus the degrees of freedom of model residuals, thereby influencing statistical inference (Dormann *et al.*, 2007b) by inflating the statistical significance of measured spatial relationships and consequently increasing the likelihood of type I errors (Segurado *et al.*, 2006). Further, species' occurrences tend to be aggregated at most spatial scales, and the more aggregated species' occurrences are, the more likely it is that environmental variables will show some explanatory power simply because of the fact that environmental conditions tend to be more similar at neighbouring sites (Segurado *et al.*, 2006).

Therefore, when methods such as variable selection in stepwise logistic regression are used, the inflation of explanatory power for spatially autocorrelated variables makes them, *a priori*, disproportionately likely to be selected in the final models (Segurado *et al.*, 2006). This is made at the expense of selecting potentially more important variables with lower SAC. However, Segurado *et al*. (2006) note that even if SAC inflates variable significance it does not mean that the final model configuration will exclusively include the most autocorrelated variables. If an autocorrelated variable is included in the model, it may explain a substantial fraction of the SAC in the species' pattern of occurrence. Segurado *et al*. (2006) also note that employing a method such as variable selection procedures using the Akaike information criterion (AIC), is preferable because it does not fully rely on significance thresholds. However, there has been very little research into the issue of variable selection under SAC so the effect of SAC on the identification of best-fitting models remains unclear (Dormann *et al.*, 2007b).

Some studies have reported biases due to SAC in coefficient estimation, whereas others have found no serious effects of SAC, but so far no extensive simulation study has been carried out to investigate how spatial versus non-spatial methods perform under different forms and causes of SAC (Dormann *et al.*, 2007b). The impact of SAC will likely depend on the data and species characteristics. Diniz-Filho *et al*. (2003) suggest that although spatial autocorrelation can impose problems, it is not necessarily the case that all analyses ignoring spatial autocorrelation are flawed. Nevertheless, if the effect of spatial autocorrelation is not analysed, it remains unclear whether regression coefficients would be affected (Carl and Kühn, 2007). Diniz-Filho *et al*. (2003) recommend that SAC in the residuals of multiple regression models should always be checked and spatial models used if the residuals are spatially autocorrelated (Dormann, 2007).

There are several tests, such as Moran's *I* and Geary's *c* statistics, correlograms and semi-variograms which can be used to check for SAC, where a measure of similarity (Moran's *I*, Geary's c) or variance (variogram) of data points is plotted as a function of the distance between them (Dormann *et al.*, 2007b). Moran's *I* coefficient, one of the most commonly used measures of SAC for a data set, is calculated for *N* observations on a variable at *x* locations *i, j* as (Newton, 2007):

$$I = (N/S_0)\Sigma_i\Sigma_j w_{ij}(xi-\mu)(xj-\mu)/\Sigma i(xi-\mu)^2$$

where μ is the mean of the *x* variable, wij are the elements of the spatial weights matrix, and S_0 is the sum of the elements of the weights matrix:

$$S_0 = \Sigma_{ij} wij$$

Moran's *I* usually varies between -1.0 and 1.0 for maximum negative and positive autocorrelation, respectively, with non-zero values of Moran's *I* indicating that values at sites connected at a given geographical distance are more similar (positive autocorrelation) or less similar (negative autocorrelation) than expected for randomly associated pairs of sites (Diniz-Filho *et al.*, 2003). The geographical distances can be partitioned into discrete classes, creating successive matrices and allowing computation of different Moran's *I* values for the same variable. This allows evaluation of the behaviour of autocorrelation as a function of spatial distance, in a spatial correlogram graph (Diniz-Filho *et al.*, 2003).

Many methods have been developed to help account for SAC when modelling, including *a priori* procedures at the level of sampling design, modifications at the level of model adjustment and *a posterior* procedures, such as the use of correction factors, to improve statistical accuracy of models (Segurado *et al.*, 2006). Some spatial autocorrelation in the model may be filtered out by fitting explanatory variables which are also autocorrelated; however, this does not necessarily account for all the autocorrelation (Smith, 1994). Lennon (2000) suggests that a somewhat defeatist approach to SAC is to subsample the original species' distribution usually by adopting a systematic scheme that constrains observations to be spaced far enough from each other (Segurado *et al.*, 2006). However, this method may discard information, which will be a particular problem for small data sets and given that ecological information is often expensive to gather, this is really not a desirable option (Lennon, 2000).

In their review of methods to account for spatial autocorrelation in the analysis of species distributional data, Dorman *et al*. (2007b) arrange the methods into three broad groups. The first group is autocovariate regression and spatial eigenvector mapping (SEVM), which seek to capture the spatial configuration in additional covariates, which are then added into a GLM. Autocovariate models address spatial autocorrelation by estimating how much the response variable at any one site reflects response values at surrounding sites. This is achieved through a simple extension of generalised linear models by adding a distance-weighted function (called a covariate) of neighbouring response values to the model's explanatory variables (Dormann *et al.*, 2007b). Spatial eigenvector mapping is based on the idea that the spatial arrangement of data points can be translated into explanatory variables, which capture spatial effects at different spatial resolutions. During the analysis, those eigenvectors that reduce spatial autocorrelation in the residuals best are chosen explicitly as spatial predictors (Dormann *et al.*, 2007b).

The second group includes generalised least squares (GLS) methods, which fit a variance-covariance matrix based on the non-independence of spatial observations; simultaneous autoregressive models (SAR) and conditional autoregressive models (CAR), which do the same but in different ways to GLS; and generalised linear mixed models (GLMM) which are employed for non-normal data and are a generalisation of GLS. GLMMs are generalised linear models (GLMs) in which the linear predictor may contain random effects and within-group errors may be spatially autocorrelated (Venables and Ripley, 2002; Dormann *et al.*, 2007b). Generalised additive mixed models (GAMMs) are the equivalent for GAMs. Spatial generalised estimating equations (GEE) make up the third group and work by splitting the data into smaller clusters (of sampling units, e.g. in space or time) before modelling the variance-covariance relationship (Dormann *et al.*, 2007b).

Bolker *et al*. (2009) suggest that GLMMs are the best tool for analysing non-normal data that involve random effects. Dormann *et al*. (2007b) also suggest that the most flexible methods, addressing SAC for different error distributions, are spatial GLMMs, GEEs and SEVM. The use of GLMMs in species distribution modelling is still relatively uncommon, and consequently there are very few examples available of their application. However, Milson *et al*. (2000) used GLMMs to model bird species distribution in coastal grazing marshes, because they allowed for the grouping of marshes into land-holdings (where they tended to be more similar than marshes on different land-holdings). Milson *et al.* (2000) concluded that binomial GLMMs comprised a useful tool for investigating habitat factors that affect the distribution of birds at the two nested scales used in their study. They additionally found that the GLMMs in their study performed better than autologistic models, but suggest that the relative merits of the two classes of models may vary between data sets.

GLMMs are composed of fixed-effect parameters (factors whose levels are experimentally determined or whose interest lies in the specific effects of each level) and random-effect parameters (factors whose levels are sampled from a larger population, or whose interest lies in the variation among them rather than the specific effects of each level) (Bolker *et al.*, 2009). Many modern statistical tools, fit these parameters by maximum likelihood (ML), but for GLMMs this is less appropriate (Bolker *et al.*, 2009). Instead, there are various ways to approximate the likelihood to estimate GLMM parameters, including pseudo- and penalised quasilikelihood (PQL), Laplace approximations and Gauss-Hermite quadrature (GHQ) (Bolker *et al.*, 2009).

PQL is the simplest and most widely used GLMM approximation but it yields biased parameter estimates if the standard deviations of the random effects are large, especially with binary data (Bolker *et al.*, 2009). Another disadvantage of PQL is that it computes a quasi-likelihood rather than a true likelihood and many statisticians feel that likelihood-based methods should not be used for inference (e.g. hypothesis testing, AIC ranking) with quasi-likelihoods (Bolker *et al.*, 2009).

An alternative GLMM approximation is the Laplace approximation which, as well as reducing bias, approximates the true GLMM likelihood rather than a quasi-likelihood, allowing the use of likelihood-based inference (Bolker *et al.*, 2009).

Gauss-Hermite quadrature is more accurate still, but is slower than Laplace approximation and because the speed of GHQ decreases rapidly with increasing numbers of random effects, it is not feasible for analyses with more than two or three random factors (Bolker *et al.*, 2009).

Appendix 14. Discussion of variables used for expert variable selection for GLMs and GAMs

(Refer to Tables 13 to 20 in section 3.2.2.2, Chapter 3).

C. nobile:

The two terrain variables (Aspect.flat and Slope) were selected as *C. nobile* occurs in areas that are seasonally wet (Winship and Chatters, 1994; Killick, 2002), i.e. lying water and flatter (less sloped) areas are expected to be more likely to have lying water. All of the soil type variables were included as the first three are all associated with heath (where *C. nobile* is known to occur) and S84102 is seasonally wet and associated with grassland (and *C. nobile* occurs on acid grassland). Cover of wet heath (HL2.cover) and dry heath/acid grassland mosaic (HL3.cover) to reflect a preference for wetter areas and grassy heath areas respectively. The patch area of the habitat is not referred to in the literature as being important, which was why cover of the habitat types was used. The edge density of dry heath and dry heath/acid grassland mosaic was included to reflect the fact that *C. nobile* has been found to occur on edge of heaths (such as near road verges). The Euclidean distance to this merged habitat type was also included. (All the Euclidean distance to the heathland types (individual and merged) were found to be correlated, so only one variable of these could be used). Finally, Euclidean distance to ponds (all sizes) was included due to its requirement for seasonally wet areas and due to personal observations (S. Douglas, during fieldwork for this project) of it occurring near to ponds.

(Aspect.flat and HL2.cover were removed after running the initial GLM as they had negative coefficient values, reducing the total number of selected variables from 11 to 9).

G. constrictum:

Aspect.flat and Slope were selected as *G. constrictum* occurs in wet areas (including ponds) (Brewis *et al.*, 1996; Stace, 1997; Meek, 2002) and it is expected that these would be more likely to occur in flatter areas, where water cannot drain away as easily, and where ponds are more likely to occur. The two ‘wetter’ soil types were also included. To reflect the occurrence of *G. constrictum* around margins of ponds, the % cover of ponds <0.5ha were included (cover of larger ponds were not included as examination of the data showed that most of the presence values for this variable were 0 (and were consequently removed for the GLMs); see section 3.2.1.2). Cover of wet heath and mires (HL2.AQ1.cover) and Euclidean distance to this habitat (ED.HL2.AQ1) was included as *G. constrictum* occurs in wet places, but cover of dry heath and dry heath/acid grassland mosaic (HL1.HL3.cover) was also include to reflect the fact that the ponds and ditches it favours also occur in these areas (Euclidean distance to this habitat type was no included as it was highly correlated with ED.HL2.AQ1). The patch area has not been noted to be important. The Euclidean distance to wet heath/valley mire habitats and to ponds of all sizes were included as *G. constrictum* tends to occur closer to wetter areas.

After running the GLM with the original set ED.AQ5.AQ6 had a positive coefficient value, so was removed and it was decided to re-run the GLM with ED.AQ5 instead, but this also had a positive coefficient value so was removed. S71107 also had the 'wrong' coefficient value (negative) so was removed.

G. illyricus:

Slope was selected as *G. illyricus* has been reported to occur on flat or gently sloping sites (Stokes, 1987) and the soil type S64301 was the one that best encompassed its requirements for a fine sandy soil (Stokes, 1987). The cover of continuous bracken (GL8.cover) was included as *G. illyricus* is always associated with bracken (Stokes, 1987), although the patch area does not appear to be important (edge density and patch compactness of GL8 were also correlated with GL8.cover). Cover of dry heath/acid grassland was also included as *G. illyricus* favours bracken dominated grass heath (Stokes, 1987). The Euclidean distance to continuous bracken (ED.GL8) was also included to reflect the dependence on bracken. It would have been preferable to include the Euclidean distance to dry heath and dry heath/acid grassland mosaic (ED.HL1.HL3) as bracken areas (not necessarily continuous) tend to occur near these habitat types. However, ED.GL8 and ED.HL1.HL3 (and all the other Euclidean distance heathland variables) were highly correlated. The Euclidean distance to woodland edge was also included as it has been reported that *G. illyricus* colonies tend to occur near to woodland (type not specified) edges (Stokes, 1987), and bracken is frequently found near woodland edges.

H. semele:

H. semele occurs on dry heaths, but not where there is dense heather (Oates *et al.*, 2000). Therefore, the patch area of dry heath/acid grassland was selected, as it would be expected to be more likely to occur on larger patches. The Euclidean distance to dry heath (>25% *Calluna*; HL1) and dry heath/acid grassland (<25% *Calluna*; HL3) were included, as although HL3 may be preferable, HL1 would also be suitable if the cover of *Calluna* was not too high. None of the other heathland variables could be included as they were all highly correlated. Cover of dense scrub was also included as during fieldwork (S. Douglas, personal observation) it was noted that *H. semele* often occurred where areas (such as of scrub) had been recently cleared. It is also reported in the literature that *H. semele* is abundant on heaths where heather are regenerating after burns (Oates *et al.*, 2000). None of the terrain variables were expected to be influential.

N. sylvestris:

N. sylvestris is reported to require preferably at least 30% broadleaf trees (Brouwers, 2008) and has a preference for the leaf litter occurring under oaks. Therefore, patch area of broadleaf woodland (PA.W1.W2) and patch area of mixed woodland (PA.W7.W8) were selected. As they were not correlated, the edge density of these habitat types were also selected. The edge density of coniferous woodland (Eden.W1.W2) was additionally selected, as coniferous woodland in the New Forest frequently have a few broadleaved trees on their edges, where, during fieldwork (S. Douglas, personal observation), *N. sylvestris* was found to occur. The Euclidean distance to woodland edge (any type) was included as *N. sylvestris* has a strong preference for woodland edge (where there are broadleaved trees) (Marshall and Haes, 1988; Brouwers, 2008). The soil type S64303 was included as it is associated with deciduous woodland, although it would not directly affect *N. sylvestris* occurrence. Aspect.south was included because *N. sylvestris* has a preference for warm and sunny clearings or margins of woods (Marshall and Haes, 1988), and south-orientated sites would be more likely to provide these conditions (Brouwers, 2008).

Although Eden.W1.W2 returned a negative estimate/slope value (‘wrong’ relationship), it was kept in at this stage as *N. sylvestris* has a known preference for woodland edges (Marshall and Haes, 1988; Brouwers, 2008).

P. argus:

Aspect.flat and Slope were selected because *P. argus* is reported to occur on flat or gently sloping land (Joy, 1995). It is also reported to prefer warmer south-facing slopes further north, although this preference is not apparent in the south of the UK (Joy, 1995). However, as the maximum number of variables to be included was not reached, Aspect.south was included. Although soil type is unlikely to be directly relevant, soil types S64301, S64303 and S71107 were included as they are associated with heathland habitats. *P. argus* is a heathland species and uses both dry and wet heath in the New Forest (Oates *et al.*, 2000). Therefore the patch area of both of these heathland types were included, as larger patches are better (Joy, 1995). The merged dry heath patch area layer (PA.HL1.HL3) was used as opposed to the individual (PA.HL1 or PA.HL3) as the cover of *Calluna* does not appear to be too important.

The edge density of these habitat types were also included as the borders between wet and dry heath, both of which are used by *P. argus*, may be particularly suitable. The Euclidean distance to wet heath (ED.HL2) was also included to reflect this, but the Euclidean distance to the other heathland types could not be included as they were all highly correlated. Euclidean distance to wet heath was chosen over the others as it is more likely to contain cross-leaved heath (*Erica tetralix*) (and heather, *Calluna vulgaris*), the favoured food plants of *P. argus* in the New Forest (Oates *et al.*, 2000).

P. globulifera:

As for *G. constrictum*, Aspect.flat and Slope were selected as *P.globulifera* occurs in wet areas (including ponds and wet muddy ditches) (Jermy, 1994; Brewis *et al.*, 1996; Stace, 1997; Scott *et al.*, 1999; Preston, 2002) and it is expected that these would be more likely to occur in flatter areas, where water cannot drain away as easily, and where ponds and lying water are more likely to occur. The three soil types were selected that contained clays and sands as these were reported as the associated soil type for *P. globulifera* (Plantlife International, 2006). The cover of the merged dry heath habitats (HL1 and HL3) and the wet heath and mire habitats (HL2 and AQ1) were included because ponds and other wetland features may occur in any of these habitat types, although lying water may be more likely to occur in the wetter habitat types. The Euclidean distance to these wetter heath habitats was included for that reason (the other Euclidean distance to heathland habitats variables were highly correlted). Patch area has not been reported to be important, so the cover variables were used. Cover and Euclidean distance to ponds of all sizes was included as a preference for a certain size of pond is not reported in the literature.

After running the GLM, none of the soil types had a positive coefficient value, so they were removed (as was Aspect.flat). As there were no longer the maximum number of variables allowed (based on the number of records), the GLM was run with soil type S84102, as it is a seasonally wet soil. This produced a positive coefficient value, so was kept.

P. punctata:

P. punctata is found on dry and wet heath (Cox and Pickess, 1999; Poland, 2004), with the cover of *Calluna* not reported to be a factor, although *P. punctata* does not appear to occur as frequently on just acid grassland with no heathland mosaic (Poland, 2004). The Euclidean distance to dry heath and dry heath/acid grassland mosaic was also included. (Only one of the Euclidean distance to the heathland type variables could be included as they were all highly correlated). The soil types S64301 and S64303 were included as they are associated with heathland habitats (see Appendix 6). Although it is also associated with heathland habitats, soil type S71107 was not included as it is seasonally waterlogged and Poland (2004) suggests that soils with a level of drainage impedence are suitable, but not generally waterlogged soils. None of the terrain variables are reported as relevant in the literature so none of these variables were included.

Climate variables were not selected for any of the species as it was felt that they did not show enough variation to have an influence across the New Forest and information on the preferences of species at that scale was not available.

Appendix 15. Variables entered into stepwise variable selection for GLMs and GAMs

Table A-23 below shows the variables used in the ‘full’ GLMs and GAMs which were used to run stepwise variable selection (see section 3.2.2.3, Chapter 3). For the GLMs this was the set of variables used in the GLM for each species after removal of variables with all or mostly zero values for the species presence and absence sites, and after removal of one of the pair of correlated variables or variables with ‘wrong’ direction coefficients (see section 3.2.2.1). For the GAMs, this included variables that had *P*-values less than or equal to 0.1 from the univariate GAMs run using the set of variables used for the GLMs after the correlated variables and those with all or mostly zero variables had been removed (but before removal of variables with ‘wrong’ relationships) (see section 3.2.3.3).

Model	**GLMs**								**GAMs**	
Species	**C. no**	**G. co**	**G. il**	**H. se**	**N. sy**	**P. ar**	**P. gl**	**P. pu**	**C. no**	**P. pu**
No. of presence training records	136	85	66	130	103	180	99	70	136	70
Maximum no. of variables	14	9	7	13	10	18	10	7	14	7
No. of selected variables	22	20	15	18	27	29	23	17	15	10
Selected variables:										
S64301			/		/	/	/	/		
S64303	/	/	/	/	/	/	/	/		
S71107			/	/	/	/	/			
S84102	/	/		/	/	/	/		/	
DTM				/				/		/
Slope	/	/	/	/	/	/	/	/	/	
Aspect.flat	/	/		/	/	/				
Aspect.north					/			/		
Aspect.south	/	/				/	/			
Bio12				/		/	/	/		/
Prec12	/	/	/		/					
Tmin2	/	/				/	/		/	
AQ5.cover		/								
AQ5.AQ6.cover							/			
GL8.cover				/						
HL3.cover	/		/	/					/	/
HL1.HL3.cover				/		/	/	/		/
W1.W2.cover					/	/				
Eden.AQ1						/				
Eden.GL3	/									
Eden.GL8			/			/				
Eden.HL1	/				/			/		/
Eden.HL2				/						

Eden.HL3			/		/	/		/		/
Eden.HL1.HL3	/	/	/	/		/	/		/	
Eden.HL2.AQ1		/	/			/	/			
Eden.W1.W2	/				/	/	/	/	/	
Eden.W5.W6					/	/	/			
Eden.ST1	/	/		/		/		/		/
PC.GL3									/	
PC.GL8		/								
PC.HL1		/								
PC.HL3							/			
PC.HL1.HL3	/		/						/	
PC.W1.W2		/		/	/	/				
PC.W7.W8					/					
PC.ST1							/			
PA.AQ1				/		/				
PA.GL8					/			/		
PA.HL1	/						/		/	
PA.HL2		/				/				
PA.HL3					/			/		
PA.HL1.HL3		/				/			/	
PA.HL2.AQ1	/				/		/	/		
PA.W1.W2	/	/	/		/	/				
PA.W5.W6					/					
PA.W7.W8			/		/					
ED.AQ5				/		/				
ED.AQ6		/			/	/	/	/	/	/
ED.AQ5.AQ6	/						/			
ED.GL8			/							
ED.HL1				/		/				
ED.HL3					/		/			
ED.HL1.HL3		/						/		/
ED.HL2.AQ1	/								/	
ED.Bld.edge	/								/	
ED.con.edge					/	/	/			
ED.mix.edge			/		/					
ED.wood.edge					/		/			
ED.UR.nonres	/	/		/		/		/	/	/
ED.UR.res	/				/				/	

Table A-23. Variables entered into stepwise variable selection for GLMs and GAMs. No. of training records is the number of training records in the presence or absence category (i.e. on which the maximum number of variables can be calculated). Maximum number of variables is the maximum number of suggested variables that should be in the final model (e.g. after stepwise selection) based on the rule of thumb suggested by Harrell *et al.* (1996) that no more than *m*/10 predictors should be included in the final model, where *m* is the total number of observations or the number of observations in the least represented category in the case of a binary response (shown in the number of training records row).

Note that not all variables initially used at the very start of the modelling process are listed in the table as some variables were not included for entry into any of the stepwise models).

Appendix 16. Discussion of selection of variables for the final 'best' GLMs

C. nobile

No additional significant variables were selected by expert variable selection compared to stepwise variable selection. Only 3 of the final 9 expert-selected variables were significant. Of the non-significant variables one (S64303) was selected by stepwise, where it had a significant *P*-value. Of the remaining non-significant expert-selected variables two (HL3.cover and ED.AQ5.AQ6) were available for stepwise, but were not selected. The stepwise GLM was used as the final model, although one of the variables (ED.Bld.edge) was not significant at $P<0.10$, so was removed. When the GLM was re-run, a further variable (S64303) was no longer significant, so was removed, leaving the final model with 7 variables.

The positive coefficient for the S84102 soil type reflects the preference of *C. nobile* for seasonally wet grassland and the Euclidean distance to wet heath/mire also reflects a preference for wetter areas, although all of the Euclidean distance to heathland types were correlated, so it may just reflect a preference for acid heath, as also suggested by the positive coefficient value for edge density of dry heath/acid grassland mosaic (Eden.HL1.HL3). The negative coefficient value for slope also fits with the suggestion in the expert-selection of variables, that *C. nobile* would tend to occur on flatter (i.e. lower slope values) sites. The positive coefficient value for edge density of broadleaved woodland (Eden.W1.W2) does not make as much sense, but may reflect occurrence of *C. nobile* in woodland glades (Plantlife International, 2001). The association with sites closer to residential and non-residential urban areas may indicate its occurrence on the damper New Forest lawns, notably those at settlement edges (Tubbs, 2001) and on road and track edges (Plantlife International, 2001).

G. constrictum

Only three of the final 6 expert-selected variables were significant and the expert-selected set of variables only produced one additional significant variable (ED.HL2.AQ1) to stepwise variable selection, but this was correlated with the

ED.HL1.HL3 variable which was selected by stepwise. As the ED.HL1.HL3 (stepwise-selected) variable had a more significant *P*-value, it was decided to use the stepwise set of variables, including ED.HL1.HL3, rather than the ED.HL2.AQ1 (expert-selected) variable. Two of the stepwise-selected variables (Eden.HL2.AQ1 and PC.W1.W2) were not significant at $P<0.1$, so were removed. The final model contained 6 variables which were all significant at $P<0.05$.

The association of *G. constrictum* with wetter areas is reflected by the positive coefficient value for soil type S84102 and perhaps also by the negative slope coefficient value (i.e. flatter areas where water may lie). The occurrence on sites with low Euclidean distance to dry heath/acid grassland mosaic (ED.HL1.HL3) corresponds with reports in the literature (Brewis *et al*., 1996). The association with south-facing sites it's not documented in the literature but may reflect a requirement for lighter sites and may also be linked to the association with higher minimum February temperatures (Tmin2). However, this could just be an artefact of the data, reflecting recordings of the species more in the south-east of the study area, where temperatures are slightly warmer than in the north-west. Lastly, the association with sites closer to non-residential urban areas may be due to the occurrence of *G. constrictum* in ditches, which often occur on road edges, or may also be the results of recording bias at sites which are easily accessible.

G. illyricus

Soil type S64301 and ED.GL8 were selected as significant by both expert and stepwise variable selection. Two of the expert-selected variables (Slope (not significant) and HL3.cover (significant)) were not selected by step, even though they were available to be selected. GL8.cover (an expert-selected variable) was not selected by stepwise as it was not available. However, Eden.GL8, which was correlated with GL8.cover, was available but was not selected by stepwise. The final expert-selected variable, ED.wood.edge (which was not significant), was not available to be selected by stepwise, although ED.mix.edge was, but was not selected. Therefore, the expert-selected variables did not provide any additional significant variables that were not selected by stepwise (or correlated variables), so it was decided to use the stepwise model as the final model. The set of 8 stepwise-selected variables were all significant at $P<0.10$, with the least significant variable, S64303, with a *P*-value of 0.06. However, the maximum number of permitted variables based on the sample size was 7, so the least significant variable, and the only variable that was not significant at *P*<0.05 (S64303), was removed, so that there were 7 variables in the final model.

The inclusion of soil type S64301 is a reflection of the requirements of *G. illyricus* for a fine sandy soil. However, the association with soil type S71107 is not one that was included for the expert-selection of variables, but it is a loamy soil and is wetter than S64301. However, Stokes (1987) suggests that the soil should be damp enough to prohibit bracken from becoming dominant, without being so wet that the corms rot. The association with dry heath, as reported in the literature (Stokes, 1987; Brewis *et al.*, 1996; Taylor, 2002) is also shown by the inclusion of the edge density and patch compactness of dry heath/acid grassland mosaic variables and the main association, with bracken, incorporated in the Euclidean distance to continuous bracken variable (Stokes, 1987; Brewis *et al.*, 1996; Taylor, 2002). The suggestion in the literature (Stokes, 1987) that *G. illyricus* is often found near to woodland is given some support by the positive coefficient values for patch area of deciduous woodland and mixed woodland, suggesting that *G. illyricus* tends to occur at sites where there is also woodland that is part of a larger patch of woodland present.

H. semele

One of the significant expert-selected variables for the *H. semele* model, S64303, was also selected by stepwise. The other significant expert-selected variable, ED.HL1.HL3, was not selected by stepwise, but ED.HL1, with which it was highly correlated, was (ED.HL1.HL3 had been removed during checks for collinearity). The other two expert-selected variables were not significant. ST1.cover was not available to be selected by stepwise, but Eden.ST1 (with which it was correlated) was and was selected as highly significant by stepwise. The other expert-selected variable, PA.HL3 was not available to be selected by stepwise, but HL3.cover, with which it was correlated with was, but was not selected. It was therefore, decided to use the stepwise set of variables, as the expert set did not contribute anything extra. One of the stepwise-selected set of variables, Eden.HL2, was not significant (at *P*<0.1) so was removed and the GLM re-run. The remaining 9 variables were all significant at *P*<0.1 (with only 2 not significant at *P*<0.05) so were used as a final model.

Of the two remaining variables that were not significant at $P<0.05$, ED.AQ5 did not appear to make sense ecologically as there is no association with ponds reported in the literature. The association with lower annual precipitation may reflect the preference of *H. semele* for drier sites (although this is unlikely to be evident at this scale), but it was decided to remove these two variables.

The association of *H. semele* with dry heath is demonstrated by the inclusion of the HL1.HL3.cover and Euclidean distance to dry heath variables. The soil type S64303 is also associated with heathland, and although it does not directly influence butterfly occurrence, it influences the vegetation type, which is important for *H. semele* occurrence. The other soil type selected, S71107, is a wetter soil and is associated with wet heath habitat, which does not seem to fit with the literature. It may be that *H. semele* was recorded passing through wetter areas as it is relatively mobile and the different heathland types often occur as a mosaic, or it may reflect the broadness of the soil type categories and the possible difficulty in defining a boundary of where the soil type changes from one type to another. The positive coefficient values for edge density of dense scrub (Eden.ST1) support the suggestion in Appendix 14 that *H. semele* is often found in areas where scrub may have been cleared. However, this may not be a good variable to include because it is likely to change in a relatively short time.

The association with lower elevation (DTM) values is not something that is reported in the literature, but those sites may be more sheltered and warmer, or it may just be that the areas where *H. semele* was recorded tended to be in the areas of the New Forest with lower elevation (such as the south-east), although there are not any areas of the Forest with an elevation greater than 153 metres above sea level. The positive coefficient value for Slope suggests that *H. semele* tends to occur on slopes, rather than flatter areas, which is again something that is not mentioned in the literature, but may be important, or may just be a coincidence.

N.sylvestris

Both of the two significant expert-selected variables for the *N. sylvestris* model, PA.W1.W2 and ED.wood.edge, were also selected by stepwise. Of the 5 non-significant expert-selected variables, 4 were available for stepwise-selection, but were not selected. The other non-significant expert-selected variable was not available for stepwise selection, but PC.W7.W8, with which it was correlated with, was, but was not selected. It was therefore decided to use the stepwise selected variables as the final model. Of the 11 stepwise-selected variables, one (Eden.HL3) was not significant at $P<0.10$, so the GLM was re-run with it removed. PA.GL8 then became non-significant, so the GLM was re-run without it and the remaining 9 variables were all significant at $P<0.10$, with only one (ED.wood.edge) not significant at $P<0.05$. The 9 variables were used for the final model.

Some of the final stepwise-selected variables did not make ecological sense. Patch area of wet heath/mire (PA.HL2.AQ1) in particular seemed particularly strange, as there is no mention of an association with these habitats in the literature or from personal (fieldwork) experience (*N. sylvestris* was not found in wet areas). Perhaps by chance, a number of the *N. sylvestris* woodland edge sites shared a 100 m x100 m pixel with the HL2.AQ1 habitat. However, this would suggest that the woodland may be more damp than originally it was thought that *N. sylvestris* would tolerate. The stepwise model also indicates that *N. sylvestris* tends to occur closer to large ponds (ED.AQ6). However, it could be that where *N. sylvestris* occurrence was recorded when visiting sites to look for other species included in the study, namely *P. globulifera* and *G. constrictum*, associated with ponds and wetter areas, these apparent associations have arisen through an artefact of the non-stratified ad hoc recording. This is an example of one of the issues with using such data and demonstrates that results should also be analysed ecologically, rather than just relying on mechanistic model selection.

The Euclidean distance to dry heath acid grassland mosaic (ED.HL3) may just result from the fact that *N. sylvestris* tends to occur on the edge of woodlands, which in the New Forest, frequently meet heathland and where there is often bracken present (which may help trap leaf litter) (N. Brouwers, personal communication, January 9, 2009; S. Douglas, personal observation). The negative coefficient for Euclidean distance to residential urban areas appears strange, and apart from the possibility of bias of recording to easily accessible areas, there is not an obvious explanation. However, as the *P*-values for both of these variables were very significant, they were kept.

As ED.AQ6 and PA.HL2.AQ1 seemed particularly strange results, the GLM was also run with these two variables removed. The remaining 7 variables were all significant at $P<0.05$ (including ED.wood.edge, which was not before) but the residual deviance and AIC values were higher than with the model with the 9 variables.

The inclusion of the patch area and patch compactness of broadleaved woodland variables demonstrate the association of *N. sylvestris* with broadleaved woodland. The positive coefficient value for the patch area of coniferous woodland also shows that the woodland does not have to be just broadleaved to be suitable and *N. sylvestris* can often be found under a few broadleaved trees occurring on the edge of coniferous woodland (S. Douglas, personal observation). This is also reflected by the negative coefficient values for Euclidean distance to woodland edge and mixed edge.

P. argus

All of the expert-selected variables were selected by step, apart from Eden.HL2 and ED.HL2. However, neither of these variables were available for entry into stepwise selection, but the variables that they were correlated with were. Eden.HL2.AQ1 (correlated with Eden.HL2) was not selected, but ED.HL1 (correlated with ED.HL2) was selected by stepwise. As all of the expert-selected variables were included (or available but not selected) in the stepwise model, it was decided to use the stepwise set of variables as the final model. However, the stepwise-selected variables included 4 variables that were not significant at $P<0.10$ (Aspect.south, Slope, PA.AQ1 and PA.HL2), so the GLM was re-run without these variables. Six more variables (S71107, S84102, S64303, Aspect.flat, HL1.HL3.cover and W1.W2.cover) were then no longer significant, so were removed. The resulting GLM then had a further non-significant variable (Eden.GL8), which was removed to leave a final model of 7 variables, all significant at $P<0.10$.

The selected variables are mostly heathland variables and reflect the use of both dry heath (Eden.HL1.HL3, PA.HL1.HL3 and ED.HL1) and wetter mire (Eden.AQ1) habitats (which often occur in a mosaic with other (wetter) heathland habitats) and sites close to large ponds (ED.AQ6). The soil type S64301, which although would not directly affect butterfly occurrence, is also associated with wet heath habitats.

The association with dense scrub (Eden.ST1) seems slightly strange, but may reflect recordings of *P. argus* on sites that have recently been burned (to clear the scrub), which it favours. However, this was the least significant of all the selected variables, and may not be as suitable, due to the more temporary condition of this habitat type.

P. globulifera

Four of the expert-selected variables were also selected by stepwise. Of the remaining three variables, one (HL1.HL3.cover) was available for selection by stepwise but was not selected and the other two (HL2.AQ1.cover and ED.HL2.AQ1) were not available for stepwise selection, but the variables that they were correlated with were selected (Eden.HL2.AQ1 and ED.HL3 respectively). It was therefore decided to use the stepwise set of variables. Stepwise had selected a large number of variables (14), which was more than the maximum suggested number with the number of training data available (10). However, one of the variables (S71107), was not significant (at $P<0.1$) so was removed. Once S71107 was removed, 3 more variables (S64303, Eden.HL1.HL3 and ED.AQ5.AQ6) then became non-significant (at $P<0.1$) so were removed, which led to 2 more variables (Bio12 and ED.con.edge) no longer being significant and therefore being removed, with ED.con.edge then no longer being significant. The remaining 7 variables were then all significant at $P<0.1$.

A positive association with Eden.W5.W6 did not make sense ecologically as the literature suggests that *P. globulifera* is a light-loving plant (Hill *et al.*, 2004), with no suggestion of it occurring in coniferous (or any other) woodland sites. The model was therefore run with Eden.W5.W6 removed as a comparison (6 variables).

The rest of the selected variables made ecological sense, with positive coefficient values for edge density of wet heath/mire and for soil type S84102 demonstrating the association with wet sites and ponds (AQ5.AQ6.cover). The occurrence on sites with lower (i.e. flatter) slope values may also indicate sites which are more likely to be wet or have laying water. However, where a site is sloped, the preference is shown to be for south-facing sites, perhaps reflecting the light-loving nature of *P. globulifera*. The association with sites closer to dry heath/acid grassland also reflects the occurrence in heaths (Plantlife International, 2006) and on mainly acid soils (Hill *et al.*, 2004).

P. punctata

The two significant expert-selected variables (S64303 and ED.HL1.HL3) were both selected by stepwise. The other two (non-significant) expert-selected variables were unsurprisingly not selected by stepwise even though they were available. Therefore, as the stepwise model included the significant expert-selected variables, and had the option to include the non-significant ones, it was decided to use this set. The stepwise set of variables included 10 variables, when the maximum number should not exceed 7. However, there were three non-significant ($P<0.1$) variables in the stepwise selection set (S64303, PA.HL3 and PA.HL2) so these were removed. The remaining 7 variables were all then significant at $P<0.1$, with only 2 (Eden.HL1 and ED.AQ6) not significant at $P<0.05$.

The association of *P. punctata* with dry heath is demonstrated by the Eden.HL1 and ED.HL1.HL3 variables. However, the selected variables also suggest that *P. punctata* occurs closer to large ponds. There is nothing in the literature that suggests this, but perhaps it is because ponies occur more frequrntly close to ponds, or more likely, it could be the result of a recording artefact (as discussed for *N. sylvestris* above). There also appears to be an association with distance to non-residential urban areas (such as roads). This may also reflect occurrence of *P. punctata* on the edge of dry heaths (from Eden.HL1), such as where they meet roads. However, this could also be due to recording bias of people tending to record species in more easily accessible areas. Interestingly, the results suggest that *P. punctata* tends to occur on less sloped areas (which may just reflect a preference of the ponies) and north-facing sites (perhaps because the dung does not dry out as much as on south-facing sites). Finally, a preference is also identified for areas with higher annual precipitation, possibly reflecting a requirement for slightly higher dampness. Alternatively, it may just reflect fewer records in the north-east of the New Forest where annual precipitation is slightly lower, or it could just be an artefact of the data.

Appendix 17. Results of GLMMs for final set of GLM variables

Species	**Number of variables**	**Residual deviance**	**AIC**
C. nobile	7	288	306
G. constrictum	6	131	147
G. illyricus	7	74.3	92.3
H. semele	9	260	282
N. sylvestris	9	177	199
N. sylvestris	7	199	217
P. argus	7	344	362
P. globulifera	7	169	187
P. globulifera	6	176	192
P. punctata	7	83.9	102

Table A-24. Residual deviance and AIC results of GLMMs for final set of GLM variables. Number of variables is the number of variables included in the model.

Species	**Selected variables**	**Estimate**	**Standard error**	**Z value**	**Pr(>|z|)**
C. nobile (7)	(Intercept)	1.25	0.479	2.60	0.00938
	S84102	0.00948	0.00454	2.09	0.0367
	Slope	-0.445	0.0991	-4.49	0.00000714
	Eden.HL1.HL3	0.125	0.0292	4.28	0.0000191
	Eden.W1.W2	0.0851	0.0268	3.18	0.00147
	ED.UR.nonres	0.000484	0.000191	-2.53	0.0113
	ED.UR.res	-0.000669	0.000368	-1.82	0.0694
	ED.HL2.AQ1	-0.000608	0.000244	-2.49	0.0127
G. constrictum (6)	(Intercept)	-3.12	1.58	-1.97	0.0487
	S84102	0.0138	0.00608	2.27	0.0235
	Aspect.south	0.0158	0.00642	2.46	0.0140
	Slope	-0.454	0.173	-2.62	0.00877
	Tmin2	3.56	1.02	3.51	0.000453
	ED.UR.nonres	-0.000739	0.000352	-2.10	0.0357
	ED.HL1.HL3	-0.00493	0.00105	-4.69	0.00000271
G. illyricus (7)	(Intercept)	0.0751	0.589	0.128	0.899
	S71107	0.0147	0.00676	2.17	0.0301
	S64301	0.0347	0.0133	2.61	0.00902
	Eden.HL1.HL3	0.117	0.0591	1.98	0.0473
	PC.HL1.HL3	0.126	0.0449	2.81	0.00497
	PA.W1.W2	0.00177	0.000843	2.10	0.0357
	PA.W7.W8	0.160	0.0785	2.03	0.0421
	ED.GL8	-0.0142	0.00332	-4.28	0.0000188
H. semele (9)	(Intercept)	21.9	11.9	1.84	0.0661
	S71107	0.0105	0.00439	2.40	0.0165
	S64303	0.0146	0.00461	3.17	0.00151
	DTM	-0.0168	0.00677	-2.48	0.0131
	Slope	0.271	0.0830	3.26	0.00112
	HL1.HL3.cover	0.0109	0.00505	2.17	0.0304
	Bio12	-0.0281	0.0152	-1.85	0.0646
	Eden.ST1	0.116	0.0462	2.51	0.0121
	ED.AQ5	-0.000516	0.000275	-1.88	0.0606
	ED.HL1	-0.00164	0.000466	-3.53	0.000419
N. sylvestris (9)	(Intercept)	3.12	0.645	4.84	0.00000128

	PC.W1.W2	0.0863	0.0247	3.50	0.000472
	PA.HL2.AQ1	0.0736	0.0267	2.75	0.00590
	PA.W1.W2	0.00121	0.000338	3.57	0.000357
	PA.W5.W6	0.0193	0.00746	2.59	0.00974
	ED.AQ6	-0.000587	0.000170	-3.46	0.000551
	ED.HL3	-0.00201	0.000521	-3.85	0.000117
	ED.mix.edge	-0.00299	0.000688	-4.35	0.0000135
	ED.UR.res	-0.00113	0.000514	-2.21	0.0273
	ED.wood.edge	-0.00917	0.00505	-1.82	0.0693
N. sylvestris (7)	(Intercept)	1.58	0.444	3.55	0.000391
	PC.W1.W2	0.0617	0.0224	2.76	0.00581
	PA.W1.W2	0.00126	0.000331	3.82	0.000136
	PA.W5.W6	0.0155	0.00655	2.37	0.0177
	ED.HL3	-0.00165	0.000474	-3.48	0.000494
	ED.mix.edge	-0.00193	0.000562	-3.43	0.000610
	ED.UR.res	-0.000900	0.000455	-1.98	0.0481
	ED.wood.edge	-0.0103	0.00461	-2.24	0.0250
P. argus (7)	(Intercept)	0.614	0.362	1.69	0.0902
	S64301	0.00745	0.00388	1.92	0.0550
	Eden.AQ1	0.0699	0.0381	1.83	0.0669
	Eden.HL1.HL3	0.0816	0.0285	2.86	0.00425
	Eden.ST1	0.0913	0.0525	1.74	0.0820
	PA.HL1.HL3	0.00224	0.000739	3.03	0.00244
	ED.AQ6	0.000437	0.000120	-3.64	0.000271
	ED.HL1	-0.00139	0.000342	-4.07	0.0000475
P. globulifera (7)	(Intercept)	0.979	0.435	2.25	0.0242
	S84102	0.0168	0.00654	2.56	0.0104
	Aspect.south	0.0117	0.00607	1.92	0.0545
	Slope	-0.584	0.135	-4.32	0.0000154
	AQ5.AQ6.cover	0.178	0.0993	1.79	0.0735
	Eden.HL2.AQ1	0.143	0.0436	3.27	0.00108
	Eden.W5.W6	0.183	0.0741	2.47	0.0135
	ED.HL3	-0.00470	0.00102	-4.60	0.00000430
P. globulifera (6)	(Intercept)	1.05	0.428	2.45	0.0145
	S84102	0.0151	0.00629	2.40	0.0165
	Aspect.south	0.0114	0.00593	1.93	0.0537
	Slope	-0.499	0.127	-3.94	0.0000816
	AQ5.AQ6.cover	0.178	0.104	1.72	0.086
	Eden.HL2.AQ1	0.124	0.0414	3.01	0.00263
	ED.HL3	-0.00463	0.000985	-4.70	0.00000257
P. punctata (7)	(Intercept)	-57.3	21.4	-2.68	0.00735
	Aspect.north	0.0228	0.00999	2.28	0.0226
	Slope	-0.281	0.137	-2.05	0.0406
	Bio12	0.0763	0.0272	2.81	0.00501
	Eden.HL1	0.100	0.0560	1.79	0.0736
	ED.AQ6	0.000384	0.000218	-1.76	0.0780
	ED.UR.nonres	-0.00121	0.000397	-3.05	0.00230
	ED.HL1.HL3	-0.00836	0.00232	-3.60	0.000321

Table A-25. Results of GLMMs for final set of GLM variables. Where two different versions of a model were run (i.e. with different numbers of variables) these are shown. The numbers in brackets next to the species name indicate the number of selected variables. The estimate column shows the regression coefficients for the intercept and slope (for the selected variables) and their standard errors, which are both in logits. The Z-value (Wald statistic) is the estimate value divided by the standard error (the ratio of the coefficient to its standard error) and tests the hypothesis that the regression coefficient is zero (Fox, 2002), and the *P*-value indicates whether the Z-value is significantly different from zero.

Appendix 18. GAM plots for final GAMs

The GAM plots show the smooth function fitted for each term in the additive model (Venables and Ripley, 2002). The etchings on the *x* axes (called rugs) indicate the locations of observations on that variable (Venables and Ripley, 2002).

The GAM plots for all 12 smoothed variables initially selected by backwards stepwise GAM (including the 7 variables in the final GAM) for *C. nobile* are shown in Figures A-35 to A-46 below.

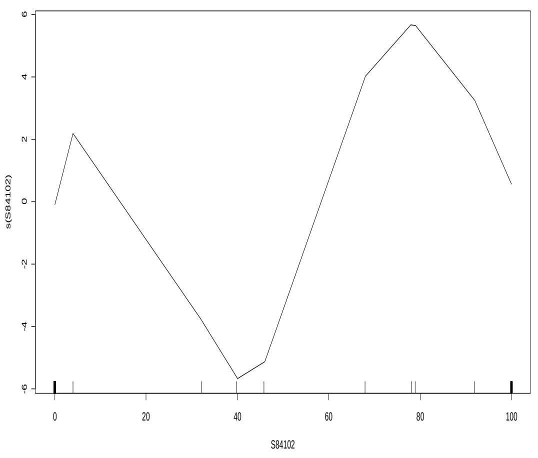


Figure A-35. GAM plot for S84102 for *C. nobile* GAM.

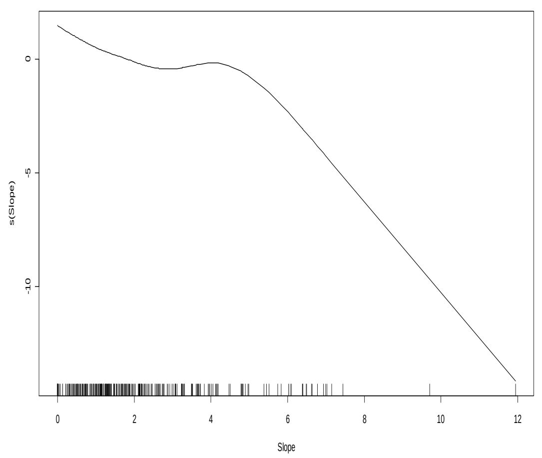


Figure A-36. GAM plot for Slope for *C. nobile* GAM.

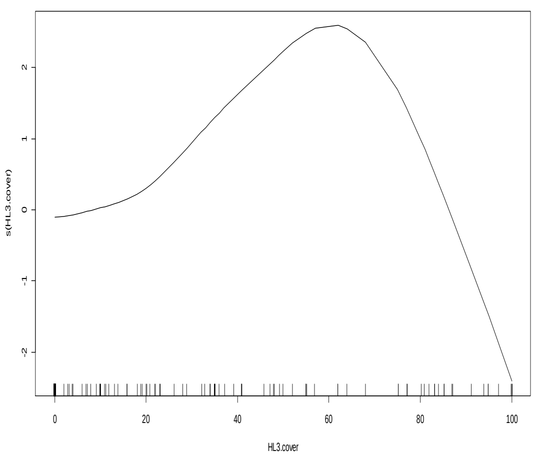


Figure A-37. GAM plot for HL3.cover for *C. nobile* GAM.

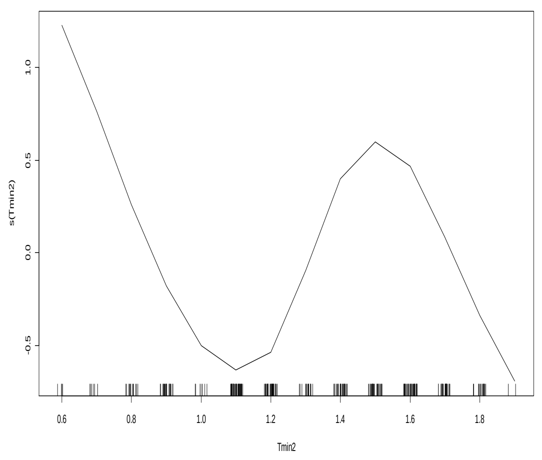


Figure A-38. GAM plot for Tmin2 for *C. nobile* GAM.

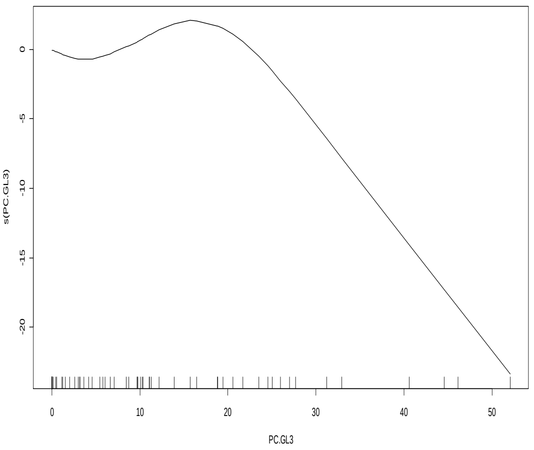


Figure A-39. GAM plot for PC.GL3 for *C. nobile* GAM.

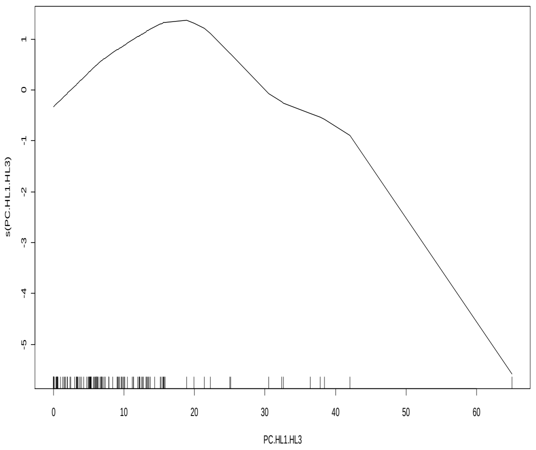


Figure A-40. GAM plot for PC.HL1.HL3 for *C. nobile* GAM.

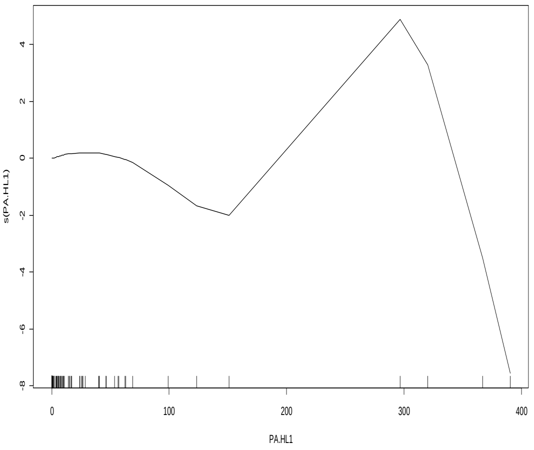


Figure A-41. GAM plot for PA.HL1 for *C. nobile* GAM.

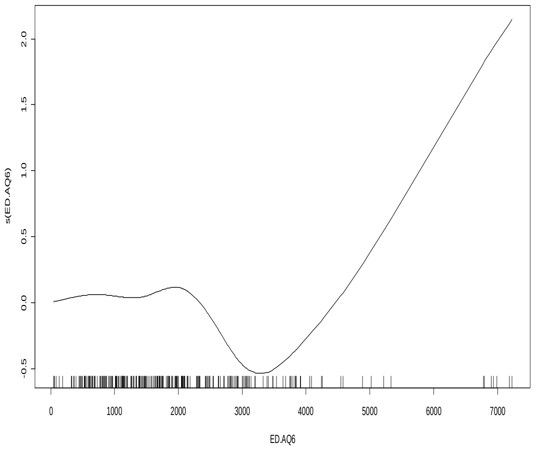


Figure A-42. GAM plot for ED.AQ6 for *C. nobile* GAM.

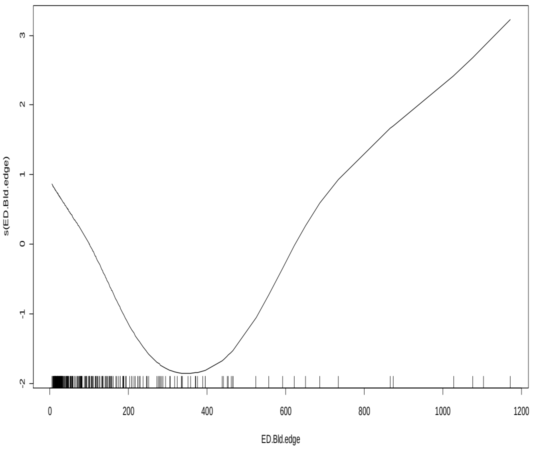


Figure A-43. GAM plot for ED.Bld.edge for *C. nobile* GAM.

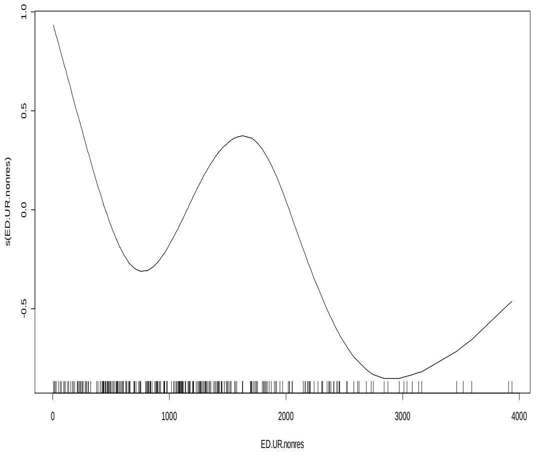


Figure A-44. GAM plot for ED.UR.nonres for *C. nobile* GAM.

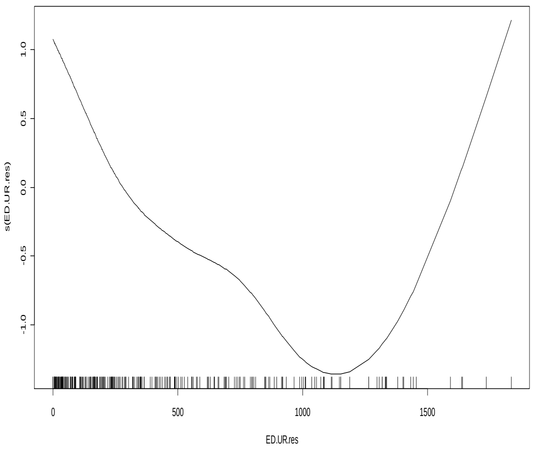


Figure A-45. GAM plot for ED.UR.res for *C. nobile* GAM.

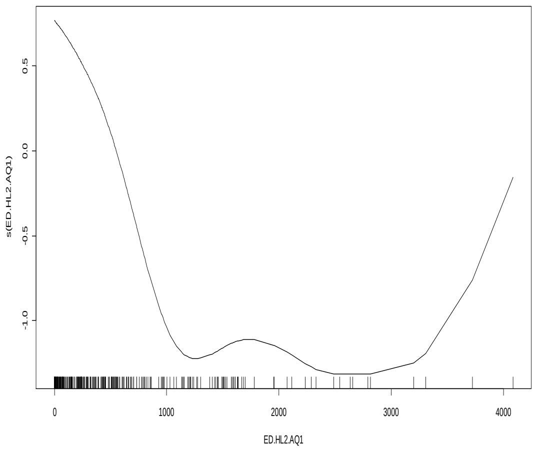


Figure A-46. GAM plot for ED.HL2.AQ1 for *C. nobile* GAM.

The graphs for the seven variables included in the 7-variable model suggested that most of the presence sites for *C. nobile* occur in approximately 10 – 90% cover of HL3 (dry heath/acid grassland mosaic) and absences occur in sites with less than 10% cover or more than 90% cover. This confirms the association of *C. nobile* with dry heath/acid grassland (Plantlife, 2001; 2007a), requiring at least 10% cover within the site. There was not a clear relationship for Tmin2 (minimum February temperature), with two bands of temperatures in which the presences occur. This is likely to just be by chance as the range of temperature values was very small and a true relationship would more likely be linear. For patch compactness of improved grassland (PC.GL3), the presences occurred mostly at sites with a value of 10 to approximately 35, although this is almost a linear relationship. There was also an association (which was almost linear) with lower values of patch compactness for dry heath and acid grassland mosaic (PC.HL1.HL3). A lot of the heathland sites may be quite patchy because they frequently occur in a mosaic with other habitats, such as wet heath. *C. nobile* also tends to occur on the edges of heaths (such as on road verges), where there is likely to be less dense cover of ericaceous species. This was also show by the occurrence of presence records within approximately 500m of non-residential urban areas (ED.UR.nonres), or between approximately 1200m and 2000m away. This relationship looks like it could potentially be linear if the sample size was larger. There was also an association of some of the presence sites with close proximity (less than 300m away) to residential areas (ED.UR.res), which may reflect occurrence on New Forest lawns occurring near settlement edges (Tubbs,

2001) and be related to the association with improved grassland (GL3), which was frequently associated with residential areas (see section 3.3.1.1). Finally, presence sites were shown to be either very close (within 100 m) to broadleaved woodland edge (ED.Bld.edge) or more than 600m from it. *C. nobile* is not generally associated with woodland, but can occur in woodland glades (Plantlife, 2001; 2007a).

For the additional variables included in the 12-variable model, the slope response was almost linear and showed that *C. nobile* presences sites all occurred on flat sites, whereas the absences ranged from the flatter to the slightly steeper (although still not very steep) values. This fits with the expectation that *C. nobile* tends to occur on flatter sites because they are more likely to be prone to winter flooding, which it favours (Plantlife, 2001). The plot for Euclidean distance to large ponds (ED.AQ6) was also nearly linear, and showed presence records mainly occurring with 2500m of large ponds, although most of the records were within this range. All of the presence records occurred within 700m of wet heath/mire (ED.HL2.AQ1), which shows the relationship with acid heath (all of the heathland types were highly correlated). In terms of patch area of dry heath (PA.HL1), presence sites generally occurred in patches less than 50 hectares or 200-300 hectares, although most of the records (presence or absence) occurred in patches less than 70 hectares. Although the plot for soil type 84102 suggests that presence records either occurred in sites with less than 10% or more than 60% of this soil type, most sites had either 0% or 100% cover of this soil type, making it difficult for the model to fit a relationship.

The GAM plots for the 3 smoothed variables selected by backwards stepwise GAM for *P. punctata* are shown in Figures A-47 to A-49 below.

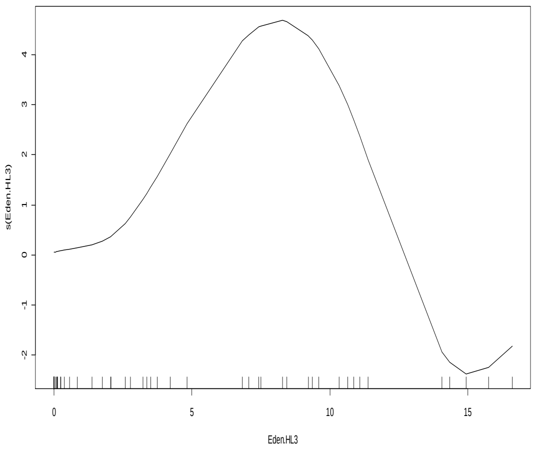


Figure A-47. GAM plot for Eden.HL3 for *P. punctata* GAM.

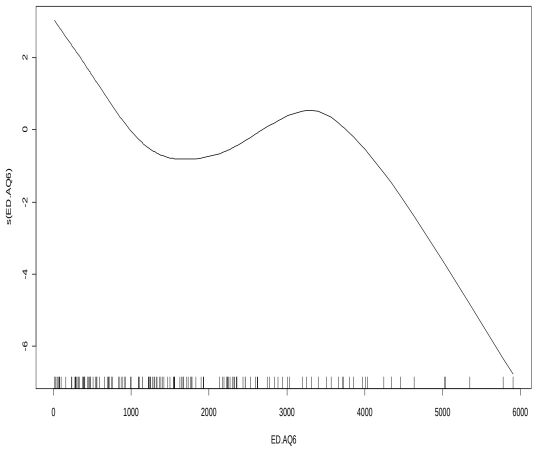


Figure A-48. GAM plot for ED.AQ6 for *P. punctata* GAM.

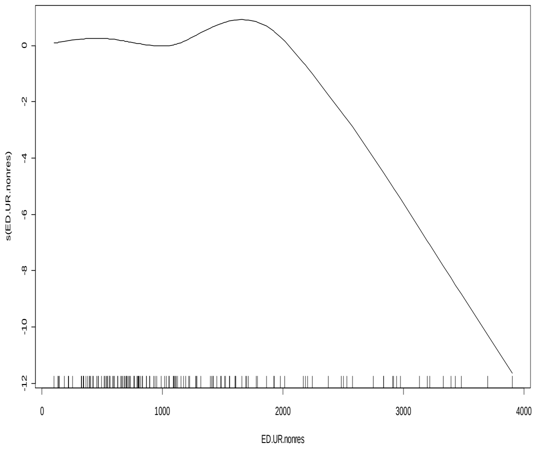


Figure A-49. GAM plot for ED.UR.nonres for *P. punctata* GAM.

Two of the graphs of smoothed variables (ED.AQ6 and ED.UR.nonres) showed an almost linear relationship of presence records for *P. punctata* tending to occur closer to large ponds and non-residential urban areas (such as roads). This may reflect recording bias or the more frequent occurrence of ponies near these habitats/land-uses. Presence locations were also more likely to be in sites with lower edge density values of dry heath/acid grassland mosaic, suggesting that it frequently occurs on heathland habitats with a lot of edge.

Appendix 19. ROC plots for final GLMs and GAMs

ROC plots for GLMs:

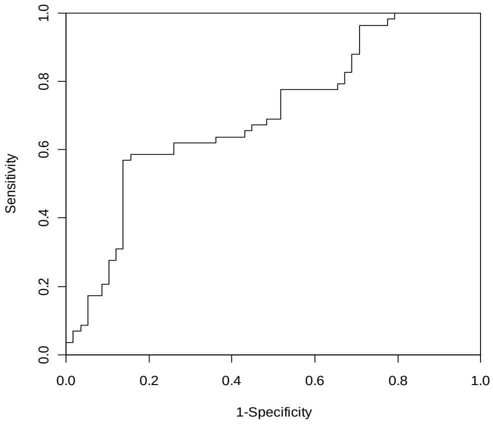


Figure A-50. ROC plot for *C. nobile* final GLM (AUC = 0.700).

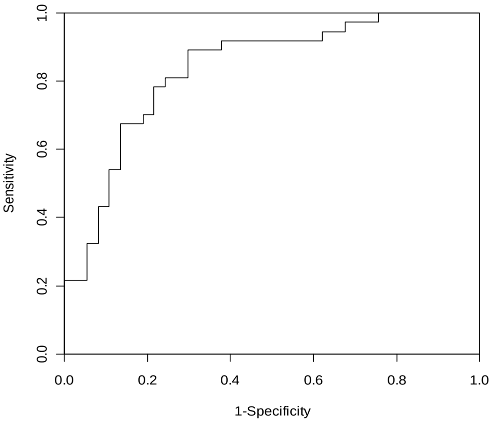


Figure A-51. ROC plot for *G. constrictum* final GLM (AUC = 0.836).

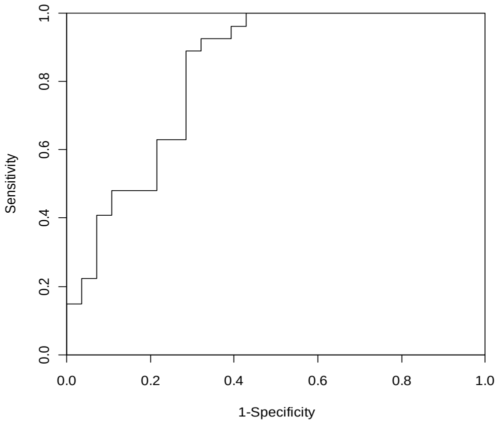


Figure A-52. ROC plot for *G. illyricus* final GLM (AUC = 0.828).

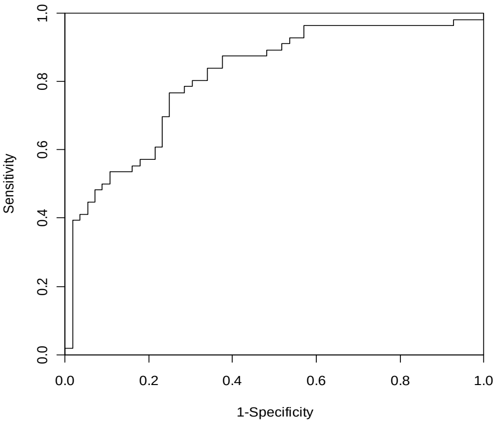


Figure A-53. ROC plot for *H. semele* final GLM (AUC = 0.812).

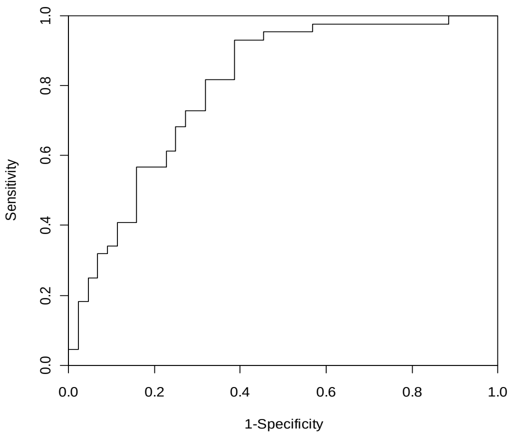


Figure A-54. ROC plot for *N. sylvestris* (9) final GLM (AUC = 0.798).

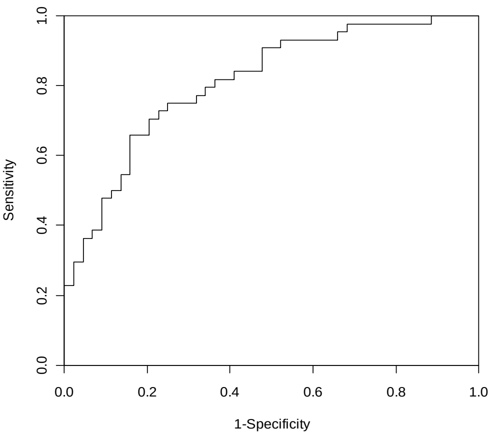


Figure A-55. ROC plot for *N. sylvestris* (7) final GLM (AUC = 0.810).

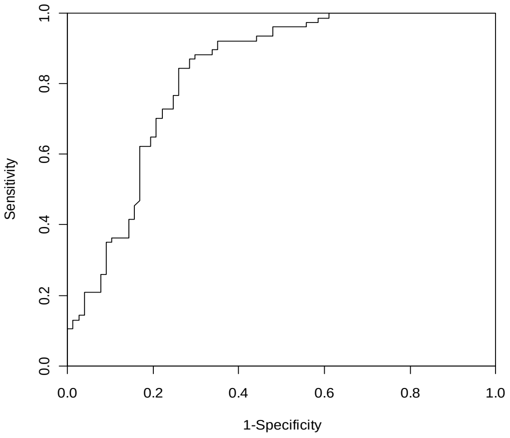


Figure A-56. ROC plot for *P.argus* final GLM (AUC = 0.824).

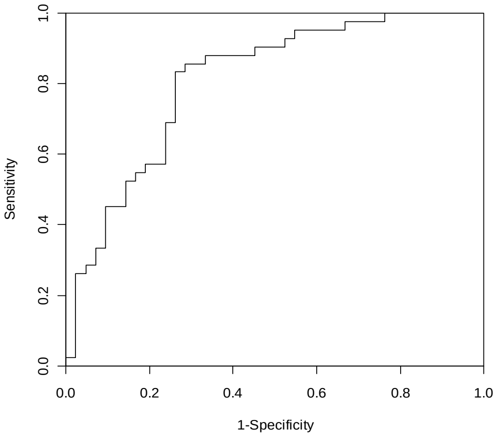


Figure A-57. ROC plot for *P. globulifera* (7) final GLM (AUC = 0.809).

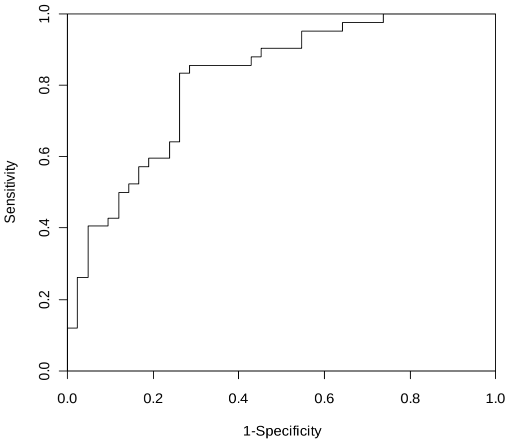


Figure A-58. ROC plot for *P. globulifera* (6) final GLM (AUC = 0.815).

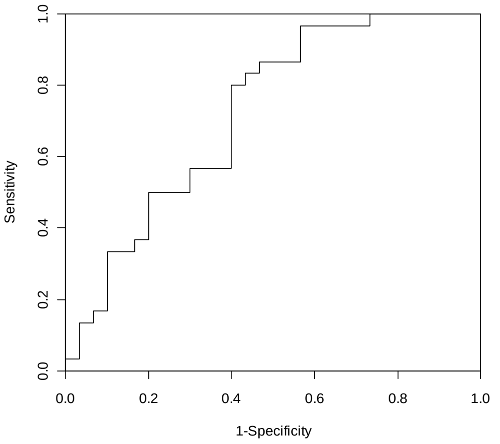


Figure A-59. ROC plot for *P.punctata* final GLM (AUC = 0.721).

ROC plots for GAMs:

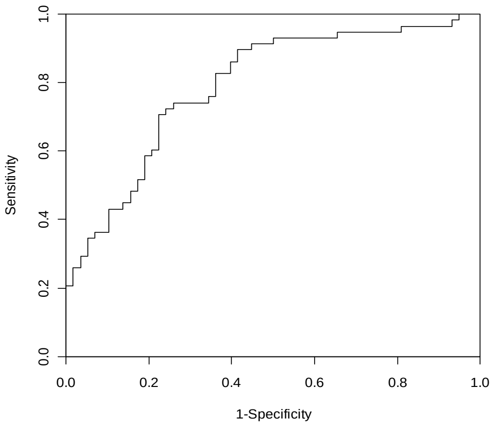


Figure A-60. ROC plot for *C. nobile* final GAM (7 variables) (AUC = 0.792).

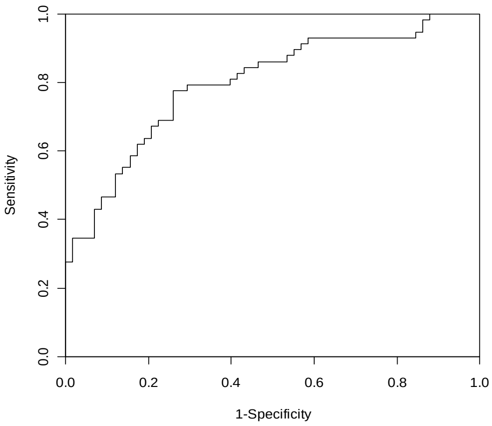


Figure A-61. ROC plot for *C. nobile* GAM (12 variables) (AUC = 0.799).

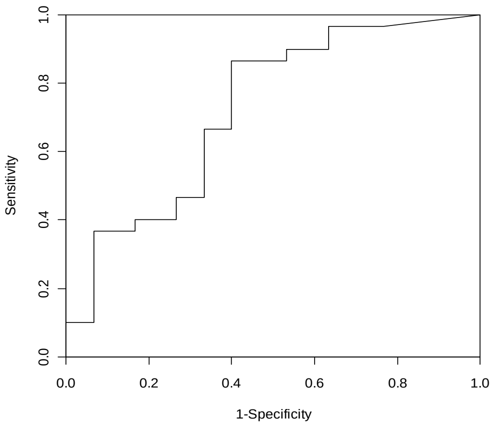


Figure A-62. ROC plot for *P.punctata* final GAM (AUC = 0.723).

Appendix 20. Tables of variables for BBNs for each species

The tables below include information gathered from the literature, as well as extracts from the expert interviews.

Wild chamomile (*C. nobile*)

Variable	Detail	References
Habitat type Presence of suitable grassland	Found in sandy heaths, sports pitches, open glades in light woodland, grazed pastures and commons, roadside and track verges. Also found on coastal cliff tops/maritime grassland where wind exposure and salt spray keep the sward short.	Plantlife (2001), Winship and Chatters (1994)
	Mesotrophic grassland.	Porley and McDonnell (1997)
	Most frequently found in herb-rich, moderately acidic, lowland grassland.	Winship and Chatters (1994)
	On cricket and football pitches where regular cutting and rolling mimic grazing pressure.	Winship and Chatters (1994)
	Occurs commonly in the damper parts of New Forest lawns, notably those at settlement edges.	Tubbs (2001)
	Sandy commons and pastures. Also in coastal grassland and on cliffs, where exposure and trampling maintain short sward.	Killick (2002)
	In short, grazed turf on dry, grassy heaths, commons, old pastures and roadsides.	Brewis *et al*. (1996)
	Acid grassland. Size of patch not too important – but larger the better in the long-term.	A. Byfield
	Circum-neutral to acidic grassland. Not in chalk grassland. In Headon-bed grasslands, shell-rich clays. In glades, road edges. Also found on grassy tussocks in mires.	C. Chatters
Canopy cover/Light	Sunny, open places - Such as sandy heaths, cricket pitches, coastal cliff-tops or open glades in light woodland	Plantlife (2007a) Plantlife (2001)

	Light-loving plant rarely found where relative illumination in summer is less than 40%. Chamomile cannot grow in shady conditions. Shading by taller plants is detrimental to Chamomile populations. It will not grow in the shade of overhanging trees or bushes, nor will it tolerate more competitive plants such as grasses or bracken. (The spread of Rhododendron (*Rhododendron ponticum*) is now a major concern in many heaths and commons, as it shades native flora, such as Chamomile, which cannot compete and soon dies. The same effects may be attributed to Bracken (*Pteridium aquilinum*)). Needs to be completely open. No canopy cover. Not in woodland. Will tolerate a little bit of dappled light partial canopy cover, short-term. But shade tends to draw up vegetation to be taller, so that could be a problem. However, grazing may counter that. Highly illuminated. (Sometimes in glades).	Plantatt (2004) Plantlife (2001) Winship (1993) G. Read A. Byfield C. Chatters
Bare ground (and disturbance)	Patches of bare ground – small-scale disturbance by animals, people or vehicles. Does not necessarily require bare ground, but is very poor competitor. Requires bare ground, open. Bare ground not that important, but would require bare ground for seedling establishment/colonisation – helps it to spread. So does not need bare ground all the time. Can survive in 100% cover lawns. Presence of grazing probably more important.	Winship and Chatters (1994) G. Read C. Chatters A. Byfield

Vegetation height	Mown or grazed grassland. Short, grazed turf. Short grassy places. Short turf is the most important variable. Never found in more than 2-3 inches height – may compete for a few years then disappears. Shorter the better. Short turf is the most important variable. Shorter the better (as long as it is still a grassland). 1 – 4 cm probably best, 5 – 7 cm not as good. In more than 10 cm it would fairly quickly die out. Not sure exactly how long it would survive in tall grass. Long continuity of short turf/grazing is important. Once lost does not easily recolonise. Very short herbage. Grass taller than Chamomile flowers not suitable. Turf height and openness important.	Plantlife (2001) Brewis *et al*. (1996) Stace (1997) G. Read A. Byfield C. Chatters
Grazing/Mowing	Adapted to life with grazing. Thrives on being trampled. Grazing keeps down taller plants and creates areas of open ground which it can colonise. Horses/ponies – suitable, Deer – suitable if not too intensive, Waterfowl – suitable if not too intensive, Sheep and Rabbits – not as good (because they create a very close cropped sward). Recommended that stocking levels reduced in September and October to allow flowering and seed- set. Closely grazed.	Winship and Chatters (1994) Brewis *et al*. (1996) Plantlife (2001) Winship and

	Flourishes in relatively high nutrient situations providing that sufficient grazing, mowing or trampling occurs to prevent grasses and taller plants from dominating the sward.	Chatters (1994) Winship and Chatters (1994)
	Grazing is an important factor for the survival of Chamomile. Animals move around choosing fresher pastures when an old stand is exhausted. This enables Chamomile to recover and perhaps to flower and set seed, whereas mowing seldom gives the plant this option. The large herbivores, particularly mixed grazing by cows and ponies, produce a very even sward, with poached areas around watering spots and fences. Poaching can have a beneficial effect, by breaking up tight swards and allowing the trampling-resistant Chamomile to colonise. Grazing by sheep and rabbits also produces short sward but does not generate the poached ground, although disturbance by rabbits burrowing, moles and earthworms may produce the same bare ground. Key grazing times are during the growing season, April to September. If allowed to flower and set seed, reproduction will ensure genetic conservation. Even under heavy grazing pressure such as on the New Forest lawns, a significant proportion of plants are found to flower.	Winship (1998)
	Continual grazing (all seasons) is important. Does not matter what animal. Mowing pretty much as good. Cannot be over-grazed. Keeps competition at bay. Unenclosed grazing is really important for Chamomile. Helps to spread it. Roots easily – ½ inch fragment can form a new plant. Grazing is good because can get poached ground. A little bit of poached ground is good – allows it to grow quickly – aggressive spreading phase. Has become rare in Dorset due to decline of grazing, but still on Corfe Common, which is grazed.	A. Byfield
	Requires highly stressed environment. Grazing, poaching. Can be recreated through mowing	C. Chatters

	(not as good). Very hard grazed. Does not matter what animal it is, just as long as the sward is kept short. Grazing is probably slightly better than mowing – closer. Too much grazing is not a problem. Mowing no more than once a week and less frequently around September and October to allow flowering and seed-set. On sports pitches the lack of soil disturbance restricts opportunist Chamomile plants and produces a different, less diverse sward. In general, the cricket pitches/sports fields in which Chamomile is found are cut and rolled weekly. Pressure from rolling ensures vegetative spread and is a necessary inclusion for management. Removal of lawn cuttings is preferable. Chamomile inhabits nutrient-rich soils and it is possible that decaying cuttings may have the same effect as dung, but the Chamomile must not be smothered. (Also see vegetation height)	 G. Read Plantlife (2001) Winship (1993) Winship (1998)
Soil type	A plant of dry, sandy or gleyed soils (but common requirement is to be seasonally wet). The sun also often bakes sandy soil hard in the summer, rendering it slowly permeable thus leading to temporary waterlogging in winter. Sandy soils. Sandy soils. In Headon-bed grasslands, shell-rich clays. Claggy clays, more than on sands. If on sands then on less freely drained sands.	Winship and Chatters (1994) Plantlife (2001) Stace (1997) C. Chatters

	Organic, loamy soils – more likely to hold moisture and nutrients. Not on bare sandy soils?	A. Byfield
Soil fertility (nitrogen)	Intermediate fertility (5).	Hill *et al*. (2004)
	Flourishes in relatively high nutrient situations providing that sufficient grazing, mowing or trampling occurs to prevent grasses and taller plants from dominating the sward.	Winship and Chatters (1994)
	Can grow in fairly rich soils, but if artificial fertilisers are heavily applied to pastures, these encourage vigorous, nitrogen-hungry plants which may out-compete it.	Plantlife (2001)
	Can occur in relatively fertile acid grasslands - interplay between grazing and fertility – can have high fertility as long as have grazing. Possible eutrophication from dung and urine?	A. Byfield
	On very productive soils but very stressed – vigorous grasses can not compete. Good competitor when in stressed environment. (Stress is natural – used to be large herbivores....)	C. Chatters
Soil pH	Indicator of moderately acid soils, only occasionally found on very acid or on neutral to basic soils (5).	Hill *et al*. (2004)
	Moderately acidic.	Winship and Chatters (1994)
	Moderately acidic.	Killick (2002)
	Confined to acid soil.	Brewis *et al*. (1996)
	Moderately acidic, but not extreme.	A. Byfield

	Around pH 5-7, but not particularly fussy.	C. Chatters
Soil moisture	Between dampness indicator (mainly on constant moist or damp, but not on wet soils) and wet-site indicator (often on water-saturated, badly aerated soils). – 8	Hill *et al*. (2004)
	Requires a degree of moisture. Found in hollows in paths, (winter) flooded lawns as well as non-wet grasslands. Will not grow in dry, parched grasslands. Has a shallow rooting system – vulnerable to drought. Long, hot dry summer might kill it off.	A. Byfield
	Tolerant of extreme parching (but short term). Occurs in slightest dips in ground. Can also occur in seasonal pond edges and in parched areas (but wetter in winter) – for short time. Fairly tolerant.	C. Chatters
	Can tolerate a fair amount of dry or wet. Not too fussy!	G. Read
Areas prone to winter flooding	A common requirement is to be seasonally wet, usually in winter.	Winship and Chatters (1994)
	Seasonally wet grassland.	Killick (2002)
	The sun also often bakes sandy soil hard in the summer, rendering it slowly permeable thus leading to temporary waterlogging in winter.	Winship and Chatters (1994)
	Can survive quite long periods of winter flooding – a few months. Not too deep – partly because it will take longer to dry out.	A. Byfield
Absence of herbicides or fertilisers	Grassland which has not been fertilised or sprayed.	Winship and Chatters (1994)
	Can grow in fairly rich soils, but if artificial fertilisers are heavily applied to pastures, these encourage vigorous, nitrogen-hungry plants which may out-compete it.	Plantlife (2001)

	Chamomile shows some resistance to certain herbicides. However, herbicide and pesticide spraying is not recommended.	Winship (1998)

Table A-26. Variables identified from the literature and from interviews with experts for BBN for *C. nobile*.

Slender marsh-bedstraw (*Galium constrictum*)

Variable	**Detail**	**References**
Habitat type	Marshy places, ditches and pondsides. Around margins of ponds which dry out in summer, as well as on New Forest ‘lawns’ and in track ruts, and locally in marl-pits and ditches in water-meadows. In grassy heaths, wet flood-hollows, drainage-channels, marshy pond margins, gravel-pits and muddy track-ruts. In ponds and on edges. Seasonal ponds. Edge of village greens. Seasonal hollows in marshy areas. Not seen in it flowing water, but not to say it is not there. Grassy edges of ponds, grassy edges of heathlands. Not sure how long-term ditch habitats are.	Stace (1997) Meek (2002) Brewis *et al*. (1996) C. Chatters G. Read
Ground moisture	Wet-site indicator, often on water-saturated, badly aerated soils (9). Seasonal ponds. Can survive being parched for part of year. But wet in winter. Seasonal hollows in marshy areas. Damp, not necessarily in water, but edges of ponds.	Hill *et al*. (2004) C. Chatters G. Read
Canopy cover	Light-loving plant rarely found where relative illumination in summer is less than 40% (8). Found in full sun or in open shrubs – will tolerate a degree of shade. Most likely to be found in open or partial shade, but not in closed canopy cover.	Hill *et al*. (2004) C. Chatters G. Read

Ground vegetation height	Grows/flowers in a few inches or a few feet of grass. Height of grass does not seem to be important, but probably cannot compete with dense vegetation.	C. Chatters
Bare ground/ disturbance	Requires open, bare ground to germinate and propagate, as it is a poor competitor. However, it is not very robust. Needs bare ground to germinate. Survives extreme grazing stress. Relatively palatable – grows back.	A. Byfield C. Chatters
Soil type	Often on water-saturated, badly aerated soils. (Also see soil pH and soil fertility below).	Hill *et al*. (2004)
Soil pH/acidity	Acidity indicator, mainly on acid soils, but exceptionally also on nearly neutral ones (3). Mildly acidic clays. Circum-neutral. Middling base-rich end of C-neutral base-rich.	Hill *et al*. (2004) A. Byfield C. Chatters
Soil fertility	Between 1 (indicator of extremely infertile sites) and 3 (Indicator of more or less infertile sites) Headon beds, white clays (southern New Forest). Village greeny soil. Base-rich clays. Gravel matrices. Slightly nutrient-enriched.	Plantatt (2004) A. Byfield C. Chatters
Water pH	Not taken pH of water for *G. constrictum* habitats, but associate it with circum-neutral with a distinct lean towards base enrichment.	C. Chatters

	(Carried out pH monitoring of a pond at Godshill - found a range of pH readings over the course of the day and the season. Think this is where in poorly buffered ponds the seasonal concentrations of dung and urine, together with the creation of dissolved oxygen by photosynthesis of vegetation on the pond floor, introduces bases that then shift the pH).	

Table A-27. Variables identified from the literature and from interviews with experts for BBN for *G. constrictum*.

Wild gladiolus (*Gladiolus illyricus*)

Variable	**Detail**	**References**
Habitat type/ Presence of bracken	Bracken scrub in the New Forest.	Stace (1997)
	All the sites were contained within bracken stands. Bracken dominated acid grass heath. Frequently on the boundary of heath and bog areas.	Stokes (1987)
	Found on grass-heaths, usually in association with *Pteridium*, which may afford the plant some protection from grazing.	Taylor (2002)
	On grass-heaths, beneath the canopy of bracken.	Brewis *et al*. (1996)
	Factors influencing the recent decline in numbers include scrub and heather encroachment and, in at least some sites, past attempts to eradicate bracken to improve grazing and, more recently, indiscriminate heath-burning, have been directly responsible for losses.	Brewis *et al*. (1996)
	Requires bracken to be present. Requires bracken with grass underneath. – low, quite thin grassland (but cannot say for sure).	A. Byfield
	Open bracken communities. Grassy, herby underneath (tends to be short, 10-15 cm, but can be 20-25 cm if patchy with shorter patches). Glades. Grazing moves in early mid-summer. So may be longer at certain times of year.	C. Chatters
	Does not require a certain amount of bracken – can be in small clumps of bracken – is the sheltering effect of bracken rather than bracken itself.	C. Chatters
Bracken density	Preferably open bracken (less vigorous), but some occurrence in closed bracken. None in sparse bracken. Bracken provides protection from grazing. It needs to be dense enough for protection, but	Stokes (1987)

	not too dense so that it ‘smothers’ the Gladioli plants.	
	Flower heads flourish wherever the bracken canopy had been damaged by frost, and so failed to close into a dense stand by mid-June. In these ‘open bracken’ areas, seed set is substantially higher, because pollinating bees and large skipper butterflies more readily find the flowers.	Stokes (2000)
	Records diminished since sessation of bracken-cutting, but... Can persist in very dense bracken (>2m).	Toone (2005)
	Not under dense bracken (requires about 60-80% cover of bracken). Needs to provide protection from grazing but still allow quite a bit of light/sun through to provide warmth for the bulb.	A. Byfield
	Needs light but protection from the bracken. Hard frost later in year is good because it keeps the bracken more open.	C. Chatters
	Not found in dense bracken – more on edges. Must have light.	G. Read
Bracken litter depth	Decrease in number of Gladioli per m^2 with increase in the thickness of the bracken litter layer (If the litter is too deep it ‘suffocates’ the Gladioli and makes it more difficult for seeds and roots to penetrate). (Open canopy bracken sites have a thinner litter layer). Stokes (1987) used the following categories: 0 – 5 cm (14.44 +/- 12 plants), 5 – 10 cm (2.8 +/- 2 plants), 10+ cm (0.5 +/- 0.5).	Stokes (1987)
	There is some evidence that many *G. illyricus* colonies have declined recently, and it may be that litter accumulation within bracken stands is having an adverse effect on this (and other plants, shade).	Tubbs (2001)
	Ideally no (bracken) litter – swamps it – will not germinate. Perennial – cannot survive long like	A. Byfield

	this. Mature plant – some litter ok. May survive under ~1 inch for ~1 year. Once litter forms, keeps getting thicker. Ideally no bracken litter. Can tolerate some litter, but weakens plants. Litter not too deep.	 C. Chatters G. Read
Grazing Bracken cutting	The survival of Gladiolus in the New Forest is likely to be threatened if a decision is made to increase or decrease the grazing levels. This could affect the balance in the bracken sites where Gladioli occur and lead to the smothering or overgrazing of the plants. Any decisions to attempt biological control of the bracken either in the New Forest or nationwide could also have a dramatic effect on the Gladioli population, reducing their protection from grazing and increasing the competition from other plants. On a smaller scale, repeated cutting of the bracken at an unsuitable period in time (e.g. during flowering) would lead to the eradication of the species from that area. Grazing is important because if the bracken communities were not grazed then they would go to oak. Need grazing to keep bracken communities, but not too much otherwise *G. illyricus* gets eaten. Grazers move over different areas. Cutting of bracken leads to thinner bracken cover and reduces litter. Grazing and trampling of bracken helps to keep it open. Pigs are good as they root around in bracken and smash it up, creating bare ground. *G. illyricus* needs bare ground for non-vegetative reproduction. Bracken cutting (in past) helped maintain *G. illyricus* habitat (as long as cutting took place in late summer (cutting in early spring wipes it out).	Stokes (1987) A. Byfield A. Byfield C. Chatters C. Chatters

Canopy cover	Semi-shade plant, rarely in full light, but generally with more than 10% relative illumination when trees are in leaf. Must have light – not found in woodlands or on edges, not found under trees. But dappled light – i.e. under bracken. Never found under trees. (Also see notes on distance to woodland). Open. Needs light.	Hill *et al*. (2004) G. Read A. Byfield C. Chatters
Soil type	Fine sandy loam (a brown earth soil), relatively high in its clay and silt fractions compared to other soils in the Forest, but low in nutrients. Stoneless and hence does not prohibit the downwards movement of the developing corms. All sites recorded on Barton sand geology. 75% of Gladioli sites were on Barton sand, 24% on Barton clay, 0.5% on Headon beds and 0.5% on silt wash from upper slopes. Brown-earth soils. On brown ranker soils, the less acid sands and clays. Brown-earth, not too fertile, acid soils. Relatively productive soils, reasonably organic. Moist, freely drained. Headon beds.	Stokes (1987) Taylor (2002) Brewis *et al*. (1996) A. Byfield C. Chatters
Soil fertility	Indicator of more or less infertile soils (3). Soils low in nutrients (% organic content ranged from 4.1 to 8.6). Provide the necessary nutrients for survival without allowing competition to become too great.	Hill *et al*. (2004) Stokes (1987)

	Relatively unpodsolised soil. (Presence of bracken canopy somewhat offsets podsolisation by retaining a high moisture level below the canopy and by prohibiting leaching). Relatively productive soils, reasonably organic. Not too fertile.	Stokes (1987) C. Chatters A. Byfield
Soil pH	Indicator of moderately acid soils, only occasionally found on very acid or on neutral to basic soils (5). Less acidic (pH ranged from 4.46 to 4.82). Acidic, brown-earth soils. Less acid sands and clays. Acid soils.	Hill *et al*. (2004) Stokes (1987) Taylor (2002) Brewis *et al*. (1996) A. Byfield
Soil/ground moisture	Between 3 (Dry-site indicator; more often found on dry ground than in moist places) and 5 (Moist-site indicator (mainly on fresh soils of average dampness)). Damp enough to prohibit bracken from becoming dominant, without being so wet that the corms rot. Moist, freely drained. Not in flood plains or flushes (needs more aerated). Not in really parched bracken sites.	Hill *et al*. (2004) Stokes (1987) C. Chatters

	Loam is relatively moisture retentive. Holds moisture and soil nutrients. Crumbly – lot of air (air porosity 25-30%?). Not sure how deep bulb is. Reasonably dry, well drained. Would not tolerate too much damp. Can tolerate quite a bit of dryness.	A. Byfield G. Read
Distance to water	Most sites were close to water (e.g. bogs, streams etc.) (within 100 m). In combination with the fact that most of the sites were also on a slope, P. Farres (personal communication, cited in Stokes (1987)) suggests that these sites are flushing areas through which water from the upper slopes moves. The constant turnover of water and nutrients ensures that the soils do not become heavily podsolised, while remaining damp throughout the year.	Stokes (1987)
Slope	Slight slope. Slope of sites ranged from flat to 7 degrees. Flat or gentle slopes. Not on steep, but Acres Down relatively steep. Slope has no effect. Flat or gentle, but no steep slopes really in the New Forest, so do not know whether it can occur on steeper slopes.	Stokes (1987) C. Chatters A. Byfield G. Read
Distance to woodland	Most of the colonies are found in bracken alongside a woodland edge, although colonies do exist up to one kilometre from the nearest woodland. These are often in places for which there is cartographic or other documentary evidence that woodland has been lost since the 18th century. Often found where there have formerly been woodlands. Leads to the kind of species-rich bracken communities that *G. illyricus* likes. More that it likes the fertile soils resulting from woodlands.	Stokes (1987) Tubbs (2001) A. Byfield

	Community associates, such as wood anemonie. Bracken acts as surrogate trees for shade. Not in woodland – just the soils it provides. Associated with woodland communities – Bluebell (*Hyacinthoides non-scripta*) etc., and Soft-leaved sedge (*Carex montana*).	C. Chatters
Pollinator	Pollinated chiefly by the Large Skipper (*Ochlodes venata*) and occasionally other insects (Gate Keeper, 2 bee species). Requirements of pollinator?	Stokes (1987)
	More grass (open sites better for *G. illyricus*) – better for grass feeding butterflies like the large skipper to survive – higher pollination in open areas (due to more insects and greater ease with which insects find plants).	Stokes (1987)
	Reproduces primarily by offsets, as flowering and seed production appear limited.	Taylor (2002)

Table A-28. Variables identified from the literature and from interviews with experts for BBN for *G. illyricus*.

Grayling (*Hipparchia semele*)

Variable	**Detail**	**References**
Habitat type	Arid places.	Oates *et al*. (2000)
	Unimproved grassland.	Green (2000)
	Almost any dry New Forest heath can hold a large population if the right habitat occurs. Occasionally, seen in unusual habitat, e.g. built-up area in Southampton. Such strong status within the New Forest that small colonies can be found along sunny rides with suitable vegetation in Inclosures, whilst individuals can be seen just about anywhere.	Oates *et al*. (2000)
	Generally, many colonies are coastal, on dunes, saltmarsh, undercliffs and clifftops. Inland, found on dry heathland, calcareous grassland, old quarries, earthworks, derelict industrial sites such as old spoil heaps, and in a few areas in open woodland on stony ground.	Asher *et al*. (2001)
	Occurs on dry heath – preferably really dry. Not in wet heath. H3 NVC community – Bristle Bent (*Agrostis curtisii*), Gorse (BB and often associated – found in similar habitats). HBIC habitat layers - HL1, HL3 – not too bothered about amount of heather – can be in heath-dominated areas, but in patches where more open and bare ground. (Not in GL11 (acid grass only) – dense/homogeneous and probably not much bare ground). Does not require a particular area/quantity of heath. But larger area likely to be less homogenous/more variety and can last longer term.	A. Barker
	Presence of heath mainly important for presence of the grass species (bristle bent).	P. Brock
Heathland growth stage	Early pioneer stage, particularly recently burnt (regular burning important to keep in pioneer stage, and grazing important). Does not require a large area of short heath – can be just firebreak areas on edge of taller heath,	A. Barker

	verges of shorter heath next to taller heath, short and mown, paths, tracks. Lays eggs on woody debris (near to food plant) – dry wood – does not get eaten and warms quickly. Often roost on the ground (helps to warm up), but if bad weather they move onto vegetation. (Also see ‘heathland management’ below)	
Heathland management/ burning	On New Forest heaths *H. semele* can be abundant where heathers are regenerating after burns, whereas in adjoining dense heather it may be missing or present in sparse numbers. Very much a species of managed heathland and is absent, or in the process of dying out, from neglected heaths. The programme of burning heathland at regular intervals obviously causes the species to move locally, as do unplanned heath fires.	Oates *et al*. (2000) Oates *et al*. (2000)
Grazing/ Disturbance	On the NE Hampshire heaths, which are ungrazed by stock, it is heavily dependent on disturbance, particularly by MoD machinery. Regular burning and grazing is required to create the favoured pioneer stage.	Oates *et al*. (2000) A. Barker
Presence of food plant/plants for ovipositing	In the New Forest, the species is strongly associated with bristle bent (*Agrostis curtisii*) but females tend to select isolated clumps for ovipositing, especially where two or three tussocks grow close together amongst bare ground. Solid carpets of bristle bent are mostly ignored, which is just as well because such areas are heavily grazed. Path-side tussocks of purple moor-grass (*Molinia caerulea*) and wavy hair-grass (*Deschampsia flexuosa*) are also used. In hot weather, adults will readily take nectar from flowers, especially bramble and heathers. *Buddleia* is also visited, whilst at Sandy Point they have been seen on sea-holly and at Blackbushe airport the species was seen taking nectar from marjoram along old runways.	Oates *et al*. (2000) Oates *et al*. (2000)

	The larvae feed on several grasses including bristle bent (*Agrostis cutisii*), sheep's fescue (*Festuca ovina*) and marram (*Ammophila arenaria).* Foodplants (general): bristle bent, barly hair-grass, tufted hair-grass, fescues (including sheep's fescue and red fescue) and marram. Tussocky sheep's fescue (*Festuca ovina*) – for larvae, and bristle bent (*Agrostis curtisii*). Does not need a particular area/quantity. But larger area likely to be less homogenous/more variety and can last longer term. Does not require a particularly large area or percentage cover of grass.	Green (2000) Bailey *et al*. (1989) A. Barker P. Brock
Presence of tussocks	Females tend to select isolated clumps for ovipositing, especially where two or three tussocks grow close together amongst bare ground. Solid carpets of bristle bent are mostly ignored. They hibernate in tussocks during the winter. Not solid grass – needs tussocks and bare ground around them. Does not require a certain quantity of tussocks (although not completely sure). Just not homogenous or solid grass. Tussocks are not too important.	Oates *et al*. (2000) Asher *et al*. (2001) A. Barker P. Brock
Presence of bare ground Open (Canopy cover)	In grass tussocks amongst bare ground. Bare pockets and along paths where vegetation is worn or heavily grazed. Plenty of bare ground in open positions. Around 10% bare ground (but difficult to estimate). Bare ground is partly an indicator of the type of community. As gets more mature the grass and heath closes in.	Oates *et al*. (2000) Asher *et al*. (2001) A. Barker

	Does not require a huge amount of bare ground – just areas where the grass is not as dense.	P. Brock
Soil type/moisture	Associated with unimproved grassland on many soil types. Occurs on wide range of soil types, but all are dry and well-drained, with sparse vegetation and plenty of bare ground in open positions. Pupal stage occurs in soil. Only occurs in very dry sites.	Green (2000) Asher *et al*. (2001) Dennis and Bardell (1996) A. Barker

Table A-29. Variables identified from the literature and from interviews with experts for BBN for *H. semele*.

Wood cricket (*Nemobius sylvestris*)

Variable	**Detail**	**References**
Presence of woodland/leaf litter	Occurs in deep leaf litter, mainly under oaks, holly and bracken. Dead leaves and perhaps associated fungi are probably the main food of this species.	Marshall and Haes (1988)
	In some places, *N. sylvestris* colonies occur in old stone walls and earthbanks, but only below scrub or on woodland margins. (In the east of the Isle of Wight, *N. sylvestris* is present in crumbling clay sea-cliffs, but always close to dense scrub or deciduous trees along the cliff top.	Marshall and Haes (1988)
	It is advised to maintain at least 30% broadleaf, preferably oak (*Quercus* spp.), tree cover within the boundaries of the woodlands to provide and maintain the necessary leaf litter layer to secure successful reproduction for *N. sylvestris*.	Brouwers (2008)
	Inclosures or Ancient & Ornamental woodland in the New Forest. Oak and beech more suitable than conifer woodlands. (Outside of New Forest, found in banks of woodland, along railway lines and in hedge banks (e.g. Royal Victoria Country Park)). Not found if there are a lot of wood ants (more in conifers?).	P. Budd
	Not known from any open or exposed localities, although, in warm summers, individuals may wander some metres from dense cover.	Marshall and Haes (1988)
	Direct hot sunshine is disliked and appears to be injurious to them. Heat they like, provided there is shelter from direct sunlight.	Richards (1952)
Depth of leaf litter	Found in deep leaf litter.	Marshall and Haes (1988)
	Well-developed leaf-litter layer. Leaf litter depth (medians: Absent = 3.00 cm, Present = 4.25 cm)	Brouwers (2008)

Leaf litter volume	Minimum depth of leaf litter in which *N. sylvestris* was found was 1 cm and maximum depth in which *N. sylvestris* was found was 16cm. Leaf litter cannot be too deep for *N. sylvestris* (it just happened that N. Brouwers did not sample any sites where *N. sylvestris* was present that had a leaf litter depth greater than 16 cm). Leaf litter volume (a function of leaf litter depth and leaf litter cover) was found to be greater in presence sites than absence sites (medians: Absent = 27500 cm^3, Present = 38400 cm^3 (using a 2 m x 2 m quadrat). Leaf litter provides protection against low winter temperatures.	Brouwers (2008) N. Brouwers Brouwers (2008) Proess and Baden (2000)
Woodland edge	Occurs on warm and sunny clearings or along sunny margins of deciduous woods. Ride and track edges, woodland peripheries and open areas within woodlands (e.g. coppice sites) were found to be the main habitat locations where *N. sylvestris* could be found. The edges provide a warm microclimate - ground-dwelling invertebrates generally favour sunlit conditions because of their thermophillic nature. Euclidean distance to nearest permanent woodland edge (medians: Absent = 19.30 m, Present = 2.700 m). Important factor was level of sunlight availability on the ground (creates warm microclimate). One of the main factors influencing this was the openness of the canopy. Main features characterised by more or less permanent open canopy conditions were found to be woodland tracks and the woodland periphery. *N. sylvestris* occurrence was strongly positively correlated with the proximity to occupied locations along these permanent features. In addition, *N. sylvestris* seemed to spread and disperse from the main source populations within the woodland by using edges of open tracks and forest peripheries.	Marshall and Haes (1988) Brouwers (2008) Brouwers (2008) Brouwers (2008)

	N. sylvestris could rapidly colonise (i.e. within 1-2 summers) new felling/coppice clearings that were created adjacent to occupied woodland tracks. Presence of woodland edge is the most important variable for habitat suitability. Preferably a minimum of 50m of woodland edge required. Found in the middle of quite dense woods, but more likely to be found on edges. Not in dense stands of conifers. In glades (but not grassy. On edge of wood/heath and sometimes in quite open heath (with perhaps just a few trees) but not in bogs. In bracken areas. On edges but also found in middle of closed woodland.	N. Brouwers P. Budd P. Brock
Euclidean distance to nearest (occupied) permanent woodland edge	Probability of wood cricket being present decreased with an increase in Euclidean distance to nearest occupied permanent woodland edge. More isolated habitat locations were more likely to be uninhabited. Median distance between occupied woodlands = 50m, Probability >40% of *N. sylvestris* being present if distance to an occupied woodland edge was between 0-100 metres. 82% of woodland fragments occupied by *N. sylvestris* were within a radius of 250m of another occupied woodland. 12% of the occupied woodland fragments were situated > 1400m away from another source population.	Brouwers (2008)
South-orientated canopy closure	Probability of wood cricket being present decreased with an increase in south-orientated canopy closure (linked with sunlight availability, which has a strong effect on microclimate conditions). For each 1% increase in canopy closure the odds dropped by a factor of 0.949. South orientated canopy closure (medians: Absent = 94.8%, Present = 75.0%).	Brouwers (2008)
Woodland size	With an increase in woodland fragment area, wood cricket was more likely to be present. For each ha increase in woodland area, the odds ration of *N. sylvestris* being present increased by a factor of 1.053. Areas providing suitable edge habitat of only 100 m^2 are thought to already provide a firm basis to	Brouwers (2008)

	sustain viable wood cricket populations. Large woodlands provide more permanent edge habitat (important for microclimate – see above).	Brouwers (2008)
Woodland age	*N. sylvestris* occurred more often in woodlands with ancient characteristics, but was also found in woodland habitat of secondary origin. *N. sylvestris* was 4 times more likely to be present in ancient woodland than in secondary woodland. Presence of a well-developed leaf litter layer was only found along mature woodland edges. It does not make a difference if it is a woodland or a plantation.	Brouwers (2008) N. Brouwers
Ground vegetation height	Probability of wood cricket being present decreased with an increase in ground vegetation height. For each cm increase in vegetation height the odds dropped by a factor of 0.954. Ground vegetation height (medians: Absent = 41.0 cm, Present = 25.0 cm). Relatively high measures of vegetation height were associated with an increase in the number of vegetation layers which would negatively influence sunlight availability at ground level, resulting in relatively lower air temperatures.	Brouwers (2008) Brouwers (2008)
Ground vegetation cover	Probability of wood cricket being present decreased with an increase in cumulative ground vegetation cover (the sum of % area covered for each individual species of ground vegetation). (Also linked with sunlight availability, which has a strong effect on microclimate conditions). For each 1% increase in vegetation cover the odds dropped by a factor of 0.986. Ground vegetation cover (medians: Absent = 90.0%, Present = 55.0%). Ground vegetation cover and height were highly correlated with each other but cover is more important.	Brouwers (2008) N. Brouwers

Ground moisture	If a site was too wet then conditions may not be suitable. *N. sylvestris* can swim (N. Brouwers carried out experiments). Always observed it in dry sites. The eggs are laid singly in the soil beneath the surface litter layers.	N. Brouwers P. Budd Brown (1978)
Management/ Disturbance/ Grazing	*N. sylvestris* demands early successional habitat conditions related to natural and anthropogenic disturbances within woodland habitat. Large woodland areas were generally found to be more heavily and frequently managed than smaller woodlands (N. Brouwers, personal observation). These activities seemed closely positively correlated with the amount of available habitat for wood cricket. In these larger managed woodlands, commercial tree harvesting and conservation activities (e.g. re-instatement of coppice rotation) not only creates more open habitat but also involves presence and maintenance of open tracks and roads for tree extraction. The Isle of Wight has also been the focus of extensive woodland restoration efforts including the restoration of planted ancient woodland sites (PAWS) by removing non-native tree species. These activities were found to increase canopy openness, and with initial low levels of ground vegetation cover, improved habitat conditions for wood cricket on the ground by the increased amount of sunlight availability. Active management of woodlands therefore has a positive effect on wood cricket populations by providing more suitable habitat. Grazing may be good way of suppressing vigorous growth of ground vegetation (which is unfavourable for *N. sylvestris* persistence). And as *N. sylvestris* is highly persistent and widespread across the New Forest where there is grazing, it suggests that this is favourable.	Brouwers (2008) Brouwers (2008) Brouwers (2008)

Table A-30. Variables identified from the literature and from interviews with experts for BBN for *N.sylvestris*.

Silver-studded blue (*Plebejus argus*)

Variable	**Detail**	**References**
Habitat type/ Presence of heathland	In Hampshire, *P. argus* breeds exclusively on heathland (dry and wet), including the edges of wet heaths and pockets of heathland in coniferous plantations. (Some specimens have been recorded miles from heaths, which occur in exceptionally hot weather (e.g. 1976 drought – seen in fair numbers on downs and in woods)).	Oates *et al*. (2000)
	3 main habitats used: lowland heathland (the most widely used); calcareous grasslands (in N Wales, Pembrokeshire, and the Isle of Portland in Dorset); and sand dunes (e.g. in Cornwall). Occasionally occurs in other habitats such as bogs.	Asher *et al*. (2001)
	Requires heathland – humid and wet. Dry too dry. (Further north in the country, where cooler – on drier heath). Wet heath (HL2) in particular - it contains more cross-leaved heath (main food plant – flowers at the time that *P. argus* is flying (heather flowers later)). Also a bit of bell heather. Does not mind it wet (with a bit of *Sphagnum*). AQ1 – probably does not breed there but may use it as a food source (nectar?), particularly tussock islands (with ericaceous species on them). Good resource for dry years. HL3 – depends how patchy the cover of ericaceous species is. Less likely to be ideal. The borders between wet and dry heath are particularly good – gradation in wetness (slopes).	A. Barker
Heathland species (Food plant/ovipositing)	It appears that *P. argus* uses heather (*Calluna vulgaris*) and cross-leaved heath (*Erica tetralix*) in the New Forest, and possibly bell heather (*Erica cinerea*).	Oates *et al*. (2000)
	A variety of ericaceous and leguminous plants are used: on heathland the most common are heather, bell heather, cross-leaved heath, gorses (*Ulex* spp.).	Asher *et al*. (2001), Joy (1995)
	Larvae feed on buds, flowers, young leaves or growing tips of the foodplants.	Asher *et al*. (2001), Joy (1995)

	Eggs are laid on wide variety of substrate types including potential larval foodplants (commonly heather, bell heather and a variey of leguminous plants), mosses, lichens, plant litter, grasses, soil and bare earth. They have also been found on bracken (*Pteridium aquilinum*). Slighlty different requirements in different areas of the country – Suffolk and Surrey: bell heather. Dorset (on humid and wet heaths): cross-leaved heath. Ideally, presence of more than 1 of the ericaceous species (e.g. cross-leaved heath and heather).	Joy (1995) Ravenscroft and Warren (1996) P. Brock
Habitat size	Should ideally be at least 5-10 (and if possible >50) ha of habitat in suitable condition for *P. argus* at any one site. Required patch size depends how isolated the patch is. Long term. Small patch (~1 ha) could support a population if others are nearby. A smaller patch is likely to all go through same successional stage at once and could all become unsuitable at same time. Whereas a larger patch is more likely to contain a range of successions, variety and therefore more able to sustain populations in the long term – *P. argus* can move within the patch as part of it becomes unfavourable.	Joy (1995) A. Barker
Soil	Heathland plants favour soils with low nutrient levels.	Joy (1995)
Heathland growth stage/ Vegetation height Bare ground	Heathers in the pioneer or early building stages. Continuous supply of pioneer heathland within one kilometre (colonisation range) of a colony so that the species can move should conditions on the site deteriorate. A flight area containing varied age structures (of heather, bell heather and/or other heaths). The older bushes may provide shelter and potential roosting sites for adults while younger established plants may be used as nectar sources.	Joy (1995) Joy (1995)

	In all habitats *P. argus* requires short or sparse vegetation, such as recently burnt heathland, or where there are thin, eroding soils (e.g. old quarries and coasts). In the south of England it is less demanding and often associated with shorter areas of wet heath dominated by cross-leaved heath.	Asher *et al*. (2001)
	Eggs are laid singly, close to the ground. On heathland, they are often laid on woody stems of the foodplants where there is sparse vegetation and patches of bare ground. Most eggs are laid in short vegetation below 10-15 cm, but even shorter vegetation (<7 cm) seems to be selected at the north of its range in Wales.	Asher *et al*. (2001)
	Warm microclimate at ground level for larvae and for ants (nestbuilding and foraging activities).	Ravenscroft and Warren (1996)
	Warm microclimate (which usually exists in areas of pioneer heathland habitat where there is a high proportion of bare ground and vegetation which is sparse and low).	Joy (1995)
	Breeds on regenerating growth in recent clearings and burnt areas, or in degenerate areas where the heather is leggy and light reaches the ground. Generally absent from mature heaths, or confined to damp hollows and along rills, where the heather grows sparsely.	Green (2000)
	Adults often congregate to roost on sheltered bushes or grass tussocks.	Asher *et al*. (2001)
	Early successional stages, and those held in check by grazing (nowadays mostly rabbits, except in the New Forest) are preferred in all habitats.	Ravenscroft and Warren (1996)
	Aim to maintain a mosaic of heathland of different stages and a continual presence of early successional vegetation, which encourages good populations of the symbiotic ants. This is less important on large expanses of habitat where suitable conditions are likely to occur through random events such as disturbance and burns.	Ravenscroft and Warren (1996), Warren and Wigglesworth

	Pioneer heath, but not first year or so because nothing flowering. Mid-sucessional. Need some building heath for shelter. Tracks in more mature heath can help make it suitable as it creates warmer areas.	(2008) A. Barker
Bare ground	Requires open ground for breeding and either bare soil or short vegetation in all habitats in Britain. Eggs usually laid along vegetation/bare ground margins where the vegetation is sparse and low. Within heathland sites *Lasius* ant nest densities tend to be highest on the least vegetated areas where high proportions of bare ground and short turf create a warm microclimate ideally suited to their needs. Bare ground is important for the ants. Ants need it dry. Does not need a huge amount of bare ground. Paths. Bare ground not too important – does not need too much.	Ravenscroft and Warren (1996) Joy (1995) Joy (1995) A. Barker P. Brock
Heathland management/ Burning/Grazing	Requires managed heathland. Colonies die out on neglected heaths as mature heathers are unsuitable. Restricted to lowland acid heath which as a result of burning or other disturbance within the previous 15 years, or from exposure, has been kept open and fresh in growth. Grazing, rotational burning or periodic disturbance are essential. The New Forest colonies which are managed only by grazing are fairly stable, threatened only by short-term effects of uncontrolled fires and long term problems such as Silver Birch and Scots	Oates *et al*. (2000) Bretherton *et al*. (1989) Oates *et al*. (2000) Oates *et al*. (2000)

	Pine invasion. However, colonies on the New Forest heaths which are managed by rotational burning do fluctuate greatly. The butterfly is reasonably mobile in the Forest and so can readily colonise new habitat when it develops.	
	Periodic, patchy/small scale burning can be beneficial. Suitable conditions may take 2-5 years to develop after burning.	Ravenscroft and Warren (1996)
	Grazing can prolong suitable conditions on heathland and the presence of large herbivores may also help by providing some local soil disturbance. Disturbance of the ground, especially if the topsoil is removed, can produce suitable conditions.	Ravenscroft and Warren (1996), Warren and Wigglesworth (2008)
	Grazing alone is unlikely to create the sort of pioneer habitat favoured by *P. argus*. Intensive cattle grazing for short periods may be more effective then sheep grazing on *P. argus* sites as their trampling may open up the turf and create sites for seedling establishment.	Joy (1995)
	Longer periods of grazing by cattle may damage the heather. So grazing should be restricted to the spring and summer months.	Joy (1995)
	Mixed grazing (by cattle, sheep and horses) may be the preferred option e.g. in the New Forest.	Joy (1995)
	Beneficial to remove/localise areas where there is grazing animal dung (removal of nutrients).	Joy (1995)
	A combination of rabbit grazing and rabbit digging may have maintained suitable conditions for *P. argus* on some sites.	Joy (1995)
Presence of *Lasius* species of ant	*P. argus* requires the presence of the correct species of ant – more work needs to be carried out on this in the Hampshire heaths, but pupae are attractive to ants of the genus *Lasius* with which	Joy (1995)

Lasius habitat requirements	larvae and pupae are strongly associated.	
	The larvae have a close relationship with ants, whereby the ants tend and protect them in return for sugar-rich liquids produced from special glands on the larvae’s bodies. Females choose to lay eggs where they detect suitable ant pheromones. On heathland, the most commonly associated species are black ants (*Lasius niger* and *L. alienus*). Larvae pupate within or close to ant nests.	Asher *et al*. (2001)
	Ants probably provide protection from parasites and predators.	Ravenscroft and Warren (1996)
	L. alienus is a species which only inhabits dry and especially warm pastures and heaths. As ants require warm conditions for foraging and for building their nests, the sandy soils of heaths provide ideal conditions. Within heathland sites ant nest densities tend to be highest on the least vegetated areas where high proportions of bare ground and short turf create a warm microclimate ideally suited to their needs. *L. alienus* prefers the least vegetated areas of a heath. *L. niger* commonly form nests in warm patches of bare soil. The ants are most likely to be found in the pioneer habitat favoured by *P. argus*.	Joy (1995)
	P. argus is able to associate with ants in general (the larvae retain some ability to enter into generalist associations with ants in genera other than *Lasius*) but has specific adaptations to the genus *Lasius*. There is some local differentiation that gives an improved behavioural interaction between specific butterfly populations and the locally abundant species of *Lasius*. Populations of *P. argus* differ in whether they are associated with *L. niger* or *L alienus*.	Jordano and Thomas (1992)
	Within habitat patches, eggs, larvae and pupae all spatially associated with *Lasius*. On larger scale, the densities of butterflies in different habitat patches and populations, and whether the butterfly was present or not, were correlated with *Lasius* ant densities.	Jordano *et al*. (1992)

Ant nest density	It may be that *L. alienus* or *L. niger* nests have to reach a certain density before a site is suitable for *P. argus*. (Ant densities should be measured as number of nests/100 metres, or as appearance at honey ant baits).	Joy (1995)
	Ants need it dry. Ants may not be in the wetter areas. Egg-laying (up the slope) where drier and better for ants. Plus pupae cannot be in water.	A. Barker
Presence of shelter	*P. argus* forms discrete colonies in areas sheltered by dykes, the local relief, or clumps of trees and bushes.	Oates *et al*. (2000)
	Requires a warm microclimate.	Joy (1995)
Slope	Flat or gently sloping land, which is sheltered [and preferably south facing [northern]].	Joy (1995)
	Slope not a factor in the New Forest.	P. Brock
(Aspect)	Will breed on slopes of all aspects in the south. (In cooler northern parts of country only on the warmer south-facing slopes).	Joy (1995)
	The borders between wet and dry heath are particularly good – gradation in wetness (slopes). Ants need it dry. Ants may not be in the wetter areas (down the slope). But may be good nectar source down the slope. But egg-laying up the slope where drier and better for ants. Plus pupae cannot be in water. Flat slopes can be all or nothing (can be too dry or too wet), whereas sloping means it is more likely that some part with be suitable (and more likely to get gradation from wet to dry heath). If irregular, get more variation. Flat areas can be fine (as long as heterogeneous structure) but valley sides probably better. Valley bottoms not suitable – mires.	A. Barker

Table A-31. Variables identified from the literature and from interviews with experts for BBN for *P.argus*.

Pillwort (*Pilularia globulifera*)

Variable	**Detail**	**References**
Habitat type/ Presence of wetland habitat	Found in shallow water on pond margins, in poached wet grassland and mires.	Jermy (1994)
	(In Scotland it seems to occupy a different niche – most records are for sides of rivers and lakes, where perhaps the scouring effect of running water creates the open conditions it requires).	Botanical Society of the British Isles (2008b)
	On silty mud by lakes, ponds and reservoirs.	Stace (1997)
	Lakes, ponds and marshlands.	UK Biodiversity Action Plan (2008)
	Silty or peaty lake and pond margins and shallow pans and pools resulting from brick-earth or gravel extraction.	Jermy (1994)
	Can occur in disused clay or gravel pits.	Preston and Croft (1997)
	Mires. Muddy flats at the edges of low-lying fresh-water lakes. Also on silty mud at the edges of ponds, slow-flowing river backwaters and river mouths, in wet sandy hollows in dunes and heaths, and sometimes invades muddy ditches and old claypit workings in shallow, wet situations.	Jermy (1994)
	Within heathland and upland grassland.	Plantlife (2006)
	Edges of non-calcareous lakes, reservoirs, ponds or slow-flowing rivers and sometimes on damp mine workings or as a submerged aquatic.	Preston (2002)
	On mud, fine gravel or wet clay, in or around shallow ponds on heaths; in old gravel pits; open bog zones and acid swamps. Locally in ditches and muddy ruts in Inclosure rides.	Brewis *et al*. (1996)
	Pillwort grows most typically on soft mud in shallow water at the edges of ponds and lakes in	Scott *et al*. (1999)

	which water levels vary regularly within a small range, but it is also found in muddy ditches, slow-flowing river backwaters, wet, sandy hollows in dunes and heaths and man-made habitats such as old clay-pit workings and sites of gravel extraction.	
	Occurs on edges of ponds, wet gravel and sand pits. Early successional stages. Open. On edge of bogs, in heaths – grassy places, where there are animals – churned. On stream edges (e.g. Crockford Bottom).	C. Chatters
	Wetland/aquatic habitats. All colonies in clean water bodies in winter (water in winter). Only sporulates when in seasonally wet areas. Ponds in grassy areas rather than heathy areas – partly because of stock – they go to grass - open. Plus heathy ponds tend to be peaty.	A. Byfield
Ground moisture	Indicator of shallow-water sites that may lack standing water for extensive periods (10).	Hill *et al*. (2004)
Water depth	Pillwort can survive in deep water, but most at home in shallow waterbodies.	Plantlife (2006)
	It can survive periods of complete immersion.	Page (1982)
	It sometimes occurs in water up to about 1m depth.	Page (1982)
	Usually grows in habitats which remain more or less permanently damp throughout the summer, and which warm rapidly whenever there is sunshine.	Page (1982)
	Shallow water.	Scott *et al*. (1999)
	Will grow in 5m of water, but preferably 1-2 feet. It can ‘cruise’ in deep water, but forms weak growing mats. It does best (long term) in temporary water bodies where it forms dense mats at the surface.	A. Byfield

	If the water is clear, it can be a couple of feet, but if it is turbid then not so deep. *P. globulifera* can survive in deeper water for short periods. In shallow water.	C. Chatters G. Read
Fluctuating water levels	Favours seasonally-dry acid pools. Ponds subject to fluctuating water levels. (Permanent flooding and control of water levels is a threat). Edges of ponds and lakes in which water levels vary regularly within a small range. It is a poor competitor and is soon ‘swamped’ by more strongly growing aquatic species unless external factors such as fluctuating water levels or trampling by animals serve to exclude these. It can withstand short periods of exposure when water levels fall, but is soon out-competed by terrestrial plants unless it is again inundated. Thrives where soft, semi-consolidated, silty mud is periodically flooded by freshwater conditions. Fluctuating water levels. Seems to regenerate quite well – if an area dries up *P. globulifera* still grows the next year. Found where water depth fluctuates. Favours seasonal wetness.	Plantlife (2006) UK Biodiversity Action Plan (2008) Scott *et al*. (1999) Page (1982) G. Read A. Byfield C. Chatters
Speed of water flow	The kind of habitats that Pillwort occurs in mostly are not flowing. But it does occur in streams (e.g. Crockford Bridge). But shallow streams which tend to be slow-flowing. (If water flow is too fast it may remove Pillwort).	A. Byfield

	Generally non-flowing water but it can occur on stream edges (e.g. Crockford Bridge) but gently flowing. (Also see habitat type)	C. Chatters
Water pH	Favours (seasonally-dry) acid pools. Mildly acid pools. Base-poor pools. Base-poor pools around pH 6.0. It usually grows in habitats which are circum-neutral to slightly acidic in character. Not taken pH of water for *P. globulifera* habitats, but in broad terms *P. globulifera* seems tolerant of a relatively broad range around circum-neutral. (Carried out pH monitoring of a pond at Godshill - found a range of pH readings over the course of the day and the season. This is possibly where in poorly buffered ponds the seasonal concentrations of dung and urine, together with the creation of dissolved oxygen by photosynthesis of vegetation on the pond floor, introduces bases that then shift the pH).	Plantlife (2006) Plantlife (2006) Preston and Croft (1997) Jermy (1994) Scott *et al*. (1999) C. Chatters
Absence of nitrate/phosphate pollution	Pollution increases pH and growth of competitive vegetation. Sensitive to pollution.	UK Biodiversity Action Plan (2008) Scott *et al*. (1999)
Bank/shore angle	Nearly level muddy flats. Creation of very steeply sloping banks on some lakes is a threat.	Page (1982) UK Biodiversity

	Gently sloping banks – have maximum mud in summer. With a steep bank you do not get much shore. Gentle is more likely to be muddy. Association with gentle slopes is partly because where water depth fluctuates (favoured by Pillwort, and is muddy), it will tend to be gentle. Gently sloping to flat.	Action Plan (2008) A. Byfield C. Chatters
Bank/shore substrate	On soft, nearly level, muddy flats... silty mud. Grows on bare mud (often a clay or clay-sand substrate). Soft mud. Requires sediment that is fine, clayey and not gritty. Fine material in gravel pits. Grows on fine sediment. Clay. Not in gravel or sands (but sands do not retain nutrients). Fine, silty material (not coarse).	Page (1982) UK Biodiversity Action Plan (2008) Scott *et al.* (1999) C. Chatters A. Byfield G. Read
Presence of bare substrate	Grows on bare mud. Requires open substrate which it will rapidly colonise, eventually to be ousted as the hydrosphere processes. It also colonises bare mud exposed by falling water levels. Its creeping rhizomes are able to colonise bare mud rapidly. Abundant on bare ground – spreads.	UK Biodiversity Action Plan (2008) Jermy (1994) Scott *et al*. (1999) C. Chatters

	Found on bare ground. Pillwort is a poor competitor, so bare ground provides areas to establish. Not a good competitor, so requires bare ground.	A. Byfield G. Read
Disturbance/ Grazing	It generally favours habitats with some disturbance. Flourishes in habitats under a long tradition of heavy grazing. In some of its sites suitable conditions are maintained by the trampling of cattle and horses. In many sites competition may be kept to a minimum, and thus Pillwort maintained, by cattle or horse trampling (poaching). It is a poor competitor and is soon 'swamped' by more strongly growing aquatic species unless external factors such as fluctuating water levels or trampling by animals serve to exclude these. Long continuity of management important – grazing and trampling. Otherwise ponds go to willows then oak woodland. Grazing and poaching leads to open ground. In summer, as dries out, stock move more to ponds – get divots from cattle/horse – wetter microtopography – means it does not really dry out. Requires grazing and disturbance – open. Hoof marks (breaks grass). High stress.	UK Biodiversity Action Plan (2008) Plantlife (2006) Preston and Croft (1997) Jermy (1994) Scott *et al*. (1999) A. Byfield C. Chatters
Soil type	Clays, sands and peaty substrates. Ponds on mildly acidic, quite fertile clays, rather than on sandy (mainly because the clay holds the water). Open clay pits. On Barton Beds. On lots of soils. Claggy soils. Clay that never really dries out. Wet, cold. Seasonally wet. Wet	Plantlife (2006) A. Byfield C. Chatters

	gravel and sand pits. Wet sands or clay. Including base-rich clays (Headon beds).	
	Mineral soils, not organic soils. No deep leaf litter. Low humus. If there is poaching then some leaf litter cover ok, as it helps to remove the litter and bring clay substrate to surface.	A. Byfield
Soil pH	Between 3 (acidity indicator, mainly on acid soils, but exceptionally also on nearly neutral ones) and 5 (indicator of moderately acid soils, only occasionally found on very acid or on neutral to basic soils).	Hill *et al*. (2004)
	Slightly acid to neutral.	UK Biodiversity Action Plan (2008)
	Usually grows in habitats which are circum-neutral to rather acidic in character.	Page (1982)
	It usually grows in habitats which are circum-neutral to slightly acidic in character.	Scott *et al*. (1999)
	Ponds on mildly acidic clays.	A. Byfield
Soil fertility	Between 1 (indicator of extremely infertile sites) and 3 (Indicator of more or less infertile sites)	Hill *et al*. (2004)
	Ponds on moderately, but not too fertile clays.	A. Byfield
Canopy cover/openness	Light-loving plant rarely found where relative illumination in summer is less than 40% (8).	Hill *et al*. (2004)
	Exposed margins of pools.	Plantlife (2006)
	Requires high light. Very rarely seen under even light shade. Open.	C. Chatters
	Needs good light. Also canopy cover means pond may clog with leaf litter (see notes below on soil fertility). Long-term – better without cover.	A. Byfield

	No canopy cover.	G. Read
Absence of competitive vegetation/non-native species	Grows on bare mud. Nitrate/phosphate pollution is a threat because it increases growth of competitive vegetation.	UK Biodiversity Action Plan (2008)
	Grows mainly in habitats which are largely competition free.	Page (1982)
	Alien species such as *Crassula helmsii* have caused losses.	Stewart *et al*. (2000)
Altitude	Most sites are at low altitudes, but it has been recorded up to 450m (in Wales).	Preston and Croft (1997)
	Occurs from sea-level to about 380m altitude.	Page (1982)

Table A-32. Variables identified from the literature and from interviews with experts for BBN for *P.globulifera*.

Nail fungus (*Poronia punctata*)

Variable	**Detail**	**References**
Habitat type (Vegetation composition, Soil type)	Unimproved (unfertilised) acid-loving vegetation. H3 humid heath (Ulex minor-Agrostis curtsii), with 18% from M16 wet heath (Erica tetralix-Sphagnum compactum), 13% from H2c dry heath (Calluna vulgaris-Ulex-minor) and only 3% found on U3 acid grassland (Agrostis curtisii). Rarely recorded on open acid grassland – supposedly a favoured habitat.	Poland (2004)
	All Dorset locations comprised acidic vegetation types. – 65% in H2 (dry heath), 17% in H3 (humid heath), 10% U3 (Bristle Bent acidic grassland), 4% in M16 (wet heath) and 4% in recent gorse-clearance areas (acidic bare ground).	Cox and Pickess (1999)
	Changes in the movements of the ponies may be one of the causes of the year-to-year differences between the numbers of *Poronia*-bearing dung deposits occurring in the different NVC communities.	Cox *et al*. (2005)
	P. punctata is found almost entirely in acid heathy areas where the grasses are largely hair-grass, *Deschampsia*, species.	Dickson (1997)
	Requires some acid grass, in dry heaths. Found on grass on edge of roads between heath. Not on grassy rides. Grass required for ponies. – Related to pony presence.	S. Skeates
	Requires absence of shade (so dung more suitable) but vegetation type may not actually be important – just as long as vegetation that ponies eat is present.	S. Skeates
	Has seen it all across the New Forest (where there are the right conditions). Thinks it will fruit wherever the dung is dropped, in any habitat, but more likely to be in habitats where ponies are more likely to be. So heathland best – where ponies are. And more grass in the heathland – more ponies. Also found in wetter areas.	A. Lucas

	Acid vegetation. Not on fertilised grass.	G. Dickson
Tree canopy cover	Not on dung dropped in woodland. Sub-shrub canopy was always open and vegetation usually did not totally surround the dungpiles, leaving them exposed to varying aspects. Open sites. Not personally found it in woodland, although has heard that it does occasionally occur in woodland. Has found it on the fringes of woodland. Suggestion that in woodland the dung does not dry out as much, and *P. punctata* only found on slightly dried out dung. Found it occasionally in woodland – where ponies have wandered in. But ponies mainly in the open and that's why there's more there. Not generally found on dung dropped in woodland, probably because it is not as warm on the ground. General absence in woodland may be a function of temperature. Ground temperature would not be as warm in woods, which is why it may not be found there. Temperature may also explain preference for short vegetation.	Whalley and Dickson (1986) Poland (2004) Webster (1999) S. Skeates A. Lucas G. Dickson A. Newton
Liming/fertilisation	Hardly ever found on areas that were limed during WWII. Never found where ponies are pasteurised in fertilised grassland. Either the fertiliser or the grass species is inimical to the fungus.	Dickson (1997)

	Large grassy areas that are known to have been well limed in the past fail to produce *P. punctata*. Has been recorded (Greenlands Farm, Studland peninsula) on grassland that had received lime and fertiliser applications after WWII, although the vegetation there is now primarily acid grassland and heathland. Not found in areas of the New Forest which were ploughed and fertilised (liming) for crops (neutral soil). (E.g. at Long Slade Bottom).	Webster (1999) Cox and Pickess (1999) G. Dickson
Vegetation height	The majority of *P. punctata* was found within vegetation with a maximum height between 10-15 cm and absent from vegetation above 30 cm. Frequency in different vegetation heights (cm): 0-5 (3%), 5-10 (8%), 11-15 (63%), 16-20 (5%), 21-25 (10%), 25-30 (13%), 31+ (absent) Semi-open environment. Mostly found on short vegetation (e.g. where recently burnt), although some occurred amongst mature *Calluna* up to 30 cm high. Mean vegetation height was 12.8 cm. However, it is more likely to have been overlooked in taller vegetation. On short vegetation, but then there is not much long grass in the New Forest. Fungi generally found in the short grass, but this could be due to a number of factors: easier to find in short grass, less competition for fungi in short grass, or short grass indicates that ponies present, so more likely to be dung and hence *P. punctata*. Vegetation height not too important – just as long there's dung. Shorter vegetation may allow for warmer ground temperature which may be important for *P.*	Poland (2004) Cox and Pickess (1999) S. Skeates A. Lucas A. Newton

	punctata.	
Management – Cutting, Burning	Many of the *Poronia*-bearing dung deposits were found in open habitats, frequently on or near mown firebreaks and on former fire sites where the vegetation was short. Often found in heathland areas which had been recently cut or burnt. Perhaps this is because it attracts the ponies to the young vegetation. And it means the vegetation height is low and the ground temperature warmer.	Cox and Pickess (1999) A. Newton
Ground moisture	Requires a level of drainage impedance but not generally waterlogged. (H2c-H3-M16 spectrum of increasing drainage impedance). Some moisture is required and it may not develop on arid sites. Excessive moisture may inhibit growth of *P. punctata* by some means – as it was rarely found on saturated dung piles and not in very wet communities – on undulating heathland sites, *P. punctata* did not occur on dung deposited in the seasonally flooded short turf hollow dominated by *Molinia* and *Carex* spp. (primarily (*C. Panacea*). *P. punctata* occurred on slightly raised ground or heathy hummocks (a height differential of c.20 cm) that remained unflooded. *P. punctata* was also absent from dung deposited in M1 *Sphagnum auriculatum* bog pools, M21 *Narthecium ossifragum-Sphagnum papillosum* valley mires or M29 *Hypericum elodes-Potamogeton polygonifolius* soakways. These communities are constantly wet with a high water table. Damp, peaty areas to drier gravelly sites. Moisture not too important. Found it in wetter areas as well as drier areas (see notes on habitat type above). Mostly on dry ground. Not sure about wetter.	Poland (2004) Webster (1999) A. Lucas S. Skeates

	Maybe not found as much on the very wet areas of heaths because the ponies do not like it as much (e.g. sink in the mud).	G. Dickson
Presence of (dung from) ponies feeding on natural vegetation (no additives)	Only occurs on pony/horse droppings in the UK. Likely that movements of ponies between suitable sites facilitates the spread of *P. punctata*. Presence of ponies ingesting the spores is the most important factor. It may only occur on dung from the wild ponies due to worming (Ivemectin?), and only on the dung from ponies that are not rounded up and wormed. This may possibly be part of the reason why it does not occur on cattle dung – because all cattle are wormed and rounded up? May also be because cattle dung is different – runnier. The species of pony does not seem to be an important factor, because it has been found on dung from other species than the New Forest pony, such as Prczwalskis horse (north of Hampshire) and Exmoor ponies (Arne, Dorset). The quantity of dung is not important.	Whalley and Dickson (1986) Cox *et al*. (2005) A. Lucas G. Dickson G. Dickson
Moisture content of dung (linked to age)	Percentage of *P. punctata*-containing dungpiles which were: Normal (65%), Desiccated (33%), Saturated (3%). The dung needs to be slightly dried out. Moisture not too important. If dung dries out too fast, that may prevent it fruiting. If dung was saturated for a long period that might be a problem. But saturated for a short while and then drying out would be ok.	Poland (2004) S. Skeates A. Lucas
Dung age/how	All *P. punctata*-bearing dung showed some signs of weathering and stromata were not observed	Cox and Pickess

recent	on fresh dung. Not on fresh dung. Not found fruiting on fresh dung, just because it takes a while to grow.	(1999) S. Skeates A. Lucas
Structure of dungpile (linked to age)	Percentage of *P. punctata*-containing dungpiles which were: Normal (43%), Disturbed (20%), Collapsing (18%), Flat (20%). Only on solid or partly broken lumps; not on dung that birds have raked through. Does not occur on old, collapsed droppings but is typically found on dung which is about 1 month old and still in lumps. The large majority of dung deposits still retained the shape of the original boluses, although a small number of examples had been weathered almost flat. Some dung had been broken open (probably by foraging birds, foxes or badgers). *P. punctata* not observed on extensively disrupted boluses. (Coprophilous beetle burrows frequently observed within or beneath dung deposits). If dung has been extensively fragmented any fungal mycelium would be more likely to desiccate before stromata could be formed. Not on very old, broken up dung, just because too broken up, so *P. punctata* gone. On dung still in lumps. Not on completely flat dung; still some structure.	Poland (2004) Webster (1999) Whalley and Dickson (1986) Webster (1999) Cox and Pickess (1999) Cox and Pickess (1999) A. Lucas S. Skeates

Pony food plants (where spores need to land) and movements	Large clumps of *Molinia caerulea* in the valley bogs and a sparser growth on moist heaths seemed to form the bulk of food throughout the summer. *Agrostis* spp. and *Festuca* spp., the dominant species of grass on many woodland and streamside 'lawns' and on patches of 'grass heath' were grazed throughout the year. Few other species of grass were of any importance except *Glyceria fluitans* (Flote grass) and other *Glyceria* spp.; these were favoured by certain individuals who spent long periods wading in ponds and slow-moving ditches, then feeding almost exclusively on them. Heather plants, *Calluna vulgaris* and *Erica* spp., seemed to be of relatively little importance in the ponies' diet, despite the large areas covered by these species. Some heather shoots were browsed in the winter and the young plants resulting from the young plants resulting from the clearing of older woody plants were also eaten, but never to a large extent.	Tyler (1972)
	The home range (the constant area normally traversed by an individual in its day to day activities) of different groups of ponies often coincided with or considerably overlapped each other and varied in size from about 200 to 2500 acres (82 to 1020 ha).	Tyler (1972)

Table A-33. Variables identified from the literature and from interviews with experts for BBN for *P.punctata*.

Appendix 21. List of experts who contributed expert knowledge in the development of BBNs

I am extremely grateful to the following people who contributed as sources of expert knowledge in the development of the BBNs (in either identification of relevant variables or filling in of CPTs, or both stages).

Plant species (*C. nobile, G. constrictum, G. illyricus, P. globulifera*)

A. Byfield – Plantlife International
C. Chatters – Hampshire and Isle of Wight Wildlife Trust
G. Read – Independent botanist

Butterfly species (*H. semele and P. argus*)

A. Barker – Butterfly Conservation
L. Barker – Butterfly Conservation
P. Brock – Natural History Museum

N. sylvestris

P. Brock – Natural History Museum
N. Brouwers (Author of thesis on *N. sylvestris* on the Isle of Wight; Brouwers (2008))
P. Budd – Southampton Natural History Society

P. punctata

P. Budd – Southampton Natural History Society
G. Dickson – Mycologist
A. Lucas – Hampshire Fungus Recording Group
A. Newton – Bournemouth University
S. Skeates – Hampshire Fungus Recording Group

Appendix 22. Main sources of bias in expert elicitation for BBNs

Table A-34 below shows the mostly likely main biases associated with the chosen elicitation procedures, identified by Meyer and Booker (1991).

Elicitation procedure	Type of bias	Source
Elicitation situation: Individual interview	Motivational	Social pressure (from data gatherer) Wishful thinking
	Cognitive	Inconsistency
Mode of communication: Face-to-face	Motivational	Social pressure (from data gatherer) Wishful thinking
	Cognitive	Underestimation of uncertainty
Response mode: Probabilities	Motivational bias	Misinterpretation by expert
	Cognitive	Inconsistency Underestimation of uncertainty

Table A-34. The main biases associated with the chosen elicitation procedures (from Meyer and Booker, 1991).

Table A-35. below describes the biases above in more detail as well as some other biases that may arise, and how to minimise them.

Type of bias	Description	How to minimise
Motivational bias	Motivational bias occurs when an expert consciously or unconsciously makes accommodations to please the interviewer, something which is assumed to be driven by our human needs, such as for approval (Meyer and Booker, 1991). Meyer and Booker (1991) report that experts may not report their actual solutions or thought processes, for example because of:	
<u>Social acceptability/ pressure</u>	Concerns over the social acceptability of their responses (including social pressure from the interviewer, such as leading questions and the reactions of the interviewer) (Meyer and Booker, 1991).	This is unlikely to be much of a problem in this context, given the nature of the work.
	In face-to-face situations, data gatherers can intentionally or unintentionally influence the expert through their body language, facial expression, intonation and word choice (Meyer and Booker, 1991).	Be aware of this and try to reduce it.
	Social pressure bias is more pronounced when the data gatherer asks leading questions. The concurrent verbal report requires experts to think aloud as they solve a problem and this process almost eliminates the chances of leading the experts because it does not involve questioning them (Meyer and Booker, 1991).	Try not to ask leading questions and ask experts to 'think aloud' as much as possible.
<u>Wishful thinking</u>	Occurs when expert's responses are influenced by their hopes or involvement in the area on which they are being questioned; what the individual thinks should happen will influence what they think will happen (Meyer and Booker, 1991).	Meyer and Booker (1991) suggest selecting those experts who have the least to gain from their response and obtaining diverse expertise by selecting experts that represent different organisations and the various theoretical stances on the subject. Therefore, selecting as diverse a range of experts as possible should be aimed for, although this is unlikely to be an issue with the nature of the work. However, the background (including affiliation organisations) should be recorded for

	In general, the effects of wishful thinking will be most pronounced when the experts do not have to explain their reasoning. The experts’ highly optimistic responses are checked by having them disaggregate the problem and explain their problem solving (Meyer and Booker, 1991). Wishful thinking is indicated if the experts were previously judged to have something to gain from their answers and if the experts appear to answer quickly and with little thought (Meyer and Booker, 1991).	each of the experts. The tendency toward wishful thinking can be countered by making it more difficult for the expert to indulge in it. If the experts must explain their answers in detail, it will become apparent whether there was any objective basis for their responses (Meyer and Booker, 1991). This may not necessarily indicate wishful thinking; the expert may just be confident in their answer because they know it. However, this is still something to be aware of during the elicitation.
Misinterpretation	Where the interviewer or knowledge engineer misinterprets the experts’ reports or the analyst or knowledge engineer misinterprets the experts’ knowledge; humans tend to perceive and interpret incoming information in a selective manner that supports what they already believe, which can, in some cases, lead to a misinterpretation, or biasing, of the information (Meyer and Booker, 1991).	Making sure that questions are as clear and unambiguous as possible and carrying out a pilot study to check the procedure should help, as well as making it clear to experts at the start of the elicitation that they may ask questions at any time if at all unsure or unclear.
Cognitive bias		
Inconsistency bias	Inconsistency bias occurs when people are inconsistent in their solving of problems, especially through time, due to limitation in information processing (Meyer and Booker, 1991). As the individuals’ thinking evolves, their current thoughts or solutions may contradict those that they expressed earlier.	Making sure that the questions are as clear and as straightforward as possible (without asking too much at once) may help to reduce confusion. In addition, recapping of definitions and assumptions will help to improve consistency (Meyer and Booker, 1991). Asking experts to verbalise their thoughts and answers will help detection of inconsistencies (Meyer and Booker, 1991).

	Inconsistency in individuals' judgements can stem from their remembering or forgetting information during the elicitation session and this natural tendency can be acerbated conditions such as memory problems, fatigue and confusion (Meyer and Booker, 1991).	Inconsistencies can be highlighted in real time (Meyer and Booker, 1991), for example by asking questions in more than one way and checking for consistency (von Winterfeldt and Edwards, 1986) and contradictions (by checking back over responses). Fatigue can be reduced by keeping the interviews to a maximum of two hours or scheduling a break if the experts appear to be becoming more inconsistent with time or exhibit signs of fatigues such as by providing briefer responses (Meyer and Booker, 1991).
Availability bias	Availability bias arises from the differing ease with which events can be retrieved from long-term memory and affects people's ability to accurately estimate frequencies (of a class membership) or probabilities and recall other aspects of some events (Meyer and Booker, 1991). The more easily examples of events are recalled, the more probable this event is judged to be. For instance, examples from large classes are usually recalled better and faster than examples from less frequent classes, and likely occurrences are easier to imagine than unlikely ones (Garthwaite *et al.*, 2005). In addition, recall is also affected by factors such as familiarity, salience and recency (Garthwaite *et al.*, 2005). A potential problem with availability bias is indicated if the experts do not mention more than one or two considerations in giving their answers. If the experts	Stimulating the expert's memory associations can counter availability bias (Meyer and Booker, 1991). For example, discussion of the variables and information obtained from the literature, as well as fieldwork experiences, may help to trigger experts' memory/thinking. In addition, participation in both stages of model development should help. Other approaches include free association (similar to brainstorming) or hierarchical structuring of the presentation of information to flow from the general to the specific (to allow the maximum number of relevant associations to enter into the expert's final judgement) (Meyer and Booker, 1991). However, some more straightforward answers may only require a few considerations. Asking experts to 'think aloud' will help to identify this.

	only consider a few things, these were probably the most easily remembered and their answers are likely to be skewed to reflect these few (Meyer and Booker, 1991).	
Anchoring	Anchoring occurs when an individual fails to adjust sufficiently from his or her first impression in solving a problem (Meyer and Booker, 1991). People often start with a first impression and integrate information in a sequential manner, making only a few minor adjustments. If individuals who have already reached initial solutions receive contradictory data, they will probably not take this data sufficiently into account when generating final answers (Meyer and Booker, 1991). Assessments acquired this way are typically biased towards the starting value due to insufficient adjustment (Renooij, 2001). Anchoring bias should be suspected when an expert receives additional information from experts or other sources during the elicitation but never waivers from his or her first impression (Meyer and Booker, 1991).	Monitoring for signs of anchoring should be carried out during the elicitation and methods such as asking experts to 'think aloud' and justify their answers as much as possible as well as reassuring them that it is ok to change their answers at any point should help. Techniques similar to those used to counter availability bias are also used to counter anchoring (Meyer and Booker, 1991). Another technique involves asking the experts for their extreme judgements before obtaining their likely ones (Meyer and Booker, 1991). A similar approach is carried out for populating CPTs, by setting the extreme values first (Marcot *et al.*, 2006).
Representativeness heuristic	The representativeness heuristic is the process where people use the similarity of two events to estimate the degree to which one event is representative of the other and is often characterised as judgements made according to the similarity between instances and expectations of those instances (O'Hagan *et al.*, 2006).	This should not be as applicable, but asking experts to think aloud and explaining their thought processes may help to identify whether this heuristic is being used.

	Cognitive biases can be introduced by the representativeness heuristic, such as: the conjunction fallacy (when a more detailed description can seem to be more representative, though the conjunction of two events can never be more likely than the probability of either event alone); the gamblers fallacy (the belief that when a series of trials all have the same outcome then soon an opposite outcome will follow); and base-rate neglect (neglect of the relative frequency with which an event occurs).	
Underestimation of uncertainty/ overconfidence bias	Occurs when people underestimate the amount of uncertainty in the answers they provide (Meyer and Booker, 1991). Morgan *et al*. (2001) suggest that experts and non-experts tend to be overconfident, where, given their knowledge, their subjective probability distributions tend to be too narrow. Overconfidence is especially a problem with extreme probabilities (i.e. those close to 0 or 1) as people find them hard to assess Renooij (2001).	This may be reduced by asking the expert to further disaggregate the parts of the question and give estimates of the quantities of interest for each small part. In this way, the experts are less likely to overlook something (Meyer and Booker, 1991). Asking experts to provide a maximum and minimum probability value, as well as a 'best judgement', will help to determine the level of uncertainty in their answers (Meyer and Booker, 1991). However, this would provide a large number of values. Asking for (and recording of) explanation and justification for answers and the level of uncertainty will also help to minimise this. For example, if an expert appears to be very confident, but is not basing this on strong evidence then this could indicate overconfidence bias.
Hindsight bias	Hindsight bias can arise when people are asked to assess their *a priori* probability of an event that has actually	As for the other biases, asking experts to justify their answers (such as providing examples) and breaking up the questions

	occurred. Knowledge of what has occurred tends to distort memory and people tend to exaggerate their *a priori* probability for an event that has occurred (Garthwaite *et al.*, 2005). This may increase the likelihood of the other biases occurring.	may help to reduce the impact of this bias.

Table A-35. Main biases likely to occur in expert elicitation and how to minimise them.

Appendix 23. Use and development of verbal-numerical probability scale

A study by Witteman and Renooij (2003) showed that presenting subjects with a response scale that included both verbal and numerical labels for their probability assessments facilitated the assessment process. The accuracy of the assessments with the double scale was comparable to that of assessments with a numerical scale and people found the double scale more comfortable to use. Results indicated that assessments with the double scale remain stable over time, implying that the verbal labels do not cause random variation in the assessments. Stating probabilistic information may be a daunting task for experts when the questions are presented in a format that makes great demands on their cognitive processes. When their response mode preferences are taken into account by presenting them with the opportunity to choose whether to state their probabilities verbally or numerically the task becomes more feasible (Witteman and Renooij, 2003). Van der Gagg *et al*. (1999) also found that experts reported that the more uncertain they were about the probability to be assessed, the more they were inclined to think in terms of words. The verbal anchors on the scale then helped them to determine the position that they felt expressed the probability they had in mind.

Several adaptations were made to the original scale developed by Renooij and Witteman (1999). The modifications included those used by van der Gaag *et al*. (1999) and Witteman and Renooij (2003) of positioning the verbal labels so that they were not right beside the numerical labels. This meant that the verbal probability labels would not, incorrectly, be taken to be exact translations of precise numbers but as a set of labels with a stable rank-ordering, covering the whole probability continuum (Witteman and Renooij, 2003).

The words for the verbal labels were also changed. Witteman and Renooij (2003) note that different groups of subjects, from different professional domains, may prefer other words than the ones they used. The actual numerical interpretation of the words, which may vary per context, is of less importance than familiarity with terms, as the continuous scale allows the assessor to correct for effects of variable interpretation.

The original scale included words such as probable, expected and uncertain. However, it was decided to change the words to likely, unlikely etc, as it was felt that using the same base objective would enable experts to have a quick familiarity with the layout and orientation of the scale and instead of having to interpret and evaluate the different words, experts could more easily consider the scale as a continuum and thereby mentally move up and down the scale with ease (Windschitl and Wells, 1996). In addition, the use of 'uncertain' was not appropriate in this context as it may imply that the expert was uncertain about the probability of the habitat being suitable; unlikely was felt to be more appropriate. The moderator "(almost)" was added to the most extreme verbal expressions (certain and impossible) to indicate the positions of very small and very large probabilities, after van der Gaag *et al.* (1999).

More markers were added to the scale to help experts mark their answers more precisely as it was felt that only having markers around the numerical and verbal markers (which were not evenly spaced) may encourage users to cluster their markings around those areas. With the same reasoning, it was also decided to remove the numerical markers for 0.15 and 0.85 from the original scale) and just leave the 0, 0.25, 0.50, 0.75 and 1.0 markers, so the markers were evenly spaced. This would help to reduce any bias towards selecting values around the areas where there were more markers and numbers.

An additional category of 'don't know' was included on the sheet with the printed probability scale as O'Hagan *et al.* (2006) report that people often use the term 'fifty-fifty chance' as a proxy for 'don't know' rather than a probability of 0.5, but excessive (and arguably inappropriate) use of this 'fifty-fifty' response falls away when people are also provided with a 'don't know' category. In addition, if an expert was unsure, rather than experts giving 'guessed' answers which would not be very useful, such gaps in knowledge should be documented. Alternatively, experts could provide an upper and a lower estimate (and a best guess), if they were unsure about a probability value. An indication of the level of uncertainty could then be determined from the size of this range (Meyer and Booker, 1991). However, after running a pilot interview, it was decided not to use this approach if possible (see section 4.3.5, Chapter 4).

Although this verbal–numerical scale makes it difficult to express fine-grained probabilities (Witteman and Renooij, 2003), for this application that is not a problem as the small differences between numbers were not too important and small fractions of numbers were unlikely to be used.

Experts had a choice of whether they would like to use the verbal-numerical probability scale or not, for any of the questions. This could involve using it as a visual guide, or actually marking on the scale. The length of the verbal-numerical probability scale was to scale so that marked values could be extracted by measuring from one end with a ruler. An individual scale for each set (i.e. summing to 1) of probabilities to be elicited was provided (after Witteman and Renooij, 2003) to avoid a spacing effect, that is: people’s tendency to evenly or aesthetically distribute different assessments on one scale (von Winterfeldt and Edwards, 1986). A copy of the scale used can be found in Appendix 24.

Appendix 24. Verbal-numerical probability scale provided for probability elicitations for BBNs

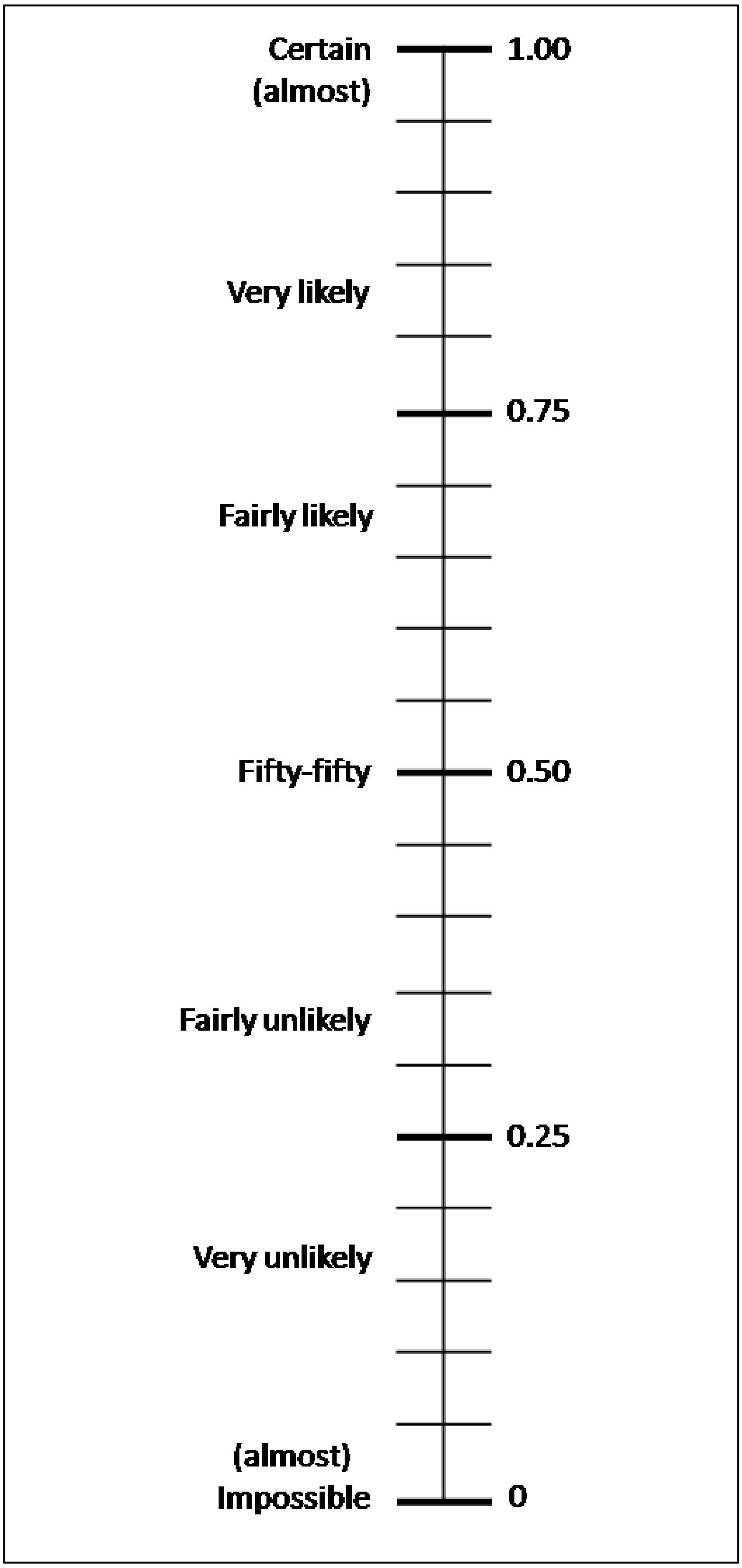


Figure A-63. Verbal-numerical probability scale.

Appendix 25. Notes on BBN structures

Several of the species occur in more than one particular habitat type, which can be very different, but the models were only developed for the main habitat in the New Forest. For example, the models did not take account of the calcareous grassland and sand dune habitats also used by *P. argus* in other parts of the country or the coastal habitat in which *C. nobile* also occurs.

C. nobile

All of the variables identified in Table A-26 (Appendix 20) were included in the BBN model apart from ‘soil fertility’ and ‘soil pH’. This was because it was felt that these variables would be more difficult to measure for the quick field surveys (to accompany the BBNs) and that the ‘soil type’ and ‘habitat type’ would partly account for these variables (and were both easily available as GIS layers). In addition, the definition of ‘suitable grassland’ includes a reference to the acidity of the grassland (which is largely indicated by the species present).

The ‘habitat type’ node has more states than the suggested 5, but not this is not a problem because it is the only parent node of the ‘suitable grassland’ node so does not make the CPT too complicated. The ‘vegetation suitability’ node also has 4 parent nodes rather than the recommended 3, but this was unavoidable and as two of the parent nodes only had 2 states and the others 3 and 4, it was not considered too much of a problem.

‘Ground vegetation height’ was the most difficult variable to assign the states to, and was based purely on expert judgement, but there was a general agreement among the experts. ‘Suitable grazing/mowing’ was initially going to be included as a node but it was decided that it was better to instantiate the ‘bare ground’ and ‘ground vegetation height’ nodes and just show ‘suitable grazing/mowing’ as an important influence on those variables, but not actually include it as having an effect in the model. In addition, population of these CPTs would not necessarily involve expert knowledge of the species, but more a knowledge of the likelihood of bare ground and of different vegetation heights occurring under different grazing regimes, which may be quite difficult to estimate. Also, suitable grazing/mowing may be more difficult to record than presence of bare ground and ground vegetation height.

The 'vegetation suitability' and 'suitable ground conditions' were included as proxy nodes to reduce the number of parent nodes for the 'habitat suitability for Chamomile' node and do not require instantiating.

G. constrictum

As for *C. nobile* (see above), it was decided not to include soil pH and soil fertility. The 'habitat type' node was included to partially take account of this (as was the 'soil type' node). Some of the habitat types were merged to reduce the number of states for this node although it was not possible to reduce the number to lower than 6. The main factor influencing the occurrence of *G. constrictum* was the presence of a wetland habitat which could occur in most of the main habitat types included in the 'habitat type' node (but would generally not occur within the GIS layer as a separate habitat type, e.g. ditches). Therefore, a separate node was created for 'wetland habitat' and 'habitat type' was still included to give an indication of the soil.

Ground vegetation height was not included as a variable because it was identified (see Table A-27, Appendix 20) that *G. constrictum* could occur in a wide range of vegetation heights). In addition, water pH was not included because in many of the types of wetland habitats in which *G. constrictum* tends to occur, laying water may not be present at the time of the survey. The 'vegetation suitability' and 'wetland habitat suitability' nodes were included as proxy nodes to reduce the number of parent nodes for the 'habitat suitability for slender marsh-bedstraw' node and do not require instantiating. The *G. constrictum* model was one of the most difficult to develop as there was the least literature available and was the least well known among experts.

G. illyricus

A few of the variables identified in Table A-28 (Appendix 20) were not included in the BBN. Distance to water was not included as the experts did not feel that this was as important. The same went for slope, particularly because the experts felt that there were not any very steep slopes in the New Forest. As for the other plant species, soil fertility and soil pH were not included, as the soil type and habitat type would partly indicate this.

The presence of suitable pollinator was also not included as the range of possible pollinators has not been fully established and *G. illyricus* can reproduce vegetatively. In addition, it is assumed that if the habitat is suitable for *G. illyricus*, then the conditions are largely suitable for its pollinators. A further issue is that recording the presence of pollinators would be more time-consuming. Grazing/bracken cutting were not included at this stage, as it was the presence of the correct density of bracken that was more important, and suitable grazing/bracken cutting could be included as a management/decision node, if required, in the future.

It was noted by several of the experts that *G. illyricus* occurs where there is bracken with grass underneath. The presence of grass was not included as a separate node as it is not directly important, but tends to be associated with the less vigorous bracken that *G. illyricus* favours and this is indicated by the bracken density categories (as vigorous has impoverished ground flora and less vigorous has a relatively well-developed ground flora) and the bracken litter depth.

Too much cover of ericaceous species (such as Heather) would decrease the likelihood of *G. illyricus* occurring. This is taken account of to some extent by the different heathland habitat types (HL1 and HL3) and also by the bracken density as ‘less vigorous’ bracken cover (the density favoured by *G. illyricus*) would be less likely if there was a lot of heather present. The ‘suitable habitat situation’, ‘suitable bracken/vegetation structure’ and ‘suitable ground conditions’ nodes were included as proxy nodes to reduce the number of parent nodes for the ‘habitat suitability for Wild Gladiolus’ node and do not require instantiating.

H. semele

All of the variables identified in Table A-29 (Appendix 20) were included in the BBN. Only ground moisture was included rather than soil type as the moisture was the most important aspect of the soil. Canopy cover was included as an additional variable, to take account of the preference of *H. semele* for open sites, but also that it does sometimes occur in Inclosures.

Burning and suitable grazing/disturbance are the most important factors for creating bare ground and for the presence of pioneer heath (along with habitat type). However, it is more important to instantiate ‘pioneer heath’ and bare ground’ (i.e. know their states accurately), so it was decided to just show ‘recent burning’ and ‘suitable grazing (or disturbance)’ as important factors influencing those variables, rather than actually include their influence in the CPTs. In addition (as discussed for *C. nobile*), populating the CPTs for these would not necessarily involve expert knowledge of the species, but more a knowledge of the likelihood of bare ground and of different heathland phases occurring under different grazing and burning regimes, which may be quite difficult to estimate. The ‘presence of suitable grass species’ does not require instantiation if ‘habitat type’ is instantiated. However, if ‘presence of suitable grass species’ is known then this can be instantiated to provide a more accurate indication, but if it is unknown then ‘habitat type’ can be used.
Habitat type also influences ‘pioneer heath’ but the probability of pioneer heath occurring cannot be based on habitat type alone (it also depends on ‘time since burning’ and ‘suitable grazing or disturbance’), so although the two are linked, ‘pioneer heath’ should be instantiated.

Also, ‘habitat type’ can be derived from the GIS layer so this may be more obtainable, and could give an assessment of the likelihood of there being suitable grass species for *H. semele* without actually visiting the site. There are also two proxy nodes, ‘grass suitability’ and ‘suitable vegetation structure’, which do not require instantiation.

N. sylvestris

‘Woodland type’ was used rather than leaf litter type because it is easier to determine from the GIS habitat layer or aerial photos, as well as at the site. ‘Woodland cover’ (i.e. percentage cover of woodland within the site) was used to indicate woodland size. Although the amount of woodland cover within a site may be low, it could form part of (or be close to) a larger woodland (or woodland edge) outside of the site. It may then be more suitable for *N. sylvestris*. However, it is the suitability within the actual site that is important, and how much woodland (and therefore edge) is present.

The ‘amount of woodland edge’ is one of the most important variables (see Table A-30, Appendix 20). Brouwers (2008) also identified the Euclidean distance to the nearest occupied (i.e. with *N. sylvestris*) permanent woodland edge as a significant variable, but after discussion with N. Brouwers (personal communication, January 9, 2009) during the expert interview, it was decided that this variable was not as important, and it was more important to focus on the edge present within the site. It would also be more difficult to record distance to an occupied edge as quickly. Although *N. sylvestris* might be present in a site without any woodland (edge), but woodland edge might be present just outside the site, the survey site would not actually be providing the suitable habitat, the adjacent site would. In addition, it is likely that if *N. sylvestris* was recorded in such situations, it would only be a few individuals who had strayed away from the woodland temporarily in good weather, and not large populations.

N. Brouwers (personal communication, January 9, 2009) also suggested that south-orientated canopy closure was not as important as just the presence of edge (in any orientation). Woodland age was also not included as it was more important that the associated features (i.e. the structure), such as good leaf litter and presence of edge, were present. This was also the case with presence of management/disturbance/grazing. However, this could be added as a decision node at a later date. Only ground vegetation cover was used as opposed to ground vegetation height as Brouwers (2008) found these two variables to be highly correlated, so it was unnecessary to include both and N. Brouwers (personal communication, January 9, 2009) suggested that cover was a better indicator of suitable habitat.

Both leaf litter depth and leaf litter volume were identified as important variables by Brouwers (2008), but leaf litter volume is a function of leaf litter depth and leaf litter cover, so it was decided to use these as the variables, rather than just leaf litter volume. This is because an area may have deep leaf litter, but only over a small area, so could still be very suitable. However, a low depth over a large area may not be so suitable. This can be taken account of in the leaf litter suitability CPT. The decision on cut-offs to use for the leaf litter depth states was one of the more difficult discretisations, and was based entirely on expert judgement.

The ‘woodland suitability’, ‘leaf litter suitability’ and ‘edge ground habitat suitability’ nodes were included as proxy nodes to reduce the number of parent nodes for the ‘habitat suitability for Wood Cricket’ node and do not require instantiating.

P. argus

The structure of the BBN for *P. argus* is fairly simple, but mainly because the most important variable for *P. argus* is the presence of pioneer heath. However, a lot of the information is contained within the definitions for each of the variables (see Appendix 26).

The heathland ericaceous species were not included as a separate variable because it appears that *P. argus* will use all three of the main species present in the New Forest (see Table A-31, Appendix 20), and this will also be largely accounted for by the ‘habitat type’ node. The soil type was not included as it was not really required because it was the presence of heathland that was most important. In addition, slope was not included as it was not as important as the other variables and is unlikely to be as relevant in the New Forest where there are very few steep slopes. The presence of shelter was also not included, as this was not as important and was partly taken account of by patch size/varied age structure.

Presence of *Lasius* ants is important for *P. argus*, so could be included as a variable affecting its habitat suitability. However, the variables included in the BBN that are important for *P. argus* are also important for *Lasius* ants as well, so that the model is also almost a model of *Lasius* ant habitat suitability. In addition, it would also take longer and be more difficult to survey for the presence (and perhaps density) of *Lasius* ants. If the model does not work well, then this is something that could potentially be added to try and improve the model.

P. globulifera

Most of the variables from Table A-32 (Appendix 20) were included in the BBN for *P. globulifera*. As for the other plant species, soil pH and soil fertility were not included. Absence of competitive vegetation/non-native species was not included as it was felt that this was largely accounted for by presence of ‘bare/open substrate’.

Altitude was also not included as *P. globulifera* can occur at all altitudes present in the New Forest. The absence of nitrate/phosphate pollution was not included, because again the presence of ‘bare/open substrate’ and ‘water pH’ were more important. However, if required, pollution could be included as a management node (or an environmental change node) later on. Just the ‘presence of wetland habitat’ was included rather than habitat type as well, as the wetland habitats could occur in any of the habitat types and the ‘soil type’ and ‘water acidity’ would largely indicate the soil acidity/fertility.

If the ‘disturbance/grazing’ and ‘water depth fluctuation’ nodes are instantiated it is not necessary to instantiate the ‘bare/open substrate’ node. However, it is actually the presence of bare/open substrate that is important, and although the presence of disturbance/grazing and water depth fluctuation are important, they do not mean that bare/open substrate will definitely be presence or absent (and to instantiate the ‘bare/open substrate’ CPT would not necessarily require expert knowledge of the species). It may be better to just instantiate ‘bare/open substrate’, in which case the ‘disturbance/grazing’ node is not required, other than to show how bare/open substrate might arise. Water depth fluctuation is also associated with water flow suitability but would not need to be linked to bare/open substrate.

‘Soil suitability’, ‘bank/shore suitability’, ‘water flow suitability’, ‘water suitability’ and ‘wetland habitat suitability’ were all included as proxy nodes to reduce the number of parent nodes into child nodes, as this model included a lot of nodes.

P. punctata

There were several variables from Table A-33 (Appendix 20) that were not included in the BBN for *P. punctata*. Liming/fertilisation was not included as it was felt that ‘habitat type’ would reflect this and was sufficient. Likewise, soil type was not included. Management/cutting/burning was not included as it is mainly important through its effect on ground vegetation height and it could potentially be included at a later date as a management node if required.

Dung structure, moisture and age were not included as if pony dung is present within a site there are likely to be different stages present (in terms of age and structure in particular) so the dung is likely to be suitable at some stage. In addition, *P. punctata* could be present in fresh dung but not fruiting because it has not had a chance to or the dung has not dried out enough. The dung moisture will be largely influenced by ground moisture. Canopy cover may also have an influence, but is not linked directly into dung suitability.

Rather than link ‘habitat type’ directly into vegetation characteristics suitability, where it would create a very complicated CPT (as there were more than 5 states for habitat type), it was decided to link ‘habitat type’ into a proxy node called ‘vegetation composition suitability’, so that probability of the site being suitable for *P. punctata* could be assigned just based on the habitat type, which was much more simple. ‘Vegetation structure suitability’ was used to reduce the number of parent nodes into ‘vegetation characteristics suitability’ and ‘vegetation characteristics suitability’ and ‘dung suitability’ were used as proxy nodes to reduce the number of parent nodes into ‘habitat suitability for Nail fungus’ and therefore do not require instantiating.

The BBN for *P. punctata* is partly indicating the preferences of ponies in where they are likely to ingest *P. punctata* spores and where they deposit their dung. Although these are likely to be slightly different locations (as the ponies move about), it will generally be similar habitat.

Appendix 26. Additional information/survey sheets (to accompany BBNs)

The likelihood function within Hugin (Hugin Expert A/S, 2007) can be used if there is more than one state present at the site for a variable, e.g. if there are two different habitat types, the proportion of each can be used as a likelihood.

Habitat suitability for *C. nobile*

VEGETATION (STRUCTURE) SUITABILITY - Proxy

HABITAT TYPE

Definition/Description: Habitat type(s) based on the HBIC GIS habitat layer classifications (see Appendix 5).

Why: *C. nobile* favours acid grassland found in a variety of habitats (Plantlife, 2001; Killick, 2002; Plantlife, 2007a).

States: (From HBIC GIS habitat layer – see Appendix 5)

- Unimproved acid grassland with 0% *Calluna* (GL11)
- Dry heathland/acid grassland mosaic (<25% *Calluna*) (HL3)
- Dry heathland (including humid heath) (>25% *Calluna*) (HL1)
- Wet heath (including wet heath/acid grassland mosaic) (HL2)
- Valley mire (AQ1)
- Unimproved neutral grassland (GL12)
- Semi-improved neutral grassland (GL13)
- Improved grassland (GL3)
- Woodland – Broadleaved (W1, W2), Coniferous (W5, W6), Mixed (W7, W8)
- Other – Any other habitat not included in any of the other categories

Measurement: Use GIS habitat layer or carry out a visual assessment at site (e.g. look for species that are characteristic of the different habitat types).

SUITABLE GRASSLAND

Definition/Description: Presence of open acidic (moderately acidic, to circum-neutral (C. Chatters, personal communication, March 13, 2009)) grassland within site, with not more than approximately 50% cover of shrubs (such as ericaceous species, gorse or bracken). The size of the area of grassland is not too important as *C. nobile* can occur in small patches (although the larger the area the better (particularly in the long-term) as it is more likely to persist (A. Byfield, personal communication, February 27, 2009)).

This sort of grassland favoured by *C. nobile* can be found in areas such as on sandy heaths, roadside and track verges (such as on the edge of heaths), grazed pastures and commons and also in open glades in (light) woodland, grassy tussocks in mires and on sports pitches (particularly cricket pitches, with regular cutting and rolling (Plantlife, 2001; C. Chatters, personal communication, March 13, 2009). It commonly occurs in the damper parts of New Forest lawns, notably those at settlement edges (Tubbs, 2001). It can also occur on coastal cliff tops, but this habitat type does not form part of this model.

Why: *C. nobile* favours acid grassland in the situations outlined above (Winship and Chatters, 1994; Plantlife, 2001; Tubbs, 2001; A. Byfield, personal communication, February 27, 2009; C. Chatters, personal communication, March 13, 2009).

States:

Present = Presence of suitable grassland
Absent = Absence of suitable grassland (no grassland present, or grassland present, but dominated by shrubs)

Measurement: Visual assessment of the site and assignment to the most appropriate category. This can be based on the habitat type and also on the presence of (grass) species characteristic of more acidic sites. Such aharacteristic species may include

Common Bent (*Agrostis capillaries*), Sweet Vernal grass (*Anthoxanthum odoratum*), Tormentil (*Potentilla erecta*), Wavy Hair grass (*Deschampsia flexuosa*) and Purple Moor-grass (*Molinia caerulea*) (A. Diaz, personal communication, May 14, 2008).

TREE CANOPY COVER

Definition/Description: Tree or large shrub (such as *Rhododendron ponticum*) canopy cover over the (majority of the) grassland (recorded for presence of suitable grassland) if present, or else over the site in general.

Why: *C. nobile* favours sunny, open places (Plantlife International, 2007a) and is rarely found where relative illumination in summer is less than 40% (Plantatt, 2004).

States:

Open = No tree canopy cover (or very scattered trees, with less than 30% cover)
Partially open = 30 – 70% tree canopy cover
Closed = More than 70% tree canopy cover

Measurement: Visual assessment, aided by the use of a spherical densityometer if required, and assignment to most appropriate category.

BARE GROUND

Definition/Description: Presence of patches of bare ground (in the open grassland, or else in general in the site) caused, for example, by trampling and other small-scale disturbance (such as by people, animals, and vehicles (Winship and Chatters, 1994)).

Why: *C. nobile* is a poor competitor and requires bare ground for establishment and colonisation (A. Byfield, personal communication, February 27, 2009). Poaching by animals helps to break up tight swards and allow the trampling-resistant *C. nobile* to colonise (Winship, 1998).

States:

Present = Presence of bare ground
Absent = No bare ground present in the site

Measurement: Visual assessment at the site.

GROUND VEGETATION HEIGHT

Definition/Description: Average height of the sward/ground vegetation (grasses, herbs etc.) in the open grassland. (If there are a mixture of heights, the proportion of each can be entered as likelihoods).

Why: *C. nobile* is unable to compete if taller plants dominate the sward (Winship and Chatters, 1994).

States:

< 4 cm = Average height of sward is less than 4 cm
4 – 8 cm = Average height of sward is between 4 cm and 8 cm
> 8 cm = Average height of sward is greater than 8 cm
Grass absent = No grassland present within the site.

Measurement: Measure average height of the open grassland sward (at several locations if required) using a measuring rule.

SUITABLE GRAZING/MOWING

(Decision node: as currently configured, this node does not have an effect on the values of its child nodes).

Definition/Description: Presence of regular/continual grazing or mowing at the site.

Why: Grazing or mowing help to keep down taller plants with which *C. nobile* cannot compete and help to create areas of open ground which it can colonise (Plantlife, 2001). *C. nobile* is adapted to a life with grazing and can flourish in relatively high nutrient situations providing that sufficient grazing, mowing or trampling occurs to prevent grasses and taller plants from dominating the sward (Winship and Chatters, 1994).

Unenclosed grazing also helps to spread *C. nobile*, as it roots easily from small fragments to form new plants (A. Byfield, personal communication, February 27, 2009). Grazing by large herbivores also helps to create poached ground, which can have a beneficial effect, by breaking up tight swards and allowing the trampling-resistant *C. nobile* to colonise (Winship, 1998; A. Byfield, personal communication, February 27, 2009). Grazing by sheep and rabbits also produces short sward but does not generate the poached ground, although disturbance by rabbits burrowing, moles and earthworms may produce the same bare ground (Winship, 1998).

Mowing helps to recreate the grazing environment required by *C. nobile* (C. Chatters, personal communication, March 13, 2009). In general, the cricket pitches and sports fields in which *C. nobile* is found are cut and rolled weekly; pressure from rolling ensures vegetative spread (Winship, 1998).

States:

Suitable = Presence of regular/continual grazing or mowing within the site

Unsuitable = Absence of, or infrequent (or very light), grazing or mowing within the site.

Measurement: Visual assessment; presence of grazing animals or evidence of grazing animals (e.g. dung, browsing of tree/shrub seedling/sapling shoots etc.) or mowing.

<u>SUITABLE SOIL</u> – Proxy

<u>SOIL TYPE</u>

Definition/Description: Main soil type within the site, according to the GIS layer categories (or use likelihood if more than one soil type within the site).

Why: *C. nobile* is a plant of dry, sandy or gleyed, moderately acidic soils, but a common requirement is to be seasonally wet (Winship and Chatters, 1994).

States: From soil GIS layer (see Appendix 6 for more details)

64301 – Deep sandy to clay
64303 – Deep loam to clay
71107 – Seasonally wet loam to clayey over shale
84102 – Seasonally wet deep loam
Other – None of the above soil types present.

Measurement: Determine from soil GIS layer (using an SPLUSIDIS extraction as carried out for the GLMs and GAMs (see section 3.2.1.2)).

GROUND MOISTURE

Definition/Description: Ground moisture in general (throughout most of the year) in the site.

Why: *C. nobile* can occur on dry soils as long as they are seasonally wet (Winship and Chatters, 1994) as well as on constantly moist or damp, to almost water-saturated, badly aerated soils (Hill *et al.*, 2004).

States: Based on the Ellenberg scale in Hill *et al*. (2004). (The moist categories (5 and 6) were combined with the damp categories (7 and 8) as there was unlikely to be much difference in suitability between them, and this created fewer states).

Dry = PLANTATT categories 1, 2, 3 and 4.
Moist/Damp = PLANTATT categories 5, 6, 7 and 8.
Wet = PLANTATT categories 9, 10, 11, 12.
Refer to Ellenberg scale in PLANTATT (2004) for more details.

Measurement: Visual assessment and assignment of most appropriate category. Vegetation may also be used to provide an indication, as well as topography. Obviously the time of year and recent weather events will influence ground moisture, but the aim is to determine the average ground moisture throughout the year (the ‘prone to winter flooding’ node is used to account for wetter winter ground moisture).

PRONE TO WINTER FLOODING

Definition/Description: Lying water for prolonged (at least several weeks) periods of time during the winter months.

Why: A common requirement of *C. nobile* is for seasonally wet grassland (Winship and Chatters, 1994; Killick, 2002).

States:

Yes = At least part of the site is prone to winter flooding

No = No part of the site floods for a prolonged period of time during the winter.

Measurement: Visual assessment at the site. However, this would require visits in winter. For summer visits, look for evidence that the site may be flooded in winter (such as for gorund moisture).

APPLICATION OF HERBICIDE

Definition/Description: Application of herbicide at the site either recently or not recently/never.

Why: Certain herbicides may kill *C. nobile* (Winship and Chatters, 1994; Plantlife, 2001).

States:

Recent = Application of herbicide within the last year, or evidence of herbicide application apparent

Not recent = No evidence of herbicide application.

Measurement: Visual assessment at the site. Look for evidence of death of plants and the type of habitat (open Forest is unlikely to have been sprayed).

Habitat suitability for *G. constrictum*

VEGETATION SUITABILITY – Proxy

TREE CANOPY COVER

Definition/Description: Tree or large shrub (such as *Rhododendron ponticum*) canopy cover over the (majority of the) wetland (recorded for presence of wetland habitat) if present, or else over the site in general.

Why: *G. constrictum* is a light-loving plant rarely found where relative illumination in summer is less than 40% (Hill *et al.*, 2004).

States:

Open = No tree canopy cover (or very scattered trees, with less than 30% cover)
Partially open = 30 – 70% tree canopy cover
Closed = More than 70% tree canopy cover

Measurement: Visual assessment, aided by the use of a spherical densityometer if required, and assignment to most appropriate category.

BARE GROUND

Definition/Description: Presence of patches of bare ground in or around (within a few metres) of the wetland habitat (recorded for presence of wetland habitat), if present, or else in general in the site caused, for example by grazing animals. (There does not need to be a minimum quantity of bare ground, and the patches can be small).

Why: *G. constrictum* requires open, bare ground to germinate and propagate, as it is a poor competitor (A. Byfield, personal communication, February 27, 2009; C. Chatters, personal communication, March 13, 2009).

States:

Present = Presence of bare ground
Absent = No bare ground present in the site.

Measurement: Visual assessment at the site.

WETLAND HABITAT

Definition/Description: Presence of any wet (or moist) 'wetland' habitat in which *G. constrictum* may occur in the site. This includes:

Ponds/pond margins (particularly those which dry out in summer)
Marshy places
Ditches and drainage channels (such as along roadsides)
Wet flood hollows
Wet New Forest lawns/edge of village greens
Wet/damp (edges of) grassy heaths
Disused marl/gravel pits

The 'wetland' habitat does not necessarily have to contain water at the time of the survey, but show evidence that it does sometimes contain water (for example, the roadside ditches often dry up in the summer).

Why: *G. constrictum* is a wet-site indicator (Hill *et al.*, 2004), occurring in wetland habitats such as those described above (Brewis *et al.*, 1996; Stace, 1997; Meek, 2002; C. Chatters, personal communication, March 13, 2009; G. Read, personal communication, February 13, 2009).

States:

Present = Presence of a wetland habitat such as those described above
Absent = Absence of any wetland habitat in the site.

Measurement: Visual assessment at the site. Although the presence of a wetland habitat may be less clear in the summer, when many wetland habitats may dry out, evidence of such habitats should be sought, such as bare muddy areas, or the presence of species characteristic of wetland habitats (rushes, sedges etc.).

GROUND MOISTURE

Definition: Ground moisture in general (throughout most of the year) in the site.

Why: *G. constrictum* is a wet-site indicator, often found on water-saturated, badly aerated soils (Hill *et al.*, 2004).

States: Based on the Ellenberg scale in Hill *et al*. (2004). (The moist categories (5 and 6) were combined with the damp categories (7 and 8) as there was unlikely to be much difference in suitability between them, and this created fewer states).

Dry = PLANTATT categories 1, 2, 3 and 4.
Moist/Damp = PLANTATT categories 5, 6, 7 and 8.
Wet = PLANTATT categories 9, 10, 11, 12.
Refer to Ellenberg scale in PLANTATT (2004) for more details.

Measurement: Visual assessment and assignment of most appropriate category. Vegetation may also be used to provide an indication, as well as topography. Obviously the time of year and recent weather events will influence ground moisture, but the aim is to determine the average ground moisture throughout the year.

SUITABLE SOIL – Proxy

SOIL TYPE

Definition/Description: Main soil type within the site, according to the GIS layer categories (or use likelihood if more than one soil type within the site).

Why: *G. constrictum* is associated with acidic to nearly neutral, fairly infertile, water-saturated and badly aerated soils (Hill *et al.*, 2004).

States: From soil GIS layer (see Appendix 6 for more details)

64301 – Deep sandy to clay
64303 – Deep loam to clay
71107 – Seasonally wet loam to clayey over shale
84102 – Seasonally wet deep loam
Other – None of the above soil types present.

Measurement: Determine from soil GIS layer (using an SPLUSIDIS extraction as carried out for the GLMs and GAMs (see section 3.2.1.2)).

HABITAT TYPE

Definition/Description: Habitat type(s) based on the HBIC GIS habitat layer classifications (see Appendix 5).

Why: The wetland features that *G. constrictum* favours can occur in a variety of habitats, although *G. constrictum* particularly occurs on grassy heaths, marshy areas and New Forest lawns (Brewis *et al.*, 1996; Meek, 2002).

States: (From HBIC GIS habitat layer – see Appendix 5)

Heath (including HL1 (dry and humid heath) and HL2 (wet heath, including wet heath/acid grassland mosaic)
Dry heath/acid grassland mosaic (<25% *Calluna*)/unimproved acid grassland with 0% *Calluna* (HL3, GL11)
Mire (in heathland situations) (AQ1)
Unimproved/semi-improved neutral grassland (GL12, GL13)
Woodland (W1, W2, W5, W6, W7, W8)
Other - Any other habitat not included in any of the other categories.

Measurement: Use GIS habitat layer or carry out a visual assessment at site (e.g. look for species that are characteristic of the different habitat types).

Habitat suitability for *G. illyricus*

HABITAT SITUATION SUITABILITY - Proxy

HABITAT TYPE

Definition/Description: Habitat type(s) based on the HBIC GIS habitat layer classifications (see Appendix 5).

Why: *G. illyricus* occurs on bracken-dominated acid grass heath and frequently on the boundary of heath and bog areas (Stokes, 1987).

States: (From HBIC GIS habitat layer – see Appendix 5)

- Dry heathland (including humid heath) (>25% *Calluna*) (HL1)
- Dry heathland/acid grassland mosaic (<25% *Calluna*) (HL3)
- Wet heath (including wet heath/acid grassland mosaic) (HL2)
- Mire (AQ1)
- Continuous bracken (GL8)
- Unimproved acid grassland with 0% *Calluna* (GL11)
- Unimproved neutral grassland (GL12)
- Semi-improved neutral grassland (GL13)
- Woodland – Broadleaved (W1, W2), Coniferous (W5, W6), Mixed (W7, W8)
- Other – Any other habitat not included in any of the other categories

Measurement: Use GIS habitat layer or carry out a visual assessment at site (e.g. look for species that are characteristic of the different habitat types).

DISTANCE TO WOODLAND

Definition/Description: Distance from (main bracken stand in) the site to the nearest woodland (of any type). If there is not bracken in the site, then the distance from the centre of the site to the nearest woodland.

Why: *G. illyricus* is often found near woodland edges (Stokes, 1987), or where there have formerly been woodlands (Tubbs, 2001) as these tend to result in the kind of species-rich bracken communities that *G. illyricus* favours and the fertile soils resulting from woodlands (A. Byfield, personal communication, February 27, 2009). *G. illyricus* is frequently found with other woodland community species, such as bluebell (*Hyacinthoides non-scripta*) (A. Byfield, personal communication, February 27, 2009; C. Chatters, personal communication, March 13, 2009).

States: Based on reports from Stokes (1987).

< 100 m = Woodland less than 100 metres from bracken in the site, or from the centre of the site if no bracken present (woodland within or adjacent to the site)

100 – 1000 m = Woodland between 100 and 1000 metres from bracken in the site, or from the centre of the site

> 1000 m = Woodland more than 1000 metres from bracken in the site, or from the centre of the site.

Measurement: Estimate the approximate distance to the nearest woodland (of any type) from the bracken stand (used to measure bracken density) in the site, or the centre of the site if there is no bracken present (or it is not in a stand). Measure from edge to edge. An OS map can be used to do this.

BRACKEN/VEGETATION STRUCTURE SUITABILITY - Proxy

TREE CANOPY COVER

Definition/Description: Tree or large shrub (such as *Rhododendron ponticum*) canopy cover over the (majority of the) bracken (recorded for ‘bracken density’) if present, or else over the site in general.

Why: *G. illyricus* does not occur in woodland under trees (A. Byfield, personal communication, February 27, 2009; C. Chatters, personal communication, March 13, 2009; G. Read, personal communication, February 13, 2009).

States:

Open = No tree canopy cover (or very scattered trees, with less than 30% cover)

Partially open = 30 – 70% tree canopy cover

Closed = More than 70% tree canopy cover

Measurement: Visual assessment, aided by the use of a spherical densityometer if required, and assignment to most appropriate category.

BRACKEN DENSITY

Definition/Description: Dominant density of bracken within the site (or proportion of each density if more than one density occurs within the site).

Why: Dense bracken deters over-intensive grazing by deep and ponies, as well as slugs, rabbits and cows (Stokes, 1987; 2000). This protection from grazing is important as *G. illyricus* is not well adapted to survive grazing (it is not unpalatable and does not have a basal rosette of leaves (A. Byfield, personal communication, February 27, 2009). *G. illyricus* appears to favour sites with less vigorous bracken (with grass underneath) as this provides protection from grazing but still allows quite a bit of sunlight through, providing warmth for the bulb (Stokes, 1987; A. Byfield, February 27, 2009). In addition, in these ‘open bracken’ areas, seed set is substantially higher, because pollinating bees and large skipper butterflies more readily find the flowers (Stokes, 2000). However, sex is rarely a successful strategy for this plant; most new recruitment to the population comes from vegetative reproduction (Stokes, 2000). If there’s not enough bracken, *G. illyricus* is more likely to get grazed, but if there’s too much then it may be too dark and *G. illyricus* may stop flowering (A. Byfield, personal communication, February 27, 2009). Flower heads flourish wherever the bracken canopy had been damaged by frost, and so failed to close into a dense stand by mid-June (Stokes, 2000).

States: Based on the three categories used by Stokes (1987) from Nicholson and Patterson (1976).

Vigorous = Bracken with a complete canopy cover and impoverished ground flora

Less vigorous = Bracken with an incomplete canopy cover and a relatively well-developed ground flora

Sparse = Bracken with an open canopy and a ground flora essentially the same as in adjacent areas where the bracken is absent

Absent = No bracken present within the site.

Measurement: Visual assessment at the site and assignment of the most appropriate state according to the descriptions. (If there is more than one density of bracken present within the site then record the proportion of each and enter as a likelihood). Record bracken with no tree canopy cover first, but if this is not present, then record density of bracken under tree cover. (Bracken growth peaks during late spring/summer and dies down during the winter months, so it may be more difficult to assess the density at that time. It would therefore be preferable to survey when *G. illyricus* is showing).

BRACKEN LITTER DEPTH

Definition/Description: Average depth of bracken litter under bracken (the same bracken used to record bracken density).

Why: Too much bracken litter suffocates *G. illyricus* and makes it more difficult for seeds and roots to penetrate and germinate. A mature plant may be able to tolerate some litter, but once the litter forms it tends to keep getting thicker (A. Byfield, personal communication, February 27, 2009). Stokes (1987) found a decrease in the number of *G. illyricus* plants per square metre with increasing thickness of the bracken litter layer. Bracken is also more prone to frost if there is an absence of bracken litter, which tends to create a more open bracken canopy (Stokes, 1987).

States: From the categories used by Stokes (1987):

Absent = No bracken litter layer present

< 5 cm = Bracken litter layer present, but less than 5 cm in depth

5 – 10 cm = Bracken litter layer depth between 5 and 10 cm

> 10 cm = Bracken litter layer greater than 10 cm deep.

Measurement: Use a measuring stick/probe to record the average depth of the litter under the bracken (take several measurements if required). There may be different depths of litter within the site (for example, under different densities of bracken or at the edge of bracken stands compared to the middle), so the proportion of each can be recorded and entered into the model as likelihood.

GROUND CONDITIONS SUITABILITY – Proxy

SOIL TYPE

Definition/Description: Main soil type within the site, according to the GIS layer categories (or use likelihood if more than one soil type within the site).

Why: Stokes (1987) reported that *G. illyricus* was found on fine sandy loam (a brown earth soil), relatively high in its clay and silt fractions compared to other soils in the Forest, but low in nutrients and less acidic .

States: From soil GIS layer (see Appendix 6 for more details)

64301 – Deep sandy to clay
64303 – Deep loam to clay
71107 – Seasonally wet loam to clayey over shale
84102 – Seasonally wet deep loam
Other – None of the above soil types present.

Measurement: Determine from soil GIS layer (using an SPLUSIDIS extraction as carried out for the GLMs and GAMs (see section 3.2.1.2)).

SOIL/GROUND MOISTURE

Definition/Description: Ground moisture in general (throughout most of the year) in the site.

Why: *G. illyricus* tends to occur on soils that are dry to moist (Hill *et al.*, 2004). Stokes (1987) suggests that the ground needs to be damp enough to prohibit bracken from becoming dominant, without being so wet that the corms rot.

States: Based on the Ellenberg scale in Hill *et al*. (2004). (The moist categories (5 and 6) were combined with the damp categories (7 and 8) as there was unlikely to be much difference in suitability between them, and this created fewer states).

Dry = PLANTATT categories 1, 2, 3 and 4.
Moist/Damp = PLANTATT categories 5, 6, 7 and 8.
Wet = PLANTATT categories 9, 10, 11, 12.
Refer to Ellenberg scale in PLANTATT (2004) for more details.

Measurement: Visual assessment and assignment of most appropriate category. Vegetation may also be used to provide an indication, as well as topography. Obviously the time of year and recent weather events will influence ground moisture, but the aim is to determine the average ground moisture throughout the year.

Habitat suitability for *H. semele*

GROUND MOISTURE

Definition/Description: Ground moisture in general (throughout most of the year) in the site.

Why: *H. semele* favours dry sites (Oates *et al.*, 2000; A. Barker, personal communication, March 6, 2009).

States: Based on the Ellenberg scale (as used for the plant species) in Hill *et al.* (2004). (The moist categories (5 and 6) were combined with the damp categories (7 and 8) as there was unlikely to be much difference in suitability between them, and this created fewer states).

Dry = PLANTATT categories 1, 2, 3 and 4.
Moist/Damp = PLANTATT categories 5, 6, 7 and 8.
Wet = PLANTATT categories 9, 10, 11, 12.
Refer to Ellenberg scale in PLANTATT (2004) for more details.

Measurement: Visual assessment and assignment of most appropriate category. Vegetation may also be used to provide an indication, as well as topography. Obviously the time of year and recent weather events will influence ground moisture, but the aim is to determine the average ground moisture throughout the year.

GRASS SUITABILITY – Proxy

PRESENCE OF TUSSOCKS (Present, Absent)

Definition/Description: Presence of grass tussocks in the site. There does not need to be a certain quantity, but just not homogenous/solid carpet of grass (A. Barker, personal communication, March 6, 2009).

Why: Female *H. semele* tend to select isolated clumps of grass for ovipositing, especially where two or three tussocks grow close together amongst bare ground. Solid carpets of bristle bent (*Agrostis curtisii*) are mostly ignored (Oates *et al.*, 2000).

States:

Present = Grass tussocks present in the site
Absent = No grass tussocks present in the site.

Measurement: Visual assessment at the site.

PRESENCE OF SUITABLE GRASS SPECIES

Definition/Description: Presence of grass species used as food plants or for ovipositing, including bristle bent (*Agrostis curtisii*) (the most commonly used as this tends to be present in the dry heaths that *H. semele* favours), purple moor-grass (*Molinia caerulea*) and wavy hair-grass (*Deschampsia flexuosa*), as well as sheep's fescue (*Festuca ovina*) for larvae (A. Barker, personal communication, March 6, 2009). (A particular quantity is not required but a larger area is likely to be less homogenous and have more variety, so would be better in the long term (A. Barker, personal communication, March 6, 2009).

Why: In the New Forest *H. semele* is strongly associated with bristle bent, but path-side tussocks of purple moor-grass and wavy hair-grass are also used (Oates *et al.*, 2000). The larvae feed on several grasses including bristle bent and sheep's fescue (Green, 2000).

States:

Present = Presence of suitable grass species (bristle bent, purple moor-grass, wavy hair-grass or sheep's fescue) within the site
Absent = None of the suitable grass species present within the site.

Measurement: Identify whether any of the suitable grass species present. (The habitat type will give an indication of this. For example, at least one of these species is likely to be present in dry heathland).

HABITAT TYPE

Definition/Description: Habitat type(s) based on the HBIC GIS habitat layer classifications (see Appendix 5).

Why: *H. semele* favours unimproved grassland in dry heaths, containing its food plants (in particular bristle bent) (Green, 2000; A. Barker, personal communication, March 6, 2009). Almost any dry New Forest heath can hold a large population of *H. semele* if the right habitat occurs and *H. semele* has such a strong status within the New Forest that small colonies can be found along sunny rides with suitable vegetation in Inclosures (Oates *et al.*, 2000). It does not appear to require a minimum area of this habitat, and can be found in patches of more open heath with bare ground amongst heath-dominated areas, but a larger area is likely to contain more variety and be less homogenous, so remain suitable in the longer-term (A. Barker, personal communication, March 6, 2009).

States: (From HBIC GIS habitat layer – see Appendix 5)

Dry/humid heath (> 25% *Calluna*) (HL1)
Dry heath/acid grassland mosaic (<25% *Calluna*) (HL3)
Wet heath (including wet heath/acid grassland mosaic) (HL2)
Mire (in heathland situations) (AQ1)
Unimproved acid grassland with 0% *Calluna* (GL11)
Woodland – Broadleaved (W1, W2), Coniferous (W5, W6), Mixed (W7, W8)
Other – Any other habitat not included in any of the other categories

Measurement: Use GIS habitat layer or carry out a visual assessment at site (e.g. look for species that are characteristic of the different habitat types).

SUITABLE VEGETATION STRUCTURE – Proxy

PIONEER HEATH

Definition/Description: Presence of heathland ericaceous species (heather (*Calluna vulgaris*), bell heather (*Erica cinerea*), cross-leaved heath (*Erica tetralix*)) in the pioneer phase of growth. The pioneer phase is defined by the JNCC (Joint Nature Conservation Committee, 2004) as: 'The establishment phase in which heather develops from seed into small pyramid-shaped plants. The height is usually less than 10-15 cm. Short (mown, burnt or grazed) swards can be included as 'pseudo-pioneer'.

Why: *H. semele* occurs in heaths in the early pioneer stage, particularly those which have been recently burnt (A. Barker, personal communication, March 6, 2009). It does not require a large area of pioneer heath; firebreak areas on the edge of taller heathland, verges of shorter heath next to taller heath and next to paths and tracks can be suitable.

States:

Present = Heathland ericaceous species present in the pioneer phase of growth.

Only later phases = Only later phase of heathland growth present – e.g. building, mature or degenerate (as defined by the JNCC (Joint Nature Conservation Committee, 2004)).

Absent = No heathland present within the site.

Measurement: Visual assessment of the site and assignment to the most appropriate category using the guidelines above.

BARE GROUND

Definition/Description: Presence (at least approximately 10% cover within the open pioneer heath/grassland) of bare ground, including bare pockets and along paths where vegetation is worn or heavily grazed (Oates *et al.*, 2000). Bare ground often occurs when heaths have been recently burned, but mature heather can also create bare ground and the edges can then be suitable (A. Barker, personal communication, March 6, 2009). Bare ground may also occur to a certain extent where grasses are not as dense (P. Brock, personal communication, July 17, 2009). For example, where grasses such as purple moor grass do not form a dense layer and bare ground can be clearly seen.

Why: *H. semele* often roosts on bare ground (to warm up, as bare ground tends to warm up more) (A. Barker, personal communication, March 6, 2009). Bare ground is also partly an indicator of the type of community; as the heath gets more mature the grass and heather closes in (A. Barker, personal communication, March 6, 2009).

States:

Present = Presence of bare ground (at least approximately 10% cover) in the open heath/grassland

Absent = No bare ground present in the open heath/grassland, or substantially less than 10% cover of bare ground.

Measurement: Visual assessment at the site and assignment of the most appropriate category. Quadrats may be used if required.

TIME SINCE BURNING

(Decision node: as currently configured, this node does not have an effect on the values of its child nodes).

Definition: Time since the heathland present in the site was burnt.

Why: *H. semele* is very much a species of managed heathland and is absent, or in the process of dying out, from neglected heaths (Oates *et al.*, 2000). On New Forest heaths it can be abundant where heathers are regenerating after burns, whereas in adjoining dense heather *H. semele* may be missing or present in sparse numbers. Regular burning (and grazing) helps to create the early pioneer stage heathland that *H. semele* favours and helps to create bare ground (where *H. semele* frequently roosts) as well as creating the dry woody debris on which *H. semele* often lays its eggs (A. Barker, personal communication, March 6, 2009). The programme of burning heathland at regular intervals obviously causes the species to move locally, as do unplanned heath fires (Oates *et al.*, 2000), but *H. semele* can return fairly quickly, once the grass starts to shoot.

States:

Recent = Burning of heathland within the site within approximately the last 5 years.
Not recent = No evidence of burning within approximately the last 5 years.
Absent = No heathland present within the site.

Measurement: Visual assessment at the site; evidence of recent burning.

SUITABLE GRAZING (OR DISTURBANCE)

(Decision node: as currently configured, this node does not have an effect on the values of its child nodes).

Definition: Presence of regular grazing by large herbivores or disturbance (such as by heavy machinery) at the site.

Why: Regular grazing (and burning) helps to create the early pioneer stage heathland and the bare ground that *H. semele* favours (A. Barker, personal communication, March 6, 2009). On the north-east Hampshire heaths, which are ungrazed by stock, it is heavily dependent on disturbance, particularly by MoD machinery (Oates *et al.*, 2000).

States:

Suitable = Presence of regular grazing by large herbivores or disturbance (if no or little grazing) at the site
Unsuitable = Absence of regular grazing or disturbance at the site.

Measurement: Visual assessment at the site; evidence of regular grazing (such as presence of ponies and cattle, dung from those animals, browsing of tree/shrub seedling/sapling shoots etc.) or disturbance (such as disturbed ground, such as track ruts).

TREE CANOPY COVER

Definition/Description: Tree or large shrub (such as *Rhododendron ponticum*) canopy cover over the (majority of the) heathland/grassland (recorded for presence of pioneer heath/suitable grass species) if present, or else over the site in general.

Why: *H. semele* occurs in habitats in open positions (Asher *et al.*, 2001), as well as in pockets of suitable vegetation along sunny woodland rides (Oates *et al.*, 2000).

States:

Open = No tree canopy cover (or very scattered trees, with less than 30% cover)
Partially open = 30 – 70% tree canopy cover
Closed = More than 70% tree canopy cover

Measurement: Visual assessment, aided by the use of a spherical densityometer if required, and assignment to most appropriate category.

Habitat suitability for *N. sylvestris*

AMOUNT OF WOODLAND EDGE

Definition/Description: Total amount of woodland (any type, although not just a very sparse scattering of trees) edge within the site, including the periphery of woodlands, edge of tracks through woodland, or areas of clearance. Edge is where woodland meets a different vegetation type, or where there is a clear opening in the tree canopy (where there is a clear view of the sky directly above when trees are in leaf) (N. Brouwers, personal communication, January 14, 2009).

Why: The edges of woodland have a warmer microclimate, which is favoured by *N. sylvestris* (N. Brouwers, personal communication, January 9, 2009).

States: (Largely based on suggestions from N. Brouwers (personal communication, January 9, 2009)).

< 50 m = Less than 50 metres of woodland edge within the site, but some woodland edge present
50 – 100 m = Between 50 and 100 metres of woodland edge within the site
> 100 m = More than 100 metres of woodland edge within the site
Absent = No woodland edge within the site.

Measurement: Visual assessment at the site, aided by using a measuring tape, or pacing out the edge(s) and then assigning the most relevant state or by using an OS map. Alternatively, the amount of edge in a site could be measured from (recent) aerial photos (such as Google Earth (http://earth.google.co.uk/intl/en_uk/), which also contains a measurement tool, which could be used to estimate the amount of edge). This would give an indication of whether it would be worth visiting a site to check the other habitat suitability variables. Tracks have edge on each side, but only one side should be included in the calculation.

WOODLAND SUITABILITY – Proxy

WOODLAND TYPE

Definition/Description: The main woodland type (tree composition, not including understorey layer, such as holly and hawthorn) according to the definitions of the states. It does not make a different if it is a woodland, coppiced woodland or a plantation. A minimum size is not required (and this will be accounted for by the 'woodland cover' variable), just a few scattered trees does not count as a woodland.

Why: *N. sylvestris* occurs on the edges of woodland in leaf litter, with a preference for broadleaved tree leaf litter, as dead leaves and perhaps associated fungi are the main food of this species (Marshall and Haes, 1988). Also, keeping some broadleaved trees on the edges of conifer stands appears to be beneficial for *N. sylvestris* can also be found under broadleaved trees on the edge of coniferous plantations in the New Forest (S. Douglas, personal observation).

States: Based on the definitions from the National Inventory of Woodland and Trees (Smith and Gilbert, 2001), as used by Brouwers (2008).

Broadleaved = Woodland containing more than 80% by area of broadleaved tree species (W1 and W2 from GIS habitat layer)

Mixed = Woodland containing a combination of broadleaved and coniferous tree species where each category occupies at least 20% of the canopy (W7 and W8 from GIS habitat layer)

Coniferous = Woodland containing more than 80% by area of coniferous tree species (W5 and W6 from GIS habitat layer)

Absent = No woodland present within the site.

Measurement: Visual assessment at the site. Alternatively, this can be determined from the HBIC GIS habitat layer (see Appendix 5) or from aerial photos. If more than one woodland type is present, the proportion of each should be recorded and entered into the model as likelihood values.

WOODLAND COVER

Definition: Percentage cover of any woodland habitat type (total of broadleaved, mixed or deciduous, of any stand density, apart from sparsely scattered trees that do not form a woodland) within the 100 m x 100 m site. (This is not a measure of canopy cover).

Why: Brouwers (2008) found that with an increase in woodland area, *N. sylvestris* was more likely to be present. A greater amount of woodland within a site is more likely to be suitable for wood cricket mainly because it is likely to have more suitable leaf litter and more edge and will be able to support larger populations.

States: Based on expert discussion with N. Brouwers (personal communication, January 14, 2009).

Low = Less than 20% woodland cover within site
Medium = 20 – 50% woodland cover within the site
High = Greater than 50% woodland cover within the site
Absent = No woodland present within the site.

Measurement: Visual assessment at the site to estimate the percentage cover and assign the most appropriate category. Alternatively this could be derived from the HBIC GIS habitat layer or from aerial photos (Google Earth). However, these may not be as up-to-date, but could be used to indicate whether it is worth visiting the site to examine the potential habitat suitability in more detail.

(EDGE) GROUND HABITAT SUITABILITY – Proxy

GROUND VEGETATION COVER

Definition: Average (throughout the site) percentage cover of ground vegetation (up to approximately dbh/130 cm, such as bracken (*Pteridium aquilinum*), brambles (*Rubus* spp.) etc., including different layers (e.g. bracken, grass species etc.)) at the woodland edge. (If there is not woodland edge present, then in general throughout the woodland, and if there is no woodland present, then in general throughout the site).

Why: Brouwers (2008) found that relatively high measures of vegetation height (and consequently vegetation cover, which was correlated) negatively influence sunlight availability at ground level, resulting in relatively lower air temperatures, which provides less suitable conditions for *N. sylvestris* which favours a warmer microclimate.

States: Based on the expert opinion of Brouwers (personal communication, January 14, 2009).

Low = 0 – 75% cover of ground vegetation at the woodland edge
High = Greater than 75% cover of ground vegetation at the woodland edge

Measurement: Visual assessment at the edge of the tree canopy and assignment of the most appropriate category. A 1 x 1m quadrat may be used to take measurements at the woodland edge, every 10m, if required, and the average recorded. If there is a substantial difference along different parts of the woodland edge then the proportion of each category can be recorded and entered into the model as likelihood values. (Soft (gradual) edges may be more difficult to determine, but measurements should be taken where there is a more dense line, rather than odd trees.)

GROUND MOISTURE

Definition/Description: Ground moisture in general (throughout most of the year) in the site.

Why: N. sylvestris is generally found in dry sites (P. Budd, personal communication, March 5, 2009; S. Douglas, personal observation). If the leaf litter was too wet for a prolonged period of time then the conditions may not be suitable (N. Brouwers, personal communication, January 14, 2009), for example, it may not provide the warm microclimate that *N. sylvestris* favours and may be detrimental to larvae survival (which are laid in the soil beneath the surface litter layers (Brown, 1978)).

States: Based on the Ellenberg scale in Hill *et al*. (2004). (The moist categories (5 and 6) were combined with the damp categories (7 and 8) as there was unlikely to be much difference in suitability between them, and this created fewer states).

Dry = PLANTATT categories 1, 2, 3 and 4.
Moist/Damp = PLANTATT categories 5, 6, 7 and 8.
Wet = PLANTATT categories 9, 10, 11, 12.
Refer to Ellenberg scale in PLANTATT (2004) for more details.

Measurement: Visual assessment and assignment of most appropriate category. Vegetation may also be used to provide an indication, as well as topography. Obviously the time of year and recent weather events will influence ground moisture, but the aim is to determine the average ground moisture throughout the year.

LEAF LITTER SUITABILITY – Proxy

LEAF LITTER DEPTH

Definition/Description: Average depth of leaf litter (broadleaved or herbaceous, including bracken (*Pteridium aquilinum*) litter) at the edge of the woodland. (If there is no woodland edge present, then in general throughout the woodland, and if there is no woodland present, then in general throughout the site).

Why: *N. sylvestris* occurs in a well-developed leaf litter (Brouwers, 2008) and dead leaves and perhaps associated fungi are the main food of this species (Marshall and Haes, 1988). Deeper leaf litter also provides protection against low winter temperatures (Proess and Baden, 2000).

States: Based on the expert opinion of N. Brouwers, guided by the findings of his thesis (Brouwers, 2008).

< 1 cm/Absent = No leaf litter layer present
1 – 5 cm = Average leaf litter layer between 1 cm and 5 cm deep
> 5 cm = Average leaf litter layer greater than 5 cm deep.

Measurement: Take measurements of the leaf litter depth, using a marked probe or measuring stick, every 10 metres along the woodland edge(s) (directly below the tree canopy edge) and take the average (mean) of these measurements. The frequency with which measurements are made can be reduced to every 20 metres if the edge is greater than 100 m, or increased if there is very little edge. Soft (gradual) edges may be more difficult to determine, but measurements should be taken where there is a more dense line, rather than odd trees (as for ground vegetation cover).

If there is low cover of leaf litter then measurements may need to be concentrated within a smaller area where the leaf litter is present and if there is a substantial difference in leaf litter depth along different parts of the woodland edge then the proportion of each category can be recorded and entered into the model as likelihood values. If there is no woodland edge, then record the average depth of leaf litter throughout the woodland should be recorded (take measurements as for the edge, but through the centre of the woodland, from one end of the site to the other).

LEAF LITTER COVER

Definition/Description: Average cover of leaf litter (broadleaved or herbaceous, including bracken (*Pteridium aquilinum*) litter), of any depth, at the edge of the woodland. (If there is not woodland edge present, then in general throughout the woodland, and if there is no woodland present, then in general throughout the site).

Why: Dead leaves and perhaps associated fungi are the main food of this species (Marshall and Haes, 1988) and a greater quantity of leaf litter will be able to support a greater number of *N. sylvestris*. Leaf litter cover is also be frequently correlated with leaf litter depth (Brouwers, 2008) and deeper leaf litter provides more protection against low winter temperatures (Proess and Baden, 2000).

States: Based on the expert opinion of N. Brouwers, guided by the findings of his thesis (Brouwers, 2008).

Absent = No leaf litter cover
Low = Less than 25% cover of leaf litter
Medium = 25% - 75% cover of leaf litter

High = Greater than 75% cover of leaf litter

Measurement: Make a visual assessment, whilst walking along the woodland edge(s), and assign the leaf litter cover to the most appropriate state, if it is obvious which category is most appropriate. If there is no woodland edge, then record the average leaf litter cover throughout the woodland (through the centre of the woodland, from one end of the site to the other).

If it is not clear which category is most appropriate, then measurements should be taken, using a 1m x 1m quadrat, every 10m along the woodland edge(s) (at the tree canopy edge, directly below) and take the average (mean (NB used mean)) of these measurements. The frequency with which measurements are made can be reduced to every 20 m if the edge is greater than 100 m, or increased if there is very little edge. Soft (gradual) edges may be more difficult to determine, but measurements should be taken where there is a more dense line, rather than odd trees (as for ground vegetation cover). If there is a substantial difference in leaf litter cover along different parts of the woodland edge then the proportion of each category can be recorded and entered into the model as likelihood values.

Habitat suitability for *P. argus*

This model does not account for the calcareous grassland habitats or coastal/sand dune habitats used by *P. argus* in other parts of the country, but instead just focuses on the heathland habitat on which it occurs in the New Forest.

SUITABLE ERICACEOUS SPECIES IN PIONEER PHASE

Definition/Description: Presence of ericaceous species composing at least one of the following species, composing at least approximately 10% cover of the heathland present within the site: cross-leaved heath (*Erica tetralix*), heather (*Calluna vulgaris*), (and possibly bell heather (*Erica cinerea*), if one of the other species was also present), in the pioneer (or early building) phase of growth. The pioneer phase is defined by the JNCC (Joint Nature Conservation Committee, 2004) as: 'The establishment phase in which heather develops from seed into small pyramid-shaped plants. The height is usually less than 10-15 cm. Short (mown, burnt or grazed) swards can be included as 'pseudo-pioneer''. The early building phase of growth occurs as the heather starts to form a more closed canopy. There should not be a high/dense cover (more than about 50%) of bracken (*Pteridium aquilinum*) or gorse (*Ulex* spp.).

Why: It appears that *P. argus* uses heather and cross-leaved heath (*Erica tetralix*) in the New Forest, and possibly bell heather (*Erica cinerea*) (Oates *et al.*, 2000). *P. argus* breeds on regenerating growth in recent clearings and burnt areas (Green, 2000). However, it may take 2-5 years for suitable conditions to develop after burning (Ravenscroft and Warren, 1996), such as for the ericaceous species to start flowering (A. Barker, personal communication, March 6, 2009). *P. argus* is generally absent from mature heaths, or confined to damp hollows and along rills, where the heather grows sparsely (Green, 2000). Tracks with shorter growth in more mature heath can help make more suitable conditions by creating warmer areas (A. Barker, personal communication, March 6, 2009).

P. argus requires heathland with a continual presence of early successional vegetation (in the pioneer or early building stages), which encourages good populations of the symbiotic *Lasius* ants (Joy, 1995; Ravenscroft and Warren, 1996). *L. alienus* is a species which only inhabits dry and especially warm pastures and heaths and as ants require warm conditions for foraging and for building their nests, the sandy soils of heaths provide ideal conditions (Joy, 1995). Within heathland sites ant nest densities tend to be highest on the least vegetated areas where high proportions of bare ground and short turf create a warm microclimate ideally suited to their needs (Joy, 1995).

States:

> Suitable = Presence of suitable ericaceous species in the pioneer (or early building) growth phase (as described above) within the site
>
> Unsuitable = Presence of only older heathland growth phases (such as mature) present within the site, with no (or less than 10% cover of) pioneer phases of the ericaceous species. Or high (>50%) cover of shrubs such as bracken or gorse. (Or no heathland present within the site).

Measurement: Visual assessment and assignment of heathland type to most appropriate category.

HABITAT TYPE

Definition/Description: Habitat type(s) based on the HBIC GIS habitat layer classifications (see Appendix 5).

Why: In Hampshire, *P. argus* breeds exclusively on heathland (dry and wet), including the edges of wet heaths and pockets of heathland in coniferous plantations (Oates *et al.*, 2000). Although mire is probably not used for breeding, *P. argus* may use it for feeding, particularly tussock islands (with ericaceous species on them), especially in drier years (A. Barker, personal communication, March 6, 2009).

States: (From HBIC GIS habitat layer – see Appendix 5)

> Dry/humid heath (> 25% *Calluna*) (HL1)
> Dry heath/acid grassland mosaic (<25% *Calluna*) (HL3)
> Wet heath (including wet heath/acid grassland mosaic) (HL2)

Mire (in heathland situations) (AQ1)

Woodland – Broadleaved (W1, W2), Coniferous (W5, W6), Mixed (W7, W8)

Other – Any other habitat not included in any of the other categories

Measurement: Use GIS habitat layer or carry out a visual assessment at site (e.g. look for species that are characteristic of the different habitat types).

SUITABLE MANAGEMENT (GRAZING/BURNING)

Definition/Description: Continuity of suitable management (regular grazing, rotation burning, periodic disturbance etc.; defined below).

Why: *P. argus* requires managed heathland and colonies die out on neglected heaths as mature heathers are unsuitable (Oates *et al.*, 2000). Grazing, rotational burning or periodic disturbance are essential (Oates *et al.*, 2000). Grazing can prolong suitable (pioneer) conditions on heathland and the presence of large herbivores may also help by providing some local soil disturbance (Ravenscroft and Warren, 1996); the New Forest colonies which are managed only by grazing are fairly stable, threatened only by short-term effects of uncontrolled fires and long term problems such as Silver Birch and Scots Pine invasion (Oates *et al.*, 2000). Although periodic, patchy/small scale burning can be beneficial, suitable conditions may take 2-5 years to develop after burning (Ravenscroft and Warren, 1996). Disturbance of the ground, especially if the topsoil is removed, can produce suitable conditions (Ravenscroft and Warren, 1996).

States:

Suitable = Regular grazing by large herbivores and rotational burning at the site

Unsuitable = No grazing, or only low-level or by small herbivores. No burning.

Measurement: Visual assessment at the site; evidence of regular grazing (such as presence of ponies and cattle, dung from those animals, browsing of tree/shrub seedling/sapling shoots etc.) and/or recent burning.

PRESENCE OF BARE GROUND

Definition/Description: Presence (not a large amount; at least approximately 5% cover within the open pioneer heath/grassland) of bare ground, including bare pockets and along paths where vegetation is worn or heavily grazed (Oates *et al.*, 2000). Bare ground often occurs when heaths have been recently burned, but mature heather can also create bare ground and the edges can then be suitable (A. Barker, personal communication, March 6, 2009). Bare ground may also occur to a certain extent where grasses are not as dense (P. Brock, personal communication, July 17, 2009). For example, where grasses such as purple moor grass do not form a dense layer and bare ground can be clearly seen.

Why: *P. argus* requires open ground for breeding and bare soil or short vegetation in all habitats in Britain (Ravenscroft and Warren, 1996). The eggs are usually laid along vegetation/bare ground margins where the vegetation is sparse and low (Joy, 1995). Bare ground is important for the *Lasius* ants with which *P. argus* is associated with, as within heathland sites *Lasius* ant nest densities tend to be highest on the least vegetated areas where high proportions of bare ground and short turf create a warm microclimate ideally suited to their needs (Joy, 1995). Bare ground is also partly an indicator of the type of community; as the heath becomes more mature the grass and heather closes in (A. Barker, personal communication, March 6, 2009).

States:

Present = Presence of bare ground (at least approximately 5% cover) in the open heath/grassland

Absent = No bare ground present in the open heath/grassland, or substantially less than 5% cover of bare ground.

Measurement: Visual assessment at the site and assignment of the most appropriate category. Quadrats may be used if required.

HEATHLAND PATCH SIZE/VARIED AGE STRUCTURE

Definition/Description: Whether heathland in the site forms part of a larger patch/tract (several hectares, at least approximately 5 ha) of heathland (containing a varied age structure of ericaceous shrubs).

Why: *P. argus* requires a flight area containing varied age structures (of heather, bell heather and/or other heaths) and there should be at least 5-10 (and if possible >50) hectares of habitat in suitable condition for *P. argus* (Joy, 1995). In smaller patches the ericaceous vegetation is likely to all go through same successional stage at once and could all become unsuitable at the same time. Whereas a larger patch is more likely to contain a range of successions (pioneer, building, mature and degenerate) and variety and is therefore more able to sustain populations in the long term, as *P. argus* can move to a more suitable area of the patch as part of it becomes unfavourable (A. Barker, personal communication, March 6, 2009).

States:

Larger patch = Site is part of (or linked to) a larger (continuous) patch/tract (at least approximately 5 hectares) of heathland.
Smaller patch = Site is part of a patch/tract of heathland that is less than approximately 5 hectares and is not linked (or close (several hundred metres) to) other heathland.
Absent = No heathland present within the site.

Measurement: Visual assessment; walk around edges of the site and outwards to get an idea of the surrounding heathland. An OS map (or aerial photos, if recent) can also be used to help.

Habitat suitability for *P. globulifera*

TREE CANOPY COVER

Definition/Description: Tree or large shrub (such as *Rhododendron ponticum*) canopy cover over the (majority of the) wetland (recorded for presence of wetland habitat) if present, or else over the site in general.

Why: *P. globulifera* is a light-loving plant rarely found where relative illumination in summer is less than 40% (Hill *et al.*, 2004).

States:

Open = No tree canopy cover (or very scattered trees, with less than 30% cover)
Partially open = 30 – 70% tree canopy cover
Closed = More than 70% tree canopy cover

Measurement: Visual assessment, aided by the use of a spherical densityometer if required, and assignment to most appropriate category.

SOIL SUITABILITY - Proxy

SOIL TYPE

Definition/Description: Main soil type within the site, according to the GIS layer categories (or use likelihood if more than one soil type within the site).

Why: *P. globulifera* occurs on clays, sands and peaty substrates (Plantlife, 2006), which tend to be acidic and infertile (Hill *et al.*, 2004).

States: From soil GIS layer (see Appendix 6 for more details)

64301 – Deep sandy to clay
64303 – Deep loam to clay

71107 – Seasonally wet loam to clayey over shale
84102 – Seasonally wet deep loam
Other – None of the above soil types present.

Measurement: Determine from soil GIS layer (using an SPLUSIDIS extraction as carried out for the GLMs and GAMs (see section 3.2.1.2)).

GROUND MOISTURE

Definition/Description: Ground moisture in general (throughout most of the year) in the site.

Why: *P. globulifera* is an indicator of shallow-water sites that may lack standing water for extensive periods (Hill *et al.*, 2004).

States: Based on the Ellenberg scale in Hill *et al*. (2004). (The moist categories (5 and 6) were combined with the damp categories (7 and 8) as there was unlikely to be much difference in suitability between them, and this created fewer states).

Dry = PLANTATT categories 1, 2, 3 and 4.
Moist/Damp = PLANTATT categories 5, 6, 7 and 8.
Wet = PLANTATT categories 9, 10, 11, 12.
Refer to Ellenberg scale in PLANTATT (2004) for more details.

Measurement: Visual assessment and assignment of most appropriate category. Vegetation may also be used to provide an indication, as well as topography. Obviously the time of year and recent weather events will influence ground moisture, but the aim is to determine the average ground moisture throughout the year.

WETLAND HABITAT SUITABILITY – Proxy

PRESENCE OF WETLAND HABITAT

Definition/Description: Presence of any wet (or moist) ‘wetland’ feature/habitat within the site. Examples include:

Ponds - in or around (shallow) pond margins (particularly those which dry out in summer)
Stream or river edges
Ditches and drainage channels
Muddy, wet track ruts (including in Inclosure rides)
Wet, churned (by animal hooves) muddy/clayey ground
Wet flood hollows
Poached wet grassland
(Open, edges of) bogs/mires, acid swamps
Old gravel and sand pits

The ‘wetland’ habitat does not necessarily have to contain water at the time of the survey, but show evidence that it does sometimes contain water (for example, the roadside ditches often dry up in the summer).

Why: *P. globulifera* occurs in wetland habitats or wet sites, such as those described above (Jermy, 1994; Brewis *et al.*, 1996; Preston and Croft, 1997; Stace, 1997; Scott *et al.*, 1999; Preston, 2002; UK Biodiversity Action Plan, 2008; A. Byfield, personal communication, February 27, 2009; C. Chatters, personal communication, March 13, 2009). (Some of these wetland feature types may be more long-lasting/sustainable than others – e.g. muddy ditch compared to pond).

States:

Present = Wetland habitat (such as described above) present in the site
Absent = No wetland habitat present in the site.

Measurement: Visual assessment at the site. Although the presence of a wetland habitat may be less clear in the summer, when many wetland habitats may dry out, evidence of such habitats should be sought, such as bare muddy areas, or the presence of species characteristic of wetland habitats (rushes, sedges etc.).

BANK/SHORE SUITABILITY – Proxy (Suitable, Unsuitable)

BARE/OPEN SUBSTRATE

Definition/Description: Presence of bare substrate (e.g. bare silt or mud, absent of vegetation) in or around (within a few metres of) the wetland feature (recorded for 'presence of wetland habitat'), for example, caused by poaching (by ponies) and other disturbance. The quanitity is not too important (small patches are sufficient, but the more the better).

Why: *P. globulifera* is a poor competitor, so bare ground provides areas to establish (A. Byfield, personal communication, February 27, 2009). Its creeping rhizomes are able to colonise bare mud rapidly (Scott *et al.*, 1999). Bare substrate may also indicate the absence of competitive vegetation or invasive/non-native species such as *Crassula helmsii* (New Zealand Pigmyweed), which is a threat. It may also indicate that the wetland feature is not clogged with organic matter, such as leaf litter.

States:

Present = Presence of bare substrate in or around (within a few metres of) the wetland feature

Absent = No bare substrate in or around the wetland feature, or no wetland feature present within the site.

Measurement: Visual assessment at the site.

DISTURBANCE/GRAZING

Definition/Description: Presence of regular/continual grazing by large herbivores at the site or disturbance, such as created by heavy vehicles.

Why: *P. globulifera* flourishes in habitats under a long tradition of heavy grazing (Plantlife, 2006) as trampling (poaching) by large herbivores helps to create the bare and open substrate it requires and reduces competition from more strongly growing aquatic species (Jermy, 1994; Scott *et al.*, 1999). The hollows/divots created by pony or cattle hooves also do not tend to dry out as much in summer, maintaining a ‘wetland microhabitat’ for *P. globulifera* (A. Byfield, personal communication, February 27, 2009). Similar disturbance may also be created by tyre tracks from heavy vehicles. Grazing also helps to keep ponds open and prevent encroachment by shrubs and woodland (A. Byfield, personal communication, February 27, 2009).

States:

Present = Presence of regular grazing by large herbivores at the site, or suitable disturbance (as described above)

Absent = Absence of regular grazing by large herbivores or suitable disturbance at the site.

Measurement: Visual assessment at the site; presence of large grazing animals or evidence of their presence (for example, dung) as well as evidence of trampling and disturbance, such as hoof marks and/or ruts in the ground (particularly arounf the wetland feature, if present).

BANK/SHORE SUBSTRATE

Definition/Description: The main type of substrate on the bank/shore (or edges) of the wetland feature, if present. (If there is more than one wetland feature present within the site with different substrates or there are different substrates present within the same wetland feature, record ‘fine’ first, or record the proportion of each (to enter as a likelihood), respectively).

Why: *P. globulifera* occurs on soft silty mud as coarse material would make it difficult to establish and spread as it is fairly delicate with runners (A. Byfield, personal communication, February 27, 2009).

States:

Fine = Soft, silty mud, fine particles

Coarse = Coarse, hard particles, gravelly. This includes substrates that contain a mix of coarse and fine particles. This also includes artificial shore lines (e.g. concrete)

Absent = No wetland feature present within the site.

Measurement: Visual assessment and assignment of most appropriate classification.

BANK/SHORE ANGLE

Definition/Description: The main angle of the bank/shore/edge of the wetland feature. (If there is more than one wetland feature present within the site with different bank/shore angles or there are different bank/shore angles present within the same wetland feature, record ‘near/ flat/gentle’ first, or record the proportion of each (to enter as a likelihood), respectively. Or record for the wetland feature for which bank/shore substrate was recorded).

Why: *P. globulifera* occurs on flat to gently sloping banks (Page, 1982) as water depth fluctuation will tend to occur more in water bodies with gentle banks/edges, exposing areas of bare mud, whereas a steep bank will be unlikely to have as much shore (A. Byfield, personal communication, February 27, 2009).

States: (Angles are a guide, and are partly based on the Common Standards Monitoring guidance for standing waters (Joint Nature Conservation Committee, 2005)).

Near flat/Gentle = Bank/shore angle less than 30° (This would include muddy track ruts or areas of churned ground)
Sloped = Bank/shore angle between 30° and 50°
Steep/near vertical/undercut = Bank/shore angle greater than 50° or undercut
Absent = No wetland feature.

Measurement: Visual assessment at the site. Use printed protractor as a guide to assign the most appropriate classification.

WATER SUITABILITY - Proxy

WATER ACIDITY

Definition/Description: Acidity (measured by pH) of the water in the wetland feature (if present).

Why: *P. globulifera* favours mildly-acidic water (Plantlife, 2006).

States:

High acidity = pH 1 – 3
Moderate acidity = pH 4 – 5
Low acidity/Circum-neutral = pH 6 – 7
Alkaline = pH 8 – 9
Absent = No wetland feature present.

Measurement: Use a pH meter or universal indicator paper to record the pH of the water. If there is no laying water present in the wetland feature at the time of the survey (e.g. an ephemeral pool that has dried up) and it is not possible to visit at another time, assume the low acidity category.

Water pH may vary slightly over the day or year (which may, in poorly buffered ponds, be due to the seasonal concentrations of dung and urine, together with the creation of dissolved oxygen by photosynthesis of vegetation on the pond floor, introducing bases that then shift the pH (C. Chatters, personal communication, March 13, 2009)), but as the categories include a range of pH values, just assign the category which is most appropriate.

WATER FLOW SUITABILITY - Proxy

SPEED OF WATER FLOW

Definition/Description: Speed of water flow in the wetland feature. (If there is more than one wetland feature present in the site, then record the slowest, or the proportion of each category to enter as likelihood values. Or record for the same wetland feature as for the bank/shore variables).

Why: The kind of habitats (such as ephemeral ponds) in which *P. globulifera* occurs are mostly non-flowing. In addition, fast-flowing water may remove *P. globulifera* (A. Byfield, personal communication, February 27, 2009).

States:

Still/Slow = No water flow (e.g. ponds or wet churned ground) or very gently flowing (such as shallow streams)

Steady = A steady flow of water (likely to occur in deeper streams or slower-flowing rivers)

Fast = Fast flowing water (such as in larger rivers)

Absent = No wetland feature (not even wet churned ground) present in the site.

Measurement: Visual assessment and assignment of the most appropriate classification for the most common speed of flow in the wetland feature. The speed of flow will depend to some extent on the time of year and recent rainfall, but the wetland feature will give an indication of the likely flow, or if it is likely that the flow differs greatly at different times then this can be entered as a likelihood into the model.

WATER DEPTH

Definition/Description: Average depth of water in the wetland feature. If there is a large variation in depth within the feature, then record the depth towards the edges of the wetland feature.

Why: *P. globulifera* favours seasonally dry shallow pools (Scott *et al.*, 1999; Plantlife, 2006). It cannot survive long-term in deep water, and although it can ‘cruise’ in deeper water, it forms weak growing mats (A. Byfield, personal communication, February 27, 2009).

States:

Shallow = Water less than 60 cm deep for most of the year (including wet muddy areas)

Deep = Water more than 60 cm deep for most of the year
Absent = No wetland feature present in the site.

Measurement: Visual assessment, aided by use of a measuring stick if required. Measurement should be taken within approximately one metre of the edge of the wetland feature. The edge of the wetland feature may change (during the year or due to recent weather conditions), but this is more likely to be for the shallower water bodies, which are likely to be <60 cm depth anyway. If more than one wetland feature is present within the site then record the shallowest, or record the proportion of each depth and enter as a likelihood into the model.

WATER DEPTH FLUCTUATION

Definition/Description: Whether the wetland feature is subject to fluctuating water levels, such as seasonally-dry (ephemeral) pools.

Why: *P. globulifera* favours shallow water, and fluctuating water levels tend to produce the fine, silty bare mud that it favours (Page, 1982; Plantlife, 2006). *P. globulifera* is a poor competitor and is soon 'swamped' by more strongly growing aquatic species unless external factors such as fluctuating water levels (or trampling by animals) serve to exclude these (Scott *et al.*, 1999). Although *P. globulifera* can withstand short periods of exposure when water levels fall, it is soon out-competed by terrestrial plants unless it is again inundated (Scott *et al.*, 1999).

States:

Yes = (Seasonal) fluctuation of water levels in the wetland feature
No = No (or not noticeable/significant) seasonal fluctuation of water levels in the wetland feature
Absent = No wetland feature present within the site.

Measurement: Visual assessment at the site. Look for evidence of water depth fluctuation, such as silty mud at the edges of pools. (Small pools and ponds are more likely to be subject to this).

Habitat suitability for *P. punctata*

VEGETATION CHARACTERISTICS SUITABILITY – Proxy

VEGETATION COMPOSITION SUITABILITY – Proxy

HABITAT TYPE

Definition/Description: Habitat type(s) based on the HBIC GIS habitat layer classifications (see Appendix 5).

Why: *P. punctata* tends to be found on dung on unimproved acid vegetation (Cox and Pickess, 1999; Poland, 2004). It will only occur where ponies are present and depositing dung, so is therefore more likely to occur where the ponies are more likely to be present. *P. punctata* may not generally be found in woodland as much because the ponies do not spend as much time there or because of the cooler, more moist microclimate conditions (A. Newton, personal communication, December 8, 2008; G. Dickson, July 21, 2009).

States: From HBIC GIS habitat layer – see Appendix 5 (woodland definitions are based on the National Inventory of Woodland and Trees (Smith and Gilbert, 2001) classifications).

Dry heathland (including humid heath) (>25% *Calluna*) (HL1)
Dry heathland/acid grassland mosaic (<25% *Calluna*) (HL3)
Unimproved acid grassland with 0% *Calluna* (GL11)
Wet heath (including wet heath/acid grassland mosaic) (HL2)
Mire (AQ1)
Unimproved neutral grassland (GL12)
Semi-improved neutral grassland (GL13)
Improved grassland (GL3)
Continuous bracken (GL8)
Broadleaved woodland (W1, W2) (>80% by area of broadleaved tree species)
Coniferous woodland (W5, W6) (>80% by area of coniferous tree species)

Mixed woodland (W7, W8) (A combination of broadleaved and coniferous tree species where each category occupies at least 20% of the canopy)
Other – Any other habitat not included in any of the other categories

Measurement: Use GIS habitat layer or carry out a visual assessment at site (e.g. look for species that are characteristic of the different habitat types).

SUITABLE VEGETATION STRUCTURE - Proxy

TREE CANOPY COVER

Definition/Description: Tree or large shrub (such as *Rhododendron ponticum*) canopy cover over the majority of the site in general. Or if there are open and partially open or closed areas, then the proportion of each.

Why: *P. punctata* may not be found in woodland as much because the ponies do not spend as much time in there (depositing dung) (A. Newton, personal communication, December 8, 2008; A. Lucas, personal communication, March 5, 2009; G. Dickson, personal communication, July 21, 2009) or because of the microclimate conditions (i.e. cooler and more moist (A. Newton, personal communication, December 8, 2008; G. Dickson, personal communication, July 21, 2009).

States:

Open = No tree canopy cover (or very scattered trees, with less than 30% cover)
Partially open = 30 – 70% tree canopy cover
Closed = More than 70% tree canopy cover

Measurement: Visual assessment, aided by the use of a spherical densityometer if required, and assignment to most appropriate category.

GROUND VEGETATION HEIGHT

Definition/Description: Average height of the ground vegetation (grasses, heather etc.) in the sites. (If there are a mixture of heights, the proportion of each can be entered as likelihoods).

Why: Poland (2004) found that the majority of *P. punctata* was found within vegetation with a maximum height between 10-15 cm and absent from vegetation above 30 cm. Cox and Pickess (1999) also mostly found *P. punctata* on short vegetation (e.g. where recently burnt), although some occurred amongst mature *Calluna* up to 30 cm high. However, they suggest that it is more likely to have been overlooked in taller vegetation. Ponies are also more likely to be found in shorter vegetation so this is where there is more likely to be dung and hence *P. punctata* (S. Skeates, personal communication, February 25, 2009). (Longer vegetation is also more likely to have a cooler and more moist microclimate).

States: Based on findings of the Poland (2004) study.

< 15 cm = Average ground vegetation height less than 15 cm
15 – 30 cm = Average ground vegetation height between 15 cm and 30 cm
> 30 cm = Average ground vegetation height greater than 30 cm.

Measurement: Measure average height of the ground vegetation (at several locations if required) using a measuring rule.

DUNG SUITABILITY – Proxy

GROUND MOISTURE

Definition/Description: Ground moisture in general (throughout most of the year) in the site.

Why: It appears that *P. punctata* may favour sites with a level of drainage impedance but not generally waterlogged; some moisture is required and it may not develop on arid sites (Poland, 2004). This may also be related to the greater occurrence of ponies on drier, rather than very waterlogged sites (G. Dickson, personal communication, July 21, 2009). The ground moisture will influence the dung moisture and how quickly it dries out.

States: Based on the Ellenberg scale in Hill *et al*. (2004). (The moist categories (5 and 6) were combined with the damp categories (7 and 8) as there was unlikely to be much difference in suitability between them, and this created fewer states).

Dry = PLANTATT categories 1, 2, 3 and 4.
Moist/Damp = PLANTATT categories 5, 6, 7 and 8.
Wet = PLANTATT categories 9, 10, 11, 12.
Refer to Ellenberg scale in PLANTATT (2004) for more details.

Measurement: Visual assessment and assignment of most appropriate category. Vegetation may also be used to provide an indication, as well as topography. Obviously the time of year and recent weather events will influence ground moisture, but the aim is to determine the average ground moisture throughout the year.

PRESENCE OF (DUNG FROM) PONIES FEEDING ON NATURAL VEGETATION

Definition/Description: Presence of ponies feeding on natural vegetation ('free-roaming' New Forest ponies) within the site. A minimum quantity of dung is not required, but regular use of the site by ponies will allow *P. punctata* to persist.

Why: *P. punctata* only occurs on the dung from wild 'free-roaming' ponies feeding on natural vegetation (Webster, 1999; Poland, 2004). These ponies are more likely to have ingested *P. puncata* spores from vegetation.

States:

Present = Ponies present at the site or evidence that ponies use the site frequently (e.g. presence of dung)
Absent = No ponies present at the site or no evidence that ponies use the site (e.g. dung).

Measurement: Visual assessment of whether ponies present within the site, or evidence that ponies occur on the site (i.e. presence of dung). (Ponies may not necessarily be present all the time as they move about daily and seasonally, so may not be present at the time of the survey).

Appendix 27. Conditional probability tables (CPTs) for BBNs with aggregated (average) values from expert probability elicitation

C. nobile

Suitable grassland

Habitat type	Unimproved acidic grassland	Dry heathland/ acid grassland	Dry heathland	Wet heathland	Valley mire	Unimproved neutral grassland	Semi-improved neutral grassland	Improved grassland	Woodland	Other
Present	1.00	0.75	0.30	0.27	0.00	0.33	0.23	0.02	0.02	0.02
Absent	0.00	0.25	0.70	0.73	1.00	0.67	0.77	0.98	0.98	0.98

Vegetation suitability

Grassland	Present							
Canopy cover	Open							
Vegetation height	< 4cm		4-8 cm		> 8cm		Grass absent	
Bare ground	Present	Absent	Present	Absent	Present	Absent	Present	Absent
Suitable	1.00	0.90	0.65	0.50	0.25	0.13	0.00	0.00
Unsuitable	0.00	0.10	0.35	0.50	0.75	0.87	1.00	1.00

Grassland	Present							
Canopy cover	Partially open							
Vegetation height	< 4cm		4-8 cm		> 8cm		Grass absent	
Bare ground	Present	Absent	Present	Absent	Present	Absent	Present	Absent
Suitable	0.68	0.52	0.47	0.33	0.13	0.10	0.00	0.00
Unsuitable	0.32	0.48	0.53	0.67	0.87	0.90	1.00	1.00

Grassland	Present							
Canopy cover	Closed							
Vegetation height	< 4cm		4-8 cm		> 8cm		Grass absent	
Bare ground	Present	Absent	Present	Absent	Present	Absent	Present	Absent
Suitable	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Unsuitable	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Grassland	Absent							
Canopy cover	Open							
Vegetation height	< 4cm		4-8 cm		> 8cm		Grass absent	
Bare ground	Present	Absent	Present	Absent	Present	Absent	Present	Absent
Suitable	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Unsuitable	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Grassland	Absent							
Canopy cover	Partially open							
Vegetation height	< 4cm		4-8 cm		> 8cm		Grass absent	
Bare ground	Present	Absent	Present	Absent	Present	Absent	Present	Absent
Suitable	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Unsuitable	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Grassland	Absent							
Canopy cover	Closed							
Vegetation height	< 4cm		4-8 cm		> 8cm		Grass absent	
Bare ground	Present	Absent	Present	Absent	Present	Absent	Present	Absent
Suitable	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Unsuitable	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Soil/ground suitability

Winter flooding	Yes									
Ground moisture	Dry					Moist/Damp				
Soil type	64301	64303	71107	84102	Other	64301	64303	71107	84102	Other
Suitable	0.68	0.72	0.85	0.37	0.25	0.87	0.90	1.00	0.47	0.30
Unsuitable	0.32	0.28	0.15	0.63	0.75	0.13	0.10	0.00	0.53	0.70

Winter flooding	Yes					No				
Ground moisture	Wet					Dry				
Soil type	64301	64303	71107	84102	Other	64301	64303	71107	84102	Other
Suitable	0.57	0.60	0.72	0.30	0.15	0.37	0.48	0.58	0.28	0.05
Unsuitable	0.43	0.40	0.28	0.70	0.85	0.63	0.52	0.42	0.72	0.95

Winter flooding	No									
Ground moisture	Moist/Damp					Wet				
Soil type	64301	64303	71107	84102	Other	64301	64303	71107	84102	Other
Suitable	0.48	0.58	0.73	0.32	0.25	0.40	0.45	0.55	0.28	0.13
Unsuitable	0.52	0.42	0.27	0.68	0.75	0.60	0.55	0.45	0.72	0.87

Habitat suitability for Chamomile

Vegetation suitability	Suitable				Unsuitable			
Soil/ground suitability	Suitable		Unsuitable		Suitable		Unsuitable	
Herbicide	Recent	Not recent	Recent	Not recent	Recent	Not recent	Recent	Not recent
Suitable	0.00	1.00	0.00	0.03	0.00	0.03	0.00	0.00
Unsuitable	1.00	0.00	1.00	0.97	1.00	0.97	1.00	1.00

G. constrictum

Vegetation suitability

Bare ground	Present			Absent		
Canopy cover	Open	Partially open	Closed	Open	Partially open	Closed
Suitable	1.00	0.60	0.05	0.60	0.30	0.00
Unsuitable	0.00	0.40	0.95	0.40	0.70	1.00

Wetland habitat suitability

Ground moisture	Dry				Moist/Damp				Wet			
Vegetation suitability	Suitable		Unsuitable		Suitable		Unsuitable		Suitable		Unsuitable	
Wetland habitat	Present	Absent	Present	Absent	Present	Absent	Present	Absent	Present	Absent	Present	Absent
Suitable	0.60	0.00	0.05	0.00	0.95	0.15	0.00	0.00	1.00	0.30	0.00	0.00
Unsuitable	0.40	1.00	0.95	1.00	0.05	0.85	1.00	1.00	0.00	0.70	1.00	1.00

Habitat suitability for Slender Marsh-bedstraw

Soil type	64301 - Deep sandy to clay											
Habitat type	Heath (dry/wet)		Dry heath/acid grassland mosaic/acid grassland		Mire		Unimproved/semi-improved neutral grassland		Woodland		Other	
Wetland suitability	Suitable	Unsuitable	Suitable	Unsuitable	Suitable	Unsuitable	Suitable	Unsuitable	Suitable	Unsuitable	Suitable	Unsuitable
Suitable	0.65	0.00	0.95	0.00	0.30	0.00	0.65	0.00	0.15	0.00	0.10	0.00
Unsuitable	0.35	1.00	0.05	1.00	0.70	1.00	0.35	1.00	0.85	1.00	0.90	1.00

Soil type	64303 - deep loam to clay											
Habitat type	Heath (dry/wet)		Dry heath/acid grassland mosaic/acid grassland		Mire		Unimproved/semi-improved neutral grassland		Woodland		Other	
Wetland suitability	Suitable	Unsuitable	Suitable	Unsuitable	Suitable	Unsuitable	Suitable	Unsuitable	Suitable	Unsuitable	Suitable	Unsuitable
Suitable	0.65	0.00	0.95	0.00	0.30	0.00	0.60	0.00	0.20	0.00	0.20	0.00
Unsuitable	0.35	1.00	0.05	1.00	0.70	1.00	0.40	1.00	0.80	1.00	0.80	1.00

Soil type	71107 - seasonally wet loam to clayey over shale											
Habitat type	Heath (dry/wet)		Dry heath/acid grassland mosaic/acid grassland		Mire		Unimproved/semi-improved neutral grassland		Woodland		Other	
Wetland suitability	Suitable	Unsuitable	Suitable	Unsuitable	Suitable	Unsuitable	Suitable	Unsuitable	Suitable	Unsuitable	Suitable	Unsuitable
Suitable	0.70	0.00	1.00	0.00	0.35	0.00	0.65	0.00	0.20	0.00	0.20	0.00
Unsuitable	0.30	1.00	0.00	1.00	0.65	1.00	0.35	1.00	0.80	1.00	0.80	1.00

Soil type	84102 - seasonally wet deep loam											
Habitat type	Heath (dry/wet)		Dry heath/acid grassland mosaic/acid grassland		Mire		Unimproved/semi-improved neutral grassland		Woodland		Other	
Wetland suitability	Suitable	Unsuitable	Suitable	Unsuitable	Suitable	Unsuitable	Suitable	Unsuitable	Suitable	Unsuitable	Suitable	Unsuitable
Suitable	0.40	0.00	0.50	0.00	0.20	0.00	0.35	0.00	0.20	0.00	0.20	0.00
Unsuitable	0.60	1.00	0.50	1.00	0.80	1.00	0.65	1.00	0.80	1.00	0.80	1.00

Soil type	Other											
Habitat type	Heath (dry/wet)		Dry heath/acid grassland mosaic/acid grassland		Mire		Unimproved/semi-improved neutral grassland		Woodland		Other	
Wetland suitability	Suitable	Unsuitable	Suitable	Unsuitable	Suitable	Unsuitable	Suitable	Unsuitable	Suitable	Unsuitable	Suitable	Unsuitable
Suitable	0.35	0.00	0.50	0.00	0.20	0.00	0.35	0.00	0.20	0.00	0.10	0.00
Unsuitable	0.65	1.00	0.50	1.00	0.80	1.00	0.65	1.00	0.80	1.00	0.90	1.00

G. illyricus

Suitable habitat situation

Distance to woodland	< 100m									
Habitat type	Dry heathland	Dry heath/ acid grass	Wet heathland	Mire	Continuous bracken	Unimproved acidic grass	Unimproved neutral grass	Semi-impr. neutral grass	Woodland	Other
Suitable	0.58	0.77	0.03	0.00	0.80	0.80	0.30	0.03	0.00	0.00
Unsuitable	0.42	0.23	0.97	1.00	0.20	0.20	0.70	0.97	1.00	1.00

Distance to woodland	100 - 1000m									
Habitat type	Dry heathland	Dry heath/ acid grass	Wet heathland	Mire	Continuous bracken	Unimproved acidic grass	Unimproved neutral grass	Semi-impr. neutral grass	Woodland	Other
Suitable	0.57	0.75	0.02	0.00	0.78	0.78	0.28	0.03	0.00	0.00
Unsuitable	0.43	0.25	0.98	1.00	0.22	0.22	0.72	0.97	1.00	1.00

Distance to woodland	> 1000m									
Habitat type	Dry heathland	Dry heath/ acid grass	Wet heathland	Mire	Continuous bracken	Unimproved acidic grass	Unimproved neutral grass	Semi-impr. neutral grass	Woodland	Other
Suitable	0.48	0.67	0.00	0.00	0.70	0.47	0.27	0.03	0.00	0.00
Unsuitable	0.52	0.33	1.00	1.00	0.30	0.53	0.73	0.97	1.00	1.00

Suitable bracken/vegetation structure

Canopy cover	Open							
Bracken density	Vigorous				Less vigorous			
Bracken litter depth	Absent	< 5cm	5-10cm	> 10cm	Absent	< 5cm	5-10cm	> 10cm
Suitable	0.20	0.18	0.02	0.00	1.00	0.78	0.37	0.10
Unsuitable	0.80	0.82	0.98	1.00	0.00	0.22	0.63	0.90

Canopy cover	Open							
Bracken density	Sparse				Absent			
Bracken litter depth	Absent	< 5cm	5-10cm	> 10cm	Absent	< 5cm	5-10cm	> 10cm
Suitable	0.70	0.50	0.23	0.03	0.00	0.00	0.00	0.00
Unsuitable	0.30	0.50	0.77	0.97	1.00	1.00	1.00	1.00

Canopy cover	Partially open							
Bracken density	Vigorous				Less vigorous			
Bracken litter depth	Absent	< 5cm	5-10cm	> 10cm	Absent	< 5cm	5-10cm	> 10cm
Suitable	0.00	0.00	0.00	0.00	0.25	0.13	0.05	0.00
Unsuitable	1.00	1.00	1.00	1.00	0.75	0.87	0.95	1.00

Canopy cover	Partially open							
Bracken density	Sparse				Absent			
Bracken litter depth	Absent	< 5cm	5-10cm	> 10cm	Absent	< 5cm	5-10cm	> 10cm
Suitable	0.20	0.10	0.03	0.00	0.00	0.00	0.00	0.00
Unsuitable	0.80	0.90	0.97	1.00	1.00	1.00	1.00	1.00

Canopy cover	Closed							
Bracken density	Vigorous				Less vigorous			
Bracken litter depth	Absent	< 5cm	5-10cm	> 10cm	Absent	< 5cm	5-10cm	> 10cm
Suitable	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Unsuitable	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Canopy cover	Closed							
Bracken density	Sparse				Absent			
Bracken litter depth	Absent	< 5cm	5-10cm	> 10cm	Absent	< 5cm	5-10cm	> 10cm
Suitable	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Unsuitable	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Suitable ground conditions

Soil type	64301 - deep sandy to clay			64303 - deep loam to clay			71107 - seasonally wet loam to clayey over shale		
Ground moisture	Dry	Moist/Damp	Wet	Dry	Moist/Damp	Wet	Dry	Moist/Damp	Wet
Suitable	0.42	0.37	0.00	0.72	0.67	0.00	0.55	0.27	0.00
Unsuitable	0.58	0.63	1.00	0.28	0.33	1.00	0.45	0.73	1.00

Soil type	84102 - Seasoanlly wet deep loam			Other		
Ground moisture	Dry	Moist/Damp	Wet	Dry	Moist/Damp	Wet
Suitable	0.20	0.18	0.00	0.10	0.12	0.00
Unsuitable	0.80	0.82	1.00	0.90	0.88	1.00

Habitat suitability for Wild Gladiolus

Situation suitability	Suitable				Unsuitable			
Ground suitability	Suitable		Unsuitable		Suitable		Unsuitable	
Structure suitability	Suitable	Unsuitable	Suitable	Unsuitable	Suitable	Unsuitable	Suitable	Unsuitable
Suitable	1.00	0.07	0.07	0.00	0.27	0.00	0.00	0.00
Unsuitable	0.00	0.93	0.93	1.00	0.73	1.00	1.00	1.00

H. semele

Presence of suitable grass species

Habitat type	Dry/humid heathland	Dry heath/ acid grassland mosaic	Wet heathland	Mire	Unimproved acidic grassland	Woodland	Other
Present	0.90	1.00	0.05	0.00	0.90	0.23	0.20
Absent	0.10	0.00	0.95	1.00	0.10	0.77	0.80

Grass suitability

Suitable grass species	Present		Absent	
Tussocks	Present	Absent	Present	Absent
Suitable	1.00	0.40	0.17	0.00
Unsuitable	0.00	0.60	0.83	1.00

Vegetation structure suitability

Canopy cover	Open					
Pioneer heath	Present		Only later phases		Heathland absent	
Bare ground	Present	Absent	Present	Absent	Present	Absent
Suitable	1.00	0.37	0.40	0.07	0.20	0.02
Unsuitable	0.00	0.63	0.60	0.93	0.80	0.98

Canopy cover	Partially open					
Pioneer heath	Present		Only later phases		Heathland absent	
Bare ground	Present	Absent	Present	Absent	Present	Absent
Suitable	0.36	0.12	0.10	0.00	0.10	0.00
Unsuitable	0.64	0.88	0.90	1.00	0.90	1.00

Canopy cover	Closed					
Pioneer heath	Present		Only later phases		Heathland absent	
Bare ground	Present	Absent	Present	Absent	Present	Absent
Suitable	0.00	0.00	0.00	0.00	0.00	0.00
Unsuitable	1.00	1.00	1.00	1.00	1.00	1.00

Habitat suitability for Grayling

Grass suitability	Suitable					
Ground moisture	Dry		Moist/Damp		Wet	
Vegetation strucuture	Suitable	Unsuitable	Suitable	Unsuitable	Suitable	Unsuitable
Suitable	1.00	0.27	0.30	0.10	0.00	0.00
Unsuitable	0.00	0.73	0.70	0.90	1.00	1.00

Grass suitability	Unsuitable					
Ground moisture	Dry		Moist/Damp		Wet	
Vegetation strucuture	Suitable	Unsuitable	Suitable	Unsuitable	Suitable	Unsuitable
Suitable	0.00	0.00	0.00	0.00	0.00	0.00
Unsuitable	1.00	1.00	1.00	1.00	1.00	1.00

N. sylvestris

Woodland suitability

Woodland type	Broadleaved Woodland				Mixed Woodland			
Woodland cover	0/Absent	Low	Medium	High	0/Absent	Low	Medium	High
Suitable	0.00	0.55	0.86	1.00	0.00	0.38	0.66	0.80
Unsuitable	1.00	0.45	0.14	0.00	1.00	0.62	0.34	0.20

Woodland type	Coniferous Woodland				Woodland Absent			
Woodland cover	0/Absent	Low	Medium	High	0/Absent	Low	Medium	High
Suitable	0.00	0.13	0.23	0.33	0.00	0.00	0.00	0.00
Unsuitable	1.00	0.87	0.77	0.67	1.00	1.00	1.00	1.00

Leaf litter suitability

Leaf litter cover	0%/Absent			Low (< 25%)		
Leaf litter depth	< 1cm/ Absent	1 - 5 cm	> 5cm	< 1cm/ Absent	1 - 5 cm	> 5cm
Suitable	0.00	0.00	0.00	0.00	0.40	0.58
Unsuitable	1.00	1.00	1.00	1.00	0.60	0.42

Leaf litter cover	Medium (25 - 75%)			High (> 75%)		
Leaf litter depth	< 1cm/ Absent	1 - 5 cm	> 5cm	< 1cm/ Absent	1 - 5 cm	> 5cm
Suitable	0.00	0.73	0.89	0.00	0.89	1.00
Unsuitable	1.00	0.27	0.11	1.00	0.11	0.00

Edge ground habitat suitability

Ground moisture	Dry				Moist/Damp			
Leaf litter suitability	Suitable		Unsuitable		Suitable		Unsuitable	
Ground veg. cover	Low	High	Low	High	Low	High	Low	High
Suitable	1.00	0.57	0.12	0.00	0.47	0.20	0.00	0.00
Unsuitable	0.00	0.43	0.88	1.00	0.53	0.80	1.00	1.00

Ground moisture	Wet			
Leaf litter suitability	Suitable		Unsuitable	
Ground veg. cover	Low	High	Low	High
Suitable	0.00	0.00	0.00	0.00
Unsuitable	1.00	1.00	1.00	1.00

Habitat suitability for Wood Cricket

Woodland edge	Absent				< 50m			
Woodland suitability	Suitable		Unsuitable		Suitable		Unsuitable	
Edge ground suitability	Suitable	Unsuitable	Suitable	Unsuitable	Suitable	Unsuitable	Suitable	Unsuitable
Suitable	0.45	0.09	0.10	0.00	0.73	0.16	0.23	0.00
Unsuitable	0.55	0.91	0.90	1.00	0.27	0.84	0.77	1.00

Woodland edge	50 - 100m				> 100m			
Woodland suitability	Suitable		Unsuitable		Suitable		Unsuitable	
Edge ground suitability	Suitable	Unsuitable	Suitable	Unsuitable	Suitable	Unsuitable	Suitable	Unsuitable
Suitable	0.94	0.26	0.30	0.00	1.00	0.30	0.35	0.00
Unsuitable	0.06	0.74	0.70	1.00	0.00	0.70	0.65	1.00

P. argus

Suitable ericaceous species in pioneer phase

Habitat type	Dry/humid heath		Dry heath/acid grassland mosaic		Wet heath	
Management /grazing	Suitable	Unsuitable	Suitable	Unsuitable	Suitable	Unsuitable
Suitable	0.90	0.45	0.33	0.08	1.00	0.70
Unsuitable	0.10	0.55	0.67	0.92	0.00	0.30

Habitat type	Mire		Woodland		Other	
Management /grazing	Suitable	Unsuitable	Suitable	Unsuitable	Suitable	Unsuitable
Suitable	0.03	0.00	0.20	0.08	0.00	0.00
Unsuitable	0.97	1.00	0.80	0.92	1.00	1.00

Bare ground

Management /grazing	Suitable	Unsuitable
Present	1.00	0.30
Absent	0.00	0.70

Habitat suitability for Silver-studded Blue

Ericaceous/ pioneer	Present					
Heathland patch size	Larger		Smaller		Absent	
Bare ground	Present	Absent	Present	Absent	Present	Absent
Suitable	1.00	0.80	0.85	0.65	0.00	0.00
Unsuitable	0.00	0.20	0.15	0.35	1.00	1.00

Ericaceous/ pioneer	Absent					
Heathland patch size	Larger		Smaller		Absent	
Bare ground	Present	Absent	Present	Absent	Present	Absent
Suitable	0.20	0.10	0.00	0.00	0.00	0.00
Unsuitable	0.80	0.90	1.00	1.00	1.00	1.00

P. globulifera

Soil suitability

Ground moisture	Dry					Moist/Damp				
Soil type	64301	64303	71107	84102	Other	64301	64303	71107	84102	Other
Suitable	0.07	0.10	0.17	0.10	0.02	0.72	0.62	0.75	0.42	0.35
Unsuitable	0.93	0.90	0.83	0.90	0.98	0.28	0.38	0.25	0.58	0.65

Ground moisture	Wet				
Soil type	64301	64303	71107	84102	Other
Suitable	0.93	0.83	0.97	0.67	0.50
Unsuitable	0.07	0.17	0.03	0.33	0.50

Bare/open substrate

Disturbance/ grazing	Present		Absent	
Water depth fluctuation	Yes	No	Yes	No
Present	1	0.8	0.5	0.2
Absent	0	0.2	0.5	0.8

Bank/shore suitability

Bare/open substrate	Present					
Bank/shore angle	Near flat/gentle			Sloped		
Bank/shore substrate	Fine	Coarse	Absent	Fine	Coarse	Absent
Suitable	1.00	0.10	0.00	0.70	0.05	0.00
Unsuitable	0.00	0.90	1.00	0.30	0.95	1.00

Bare/open substrate	Present					
Bank/shore angle	Steep/near vertical/undercut			Absent		
Bank/shore substrate	Fine	Coarse	Absent	Fine	Coarse	Absent
Suitable	0.03	0.00	0.00	0.00	0.00	0.00
Unsuitable	0.97	1.00	1.00	1.00	1.00	1.00

Bare/open substrate	Absent					
Bank/shore angle	Near flat/gentle			Sloped		
Bank/shore substrate	Fine	Coarse	Absent	Fine	Coarse	Absent
Suitable	0.00	0.00	0.00	0.00	0.00	0.00
Unsuitable	1.00	1.00	1.00	1.00	1.00	1.00

Bare/open substrate	Absent					
Bank/shore angle	Steep/near vertical/undercut			Absent		
Bank/shore substrate	Fine	Coarse	Absent	Fine	Coarse	Absent
Suitable	0.00	0.00	0.00	0.00	0.00	0.00
Unsuitable	1.00	1.00	1.00	1.00	1.00	1.00

Water flow suitability

Speed of water flow	Still/slow					
Water depth fluctuation	Yes			No		
Water depth	Shallow	Deep	Absent	Shallow	Deep	Absent
Suitable	1.00	0.33	0.00	0.80	0.22	0.00
Unsuitable	0.00	0.67	1.00	0.20	0.78	1.00

Speed of water flow	Steady					
Water depth fluctuation	Yes			No		
Water depth	Shallow	Deep	Absent	Shallow	Deep	Absent
Suitable	0.50	0.07	0.00	0.33	0.03	0.00
Unsuitable	0.50	0.93	1.00	0.67	0.97	1.00

Speed of water flow	Fast					
Water depth fluctuation	Yes			No		
Water depth	Shallow	Deep	Absent	Shallow	Deep	Absent
Suitable	0.00	0.00	0.00	0.00	0.00	0.00
Unsuitable	1.00	1.00	1.00	1.00	1.00	1.00

Speed of water flow	Absent					
Water depth fluctuation	Yes			No		
Water depth	Shallow	Deep	Absent	Shallow	Deep	Absent
Suitable	0.00	0.00	0.00	0.00	0.00	0.00
Unsuitable	1.00	1.00	1.00	1.00	1.00	1.00

Water suitability

Water acidity	High		Moderate		Low	
Water flow suitability	Suitable	Unsuitable	Suitable	Unsuitable	Suitable	Unsuitable
Suitable	0.00	0.00	0.38	0.03	1.00	0.03
Unsuitable	1.00	1.00	0.62	0.97	0.00	0.97

Water acidity	Very low / alkaline		Absent	
Water flow suitability	Suitable	Unsuitable	Suitable	Unsuitable
Suitable	0.01	0.00	0.00	0.00
Unsuitable	0.99	1.00	1.00	1.00

Wetland habitat suitability

Wetland habitat	Present				Absent			
Water suitability	Suitable		Unsuitable		Suitable		Unsuitable	
Bank/shore suitability	Suitable	Unsuitable	Suitable	Unsuitable	Suitable	Unsuitable	Suitable	Unsuitable
Suitable	1.00	0.07	0.00	0.00	0.00	0.00	0.00	0.00
Unsuitable	0.00	0.93	1.00	1.00	1.00	1.00	1.00	1.00

Habitat suitability for Pillwort

Soil suitability	Suitable					
Canopy cover	Open		Partially open		Closed	
Wet. habitat suitability	Suitable	Unsuitable	Suitable	Unsuitable	Suitable	Unsuitable
Suitable	1.00	0.00	0.30	0.00	0.00	0.00
Unsuitable	0.00	1.00	0.70	1.00	1.00	1.00

Soil suitability	Unsuitable					
Canopy cover	Open		Partially open		Closed	
Wet. habitat suitability	Suitable	Unsuitable	Suitable	Unsuitable	Suitable	Unsuitable
Suitable	0.23	0.00	0.03	0.00	0.00	0.00
Unsuitable	0.77	1.00	0.97	1.00	1.00	1.00

P. punctata

Vegetation composition suitability

Habitat type	Dry/humid heath	Dry heath/acid grassland mosaic	Unimproved acid grassland	Wet heathland	Mire	Unimproved neutral grassland	Semi-improved neutral grassland
Suitable	0.96	1.00	0.68	0.58	0.28	0.26	0.00
Unsuitable	0.04	0.00	0.32	0.42	0.72	0.74	1.00

Habitat type	Improved grassland	Continuous bracken	Broadleaved woodland	Coniferous woodland	Mixed woodland	Other habitat
Suitable	0.00	0.13	0.09	0.03	0.05	0.02
Unsuitable	1.00	0.87	0.91	0.97	0.95	0.98

Vegetation structure suitability

Vegetation height	< 15cm			15 - 30cm			> 30cm		
Canopy cover	Open	Partially open	Closed	Open	Partially open	Closed	Open	Partially open	Closed
Suitable	1.00	0.47	0.16	0.68	0.28	0.08	0.42	0.14	0.00
Unsuitable	0.00	0.53	0.84	0.32	0.72	0.92	0.58	0.86	1.00

Vegetation characteristics suitability

Vegetation structure	Suitable		Unsuitable	
Vegetation composition	Suitable	Unsuitable	Suitable	Unsuitable
Suitable	1.00	0.12	0.19	0.00
Unsuitable	0.00	0.88	0.81	1.00

Dung suitability – average from all experts

Ground moisture	Dry		Moist/damp		Wet	
Presence of ponies/dung	Present	Absent	Present	Absent	Present	Absent
Suitable	0.75	0.00	0.86	0.00	0.40	0.00
Unsuitable	0.25	1.00	0.14	1.00	0.60	1.00

Dung suitability – average from only experts who thought ‘dry’ ground moisture was more suitable

Ground moisture	Dry		Moist/damp		Wet	
Presence of ponies/dung	Present	Absent	Present	Absent	Present	Absent
Suitable	1.00	0.00	0.71	0.00	0.03	0.00
Unsuitable	0.00	1.00	0.29	1.00	0.97	1.00

Dung suitability – average from only experts who thought ‘moist’ ground moisture was more suitable

Ground moisture	Dry		Moist/damp		Wet	
Presence of ponies/dung	Present	Absent	Present	Absent	Present	Absent
Suitable	0.63	0.00	1.00	0.00	0.58	0.00
Unsuitable	0.37	1.00	0.00	1.00	0.42	1.00

Habitat suitability for Nail fungus

Vegetation characteristics	Suitable		Unsuitable	
Dung suitability	Suitable	Unsuitable	Suitable	Unsuitable
Suitable	1.00	0.02	0.35	0.00
Unsuitable	0.00	0.98	0.65	1.00

Appendix 28. Results of testing BBNs with fieldwork data – habitat suitability values for each species for each presence and absence site

	Presence sites	**Absence sites**
1	70.18	38.89
2	87.39	49.56
3	86.42	25.67
4	43.74	16.81
5	33.71	0.84
6	81.57	5.72
7	17.55	3.60
8	78.18	16.12
9	78.18	10.56
10	70.79	10.71
11	92.72	1.74
12	7.08	49.56
13	90.30	2.58
14	59.26	2.07
15	39.78	30.16
16	26.80	38.89
17	24.83	0.45
18	97.84	25.50
19	90.30	8.16
20	15.84	12.75
Average	59.62	17.52
Range	7.08 – 97.84	0.45 – 49.56

Table A-36. Habitat suitability values for the fieldwork data applied to the *C. nobile-1* BBN.

	Presence sites	**Absence sites**
1	35.65	11.39
2	25.50	37.53
3	82.02	19.53
4	43.74	16.81
5	11.24	0.88
6	42.79	5.72
7	17.55	3.60
8	42.15	3.92
9	59.21	4.28
10	70.79	2.92
11	12.13	1.74
12	2.41	15.88
13	67.00	6.60
14	24.72	0.69
15	31.45	9.64
16	1.62	14.99
17	6.54	0.45
18	74.13	10.92
19	80.89	1.62
20	12.07	1.83
Average	37.18	8.55
Range	1.62 – 82.02	0.45 – 37.53

Table A-37. Habitat suitability values for the fieldwork data applied to the *C. nobile-2* BBN.

	Presence sites	**Absence sites**
1	62.98	40.28
2	66.50	0.00
3	92.62	0.00
4	47.50	0.00
5	11.40	0.00
6	30.00	0.00
7	15.79	0.00
8	61.75	32.94
9	66.30	62.00
10	32.30	1.66
11	30.00	39.00
12	25.50	0.00
13	39.00	0.00
14	38.00	3.73
15	47.50	4.50
16	21.87	0.00
17	19.00	0.00
18	42.40	5.38
19	47.50	0.00
20	33.01	2.95
Average	41.55	9.62
Range	11.40 – 92.62	0 – 62.00

Table A-38. Habitat suitability values for the fieldwork data applied to the *G. constrictum* BBN.

	Presence sites	**Absence sites**
1	3.61	12.11
2	23.35	0.07
3	35.14	23.77
4	60.33	1.32
5	2.66	3.78
6	55.16	28.95
7	34.40	0.00
8	12.13	0.47
9	3.54	0.56
10	5.41	0.00
11	11.30	16.91
12	13.69	0.67
13	12.53	2.04
14	32.07	42.00
15	33.02	1.04
16	42.28	2.42
17	38.09	0.50
18	38.03	14.39
19	35.74	0.88
20	23.51	32.09
Average	25.80	9.20
Range	2.66 – 60.33	0 – 42.00

Table A-39. Habitat suitability values for the fieldwork data applied to the *G. illyricus* BBN.

	Presence sites	**Absence sites**
1	100.00	65.00
2	65.00	0.00
3	70.00	30.00
4	70.00	16.64
5	65.00	1.19
6	100.00	0.00
7	36.60	0.00
8	100.00	0.00
9	40.00	15.00
10	70.00	0.00
11	40.00	21.60
12	40.00	17.40
13	59.59	12.00
14	64.06	0.00
15	70.00	0.00
16	36.20	37.10
17	65.00	2.69
18	40.00	0.00
19	36.60	2.80
20	65.00	0.00
Average	61.65	11.07
Range	36.20 – 100.00	0 – 65.00

Table A-40. Habitat suitability values for the fieldwork data applied to the *H. semele-1* BBN.

	Presence sites	**Absence sites**
1	100.00	18.29
2	42.41	2.48
3	63.85	6.35
4	67.40	12.49
5	59.60	1.90
6	91.70	0.49
7	36.60	2.45
8	89.33	2.51
9	36.00	3.17
10	63.85	2.48
11	36.00	21.60
12	37.80	14.31
13	36.60	0.71
14	63.56	6.99
15	70.00	1.70
16	32.69	15.44
17	30.58	6.29
18	36.00	8.87
19	32.69	0.02
20	44.01	1.70
Average	53.53	6.51
Range	30.58 - 100	0.02 – 21.60

Table A-41. Habitat suitability values for the fieldwork data applied to the *H. semele-2* BBN.

	Presence sites	**Absence sites**
1	34.14	22.36
2	48.44	5.70
3	18.31	3.31
4	70.15	2.28
5	84.60	18.89
6	68.69	9.40
7	68.69	0.00
8	83.37	0.00
9	83.37	0.00
10	38.32	0.00
11	39.25	2.28
12	83.37	0.00
13	83.37	16.64
14	2.28	2.28
15	18.31	0.00
16	15.22	0.00
17	83.37	0.00
18	41.01	2.28
19	70.93	0.00
20	36.65	2.08
Average	53.59	4.38
Range	2.28 – 84.60	0 – 22.36

Table A-42. Habitat suitability values for the fieldwork data applied to the *N. sylvestris* BBN.

	Presence sites	**Absence sites**
1	100.00	0.00
2	100.00	0.00
3	100.00	0.00
4	100.00	0.00
5	100.00	20.00
6	100.00	21.68
7	100.00	13.63
8	100.00	13.63
9	100.00	24.15
10	100.00	0.00
11	100.00	80.00
12	100.00	0.00
13	100.00	20.00
14	100.00	0.00
15	100.00	0.00
16	85.00	0.62
17	100.00	0.00
18	100.00	0.00
19	26.17	24.86
20	100.00	10.00
Average	95.56	11.43
Range	26.17 - 100	0 – 80.00

Table A-43. Habitat suitability values for the fieldwork data applied to the *P.argus-1* BBN.

	Presence sites	**Absence sites**
1	42.03	0.00
2	63.73	0.00
3	46.40	0.00
4	46.40	0.00
5	80.41	42.03
6	92.00	43.16
7	92.21	46.40
8	58.40	46.40
9	80.00	46.40
10	53.11	0.00
11	92.00	60.00
12	100.00	0.00
13	92.00	29.39
14	92.24	0.00
15	92.00	0.00
16	76.50	5.68
17	92.00	0.00
18	66.56	0.00
19	29.86	76.64
20	35.60	22.79
Average	71.17	20.94
Range	29.86 – 100.00	0 – 76.64

Table A-44. Habitat suitability values for the fieldwork data applied to the *P. argus-2* BBN.

	Presence sites	**Absence sites**
1	70.74	43.02
2	30.70	0.00
3	43.02	2.61
4	89.22	0.00
5	50.72	0.00
6	36.09	0.00
7	86.52	0.00
8	14.34	0.00
9	70.74	0.00
10	80.75	0.00
11	36.09	1.43
12	55.34	50.72
13	32.63	0.00
14	80.75	0.00
15	57.72	0.43
16	55.72	6.36
17	80.75	3.74
18	36.92	0.00
19	80.75	0.00
20	55.34	0.00
Average	57.24	5.42
Range	14.34 – 89.22	0 – 50.72

Table A-45. Habitat suitability values for the fieldwork data applied to the *P. globulifera-1* BBN.

	Presence sites	**Absence sites**
1	70.74	43.02
2	30.70	0.00
3	43.02	37.25
4	89.22	0.00
5	50.72	0.00
6	36.09	0.00
7	86.52	0.00
8	14.34	0.00
9	70.74	0.00
10	80.75	0.00
11	36.09	8.80
12	55.34	50.72
13	32.63	0.00
14	80.75	0.00
15	57.72	0.71
16	55.72	6.36
17	80.75	43.48
18	36.92	0.00
19	80.75	0.00
20	55.34	0.00
Average	57.24	9.52
Range	14.34 – 89.22	0 – 50.72

Table A-46. Habitat suitability values for the fieldwork data applied to the *P. globulifera-2* BBN.

	Presence sites	**Absence sites**
1	49.10	39.78
2	67.60	26.33
3	67.60	42.29
4	73.77	44.08
5	64.50	39.01
6	61.73	32.99
7	32.10	52.94
8	61.44	44.77
9	67.83	42.45
10	56.46	55.49
11	55.07	61.44
12	74.60	45.01
13	51.45	61.71
14	73.77	28.88
15	61.44	49.35
16	57.36	55.61
17	65.64	40.57
18	61.44	67.66
19	73.77	55.31
20	43.43	15.25
Average	61.01	45.05
Range	32.10 – 74.60	15.25 – 67.66

Table A-47. Habitat suitability values for the fieldwork data applied to the *P. punctata-all* BBN.

	Presence sites	Absence sites
1	65.16	23.71
2	89.58	15.55
3	89.58	25.24
4	97.71	26.33
5	68.42	41.41
6	81.83	43.90
7	34.08	70.22
8	81.44	26.74
9	89.88	56.38
10	59.90	73.60
11	58.43	81.44
12	98.81	59.76
13	54.59	65.46
14	97.71	38.47
15	81.44	52.37
16	76.06	59.00
17	86.99	24.19
18	81.44	58.84
19	97.71	45.82
20	57.67	9.12
Average	77.42	44.88
Range	34.08 – 98.81	9.12 – 81.44

Table A-48. Habitat suitability values for the fieldwork data applied to the *P.punctata-dry* BBN.

	Presence sites	Absence sites
1	41.39	49.66
2	57.05	32.97
3	57.05	52.79
4	62.27	55.01
5	65.29	39.49
6	52.09	27.76
7	32.50	44.64
8	51.83	55.87
9	57.25	35.76
10	57.15	46.81
11	55.74	51.83
12	62.98	37.93
13	52.08	62.46
14	62.27	24.28
15	51.83	49.95
16	48.39	56.29
17	55.39	50.64
18	51.83	74.19
19	62.27	64.16
20	36.59	19.03
Average	53.66	46.58
Range	32.50 – 65.29	19.03 – 74.19

Table A-49. Habitat suitability values for the fieldwork data applied to the *P. punctata–moist* BBN.

Explanations of the different models can be found in section 4.3.11.2 (Chapter 4).

Appendix 29. ROC plots for BBNs

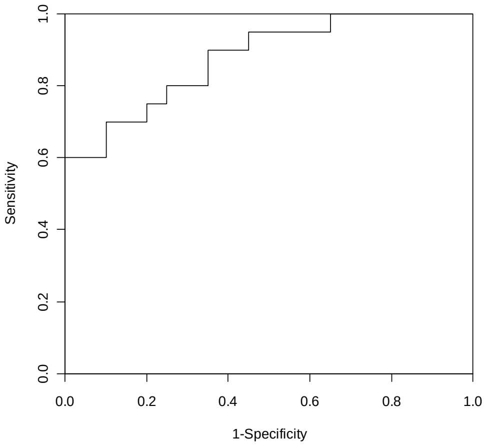


Figure A-64. ROC plot for *C. nobile* (1) BBN (AUC = 0.878).

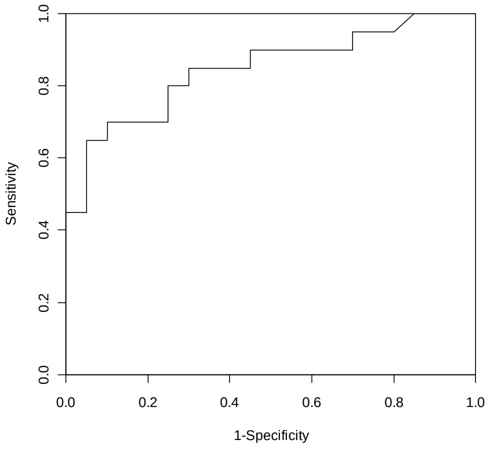


Figure A-65. ROC plot for *C. nobile* (2) BBN (AUC = 0.846).

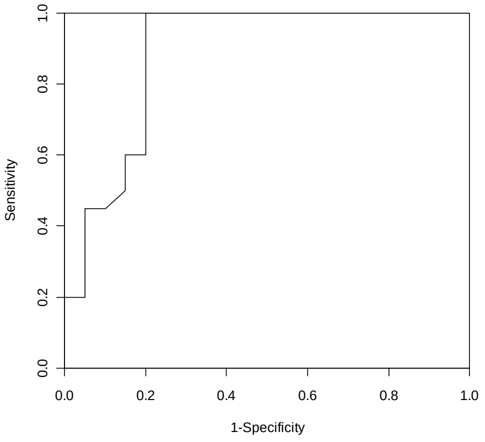


Figure A-66. ROC plot for *G. constrictum* BBN (AUC = 0.886).

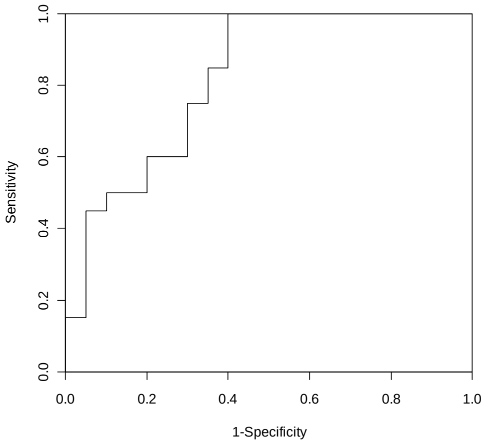


Figure A-67. ROC plot for *G. illyricus* BBN (AUC = 0.820).

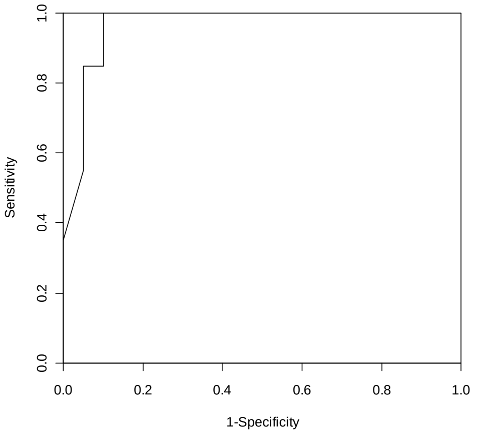


Figure A-68. ROC plot for *H. semele* (1) BBN (AUC = 0.965).

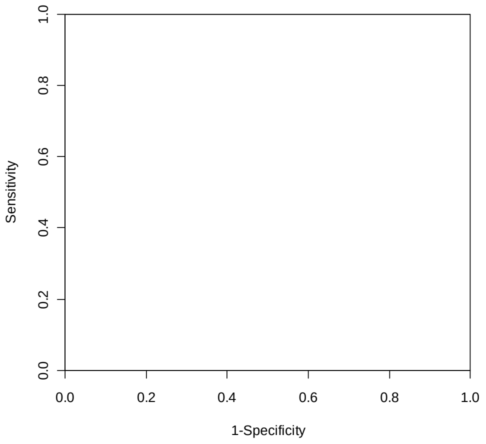


Figure A-69. ROC plot for *H. semele* (2) BBN (AUC = 1.00).

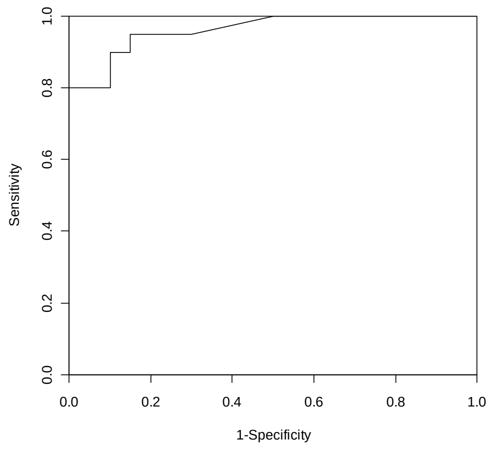


Figure A-70. ROC plot for *N. sylvestris* BBN (AUC = 0.963).

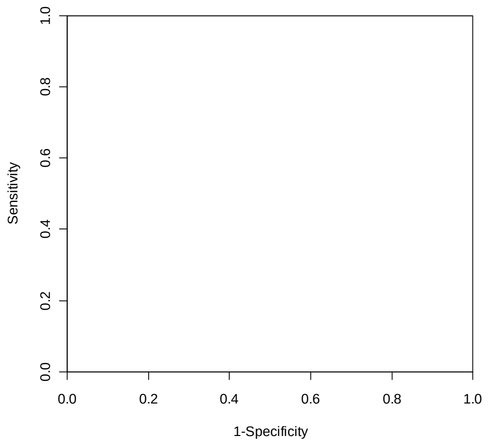


Figure A-71. ROC plot for *P. argus* (1) BBN (AUC = 0.999).

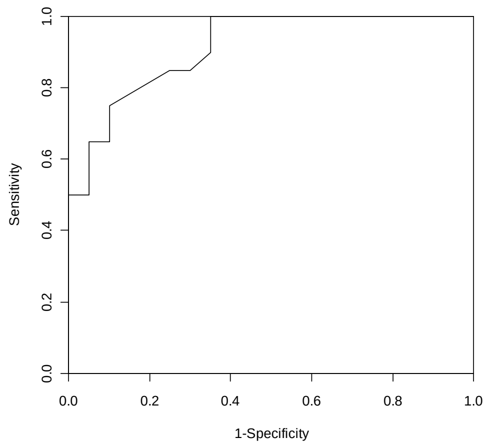


Figure A-72. ROC plot for *P. argus* (2) BBN (AUC = 0.914).

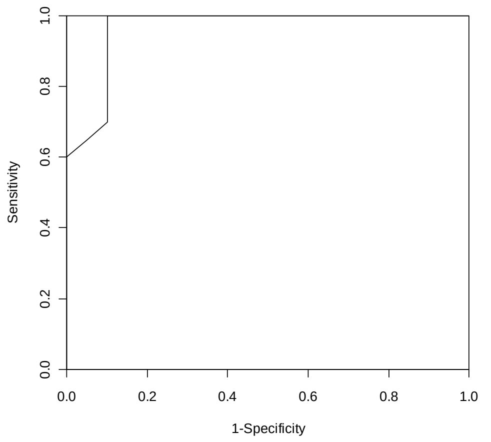


Figure A-73. ROC plot for *P. globulifera* (1) BBN (AUC = 0.965).

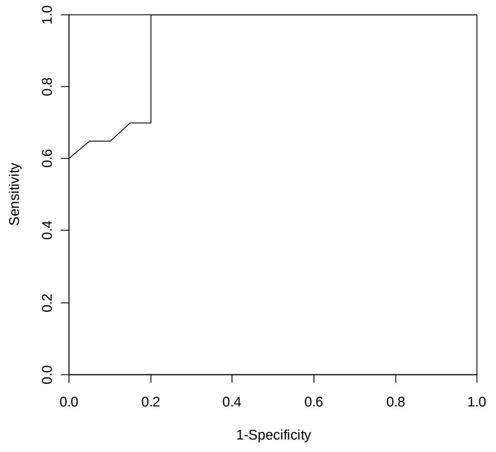


Figure A-74. ROC plot for *P. globulifera* (2) BBN (AUC = 0.933).

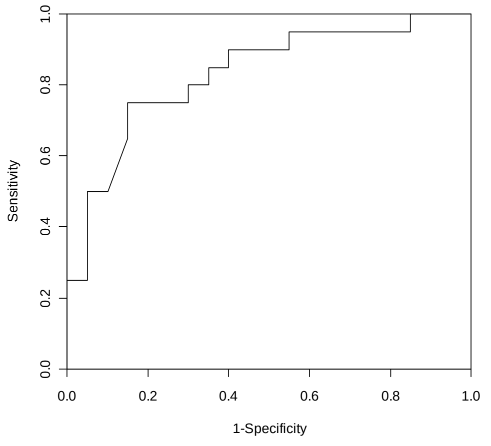


Figure A-75. ROC plot for *P. punctata* (all) BBN (AUC = 0.831).

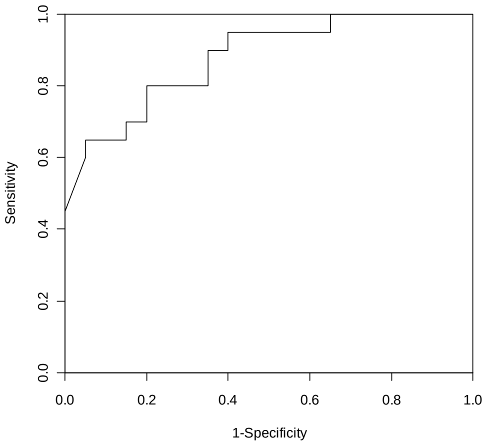


Figure A-76. ROC plot for *P. punctata* (dry) BBN (AUC = 0.879).

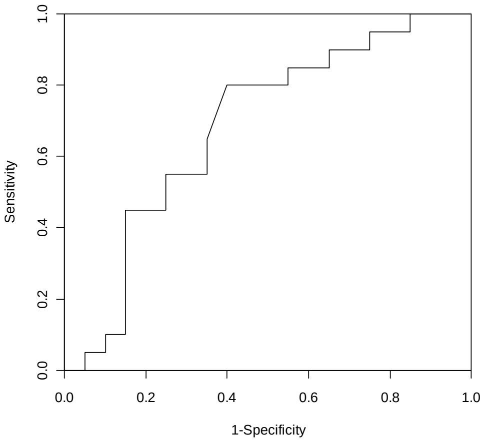


Figure A-77. ROC plot for *P. punctata* (moist) BBN (AUC = 0.684).

Appendix 30. Potential effects of climate change on habitats of the New Forest

30.1. Woodland

Changes in growth and phenology

The general view for temperate forests is that the combined effects of a rise in temperature, a lengthened growing and raised carbon dioxide levels are likely to result in an overall increase in tree growth, although it is not clear whether this would be sustained in the long-term (Broadmeadow, 2000; Nisbet, 2002). It would also depend on sufficient water availability (M. Broadmeadow, personal communication, cited in Diack, 1999). It is suggested that responses to warmer temperatures will be seen through changes in phenology, and such changes have already been observed. Menzel and Fabian (1999) found that tree and shrub species from the European network of International Phenological Gardens show a clear tendency for the average date of spring events, such as leaf unfolding, to have advanced (typically by about six days, particularly in northern Europe), whereas a typical autumn event, such as leaf colouring, was delayed by about five days (although less so in northern Europe than in central and southern Europe). When these two effects were combined, the growing season was about 11 days longer in the mid-1990s compared with the mid-1960s.

Species which require a high degree of winter chilling to break dormancy (e.g. beech, *Fagus sylvatica*) will flush later in the spring after warmer winters, while those that currently flush early (e.g. hawthorn, *Crataegus monogyna*) are likely to break dormancy even earlier (DEFRA, 2001). However, native tree species that rely on day length to trigger leaf fall and dormancy may be unable to extend their growing season into the warmer autumn (DEFRA, 2001).

Changes in the timing of individual responses to climate change will have implications for the composition and species dominance within forest habitats in the future, with knock-on impacts for other dependent species (DEFRA, 2001). For example, Diack (1999) suggests that earlier and greater tree canopy coverage would reduce the amount of light and water reaching the ground, which may affect the ground flora and other ground dwelling organisms; spring flowering herbs may find their growing season compressed as a result. It is expected that herbs and other species might undergo range shifts but more rapidly than trees (DEFRA, 2001).

Changes in species distributions and composition

Increases in temperatures are likely to change the boundaries of many tree species' distributions in the UK, leading to a change in the community composition of current woodlands (DEFRA, 2001). In particular there may be an increase in introduced and southerly distributed species that are favoured by warmer conditions (Hossell *et al*., 2000). However, although woodland is expected to be affected by climate change, some of the effects may not be evident for 100-150 years (Davis, 1989; Hossell *et al.*, 2000). For example, many of the tree species have slow reproductive and dispersal rates, and are therefore unlikely to adapt without assistance (Hossell *et al*., 2000). The main response to climatic change will likely be through migration, but the speed of migration is unlikely to be sufficient to reach suitable climatic areas in the future and will be additionally hampered by fragmented habitat as a result of human disturbance (Davis, 1989; DEFRA, 2001). Climate change impacts on forests will result not only through changes in mean climate, but also through changes in seasonal and diurnal rainfall and temperature patterns (Fischlin *et al*., 2007).

Die-back/changes in community composition and habitat structure

Increased occurrences of the death of trees may be caused by several factors related to climate change, including soil moisture stress as a result of drought, increased fire risk, increases in woodland pests and pathogens and increases in the frequency and severity of storms. In the long-term, climate change is likely to affect woodland tree composition and balance, although the full implications are difficult to predict (Diack, 1999).

Soil moisture stress/drought

Soil moisture stress, through reduced precipitation and increased temperature interactions may reduce tree growth and as climate change progresses, severe summer droughts may kill increasing numbers of trees (die-back), particularly species such as beech (DEFRA, 2001; Hossell and Rowe, 2006). Alder, willows and poplars may also suffer as they have a greater requirement for water (Diack, 1999).

Winter waterlogging, as a result of increased winter precipitation, could present an added problem of restricting rooting depth, rendering trees more liable to summer drought stress (Nisbet, 2002). There is also uncertainty over whether an increase in atmospheric carbon dioxide could increase the water use efficiency of trees by reducing stomatal opening, or whether this will be offset by an increase in leaf area as a result of increased productivity (Nisbet, 2002).

However, an indication of the potential impact of droughts on the New Forest woodlands is provided by accounts of the effects of past droughts. Tubbs (2001) reports that during a decade of particularly hot dry summers from 1974 – 1984, many hundreds of beech trees died (as beech has shallow root systems). In 1974 and for a few years afterwards, the deaths of beech trees were obvious and dramatic. However, beech deaths have continued at a diminishing rate since then, with thinning canopies and dieback in both old and young trees, emphasising perhaps that the species is more usually associated with the high rainfall of oceanic regions and high mountain slopes and that the trees are readily stressed in New Forest conditions (Tubbs, 2001).

Tubbs (2001) also notes how the effects of drought on the beech population were compounded by the successive near-hurricanes of October 1987 and January 1990, which flattened many areas of woodland across southern England. In the New Forest they flattened extensive areas of conifer plantation and hundreds more old beeches. Drought-weakened trees were the most vulnerable. However, Tubbs (2001) suggests that the near-hurricanes did have compensations as they refreshed the resource of dead wood and provided an opportunity for new generations of trees to arise in new gaps in the canopy. The effects of drought and storms in the New Forest are also reported by Mountford *et al.* (1999) and are discussed below.

Diack (1999) suggests that the greatest and most immediate impact of climate change may be on wet woodlands where drier, warmer summers will have a potentially serious negative effect by reducing damp microclimate conditions, which are important for damp, rotten-wood specialists. There may also be a negative impact of drier summer conditions on globally important Atlantic-fringe bryophyte and lower plant communities in old pasture woodlands, particularly in the New Forest (Diack, 1999). Therefore, these impacts could lead to changes in community composition and habitat structure (Hossell and Rowe, 2006).

Increased fire risk

An additional impact of drier summer conditions is an increased fire risk, which could cause significant damage to woodlands (Broadmeadow and Ray, 2005). Diack (1999) suggests that fire risk is further exacerbated by an increase in recreational use of forests, which the warmer summers are projected to encourage, but that there is also some anecdotal evidence that the increase in recreational use of forests, combined with an increase in mobile phone use has helped to reduce fire damage as fires can be reported more quickly.

Increase in woodland pests and pathogens

An effect of warmer winters is likely to be through their impact on populations of woodland pests and pathogens and through favouring species such as grey squirrels, *Sciurus carolinensis* (which preferentially strip the bark of beech) (Hossell and Rowe, 2006). Non-indigenous pests (such as gypsy moths (*Lymantria dispar*), and Asian longhorn beetle (*Anoplophora glabripennis)*) may also become a greater threat in the UK, particularly in the South East (Hossell and Rowe, 2006). Tubbs (2001) reported that during 1980-83 in the New Forest (during a decade of particularly hot, dry summers) there were population explosions of a number of moths, but especially the mottled umber (*Erannis defoliaria*) and the oak leaf roller moth (*Tortrix viridana*), which resulted in widespread defoliation of the oak woods.

The *Phytophthora* diseases of trees are promoted by fluctuating water tables (Lonsdale and Gibbs, 2002) so would be expected to become more prevalent (Broadmeadow and Ray, 2005). Warmer winters and summers and summer drought are expected to favour fungal diseases and pathogens, but the protective effects of fungal mychorrizas against various root diseases may be affected under altered soil temperatures and moisture regimes (Lonsdale and Gibbs, 2002). Further, pests and pathogens may have an even greater impact on trees that are already stressed or damaged as a result of drought or storms (Mountford *et al.*, 1999). However, it is difficult to accurately predict the impact of climate change on insect damage to forests because it is likely to alter the balance between insect pests, their natural enemies and their hosts, as well as altering the synchrony between host and pest development, and the impact of increased atmospheric carbon dioxide concentration will also have an effect (Evans *et al.*, 2002). Nonetheless, it is likely that the increase in pests and pathogens, combined with summer droughts, may cancel out any benefits of increased carbon dioxide and warmer temperatures (Diack, 1999).

Increased number and severity of storms

An increase in violent storms (although predictions of future windiness are more uncertain (Hossell and Rowe, 2006)) would add to the risk of damage to trees already weakened or damaged by drought (Hossell *et al.*, 2000). Shallow rooted species such as beech may be particularly vulnerable (Diack, 1999). A resulting increase in gaps in woodland (especially beech) may allow more understorey growth (Hossell *et al.*, 2000) or allow new species to invade (DEFRA, 2001).

An indication of the potential impacts of severe weather events for woodland in the New Forest is provided by Mountford *et al.* (1999) who recorded long-term change in Denny Wood in the New Forest. They reported that the severe drought in 1976 led to the death and severe weakening of many of the very old A-generation beech trees, although this was not apparent until several years after (1984). The storm events of 1987 and particularly 1990 caused further damage, either blowing over the dead and weakened trees, or expanding gaps already opened by the drought. At least 6 large oaks may have died as the result of the drought and several others appeared debilitated by drought, but the oaks were not as affected as the beech trees (with the death of at least 21 trees). Storm-damage also resulted in allowing the entry of fungal infections, which added to the weakening of the trees. The study also provided an example of the impact of an introduced species, the grey squirrel, which caused severe damage (through bark-stripping) to many of the younger C-generation trees, which is detrimental to regeneration. High grazing pressure is also a factor which can limit the regeneration of trees.

Largely due to its shallow root system, making it more vulnerable to drought and tree-throw (in storms), it appears that beech may be one of the woodland species most seriously affected by climate change (Mountford *et al*., 1999). This may have a particular effect in the New Forest, which contains the largest area of mature, semi-natural beech woodland in Britain and represents Atlantic acidophilous beech at the most southerly part of its UK range (Grant and Edwards, 2005). The communities associated with beech woodlands (such as lichens, fungi and deadwood invertebrates) could therefore also suffer as a result (Grant and Edwards, 2005). Grant and Edwards (2005) suggest that because beech is at the limits of its climatic range in the New Forest, factors such as pathogens and competition will play a large role in its ability to remain dominant within the community. In comparison, oak has maintained dominance for many millennia against the background of both natural and anthropogenic pressures and should remain dominant into the near future.

30.2. Heathland

Climate change is likely to affect European shrubland key ecosystem functions such as carbon storage, nutrient cycling, and species composition (Wessel *et al.*, 2004; Alcamo *et al.*, 2007). However, it is likely that the response to warming and drought will depend on the current conditions. For example in two heathland species (*Calluna vulgaris* and *Vaccinium myrtilus*), a greater response to warming was found in colder, moister environments which are more limited by temperature, whereas warmer and drier (more southern) sites are limited by water availability and were therefore more sensitive to drought (Peñuelas *et al*., 2004).

Changes in composition (wet heath replaced by dry heath)

There have been relatively few assessments of the impacts of climate change on lowland heath (Hossell and Rowe, 2006). However, it is likely that a decrease in summer precipitation will cause many of the wet heaths to dry up and likely to revert to dry heath or be replaced by an expansion of the region’s acid grasslands (Cook and Harrison, 2001). Hossell *et al*. (2005) suggest that it is likely that increased drying and its effects on the water table will cause the boundaries of the wet heath communities and their composition to change, with a decrease in species such as cross-leaved heath (*Erica tetralix*), which occurs in moister soils than the other heathland species (Webb, 1986) and has a shallow rooting system, which makes it highly susceptible to drought (Hossell *et al.*, 2005).

Instead, with the expansion of dry heath or acid grassland, species such as purple moor grass (*Molinia caerulea*) and heather (*Calluna vulgaris*) would be expected to dominate over *E. tetralix* (Hossell and Rowe, 2006). In contrast to *E. tetralix*, *M. caerulea* can occur in drier and wetter areas and has a deep root system, which may allow it to cope more effectively than *E. tetralix* with low soil moisture levels and a low water table during the summer months, which may occur as a result of climate change (Hossell *et al.*, 2005). Heather (*C. vulgaris*) is also better able to withstand soil drying than *E. tetralix* (Hossell *et al.*, 2005). Species such as bristle bent (*Agrostis curtisii*) and bell heather (*Erica cinerea*), which are associated with drier soil conditions are also likely to increase in abundance (Hossell *et al.*, 2005).

It has been suggested that in heathlands, a change in soil nitrogen levels from increased decomposition rates in warmer conditions is expected to favour grass growth, so the competitive balance may change at the interface between heath and acid grassland habitats (Hossell *et al*., 2000). Based on an assessment (using field manipulation experiments) of four European shrublands (including a site in the UK), Wessel *et al*. (2004) report that the adverse effects of increased temperature on heathlands operate mostly through the increased availability of nutrients (including increased primary production, decomposition and nutrient cycling) which may aid in the encroachment of grassland into heathland. However, drought can decrease productivity. Therefore the results of climate change will depend on how the differing effects of warming and drought interact. Wessel *et al*. (2004) also note the increased risk of fire as a result of decreased water availability and suggest that more frequent fires would contribute to a reduction in the productivity of the vegetation and the organic matter content of the soil. Together these processes would aggravate the direct effects of drought and push the ecosystem toward a more arid type (Wessel *et al*., 2004).

However, Tubbs (2001) reports that during a period of hot, dry summers in the New Forest during 1974-1984 there was a noticeable invasion of *Calluna* into many dry grasslands and the New Forest National Park Authority (2007a; Trotter, 2007) also anticipate a loss of acid grassland to heather and heath. Despite the uncertainties, it is clear that the heathlands will become drier, with an impact on the composition of their species.

Although drought will largely determine the distribution and abundance of the dominant heathland species in this habitat, other differences between them, for example in nutrient availability and resistance to burning, may also affect competition and hence abundance of these important species (Hossell *et al.*, 2005). Wekman and Callaghan (2002) also suggest that the influence of other species will be important, particularly for slow-growing perennial shrubs such as heather, because other species may respond more to a changing environment, increasing their competitive potential to the detriment of heather. Another factor will be the impact of herbivores or pests. For example, Tubbs (2001) reported that during hot dry summers in the New Forets in the past, extensive tracts of heather were completely stripped of leaf by a population explosion of heather beetles.

If *M. caerulea* becomes more abundant relative to *C. vulgaris* and *E tetralix*, this may have a significant impact on the habitat, as *M caerulea* is a tussocky grass which provides a different habitat structure to that of dwarf shrubs (Hossell *et al.*, 2005; Hossell and Rowe, 2006). Further, drier conditions could also favour the spread of more dense gorse scrub and woodland, which could produce a considerable change in the look of the landscape and produce a more closed habitat structure (Hossell and Rowe, 2006).

Certain species and species groups may benefit from the predicted changes, while others may suffer. Diack (1999) suggests that it appears that those species characteristic of the drier, more parched areas of heathland may benefit, while species more characteristic of humid heath may suffer as a result of the drier and hotter summer weather. Considering that most of the drier heaths of the New Forest are humid heathland (with the very dry heaths making up the smallest proportion of heathland) as well as there being a considerable amount of wet heathland (Forestry Commission, 2002a), there could be considerable changes in the landscape in the future.

Berry *et al*. (2007a) examined potential climate space for heathland species in Hampshire from models based on the UKCIP02 scenarios (which although are not the most recent scenarios, show similar trends to the UKCIP09 scenarios). They found that five of the selected species which are dominant or sub-dominant in lowland heathland: cross-leaved heath (*E. tetralix*), bell heather (*E. cinerea*), heather (*C. vulgaris*), wavy hair grass (*Deschampsia flexuosa*) and purple moor grass (*M. caerulea*), showed little or no loss of potential suitable climate space, except under the 2080s High scenario, where the loss was severe for southern England as the climate conditions become outside the tolerance range of those species.

Although the future of some heathland plants could be under threat, this may not necessarily lead to their extirpation if they are able to adapt to the conditions (Berry *et al.*, 2007a). Further, the models only take account of suitable climate space for those species and not the impact of other factors which may influence their distributions. For example, species that are predicted to still have suitable climate space by the 2080s may still suffer because other species on which they rely have disappeared. For example, the silver-studded blue butterfly (*Plebejus argus*) could see gains in potential suitable climate space but could be adversely affected by the potential loss of ericaceous species (Berry *et al.*, 2007a). The BRANCH Partnership (2007) therefore suggest that new species assemblages may be seen and lowland heath in Hampshire could change significantly in composition.

Competition with bracken

The competitive dynamics between bracken and heather may change as a result of climate change. This is discussed in the bracken section (Appendix 30.5) below.

Increased risk of accidental fires

Dry heaths may be particularly at risk from reduced summer precipitation leading to an increase in accidental fires (DEFRA, 2001). A possible increase in visitor usage may also increase the risk of fires being started (deliberately or accidentally), although they may also be reported more quickly (see woodland section; Appendix 30.1). Such fires can be detrimental to the ecology of these heaths, removing vegetation, releasing nutrients into the system, and exacerbating erosion (DEFRA, 2001; Hossell and Rowe, 2006). If fires become more frequent the potential beneficial effects of climate change on reptiles and other species are likely to be nullified or reversed (Diack, 1999). Small fragments of heath are particularly at risk under such a scenario as their isolation in the landscape reduces the likelihood of their being recolonised by lost species (Diack, 1999).

30.3. Terrestrial wetland

Wetlands support a large number of species (many of which are nationally rare or scarce) which are adapted to life in saturated conditions and are particularly vulnerable to climate change (particularly changes in precipitation) because of their sensitivity to water levels and the delicate balance between rainfall, temperature and evapotranspiration (Dawson *et al*., 2003). Changes in precipitation are expected to alter water availability and streamflows affecting ecosystem productivity, with lower summer water availability reducing water quality (Dawson *et al*., 2003). The increases in winter precipitation in much of central and southern England will not compensate for the deficits in summer (Dawson *et al*., 2003). Problems may also be exacerbated by the likely increased demand for water for abstraction (Diack, 1999).

Dawson *et al*. (2003) also note that although the broad-scale climate space (from models such as the SPECIES model, developed for MONARCH (Berry *et al.*, 2007b)) may appear to be suitable for a species, the water table regime may indicate significant seasonal stresses.

Changes in flow/ increased drying out

Changes in total precipitation, extreme rainfall events, and seasonality (such as increased autumn/winter precipitation and decreased summer precipitation) will affect the amount, timing and variability of flow in freshwater ecosystems (Carpenter *et al*., 1992). It is possible that changes in magnitude and temporal distributions of extreme events may disrupt ecosystems more than changes in mean conditions (Carpenter *et al*., 1992).

The ‘flashiness’ of flows is likely to increase, with the potential for more frequent winter flooding and summer drought (DEFRA, 2001). Increased winter precipitation could lead to increased ponding and flooding (Harrison *et al*., 2001). It is likely that flash flooding in winter may increase as the greater intensity of rainfall increases runoff into rivers and if flood events are severe enough these may scour river beds, washing away much aquatic life (Diack, 1999). This will have an effect on the ecology of the rivers, perhaps altering stream-bed characteristics, and niche development for freshwater species, but the long-term impact will depend upon the frequency of extreme events; year on year effects may be significant (Diack, 1999).

Cook and Harrison (2001) suggest that summer drought (with less rainfall, exacerbated by increased rates of evaporation in the warmer air) is likely to dry out the few remaining bogs and result in widespread desiccation of many wetland habitats in southern and central England, although there will be better conditions for bog growth further north thanks to increased rainfall, particularly in winter. Areas of open water are expected to decline as a result of increased evapotranspiration and wetland areas where water levels are kept high by a continuous supply of water and are dependent upon heavy rainfall to maintain moist conditions, may be particularly at risk from reduced rainfall and persistent drought in the summer, especially small sites in lowland areas (Hossell *et al.*, 2000; DEFRA, 2001). Many ephemeral ecosystems may disappear. Indeed, during droughts in the New Forest during 1974-1984, Tubbs (2001) observed that a high proportion of ponds and ephemeral wetlands and many streams dried out in summer, with unknown effects on invertebrates and amphibians.

Wetland habitats are often partly defined by topography (e.g., valley fens) and high water tables, which results in restricted availability of suitable sites even without the effects of climate change (Hossell *et al.*, 2000; DEFRA, 2001). Therefore, the ability of species of these habitats to migrate is severely limited by these factors and hampered by the lack of suitable migration routes and limited suitable alternative locations (Hossell *et al*., 2000). Certain wetland habitat species which rely on a period of low water for reproduction will be unable to move in response to climate change if open water is limited (Hossell *et al*., 2000). However, Diack (1999) suggests that valley mires in a landscape with a high proportion of other semi-natural habitat, for example the New Forest mires, are likely to be more resilient than those surrounded by development or intensive agriculture. Nonetheless, with the multiple stresses upon wetland habitats there is still a lot of uncertainty about the survival of these habitats.

It is likely that drought may exacerbate pollution problems in wetland habitats by concentrating pollutants in periods of low flow (Carpenter *et al.*, 1992; Hossell *et al.*, 2000). Changes in the timing, intensity and amount of rainfall could also affect water quality by increasing the leaching of fertilisers, herbicides and pesticides from surrounding land, as well as increasing levels of sedimentation (Hossell *et al.*, 2000). Nutrient status and acidity of wetland habitats may alter as surrounding vegetation patterns are affected by climate change (DEFRA, 2001). In particular, increases in nitrogen mineralisation in soils as a result of increases in temperatures could raise the concentration of nitrogen entering freshwater habits, leading to eutrophication (DEFRA, 2001). Low flows may also concentrate nutrients, encouraging eutrophication (Diack, 1999). It has also been suggested that an increase in temperature will increase decomposition of organic soils and weathering of rock, which may produce an increase in pH in some water bodies (Eyre *et al.*, 1993; Hossell *et al.*, 2000; DEFRA, 2001).

Changes in species composition/invasive species

Climate change is expected to potentially cause changes in species composition in wetland habitats, both as a result of warmer temperatures and changes in water availability and flow. Projected increases in air temperature will be transferred, with local modification, to groundwaters, resulting in elevated temperatures and reduced oxygen concentrations (Carpenter *et al.*, 1992). The density, growth patterns and phenology of aquatic invertebrate communities are therefore likely to be affected by temperature increases (Hossell *et al.*, 2000) and there may also be a shift in the distribution of fauna and flora of these habitats especially in response to a switch from permanent to temporary standing water bodies (Eyre *et al.*, 1993; DEFRA, 2001). In permanent water the species composition may alter to favour those that germinate and resume growth in spring in response to rising temperatures, such as Canadian waterweed (*Elodea canadensis*) (DEFRA, 2001).

It has been suggested that non-native aquatic macrophytes (such as New Zealand Pigmyweed, *Crassula helmsii*) may spread more rapidly if winters become warmer and frost events are less frequent and non-native fauna, such as bullfrog (*Rana catesbeiana*), may also benefit from warmer weather, with possibly serious implications for native amphibians (Diack, 1999). Therefore, as well as the direct effects on aquatic species, climate change will also have indirect effects on these species, such as through the facilitation of invasive species (i.e. non-native taxa that increase in abundance to the point where they have negative impacts on native species and ecosystem function and may cause economic damage) (Rahel *et al*., 2008). The mechanisms by which invasive species affect species include predation, competition, and diseases (Rahel *et al*., 2008). The virulence of non-native parasites and pathogens to native species may also be increased by climate warming, such as by allowing them to complete their life cycle more rapidly (Rahel and Olden, 2008). As native species have evolved adaptations to historic flow regimes, alteration of flow regimes (such as periods of prolonged low flows and stream drying and changes in the magnitude and frequency of floods) is most likely to benefit non-native species adapted to the changed conditions (Rahel and Olden, 2008). Climate change and invasive species may act synergistically if native species that are stressed by changes in temperature or flow regimes must also face challenges from invasive species (Rahel *et al*., 2008).

Another indirect effect of climate change will be the alteration of biotic interactions between species of conservation concern and other native species. Changes in thermal regimes, flow regimes, or salinity could alter the competitive interactions or predator–prey relations among aquatic species in ways that are detrimental to species of conservation concern (Rahel *et al*., 2008). Climate change could allow species considered native to a region to spread to new habitats, increase in abundance, and harm other native species—in essence mimicking the negative effects associated with invasive species (Rahel *et al*., 2008). Rahel and Olden (2008) suggest that climate change may therefore force a redefinition of invasive species.

From examination of peat cores at two bog ecosystems (one in England and one in Denmark), Heijmans *et al*. (2008) concluded that in the past century, the vegetation remained remarkably stable despite changes in atmospheric carbon dioxide, nitrogen deposition and temperature. They suggest that a reason for the lack of species replacements is that a warmer climate and doubling of carbon dioxide concentrations have contrasting effects. In simulations that they carried out at the sites, increasing carbon dioxide favoured the peat mosses whose growth is less nutrient limited than that of vascular plants because they can realise the increased potential growth rate. In contrast, rising temperatures favour the nitrogen-limited vascular plants by increasing nitrogen-mineralisation rates, resulting in increased nutrient availability in the peat. They therefore suggest that the future bog vegetation at the sites will not change much and that it is important that the effects of global warming should not be based on individual factors alone.

Erosion/loss of peat (in bogs)

It is likely that on disturbed bogs increased rates of decomposition, promoted by higher temperatures and drying and cracking in summer drought, on bare or partly vegetated peat surfaces may lead to increased erosion and loss of peat, exacerbated by increased winter rainfall intensity (Hossell *et al.*, 2000; DEFRA, 2001). As many protected peatland sites in the UK have some degree of disturbance, the impact of climate change may be a cause for concern (DEFRA, 2001). In addition, if species such as *Sphagnum compactum* and *S. tenellum* are lost from former bog habitats as the summers become drier, this could increase the accessibility and use of some parts (for example, in the New Forest) for recreation, which in turn may increase erosion on drying peat soils (Hossell and Rowe, 2006).

30.4. Grassland

Loss of grassland

In general, grasslands are expected to decrease in area in Europe by the end of this century, the magnitude varying depending on the emissions scenario (Alcamo *et al*., 2007). In terms of land cover in southern England, including Hampshire, grassland shows a large reduction in extent under models based on future climate change scenarios, although such models are based purely on suitable climate space (Hossell *et al*., 2005). The same models also show some redistribution and loss of neutral grasslands but a complete loss of calcareous grassland land cover under a 2050s High emissions scenario (based on the UKCIP02 scenarios, but the trends are similar to the UKCIP09 scenarios) (Hossell *et al*., 2005). Acid grasslands show a severe contraction of land cover but this may be due to the climate conditions exceeding those on which the model was trained and may not be a reflection of the fact that grassland land cover cannot exist in these areas (Hossell *et al*., 2005). However, acid grassland may also encroach into heaths as they dry out (see heathland section; Appendix 30.2).

Changes in productivity

In a review of temperate grasslands and global atmospheric change, Soussana and Luscher (2007) report that there is a general confirmation that elevated carbon dioxide concentrations stimulate photosynthesis, leading to increased plant productivity and modified water and nutrient cycles. However, the long-term response to elevated atmospheric carbon dioxide concentrations may differ substantially from the short-term response. They also suggest that the relative enhancement in growth due to elevated atmospheric carbon dioxide concentrations might be greater under drought conditions than in wet soil conditions because photosynthesis would be operating in a more carbon dioxide-sensitive region of the response curve of growth (although the effect of temperature is not taken into account). There is therefore a general view that elevated atmospheric carbon dioxide concentration reduces the sensitivity to low precipitation in grassland ecosystems, although it may also induce progressive nitrogen limitations on plant growth (Soussana and Luscher, 2007). A longer growing season is also expected to extend grassland productivity (Hossell *et al.*, 2000).

However, DEFRA (2001) reports that the response of grassland species to predicted increases in carbon dioxide and climate change are complex at a local level, and these factors, such as grazing pressure, water availability and increased carbon dioxide may offset each other, leading to no overall change in species productivity.

Change in species composition/community structure

Climate change is likely to alter the community structure of grasslands in ways specific to their location and type, although management and species richness may increase resilience to change (Alcamo *et al.*, 2007). Soussana and Luscher (2007) suggest that the diversity and botanical composition of temperate grasslands is likely to be affected by the rise in atmospheric carbon dioxide concentration, possibly through a decline in the relative abundance of grasses. Increased atmospheric carbon dioxide concentrations are predicted to increase C3-species over C4-species but the projected increase in temperature will favour C4-species, although competition will be an important moderating factor (Soussana and Luscher, 2007). Teyssonneyre *et al*. (2002) also found that elevated carbon dioxide significantly increased the proportion of dicotyledones (forbs and legumes) and reduced that of the monocotyledons (grasses) in a C3 temperate grassland community. Responses were largely species-specific and different management regimes (e.g. frequency of cutting) could moderate responses. However, the above studies did not consider the effect of temperature and water availability which will also influence plant responses to climate change.

Diack (1999) suggests that increased temperature and rainfall in winter may favour the increased growth of more competitive grasses, to the detriment of less fast growing species and that grazing regimes may have to be altered to maintain site interest. Dry summer conditions following on from winter warming can favour deep-rooted plants, as their root morphology provide greater resistance to drought, while shallow rooting grasses are more likely to suffer in the drier conditions (Hossell *et al*., 2000). Drought may eliminate drought-sensitive species which were invading shallow soils or, on deeper soils, eliminate competitive species which are sensitive to desiccation and allowing subsequent invasion by new species (Buckland *et al.*, 1997). In acidic grassland there may also be a change in the competitive ability between dry grassland and heath communities (Hossell *et al*., 2000) (also see heathland section; Appendix 30.2). There may also be potential for a reduction in bracken invasion in southern England but with increased invasion potential in the north due to warmer wetter conditions (Pakeman and Marrs, 1996; Hossell *et al.*, 2000; see bracken section (Appendix 30.5)).

Tubbs (2001) reports that during droughts in the New Forest during 1974 – 1984, over the whole decade grass production was poor and by midsummer the grasslands were desiccated and brown: there was little grass growth for two months or more. The grassland flora, however, increased in diversity with the invasion of small annual plants which were able to become established with the suppression of more vigorous grasses. There was also a noticeable invasion of *Calluna* into many dry grasslands and the effects of grazing were heightened.

In the south-east of England, Diack (1999) suggests that for grasslands, gains may include the spread and greater abundance of plant species with a continental distribution (i.e. those tolerant of drier conditions) and invertebrates requiring warmer drier microclimates. However, Diack (1999) also cautions that the ability of these species to spread will be reduced by the fragmentation of suitable habitat and the lack of migration routes. Furthermore, the impact of extreme events (such as droughts) may cancel out any benefits accrued as a result of climate change.

Although it is not entirely clear what the combined impact of different factors, such as drought, temperature, increased carbon dioxide concentration, competition between species, will be on grassland, it seems likely that there will be a shift in community composition (DEFRA, 2001). In addition, the interactions between plant and insect responses will be important (Masters *et al.*, 1998; Hossell *et al.*, 2000). Further, Hossell *et al*. (2000) suggest that changes in grazing in grassland habitats may have as great an influence as climate change and that an increase in habitat degradation may occur if patterns or levels of grazing are not adjusted to take account of the changing precipitation patterns under climate change.

30.5. Bracken

Although bracken (*Pteridium aquilinum*) is an individual species, it is included here as a habitat, in terms of bracken stands, as it is an important habitat for one of the study species, *Gladiolus illyricus*.

Potential increase or decrease

The abundance and distribution of bracken in Britain overall are expected to increase as a result of climate change (Marrs and Watt, 2006), as it likely that, overall, bracken will benefit from the currently rising temperatures in Britain (Werkman and Callaghan, 2002). Pakeman and Marrs (1996) developed a model to predict the effects of changing climate (increased summer and winter temperatures of 1.4 °C, although they used unchanged rainfall and atmospheric carbon dioxide) on bracken stands in Britain. Their model suggested that with an increase in temperature and frost-free period, there would be little change in biomass near the south and west coast of England (although there were suggested large increases in other parts of the UK). Pakeman and Marrs (1996) did not include the effect of increased carbon dioxide in their models, but it has been suggested that bracken growth is not expected to be significantly altered by increased carbon dioxide levels (Whitehead *et al.*, 1997; Caporn *et al.*, 1999), although water use efficiency has been shown to increase under elevated carbon dioxide levels (Whitehead *et al.*, 1997). However, neither of these studies considered the effects of temperature and water availability.

Despite the advantages of warmer temperatures and fewer frosts (bracken is sensitive to frost, which can cause premature senescence and death of the frond (Marrs and Watt, 2006)), it is suggested that in the south-east of England bracken biomass is restricted by low water availability (Pakeman and Marrs, 1996) and that increased drought in the south may adversely affect bracken growth (DEFRA, 2001).

Bracken is sensitive to low water supply and drought may cause emergent fronds to either be killed and shrivel back into the soil, or continue to grow up under the canopy as delicate shade forms, which may or may not survive, depending on available moisture (Marrs and Watt, 2006). However, once established, bracken can be very tolerant of dry conditions, because of its thick cuticle, rapid stomatal response and the rigidity of the pinnae, meaning drought damage is rare, especially where the rhizomes and roots are relatively deep (Marrs and Watt, 2006). Therefore, bracken may be tolerant of the forecast drier summers for a while. Indeed, the New Forest National Park Authority (2007a; Trotter, 2007) note that the expansion of invasive species, such as bracken, may be a problem, particularly as it is very difficult to eradicate and expensive to control (Forestry Commission, 2002b; Marrs and Watt, 2006). Bracken is also generally unaffected by grazing, insect herbivory and pathogens (it is remarkably resistant to disease) (Marrs and Watt, 2006). Occurrence of bracken will also depend on how it is controlled and managed in the future. However, in the longer term, bracken may decline; the results of the MONARCH climate space models (Hossell *et al.*, 2005) show a complete loss of bracken in southern England under the UKCIP02 2050s High scenario, although they do not take account of other factors influencing species' occurrence.

In terms of forecasted increases in winter precipitation, although bracken tends to be found on well-drained soils, it is neither confined to them or absent from waterlogged soil, although waterlogging is a major constraint on bracken (Marrs and Watt, 2006). Bracken occurs in flushes and gleyed soils, but frond density is often lower than on drier soils, illustrating the detrimental effects of waterlogging (Marrs and Watt, 2006). However, bracken can persist on soils with a periodically high water table (Poel, 1951). Therefore, although certain areas in the New Forest may become wetter and even waterlogged in winter, with the forecast increase in precipitation, it is unlikely to be of a problem for bracken, particularly as it would only be for a few months at a time.

Competition with heather

The effect of climate change on bracken may also be influenced by competition with heather (*Calluna vulgaris*), where they occur in combination. In experiments investigating competition for water between bracken and heather in conditions of increased temperature, drought and increased nitrogen supply, Gordon *et al*. (1999) found that heather used more water than bracken, probably owing to its greater above ground biomass and transpirational demand, and to its less efficient use of water. This led them to suggest that, with its greater water-use efficiency and ability of its rhizome to store water (which may act as a temporary buffer during drought), bracken might be more tolerant than heather to drought stress.

However, heather was the superior competitor for water as it reduced its water-use efficiency as a consequence of its roots depleting water from the bracken rooting zone, but bracken, the poorer competitor, was less stressed by low water availability and increased water-use efficiency to cope with reduced water availability owing to competition (Gordon *et al*., 1999). Therefore, Gordon *et al*. (1999) suggest that another view could be taken, that if a good competitor is defined by its ability to survive at low resource availability, bracken could be considered the superior competitor for water. This study provides a good example of how difficult it can be to predict the potential outcomes of climate change.

Gordon *et al*. (1999) found that, in general, heather was more responsive to the environmental change treatments than bracken, which can be interpreted as an advantage in favourable conditions, but a disadvantage in stressed conditions. Overall, in conditions of climate change, Gordon *et al*. (1999) suggest that the occurrence of drought, particularly in combination with high nitrogen availability, will have a greater effect on the water relations of bracken and heather than increased temperature.

Werkman and Callaghan (2002) also examined the effects of increased air temperature and additional nitrogen on bracken and heather, on both pure canopies of bracken and heather and at the boundary between the species. They report that biomass of deciduous bracken fronds (which represent within-year response potential), was significantly increased by higher growing temperatures, while biomass of the perennial rhizomes of the bracken (which represent growth over a number of years), was not altered after three years of temperature and/or nutrient treatments. Werkman and Callaghan (2002) also note that rhizome biomass was not significantly different between the pure bracken plots and the boundary plots, but frond density was significantly lower in the boundary plots. Additional nitrogen had little effect on bracken, which Werkman and Callaghan (2002) suggest was possibly due to the high efficiency with which bracken recycles nutrients. Direct effects of the temperature and nitrogen treatments on heather were small: heather in the pure canopy responded only marginally to the treatments. However, heather showed considerable reductions in vigour in the boundary plots. In this case it appears both species suffered decreases in growth when growing together, but it is important to note that the effect of water was not considered in this study and as shown by Gordon *et al*. (1999), the effect of drought is likely to be greater than increases in temperature. The affect on above- and below-ground biomass in bracken also shows the importance of considering effects in the longer-term.

In contrast to Werkman and Callaghan (2002), Whitehead *et al*. (1997) found that greater availability of nutrients (including nitrogen) stimulated bracken growth in a way which would contribute to its invasiveness (such as encroachment into heather areas where the two species compete). However, heather growth also responded to an increase in available nutrients and showed a greater overall response to rising carbon dioxide levels than bracken. They suggest that a greater availability of nutrients, which could result from a combination of gradual soil warming and atmospheric deposition, and elevated carbon dioxide levels, could significantly alter the competitive balance between bracken and heather, although precisely how is difficult to predict due to regulation by a wide range of biotic and abiotic factors and the often over-riding impact of management practice. This shows how difficult it is to determine the potential impacts of many interacting factors, which are not fully known.

Another factor, examined by Anderson and Hetherington (1999), is the potential effect of atmospheric nitrogen deposition and climate warming on the decomposition of heather and bracken litter, which can have important feedbacks on the growth balance of moorland vegetation. Anderson and Hetherington (1999) found that the decomposition rates of heather and bracken litters were higher in nitrogen-fertilised plots than in control plots, while mixtures of both heather and bracken litters showed higher again mass losses. Increased temperature also had an effect of increased microbial activity and rapid depletion of available carbohydrates for both litters. Anderson and Hetherington (1999) concluded that their results indicate that soil processes (in the moorland system) are sensitive not only to the direct effects of changes in climate and nitrogen availability, but that there are also important indirect effects of species interactions in plant and microbial communities which may enhance the rates at which the system responds to environmental change.

30.6. Soil

Consideration of the effect of environmental change on soils and soil processes is also important, as changes in temperature and precipitation, as well as other factors such as atmospheric nitrogen deposition could have significant effects on nutrient availability and rate of decomposition in soils, which will influence the community composition of habitats, via processes such as plant competition (DEFRA, 2001). The response of plant species to climate change will, therefore, in many cases be strongly affected by the impact of climate change on soils. Further, it has also been suggested that increased rates of soil mineralisation in a warmer climate will probably increase the incidence and severity of insect damage to heathland plants (DEFRA, 2001).

Climate change is also likely to impact on soil fungal diversity, with knock-on effects, such as declines in soil fungal diversity in heathland soils as a result of drought, which could affect organic matter decomposition (Toberman *et al.*, 2008), and on ectomychorrizal fungi, which could affect the plants with which they are mutualistic (Malcolm *et al.*, 2008).

However, Emmett *et al*. (2004) suggest that key soil processes will be differentially affected by predicted changes in rainfall pattern and temperature and the net effect on ecosystem functioning will be difficult to predict without a greater understanding of the controls underlying the sensitivity of soils to climate variables (Emmett *et al.*, 2004).

Appendix 31. Potential effects of climate change on species

General effects of climate change on species

The physiology, morphology, and life history of species constrains them to survive in a particular range of environmental conditions, but humans may be imposing combinations of constraints that already do, or may soon, fall outside the ranges within which many species evolved (Tilman and Lehman, 2000). Recent analyses of long-term data sets from around the world indicate that some species are already responding to climate and atmospheric change (Hughes, 2000). However, Hughes (2000) highlights some inevitable caveats in interpreting these studies. First, positive trends are more likely to be both submitted and published than negative or inconclusive ones. Second, none of the studies represents a controlled experiment and thus different possible causes of the trends are confounded. Finally, the particular time intervals chosen for the data analyses can markedly influence the apparent strength of a trend. These caveats mean that no single study can be interpreted as unequivocal evidence for human-induced change. However, it is the increasing number of examples showing trends consistent with *a priori* predictions that are convincing (Hughes, 2000).

An extensive meta-analysis by Root *et al*. (2003) of 143 studies on (a wide range of) species across the world and global warming found that 81% of the species that show changes are shifting in the direction expected on the basis of known physiological constraints of species. Root *et al*. (2003) propose that their results suggest that a significant impact of global warming is already discernible in animal and plant populations with an estimated average global warming of only 0.6 °C. Therefore, the much greater predicted temperature increases and the synergism of rapid temperature rise and other stresses, particularly habitat destruction, could easily disrupt the connectedness among species and lead to a reformulation of species communities (Root *et al.*, 2003).

Hickling *et al*. (2006) also reported that out of a total of 329 species analysed across 16 taxonomic groups in the UK, over a period of approximately 25 years during the last four decades, 275 species shifted northwards at their range margin, 52 species shifted southwards, and two species' range margins did not move, with an average northwards shift across all species of 31–60 km. Comparable findings were obtained with respect to elevation shifts: 227 species shifted to higher altitude and 102 species shifted to lower altitude, resulting in a mean increase of 25m overall. Twelve of the 16 taxonomic groups showed significant northwards shifts and shifts to higher elevation over approximately 25 years during a period of climate warming. Further, for some less well known groups, the responses may even be greater than those already observed for more widely studied groups (Hickling *et al*., 2006).

However, despite the wide range of taxonomic groups considered, the results of Hickling *et al*. (2006) showed no clear taxonomic, ecological or physiological pattern in terms of the response of groups to climate warming. They therefore suggest that the wide range of responses found among species within almost all taxonomic groupings means that within-taxon variation in ecological traits such as habitat requirements, dispersal capacity, longevity and body size may preclude broader taxonomic generalisations.

Huntley (1991) also reported that species respond individualistically to climate changes. According to Huntley (1991), this has a number of important consequences; in particular, that communities will emerge as temporary assemblages of taxa whose components will dissociate as they respond individualistically to environmental change, and will form new associations under new environmental conditions. Further, even if such qualitative changes in community composition do not take place, quantitative changes occur that change the relative abundance of the component taxa and so can change the structural and functional attributes of the community. The responses of species are also likely to differ at different sites with different conditions (Peñuelas *et al*., 2004), meaning that it is very difficult to accurately predict the consequences of future climate change on habitats and species.

Another important point made by Hickling *et al*. (2006) was that all distribution changes are taking place not only in the context of climate warming but also in the context of land use and other environmental changes. This was exemplified by the three amphibian and reptile species included in their analysis, each of them at the north–western edge of their distribution in Britain, which should, in principle, have benefited from the warming that has been experienced in recent decades. However, Hickling *et al*. (2006) report that their distributions have collapsed southwards, each species surviving in remnant populations restricted to only a small fraction of their former distribution. Parmesan and Yohe (2003) also suggest that in an absolute sense, land-use change has probably been a stronger driver of twentieth century changes in wild plants and animals than has climate change.

Phase 1 of the MONARCH study (Harrison *et al.*, 2001) characterised currently suitable climate space in the UK for fifty species associated with a range of habitats and then used climate scenarios to estimate likely changes in spatial distribution by the 2050s. The results of the study showed that climate change appears to present threats for some species, as well as opportunities for others, although the models were only based on climate and did not include other factors likely to influence changes. In general, species with northerly distributions were expected to lose suitable climate space in Britain and Ireland but some species with more southerly distributions could expand their climate space.

The vast majority of studies of terrestrial biological systems reveal notable impacts of global warming, which are consistent across plant and animal taxa (Rosenzweig *et al*., 2007). The ecological effects on climate change on plants and animals can be group broadly into several main categories (deGroot *et al.*, 1995; Hughes, 2000; Root *et al.*, 2003; Rosenzweig *et al.*, 2007): changes in phenology (such as migration or flowering), changes in morphology/physiology (such as body size and behaviour), changes in distribution and abundance, shifts in genetic frequencies, and extirpation or extinction, and these are discussed in general in Appendix 32.

Species' traits affecting vulnerability to environmental change

Although individual species are expected to respond idiosyncratically to climate change, some species that share the same ecological properties (e.g. life history traits, life forms, history) might respond in similar ways (Thuiller *et al.*, 2005; Broennimann *et al.*, 2006). Studies of extinction risk and vulnerability to environmental change (such as climate change and habitat fragmentation) consistently show that species losses are non-random, with particular factors or species traits appearing to predispose species to be more vulnerable. These attributes tend to be interlinked, with many species possessing several of these traits (for example, species with small geographic ranges are also likely to be habitat specialists). From a review of the literature (see Appendix 33), there are several traits which are frequently cited which tend to make species more vulnerable to environmental change:

- Species with small population sizes and/or low population density (Fischer and Stocklin, 1997; Saetersdal and Birks, 1997; Henle *et al.*, 2004; Lavergne *et al.*, 2006), and/or with high population fluctuations (Fischer and Stocklin, 1997; Henle *et al.*, 2004).
- Species with small/restricted geographic distributions/ranges (Saetersdal and Birks, 1997; Thuiller *et al.*, 2005; Broennimann *et al.*, 2006; Walker and Preston, 2006; Hopkins, 2007), including species with narrow climatic tolerances (Saetersdal and Birks, 1997).
- Species with specialist habitat requirements (i.e. dependent on a specific habitat), such as for resources (e.g. breeding sites or food) or ecological processes (e.g. fire or grazing) (Fischer and Stocklin, 1997; Henle *et al.*, 2004; Kotiaho *et al.*, 2005; Ewers and Didham, 2006; Walker and Preston, 2006; Hopkins, 2007), including species with a reliance on mutualists (Ewers and Didham, 2006).
- Dispersal-limited species (intermediate or low dispersal power) (Henle *et al.*, 2004; Kotiaho *et al.*, 2005; Ewers and Didham, 2006; Hopkins, 2007).
    - Intermediate powers of dispersal in butterflies (Thomas, 2000).
- Species (for which the New Forest is) at the edge of their distribution range (Walker and Preston, 2006).
- Species with poor competitive ability (Walker and Preston, 2006).
- Species with ‘slow’ life histories (i.e. reproductive rate), low reproductive potential (Henle *et al*., 2004).
- Short-lived plant species (Lavergne *et al*., 2006) (although these sorts of species may be able to adapt more quickly to change).
- Species with low storage effects (Henle *et al*., 2004).
- Species at higher trophic levels (Ewers and Didham, 2006).
- Species with a large body size (Ewers and Didham, 2006).

Appendix 32. General effects of climate change on species

The ecological effects on climate change on plants and animals can be grouped broadly into several main categories (deGroot *et al.*, 1995; Hughes, 2000; Root *et al.*, 2003; Rosenzweig *et al.*, 2007) and these are discussed below.

Changes in phenology

The life cycles of most organisms are relatively fixed in time and are often strongly determined by climatological factors such as temperature and precipitation (water availability) (deGroot *et al.*, 1995). Therefore, phenology – the timing of seasonal activities of animals and plants – is a simple process in which to observe changes in the ecology of species in response to climate change (Rosenzweig *et al.*, 2007) because when environmental conditions change, many phases in the life cycles of most plants and animals may be affected, such as the timing of (de)foliation, leaf-burst and flowering, the timing of seed-setting and ripening, the length of the growing (or breeding) season, growth and the timing of migration and appearance/emergence of butterflies (deGroot *et al.*, 1995; Rosenzweig *et al.*, 2007).

An extensive meta-analysis by Root *et al.* (2003) of studies examining shifts in spring phonologies in a wide range of temperate-zone species indicated that the estimated mean number of days changed per decade for all species showing change in spring phenology was 5.1 days earlier. However, the number of days changed per decade for a given species ranged from 24 days earlier per decade to 6.3 days later per decade. Parmesan and Yohe (2003) also carried out a global meta-analysis and documented that shifts in phenologies that have occurred (27% of the species assessed showed no trends in phenologies) were overwhelmingly (87%) in the direction expected from climate change, with a significant mean advancement of spring events by 2.3 days per decade. Although there may be some differences in the numerical values from different studies (such as due to the use of different time periods and including different species), the overall trend towards an earlier onset of spring in northern mid-latitudes across a wide range of taxa is obvious (Walther, 2004).

McCarty (2001) reports that the results from a large number of studies indicate that many species have some capacity to respond rapidly to climate changes by altering the timing of life-history events and for some species, shifting activities to a cooler time of year may be a sufficient response to climate change. Certain species will fare better than others. For example, in terms of reproduction, species with several life cycles per year (such as weedy species and aphids) will be particularly favoured by the expected increases in temperature and lengthening of the growing/breeding season as their number of generations may increase and thus lead to an increase in seed capital and offspring (deGroot *et al*., 1995).

However, McCarty (2001) cautions that it is not safe to assume that such shifts will be a general pattern because timing of life-history events depends on factors besides temperature, and a shift in phenology may disrupt important correlations with other ecological factors. For example, species that regularly move between habitats, such as migratory birds, may need to adjust to climate changes that are occurring at different rates in different areas (McCarty, 2001). Animal-animal and plant-animal interactions, such as pollination and seed dispersal, depend on synchrony between species and species depend on the appearance of specific foods at critical times (McCarty, 2001). Changes in phenology may also influence the competitive power of species (deGroot *et al*., 1995). The issue is that species will respond individually and within their areas of distribution the changes in phenology will not be the same for each site, since the changes in climate will differ for each region and because many species behave phonologically differently at their limits compared to the centre of their distribution (deGroot *et al*., 1995). Although in some systems species may respond to climate change at similar rates and maintain synchrony, there will be many other species where the loss of synchrony may have detrimental effects (McCarty, 2001).

Changes in morphology/physiology

Many physiological processes in plants and animals are sensitive to changes in climate (e.g. temperature and precipitation) and greenhouse gas concentration (notably carbon dioxide) (deGroot *et al*., 1995), which can directly affect metabolic and developmental rates in animals, and processes such as photosynthesis, respiration, growth and tissue composition in plants (Hughes, 2000). The effects may be synergistic or counteractive. For example, plants may show slowed seasonal senescence (ageing) at elevated levels of carbon dioxide at the end of the season; on the other hand, senescence may be accelerated because of higher temperatures (deGroot *et al.*, 1995).

Many plants and animals exhibit phases of reduced activity (or dormancy) during certain periods of their life cycle, usually related to climatic conditions (e.g. cold or drought) (deGroot *et al.*, 1995). For example, many insects require a dormancy period for the development of certain stages in their life cycle, the breaking of seed dormancy in plants is often dependent on temperature, and various mammals, reptiles and amphibians have a winter rest to reduce energy loss and survive this period with low food availability (deGroot *et al.*, 1995). Changes in climate could therefore disrupt physiological processes during dormancy, such as a warm spell disrupting the dormancy period and activating the animals (deGroot *et al.*, 1995). This costs energy, while at the same time food availability is probably still restricted, which influences their fitness and chances of survival (deGroot *et al.*, 1995).

Changes in climate, particularly temperature, may also affect fecundity. For example, temperature can affect butterfly egg-laying rate and microhabitat selection (see Appendix 34.2). Temperature can also affect the sex ratio of some reptile species (McCarty, 2001). Effects on fecundity can also arise indirectly, such as decreased reproductive output reported in some insects, as a result of a trade-off with evolutionary adaptations that increase dispersal ability (such as increased thorax size), which has facilitated range expansion as a response to climate change (e.g. Hughes *et al.*, 2003).

Changes in distribution and abundance

In a review of the evidence, Huntley (1991) concludes that organisms will tend to respond to the forecast climate change by means of migration, as opposed to adaptive evolutionary changes. Further, they are expected to respond individualistically (in the timing, rates and direction of their migrations), as they did so to past climate changes, resulting in changes in the composition of communities and ecosystems, with novel assemblages, and including changes in broad-scale vegetation patterns (Huntley, 1991). Plantlife (2005) suggest that climate change poses a unique challenge for plants as they have to rely on the dispersal of their seeds or spores by wind, water or other organisms (such as birds) if conditions become unsuitable. Therefore, the interaction of plants with other species will be an important factor affecting the migration of some species in response to climate change and the response of such plant species will be dependent on the effect of climate change on interdependent species (DEFRA, 2001).

It is therefore clear that the extent to which species will be able to realise a migratory response depends upon a number of factors (Huntley, 1991). In the case of trees, the maximum migration rates observed in the past are believed to approach the maximum rates achievable by these long-lived sessile organisms and it is therefore considered very unlikely that most trees could achieve the migration rates demanded by forecast climate changes that are one to two orders of magnitude faster than those during the last deglaciation (Davis, 1989; Huntley, 1991). However, some tree species, such as those with wind-borne seeds, are able to respond faster than others (DEFRA, 2001). Nonetheless, the fact that adaptation to a new climate by long-lived species, such as trees, is hampered by the fact that significant changes will occur within the lifetime of individual plants makes them particularly vulnerable (DEFRA, 2001). It has been suggested that extreme climate events are likely to have a large impact on species survival and distribution changes, and due to the long-lived nature of trees, forests may survive in a moribund state for some considerable time (DEFRA, 2001).

An important consideration in climate change is that although many species may be able to disperse, their movements may be hampered by barriers (such as mountains, lakes or human infrastructure) and further, they may have to wait until favourable habitat conditions become available (deGroot *et al*., 1995). In addition, species and ecosystems are exposed to many non-climate environmental changes simultaneously, such as pollution, land use changes and habitat loss, alteration and fragmentation, which may also be exacerbated by climate change and mean that climate change may have an even greater impact on already-stressed species (deGroot *et al.*, 1995; Rosenzweig *et al.*, 2007).

Changes in the distribution of species have been reported to have occurred across a wide range of taxonomic groups and geographical locations during the twentieth century (Rosenzweig *et al*., 2007). A global meta-analysis by Parmesan and Yohe (2003) documented significant range shifts averaging 6.1 km per decade towards the poles, matching the expected response to warming. Root *et al*. (2003), in their global meta-analysis reported a similar trend, with 81% of the species showing change were changing in the direction expected with warming, including species expanding their ranges polewards. As well as expansions of distribution, there is also evidence that climate-driven range retractions are also already widespread (Thomas *et al.*, 2006).

Physiological and phonological responses to climate change, in combination with changes in interactions between species, will influence the (relative) abundance and distribution of most species (deGroot *et al*., 1995) and recent climate change has been linked to both increases and declines in population size (McCarty, 2001). McCarty (2001) suggests that because climate acts locally, changes will be most apparent on the level of populations and metapopulations.

Although the average climate is of great importance for the distribution of different species, extreme events form the predominant force influencing the survival not only of single individuals, but also of whole populations (deGroot *et al*., 1995). Parmesan *et al*. (2000) also note that although responses to extreme weather events vary, and may even be in opposite directions among different species in the same habitat, extreme weather can have major impacts on wild populations. Further, the impacts of extreme weather on one trophic level can cascade through a food web to affect other trophic levels (Parmesan et al., 2000).

Changes in physiology, phenology and distribution of individual species will inevitably alter competitive, and other, interactions between species, with consequent feedbacks to local abundance and to geographic ranges (Hughes, 2000). Such changes in species distributions and abundance will have effects on the structure and function of ecological communities and in many parts of the world, species composition has changed (Walther *et al.*, 2002; Rosenzweig *et al.*, 2007). The future diversity and abundance of pests and pathogens (which will also be affected by climate change) will additionally have a substantial impact on the composition and functioning of ecosystems (Walther, 2004). Hughes (2000) suggests that it seems probable that at least some species will become extinct as a result of such changes, either as a direct result of physiological stress or via interactions with other species.

Shifts in genetic frequency

Evolutionary responses to environmental change are rare in comparison to other responses (Huntley, 1991; Hossell *et al.*, 2000). Evidence from Quaternary examples suggests that many species, especially longer lived species (such as trees) will be unable to evolve rapidly enough to adapt to the predicted climate changes through natural selection (Hossell *et al*., 2000). However, species with rapid life cycles, such as insects, can show genetic changes to environmental pressures over several generations in a short time period, and may therefore show greater genetic adaptability to climate change (Hossell *et al.*, 2000). Such species, however, are likely to change their geographic distribution along with evolved adaptation responses (Hossell *et al.*, 2000). For example, insects expanding their ranges have undertaken genetically-based changes in dispersal morphology, behaviour and other life-history traits, as ‘good colonists’ have been at a selective advantage (Thomas *et al.*, 2001; Hughes *et al.*, 2003; Rosenzweig *et al.*, 2007).

From a review of the literature, Parmesan (2006) reports that evolutionary adaptations to warmer conditions have occurred in the interiors of species’ ranges, and resource use and dispersal have evolved rapidly at expanding range margins. For example, in non-migratory species, the simplest explanation of northward range expansions is that individuals have always crossed the species’ boundary, and with climate warming, some of these emigrants are successful at founding new populations outside the former range. When dispersal tendency is heritable, these new populations contain dispersive individuals and higher rates of dispersal will soon evolve at the expanding boundary (Parmesan, 2006).

However, Parmesan (2006) states that overall, the empirical evidence suggests that evolution can complement, rather than supplant, projected ecological changes and there is little theoretical or experimental support to suggest that climate warming will cause the absolute climatic tolerances of a species to evolve sufficiently to allow it to conserve its geographic distribution in the face of climate change and thereby inhabit previously unsuitable climatic regimes.

Extirpation or extinction

Key indicators of a species' risk of extinction (global loss of all individuals) or extirpation (loss of a population in a given location) include the size of its range, the density of individuals within the range, and the abundance of its preferred habitat within its range (Rosenzweig *et al.*, 2007). Each of these factors can be directly affected by rapid global warming, but the causes of extinctions/extirpations are most often multi-factorial (Rosenzweig *et al.*, 2007). There are also several traits that appear to make species more vulnerable to extinction (see Appendix 31). The contribution of climate change to future extinctions depends on how quickly species can respond to change, but ongoing climate change is an additional source of stress for species already threatened by local and global environmental changes, increasing the risk of extinction (McCarty, 2001).

Evidence suggests that climate-driven extinctions and range retractions are already widespread, but have been poorly reported due, at least partly, to a failure to survey the distributions of species at sufficiently fine resolution to detect declines and to attribute such declines to climate change (Thomas *et al.*, 2006). A prominent cause of range contraction or loss of preferred habitat within a species range is invasion by non-native species (Rosenzweig *et al*., 2007).

Invasive species represent a major threat to endemic or native biodiversity in terrestrial and aquatic systems (Fischlin *et al*., 2007). The causes of biological invasions are multiple and complex and species invasions also interact with other drivers, sometimes resulting in some unexpected outcomes (Fischlin *et al*., 2007). Changes in biotic and/or abiotic disturbance regimes are recognised as primary drivers of invasive species, with communities often becoming more susceptible to invasion following extreme events, such as are projected under future climate change (Fischlin *et al*., 2007).

Hossell *et al*. (2000) suggest that species invasion in the UK is likely to be restricted by the sea, and that the species pool of the UK is unlikely to change significantly over the next century as climate changes. However, they also note that introduced species that currently have very restricted ranges within the UK, or which exist only in parks and gardens, may increase in number and range if they are favoured by climate change.

The vulnerability of ecosystems and species is partly a function of the expected magnitude and rapid rate of climate change relative to the resilience of many such systems, as well as through the impact of multiple stressors, such as from human development (for example, through blocked migration routes, fragmented habitats, reduced populations, introduction of alien species and stresses related to pollution) (Schneider *et al.*, 2007), which will affect their ability to realise their new climate space (Berry *et al.*, 2003). Vulnerability will occur, for example, where there will be little or no overlap between the current and future distribution (Berry *et al.*, 2003).

Appendix 33. Literature review of species' traits affecting vulnerability to environmental change

Studies of extinction risk and vulnerability (such as to environmental change and fragmentation) consistently show that species losses are non-random, with particular factors or species traits appearing to predispose species to be more vulnerable (Purvis *et al.*, 2000). These attributes often tend to be interlinked.

Rarity *per se* does not necessarily mean that a species is in danger of extinction, as some species are naturally rare and possess attributes that confer a greater ability to persist in small populations (Stewart and Hutchings, 1996). However, many of the most threatened species are endemic species, with a very restricted range, such as confined to one area or island (Pullin, 2002). The World Conservation Union, IUCN (2001) classifies the level of threat or vulnerability of a species using criteria based on population size, rate of decline, area of geographic distribution and habitat, and degree of population and distribution fragmentation.

In line with this, Henle *et al.* (2004), in their review of predictors of species sensitivity to fragmentation, reported that population fluctuation and small population size, (or low natural abundance/high individual area requirement) are good predictors. They suggest that this is because small populations are more vulnerable to the proximate causes of extinction, such as demographic and environmental stochasticity. Low reproductive potential, low storage effects and intermediate or low dispersal power would also make species more vulnerable to habitat fragmentation, particularly if species possessed several of the above traits.

Henle *et al.* (2004) also noted some of the reasons why specialists have a higher extinction risk than generalists, such as because; specialists are likely to have a more discontinuous distribution than generalists; habitats are likely to be affected more strongly by abiotic changes, as a result of habitat fragmentation, for specialists than for generalists; and microhabitat specialists can be unwilling or unable to cross even small gaps in habitat. Here the degree of specialisation is an important factor.

In a study of butterflies in Finland, Kotiaho *et al.* (2005) found that threatened species tend to be specialists in both larval resource requirements (restricted resource distribution) and adult habitat requirements (restricted niche breadth), have poor dispersal ability and a short flight period. However, Thomas (2000) has shown that it may be species with intermediate powers of dispersal (as also noted by Henle *et al.* (2004) above) which will be most vulnerable to extinction. Thomas (2000) suggested that butterfly species of intermediate mobility have declined the most (compared to sedentary or high mobility species) due to a combination of metapopulation (extinction and colonisation) dynamics and the mortality of migrating individuals which fail to find new habitats in fragmented landscapes. Due to these factors it is implied that species will become extinct long before the last fragment of suitable habitat is lost as declining populations become increasingly susceptible to the stochastic probability of extinction. However, butterflies in all mobility classes will decline if their habitats decline. Thomas (2000) also discusses how not all taxa will exhibit the same pattern, for example, longer-lived and perhaps more intelligent organisms may have a greater ability to avoid leaving isolated patches of habitat and be better at relocating if they do.

Henle *et al.* (2004) also point out that the dispersal process is not only influenced by dispersal power, but also by the dispersal rate and colonisation probability, and there is also an issue of scale-dependency when using dispersal power as a predictor of fragmentation sensitivity. Henle *et al.* (2004) also not that species dependent on naturally patchy resources should have evolved strong powers of dispersal and therefore may have lower extinction risks than species with limited dispersal power.

Walker and Preston (2006), in a study of ecological predictors of extinction risk in English flora in two counties, found no significant relationship between extinction and dispersal ability, or reproductive mode. It was suggested that this presumably reflected the very rapid declines that many species have suffered as a result of land use change as there was convincing evidence that habitat loss had been the main determination of extinction in the study area. Therefore it was suggested that the distribution of many species reflects more the recent patterns of human habitat destruction than an equilibrium between local extinction and colonisation.

Walker and Preston (2006) did however, find that English rare species (restricted range size) and those species at the edge of their range in the study region suffered greater declines, as well as species specialised to certain habitats (particularly bog and dwarf shrub heath). The authors do note though, that in highly modified landscapes a species' realised niche combines both ecology and the evolutionary and land use history of the habitats in which it typically occurs and it was therefore not surprising that English range size was highly correlated with habitat specialisation, indicating the loss of certain habitats. In addition, it was not necessarily habitat specialism *per se* that caused the loss of species, but specialisation within the types of habitats which have been most susceptible to modern land use change. Walker and Preston (2006) also found that competitive ability was a factor, with a significantly greater proportion of short species going extinct or than taller species.

Demonstrating the vulnerability of specialist species, Broenniman *et al*. (2006), from their study based on 975 endemic plant species in southern Africa, report that species and life form vulnerability to global changes (based on their projected distributions in 2050) can be partly explained (and therefore partially predicted) according to species' geographical distribution and ecological niche properties. They found that generally the most vulnerable species were those with a restricted niche breadth (and thus often a restricted distribution), distributed within regions most exposed to climate change (i.e. high climate anomalies), or which direction of range change hits barriers to migration like seacoasts or mountains.

In their study of 1200 European plant species, Thuiller *et al.* (2005) found that species occupying a wider range of climatic conditions across their range should be the most tolerant to climate change. In addition, they found that climate change should contribute to increased extinction risk for species with restricted ranges, due to steep decreases in suitable habitat, although such species, if present in temperate regions, may potentially gain disproportionately large amounts of habitat. Thuiller *et al.* (2005) also note how the characteristics of many species with restricted ranges (such as low dispersal ability, weak competitors and small populations) could compromise the ability of these species to increase their range by colonising geographically disjunct habitats. Although Thuiller *et al.* (2005) also showed that generally marginal species are vulnerable to climate change, they did find that such species are not always the most sensitive to climate change because of the strong relation of sensitivity to other factors such as exposure, niche breadth and range size

The results of a study by Sætersdal and Birks (1997) of Norwegian mountain plants, suggested that the species most vulnerable to climatic warming would be those with narrow temperature tolerances, characterised by small range sizes and small population sizes, namely habitat specialists with a small geographic range size.

Similarly, Lavergne *et al.* (2006) found that over a 115-year study on the Mediterranean flora (in an area experiencing dramatic modifications due to global changes), long-term rate of rare species extinction was largely predicted by initial abundance, with species with the smallest initial abundance most prone to extinction. In addition they found that annual herbaceous species exhibited a high rate of decline or regional extinction, one of the reasons for which they suggested as a factor in this was that short-lived plant species have been shown to be more sensitive to environmental and demographic stochasticity (although these sorts of species may be able to adapt more quickly to change; see Appendix 34.1). This was shown by Fischer and Stöcklin (1997) for calcareous grassland plant species in the Jura mountains of Switzerland, an area which has seen destructive changes in land use. They also found that extinction events were significantly more frequent in populations of lower local abundance and for species whose life form indicates larger fluctuations in population size. In addition, species with high habitat specificity to the area were also more prone to extinction. Purvis *et al.* (2000) noted how even if a species occurs in a large resilient population, it may still be vulnerable to extinction if other species on which it depends do not share such traits.

A reliance on other species is also reported as a vulnerability factor by Ewers and Didham (2006) in their review of confounding factors in the detection of species responses to habitat fragmentation. They noted that species at higher trophic levels, habitat specialists, species with large body size and those with poor dispersal abilities or a reliance on mutualist species are expected to go extinct first when habitat area decreases. However, species can comprise a suite of traits, which can interact to increase susceptibility to fragmentation, and interact with environmental heterogeneity so that determinants of species vulnerability in one environment will not necessarily be the same in a different environment (Henle *et al.*, 2004; Ewers and Didham, 2006).

In a review of the effects of climate change on British wildlife, Hopkins (2007) suggested that species which fail to expand their range in response to climate change and therefore are likely to be more vulnerable, tend to be rare and local species with specialised habitat requirements and low powers of dispersal. Hopkins (2007) also noted that fragmented and patchy habitat additionally hampers range expansion, particularly for those species with low powers of dispersal. Further, the exacting habitat requirements of specialists make colonisation extremely difficult even if a species is successful in reaching suitable habitat. Barriers to colonisation can include modified soil structure or chemistry (Bailey, 2007), and changes in climate could cause such modifications.

Lambeck (1997) reinforces the ideas above that species vulnerability is the result of species being either process-limited (e.g. constrained by fire regimes), dispersal-limited (where patches of suitable habitat are beyond the distance over which individuals can move or are separated by a matrix that is too hostile to permit movement), resource-limited or area-limited (where patches of suitable habitat are too small to support a breeding pair or functional social group). Area-limited species are also resource-limited species, but Lambeck (1997) suggests that species be assigned to this group if the limiting resource is not obvious or quantifiable, making area a surrogate for resources. Species may fit in more than one of these categories and, as noted by Henle *et al.* (2004), single traits alone have limited predictive powers for extinction proneness (such as to fragmentation), and the importance of the different traits will change with the environment.

Appendix 34. Potential effects of climate change on selected species of the New Forest

34.1. Effects of climate change on plant study species

There are a large number of variables that constrain abundance of plants in both terrestrial and aquatic habitats including; resource limitation (such as nutrients, water, light and carbon dioxide), recruitment limitation (all sessile plants have the potential to have their abundance limited by dispersal), predators and pathogens, disturbances (such as fires) and temperature/climate (plants have optimal growth and competitive ability at particular temperatures and are limited by temperature extremes, and may be specialised on different seasons) (Tilman and Lehman, 2000). Changes in any of these constraints could therefore change the abundance of species and genotypes in a habitat (Tilman and Lehman, 2000).

One of the expected changes as a result of warmer temperatures is a longer growing season (Walther, 2004). This may possibly benefit *C. nobile* and *G. constrictum* which have fairly long flowering times (June to August and May to August respectively), but particularly *G. constrictum*, which may potentially be able to flower earlier. However, unless its pollinators also lengthen or advance their activity, this may not confer any advantage, and its pollinators are apparently unknown. *G. illyricus* only flowers for a short period (within June to July), and may potentially flower earlier. *P. globulifera* is a fern and does not flower, but produces spores (see Appendix 2.7).

Available evidence shows an overall trend towards an earlier onset of spring in northern mid-latitudes (Walther, 2004). For example, Fitter and Fitter (2002) found that the average first flowering date (FFD) of 385 British plant species has advanced by 4.5 days during 1991-2000 compared with the previous four decades (1954-1990) in a single locality in south-central England. This coincided with marked warming in the area during that period (temperature, or possibly a climate variable correlated with temperature (such as sunshine hours) is a key determinant of flowering time). Of the selected species, 16% flowered significantly earlier in the 1990s than previously, with an average advancement of 15 days in a decade, and 3% flowered significantly later in the 1990s than previously.

Although their findings only apply to a certain area, they may provide a useful indication of potential changes elsewhere. Further, Fitter and Fitter (2002) found that there was no taxonomic pattern to their data above the genus level, so extrapolated their findings to the wider British flora to suggest that 150 to 200 species may be flowering on average 15 days earlier in Britain now than in the very recent past. Sparks *et al*. (2000) suggested that early flowering would result in increased fecundity unless a mismatch with pollinating insects, where appropriate, occurred. However, Hegland *et al*. (2009) suggest that phenological responses to climate warming may occur at parallel magnitudes in plants and pollinators, and the tendency of generalism in pollination interactions may itself ensure that most species are not severely affected by climate-driven mismatches. However, they suggest that considerable variation in responses across species should be expected.

Fitter and Fitter (2002) reported greater advancement in first flowering date (FFD) of British plants with certain characteristics. For example, annuals showed a greater advancement in FFD than perennials. The four plant study species are all perennials so are less likely to shown an advancement in this respect. However, Fitter and Fitter (2002) also found that insect-pollinated species showed a significantly greater advancement than wind pollinated ones. All of the three flowering study species (i.e. not including *P. globulifera*) are insect pollinated (Stokes, 1987; Ford *et al.*, 2008a; b; 2009), although *C. nobile* frequently reproduces by clonal spread (Winship and Chatters, 1994; Plantlife, 2001) and *G. illyricus* reproduces primarily by offsets (Stokes, 2000; Taylor, 2002), so is likely to be less affected. The full range of pollinators is not known for these species, so it is difficult to comment on the potential for asynchrony between plant and pollinator. However, it appears, for example for *G. illyricus* which is pollinated by more than one insect order (Lepidoptera and Hymenoptera), that these interactions are likely to be generalist and therefore have less potential to be disrupted.

In addition, Fitter and Fitter (2002) found that the further the centre of the distribution of the species was from the locality (of central England), the less likely it was to respond to warming by earlier flowering. None of the species are near the centre of their distribution ranges and for *G. illyricus* in particular, the New Forest is at or near the northern limit of their distributions, which tend to extend into southern Europe. *P. globulifera* and *C. nobile* also occur substantially further north in the UK, in Scotland (see Appendices 2.1 and 2.7 for details). Walker and Preston (2006) suggest that species on the edge of their range are more vulnerable to local extinction, which may suggest that *G. illyricus* and *G. constrictum* could be at a potentially greater risk of any negative effects of environmental change in the New Forest.

However, the fact that all of the species are found in warmer and drier southern parts of Europe may indicate that they may be able to tolerate to some extent the forecast climatic changes for the New Forest. Although local adaptations may make this more difficult, for example *P. globulifera* appears to occupy a slightly different niche in Scotland to in the south (Botanical Society of the British Isles, 2008b) (see Appendix 2.7). Further, there is debate as to whether *G. illyricus* in Britain is sufficiently different from its European counterparts to warrant its designation as a separate subspecies (Stokes, 1987; Toone, 2005; Botanical Society of the British Isles, 2008a; also see Appendix 2.3). However, species may have the potential for further adaptations, for example DEFRA (2001) suggest that many perennial plant species may be able to adapt *in situ*, although the ability to respond to change for some species will depend critically on the genetic variation that is available within the population.

As a general rule, it is suggested that species will migrate northwards and upwards in altitude, as their current location warms, to cooler conditions more akin to their original location (Plantlife, 2005). It may therefore be that *G. illyricus* and *G. constrictum*, which are at their northern limits in the south of England, could expand northwards. However, it may be that climate is not the factor limiting their distribution and their ability to move northwards would depend on the availability of suitable habitat and their ability to migrate/disperse. The fact that the species generally have specialist habitat requirements (e.g. *C. nobile* requires very short grazed turf, *G. constrictum* and *P. globulifera* require particular wetland habitats and *G. illyricus* requires a specific bracken habitat) is another factor that is frequently cited as a vulnerability trait (see Appendix 31) and will also hamper their ability to move northwards across fragmented landscapes (Huntley, 1991; DEFRA, 2001; Plantlife, 2005).

P. globulifera was one of the species included in the MONARCH assessments (Walmsley and Harley, 2007), at a resolution of 50 km, across the UK. The models showed an approximately 50% increase in potential suitable climate space and 100% overlap under the 2020s Low and High scenarios (based on the UKCIP02 scenarios (Hulme *et al.*, 2002)). This overlap remained for the 2050s low and High scenarios but was reduced to 75% for the 2080s High scenario, where there was also a 25% loss. However, *P. globulifera* can already be found at scattered sites throughout the UK (see Figure A-15, Appendix 2.7). Further, the MONARCH models only highlight potential changes in climate space due to direct impacts of climate and none of the many indirect impacts of climate change, such as changes in the agriculture, forestry, water resource, spatial planning, and coastal management sectors, which will also affect biodiversity. They also do not take account of other factors affecting species distributions, such as interactions with other species and dispersal ability.

In terms of potential dispersal to other sites, the other large areas of semi-natural habitat north of the New Forest include the North Wessex Downs and Cranborne Chase and the Wiltshire Downs to the north-east but they are largely chalk downland and may not necessarily contain appropriate habitat, although there are some records for *G. constrictum* in the Cranborne Chase and Wiltshire Downs area (National Biodiversity Network, 2009) and for *C. nobile* in the Chilterns to the north-east of the New Forest. However, it may be that smaller areas, such as Sites of Special Scientific Interest (SSSI's) and other nature reserves may be more significant and important in providing corridors or stepping stones of habitat, although the isolation of some habitats may make them less suitable in the long-term. Further, the specific requirements of many of the species may make their colonisation less likely.

Dispersal-limited species are also expected to be more vulnerable (see Appendix 31) and will have even more difficulty shifting or expanding their range. *G. illyricus* appears to be particularly dispersal-limited (Stokes, 1987), although it is not so clear about the other plant species. The spores of *P. globulifera* can be carried from pool to pool on the feet of livestock and waterfowl (Plantlife, 2006), which aids its dispersal in the New Forest. *C. nobile* is also able to spread in a similar way and even small fragments (which are aided in their spread by unenclosed grazing) can form a new plant (A. Byfield, personal communication, February 27, 2009). Dispersal in *G. constrictum* is less well known.

Within the New Forest, *G. illyricus* and *G. constrictum* in particular appear to have small population sizes, which is another trait suggested to increase vulnerability to environmental change (see Appendix 31). *P. globulifera* and *C. nobile* seem to have slightly larger populations (and both appear to be largely under-recorded). All of the species are also poor competitors, which is likewise highlighted as a vulnerability trait (see Appendix 31). This could therefore make them particularly vulnerable to invasive species (which environmental change is expected to favour (Walther, 2004)), such as New Zealand Pigmyweed (*Crassula helmsii*) in the case of *P. globulifera*, which requires bare substrate.

All four of the plant study species possess several of the traits that are suggested to predispose them to be more vulnerable to environmental change (see Appendix 31), which is likely to make them more susceptible to climate change (Henle *et al.*, 2004). However, the effects of environmental change on species will likely be complex, and have different impacts in the short- and long-term, and it may be that some species will be able to persist where they are for a certain amount of time. For example, perennials can often survive unfavourable periods vegetatively (deGroot *et al.*, 1995). Particular attributes of species may make them more or less vulnerable and the impacts on their habitats and other factors (such as management) will also be important, and these are discussed for each species below.

Changes in climate and atmospheric carbon dioxide concentrations can cause physiological changes in plants. For example, increasing carbon dioxide levels might have lowered stomatal density, and therefore stomatal conductance in many species (Hughes, 2000). This will also have implications for water exchange and elevated carbon dioxide levels may reduce drought stress, although these relationships are complex (and modulated by other factors) and differ between species and habitats

(DEFRA, 2001). This also demonstrates how changes in one factor can have an impact on another factor. Callaghan *et al*. (1992) suggest that the greatest uncertainties of plant responses to a changing climate relate to the duration of physiological responses because of potential acclimation, the possible degree of mismatch between the predicted rate of climate change and the potential migration rates of biota, and the long-term responses of long-lived plants to changes in environmental factors. Further, plant responses to climate change will be confounded by their responses to other environmental changes occurring concomitantly, such as the deposition of atmospheric pollutants and increased levels of UV-B (Callaghan *et al*., 1992) and the stomatal density example described above.

Plantlife (2005) suggest that plants that are fast growing, able to spread rapidly and are not specific in their habitat requirements could do particularly well as a result of climate change since they are able to make the most of any opportunities. However, these are also the characteristics of an invasive species, which could pose a threat to some of the study species, in particular *P. globulifera* (see Appendix 34.4 below).

34.1.1. *Chamaemelum nobile*

As discussed in Appendix 34.1, *C. nobile* possesses some traits that may make it potentially vulnerable to climate change, but it also appears to have some characteristics that may confer tolerance. Being able to reproduce clonally and being relatively long-lived (Winship and Chatters, 1994; Plantlife, 2001) may provide several advantages, according to Callaghan *et al*. (1992). These include: being less sensitive to short term changes in the environment (although their longevity and reduced recruitment from seed can make them vulnerable to permanent and sudden changes in the environment); the ability for rapid spread by vegetative means over short distances; the possibility for extension of developmental processes over various time scales which increases the chances of clonal plants withstanding local and short term adversities (Callaghan *et al.*, 1992).

The main habitat of *C. nobile* is (acidic) grassland and grassy heath (including wetter areas) (Winship and Chatters, 1994; Brewis *et al.*, 1996). It is not exactly clear how grassland will be affected by the different aspects of climate change and increased atmospheric carbon dioxide. However, it seems unlikely that acid grassland in the

New Forest will be under great threat, although it may become drier. There may be some encroachment of heather into dry grasslands as happened during the droughts in the New Forest during 1974 – 1984 (Tubbs, 2001), which could be a threat to *C. nobile* (G. Read, personal communication, February 13, 2009). These changes will also depend on how the habitats are managed. Winship (1993) suggests that the spread of rhododendron (*Rhododendron ponticum*) and bracken (*Pteridium aquilinium*) onto heaths and commons, which dominate and shade native flora, are also a potential threat to *C. nobile*, but this will depend on the effect of environmental change on those species and on their management. Management, particularly in terms of grazing, mowing and scrub clearance will be particularly important for *C. nobile* (Winship, 1993; 1998).

An interesting point is that during the droughts in the New Forest during 1974 – 1984, Tubbs (2001) observed that the grassland flora increased in diversity with the invasion of small annual plants which were able to become established with the suppression of more vigorous grasses. The suppression of more vigorous grasses would likely be of benefit for *C. nobile*, but the increase of small annual plants may offset any advantage. It would also depend on how drought tolerant *C. nobile* was. *C. nobile* has a shallow rooting system (A. Byfield, personal communication, February 27, 2009), but can be tolerant of short-term parching (C. Chatters, personal communication, March 13, 2009). However, the dry summers of 1991 and 1992 were reported to affect *C. nobile* populations across Britain (Winship, 1993).

It has been suggested that increased temperature and rainfall in winter may favour the increased growth of more competitive grasses, to the detriment of less-fast growing species (Diack, 1999). Further, summer droughts may increase nitrogen mineralisation and increase the nutrient status of grassland sites, encouraging invasion of more competitive species (DEFRA, 2001). However, *C. nobile* can flourish in relatively high nutrient situations providing that sufficient grazing, mowing or trampling occurs to prevent grasses and taller plants from dominating the sward (Winship and Chatters, 1994). Therefore, providing grazing by livestock is maintained in the New Forest in the future, this should not pose too much of a problem. Nonetheless, patterns and levels of grazing may need to be adjusted to take account of changing precipitation patterns under climate change (Diack, 1999;

Hossell *et al.*, 2000). The New Forest National Park Authority (2007a; Trotter, 2007) suggest that changes in agricultural practices in response to climate change may affect free-roaming grazing animals, which are dependent on the survival of pastoralism and commoning; this could have implications for *C. nobile*.

Potentially increased visitor numbers with possibly associated increases in trampling and widening of paths (Gallagher *et al.*, 2007) should also not be a problem and may even be beneficial for *C. nobile* as trampling helps to discourage competitors (Plantlife, 2001) and also creates bare ground. However, too much trampling of this kind could potentially remove *C. nobile* from some sites.

Increased winter precipitation should not be a problem in itself, and may even be beneficial for *C. nobile*, as it favours sites which are seasonally flooded (Plantlife, 2001). However, as noted above, it may increase the growth of more competitive species (Diack, 1999). Dry summer conditions are also likely to favour deep-rooted plants, with shallow rooting species, such as *C. nobile*, more likely to suffer in the drier conditions (Hossell *et al.*, 2000). A potential problem of increased winter precipitation may be an increased risk of run-off and leaching of nutrients, which will also favour the growth of more competitive species (Plantlife, 2001), although this should be less of a problem in somewhere like the New Forest.

34.1.2. *Galium constrictum*

As discussed in Appendix 34.1, *G. constrictum* possesses several traits that may make it more vulnerable to environmental change, in particular because it is largely restricted to wetland habitats, which are among the most vulnerable to climate change. Information on this species is limited and its ability to migrate and its tolerance to such changes are not well known. However, as *G. constrictum* appears to be fairly limited in its occurrence in the New Forest, and elsewhere in the UK (although it may frequently be misidentified (G. Read, personal communication, August 8, 2009)), and is restricted to a specific habitat, it is likely to be at greater risk from climate change because of an inability to migrate to more suitable locations, largely because they are not available (DEFRA, 2001; Plantlife, 2005).

Wetland areas will be particularly at risk from reduced rainfall and persistent drought in the summer (Hossell *et al.*, 2000; DEFRA, 2001), with many streams, ephemeral ponds and wetlands potentially completely drying out, as observed during droughts in the New Forest during 1974-1984 (Tubbs, 2001). *G. constrictum* can survive being parched for part of the year (C. Chatters, personal communication, March 13, 2009), as it frequently occurs in seasonal ponds. However, it is not clear how tolerant it would be of the more extreme drying out that is expected.

A further threat is that drought may exacerbate pollution problems in wetland habitats by concentrating pollutants in periods of low flow (Carpenter *et al.*, 1992; Hossell *et al.*, 2000), as well as nutrients, encouraging eutrophication (Diack, 1999). An increase in heavy rainfall and flash flooding may also increase the leaching of fertilisers, herbicides and pesticides from surrounding land, as well as increasing levels of sedimentation (Hossell *et al*., 2000), having detrimental effects on *G. constrictum*. Increases in flash floods may, if they are severe enough, scour river beds, washing away much aquatic life (Diack, 1999), potentially including plant life such as *G. constrictum,* where it occurs on the edges of rivers or streams.

It has also been suggested that an increase in temperature will increase decomposition of organic soils and weathering of rock, which may produce an increase in pH in some water bodies (Eyre *et al.*, 1993; Hossell *et al.*, 2000; DEFRA, 2001). As it is likely that *G. constrictum* favours a circum-neutral pH (C. Chatters, personal communication, April 2, 2009) too much of a difference in this may be a problem, although the pH tolerance for this species is not known. However, it is unlikely that the pH would change so much so that it was no longer tolerable for *G. constrictum*.

A key concern is the potential increase in invasive species (such as New Zealand pigmyweed, *Crassula helmsii*) which are expected to spread more rapidly if winters become warmer and frost events are less frequent (Diack, 1999). Encroachment of other species, such as bracken, rhododendron and heather as well as scrub into wetter grassy areas may also pose a threat. The height of the sward is not important for G. constrictum, although it probably cannot compete with dense vegetation (C. Chatters, personal communication, March 13, 2009), so any changes in grazing would not have as much as an effect as for *C. nobile*, although may still be detrimental, particularly if bare ground is reduced.

34.1.3. *Gladiolus illyricus*

As discussed in Appendix 34.1, *G. illyricus* possesses several traits that may make it more vulnerable to climate change. As *G. illyricus* occurs at the northern limit of its range in the New Forest, it may be plausible to suggest that it could potentially spread northwards (and there are a few older records from sites further north). However, *G. illyricus* appears to be particularly dispersal-limited, having quite a large heavy seed (A. Byfield, personal communication, February 27, 2009) and reproducing mainly by offsets (Stokes, 2000; Taylor, 2002). The means of dispersal are not yet known apart from gradual site migration (Toone, 2005). Toone (2005) also notes that as the practice of bracken cutting was common in the New Forest until the mid twentieth century, as a means of spread this is suggestive, though inconclusive without more data. The fact that *G. illyricus* would also have to move across a fragmented landscape and requires a specific habitat makes it unlikely that this would occur without human intervention.

However, the Botanical Society of the British Isles (2008a) suggest that whether *G. illyricus* is actually native in the New Forest and whether it is a separate subspecies are of key importance, because if *G. illyricus* is a southern, ruderal species, it might be expected to increase if the climate continues to warm and more waste land is created by processes such as agriculture and mineral extraction; whereas a native species might find itself endangered by changes in the climate and destruction of its specialised niche. For example, if *G. illyricus* is a southern European species, it might be expected that it would be able to tolerate warmer, drier summers. However if it is a separate subspecies, it may have specific adaptations to living in the cooler north of Europe and be less tolerant.

Tubbs (2001) notes that there is an enormous resource of *G. illyricus* corms in the New Forest sites of which only a few push up shoots each year, and they are likely to be long-lived. Stokes (1987) found that only 2% of the *G. illyricus* population in his study sites developed into above ground plants each year. Tubbs (2001) therefore suggests that the species is probably well buffered against decline in the long term. It is not clear how tolerant *G. illyricus* is likely to be of drought, but this strategy may be beneficial. However, occurrence under bracken may mean increased competition for water. Nonetheless, although the corms may allow survival through several 'bad' years, if the trend is to conditions that are detrimental to *G. illyricus* then they may eventually diminish. Being long-lived means they may also be less able to adapt (Hossell *et al*., 2000).

The impact of climate change on bracken will be key for *G. illyricus*. Although the abundance and distribution of bracken in Britain overall is expected to increase as a result of climate change (Marrs and Watt, 2006), this is unlikely to have much of an impact on *G. illyricus* due to its limited ability to spread. It is not completely clear how bracken might be affected in the New Forest, but in the short term is unlikely to be negatively affected and bracken biomass is unlikely to change much (Pakeman and Marrs, 1996). If anything, bracken may increase in the New Forest and expand into heathlands (New Forest National Park Authority, 2007a; Trotter, 2007), at least in the short-term. If bracken does increase in the New Forest, it could potentially provide more habitat for *G. illyricus* that is protected from grazing (although *G. illyricus* may not be able to benefit from it due to its limited dispersal). However, if bracken increases in density at many sites this may be detrimental to *G. illyricus* by smothering it (unless bracken is managed at the site; see below).

The decrease in frosts will likely be beneficial for bracken as frost can kill newly emerging bracken fronds (Marrs and Watt, 2006). However, it is this process which is beneficial to *G. illyricus* as it creates a more open bracken canopy later in the season (Stokes, 1987). If bracken growth becomes more vigorous earlier on in the season, as a result of less frost damage and increased temperatures, this may be detrimental to *G. illyricus* plants. Indeed, Stokes (2000) suggests that if there are too many mild winters *G. illyricus* will decline as it cannot compete with a dense bracken canopy, yet overzealous bracken control could result in the gladioli being overgrazed (Stokes, 2000). However, it may be that warmer temperatures cause *G illyricus* to flower slightly earlier, such as shown for British plants by Fitter and Fitter (2002). It was observed during fieldwork carried out as part of this project (S. Douglas, personal observation) that there were differences between flowering dates of *G. illyricus* at different sites within the New Forest, possibly due to local microclimate. For example, on the same day, most of the plants were in flower at one site, but a few miles away at another site, none of the plants were in flower, only in bud.

There have been suggestions that numbers of *G. illyricus* have diminished since the cessation of bracken cutting in about the 1960s (V. Scott, personal communication, cited in Toone (2005)), but Toone (2005) notes that *G. illyricus* can persist in very dense bracken and it is doubtful that bracken cutting can claim continuity over the last 8,000 years. However, although *G. illyricus* exists in dense bracken, that does not mean to say that the conditions are particularly favourable (in the long-term) and it may just be 'hanging-on' in sub-optimal habitat.

Stokes (1987) suggests that the survival of *G. illyricus* in the New Forest is likely to be threatened if a decision is made to increase or decrease the grazing levels (which may be altered as a result of climate change) as this could affect the balance in the bracken sites where gladioli occur and lead to the smothering or overgrazing of the plants. Further, Stokes (1987) also notes that biological control of the bracken either in the New Forest or nationwide could have a dramatic effect on the *G. illyricus* population, reducing their protection from grazing and increasing the competition from other plants, and repeated cutting of the bracken at an unsuitable period in time (e.g. during flowering), would lead to the eradication of the species from that area. Therefore, the way in which bracken is managed in the future in the New Forest, or at least at *G. illyricus* sites, is likely to have a large effect on the future status of this species.

It has been suggested that in the south-east of England bracken biomass is restricted by low water availability and bracken may therefore decrease in the longer-term as a result of drought in the south (Pakeman and Marrs, 1996; DEFRA, 2001; Hossell *et al.*, 2005). The results of the MONARCH climate space models (Hossell *et al.*, 2005) show a complete loss of bracken in southern England under the UKCIP02 2050s High scenario. However, bracken is common throughout the British Isles and occurs throughout the world except for temperate South America and the Arctic, although it is divisible into a number of geographical subspecies (Clapham *et al.*, 1990). Therefore, it appears tolerant of a large range of climatic conditions, although this appears to depend on the specific adaptations of subspecies. Further, it is able to increase its water-use efficiency and store water in the rhizome, which may act as a temporary buffer during drought (Gordon *et al.*, 1999). However, a significant decline of bracken in the New Forest would have serious consequences for *G. illyricus* (assuming that grazing levels are still high).

Although decomposition rates of bracken litter may be affected by future effects of climate change and nitrogen deposition (Anderson and Hetherington, 1999), the depth of bracken litter will still mainly be affected by the density of the bracken stand. If bracken biomass does increase this could have detrimental effects on *G. illyricus*, both in terms of the canopy and the bracken litter. However, it will also depend on how bracken is managed in the future and, as noted by Tubbs (2001), because the corms are long-lived, colonies that have apparently been lost under deep bracken litter may be induced to reappear.

Stokes (1987) suggests that *G. illyricus* soils need to be damp enough to prohibit bracken from becoming dominant, without being so wet that the corms rot, indicating that drier soils may benefit bracken, but also that the forecast increase in winter precipitation may potentially be detrimental to *G. illyricus*. At some sites (particularly flat sites where water is more likely to lay) winter flooding could be more likely, which may make the sites too wet, causing the corms to rot. Bracken is tolerant of waterlogged soil (Marrs and Watt, 2006) and is therefore unlikely to be adversely affected by this. Bracken is likely to be fairly robust to other associated impacts of climate change, such as increases in insect herbivores and pathogens and to changes in grazing (Marrs and Watt, 2006), so if it is able to cope with the direct effects of climate change in the New Forest (at least in the short-term), such as decreased summer water availability, then it is likely to fare well.

In terms of the broad habitat that *G. illyricus* occurs in, namely grassy heaths (Stokes, 1987), it has been suggested that increased drying is likely to cause wet heaths to change to drier heaths (Cook and Harrison, 2001; Hossell *et al.*, 2005) or be replaced by an expansion of the region's acid grasslands (Cook and Harrison, 2001). As *G. illyricus* often occurs on dry grassy heath adjacent to wet heath and bogs (Stokes, 1987), it may potentially be able to eventually spread further into those areas (possibly depending on suitable soil conditions), although the presence of those wetter areas help provide conditions which *G. illyricus* favours (Stokes, 1987), and the sites may become less suitable if they diminish. Hossell and Rowe (2006) also suggest that drier conditions could favour the spread of more dense gorse scrub and woodland in the New Forest which, as *G. illyricus* frequently occurs near to woodland edges (Stokes, 1987), could pose a potential threat, although this may be prevented by appropriate management. Increased temperature, drought and nitrogen supply have been found to increase the competition between bracken and heather for water (Gordon *et al.*, 1999), but it is not clear exactly how this will affect the balance between these two species.

Finally, the suggestion that potential increases in visitor numbers is likely to result in trampling of vegetation and widening of paths (Gallagher *et al*., 2007) could pose a problem for *G. illyricus* at some sites. Numerous *G. illyricus* plants occured close to paths (often where the bracken canopy is more open as a result of the path) at several of the field survey sites and therefore may suffer damage.

34.1.4. *Pilularia globulifera*

As discussed in Appendix 34.1, and as for *G. constrictum*, *P. globulifera* possesses several traits that may make it more vulnerable to environmental change, in particular because it is largely restricted to wetland habitats, which are among the most vulnerable to climate change. The problems of drying up of wetland habitats are as for *G. constrictum* (see Appendix 34.1.2), but additionally, sporocarp formation in *P. globulifera* only takes place as water levels decline and expose the plant (Plantlife, 2006), which makes it more vulnerable as it will be less able to move in response to climate change if open water is limited (Hossell *et al*., 2000). However, sporocarps are probably carried from pool to pool on the feet of livestock and waterfowl (Plantlife, 2006), which will help it to disperse, as long as suitable habitat is available.

It has been suggested that after colonies die back during years of exceptional droughts, *P. globulifera* sporocarps may persist and remain dormant in mud, perhaps for many years, but further work is required to determine this (Page, 1982; Jermy, 1994). Growth can certainly vary widely from year to year, even in the same locality, depending on climatic conditions and water levels, and reappearance of *P. globulifera* can be sporadic (Page, 1982). The fact that spores can develop to produce new sporophytes within seventeen days (Jermy, 1994), makes it a rapid opportunistic coloniser (Botanical Society of the British Isles, 2008b), and means that it can take advantage of suitable conditions quickly when they arise.

However, *P. globulifera* is a poor competitor (Scott *et al.*, 1999) and longer periods of exposure in dried pools may make it vulnerable to competition from terrestrial plants or more tolerant aquatic plants unless it is again inundated with water (Brock and van Vierssen, 1992; Scott *et al.*, 1999). It is particularly vulnerable to invasive aquatic plants such as New Zealand pigmyweed (*Crassula helmsii*), water/floating pennywort (*Hydrocotyle ranunculoides*), parrot's feather (*Myriophyllum aquaticum*). These have been reported to have contributed to losses of *P. globulifera* in Sussex (Stewart *et al.*, 2000) and are expected to spread more rapidly if winters become warmer and frost events are less frequent (Diack, 1999). Stewart *et al*. (2000) suggest that the effects of these species on *P. globulifera* are not just through direct competition but also by accumulating organic mud beneath the vegetation mats. Trampling by ponies and cattle will also be important to help exclude more strongly growing aquatic species (Scott *et al.*, 1999), so maintenance of free-roaming cattle will be important. Potential scrub encroachment and other shading species would also have a negative effect.

Potential concentration of pollutants in wetland habitats as a result of droughts (Carpenter *et al.*, 1992; Hossell *et al.*, 2000) and increased leaching of fertilisers, herbicides and pesticides from surrounding land as a result of heavy winter rainfall (leading to eutrophication) (Hossell *et al.*, 2000; DEFRA, 2001) would also be a risk to *P. globulifera*. This would be through encouraging growth of competitive vegetation (UK Biodiversity Action Plan, 2008) and because *P. globulifera* is sensitive to pollution (Scott *et al.*, 1999). Pollution may also increase the pH of the water (UK Biodiversity Action Plan, 2008), as could increase decomposition of organic soils as a result of warmer temperatures (Eyre *et al.*, 1993; Hossell *et al.*, 2000; DEFRA, 2001), although the tolerance of *P. globulifera* to such changes is not known.

Diack (1999) suggests that it is likely that flash flooding in winter may increase as the greater intensity of rainfall increases runoff into rivers and if flood events are severe enough these may scour river beds, washing away much aquatic life. This could have detrimental effects on *P. globulifera*, which could potentially get dislodged. However, in the north of Scotland most of the records are from the sides of rivers and lakes where it has been suggested that perhaps the scouring effect of running water creates the open conditions it requires (Botanical Society of the British Isles, 2008b).

A further potential threat to *P. globulifera* may be an increase in intense recreational use of wetland habitats at some sites (UK Biodiversity Action Plan, 2008), which may be more likely if visitor numbers increase as a result of warmer drier summers (New Forest Committee, 2003).

34.2. Potential effects of climate change on butterfly study species (*Hipparchia semele* and *Plebejus argus*)

As the literature is relevant to both *H. semele* and *P. argus*, and they both occur in similar habitats, they are discussed together here. Both species are able to use more than one hostplant, although they do favour particular species (*H. semele* favours bristle bent (*Agrostis curtisii*) and *P. argus* favours cross-leaved heath (*Erica tetralix*) and heather (*Calluna vulgaris*)) (Oates *et al.*, 2000). Being able to use more than one foodplant is an advantage as they can switch between these resources during environmental perturbations that affect the synchronisation between the insect and its food supply or which selectively eliminate habitat components (Dennis, 1993).

However, these foodplants occur in specific habitats, which may be altered as a result of climate change. Decreased summer precipitation is expected to cause wet heaths to dry up in south-east England and likely revert to dry heath or be replaced by an expansion of the region's acid grasslands, although some species common on these grasslands may disappear as droughts intensify (Cook and Harrison, 2001). This would favour *H. semele*, but may not be as good for *P. argus*, which tends not to occur on heaths as dry as those that *H. semele* favours, tending to use the more humid heaths as well as wet heaths (A. Barker, personal communication, March 6, 2009).

The potential for the spread of more dense gorse scrub and woodland in the New Forest as a result of drier conditions (Hossell and Rowe, 2006) and possibly also the expansion of bracken (see Appendix 30) may reduce the quality of the habitat for both species. Both *H. semele* and *P. argus* are dependent on grazing in their habitats, so any potential negative changes in grazing regimes as a result of future environmental change (New Forest National Park Authority, 2007a; Trotter, 2007) could have a significant impact. However, over-grazing is also considered to be a contributing factor to some butterfly declines in the New Forest (Green, 2000).

Drier summer conditions could also increase the risk of fire on drier heaths (see Appendix 30). Although both species favour heaths that have recently been burnt, the timing (managed fires are strategically timed in a suitable cycle to minimise damage) and the frequency (plants may find it more difficult to regenerate with repeated frequent burning) of accidental fires may be detrimental to the species. Indeed, Asher *et al*. (2001) note that accidental summer burning has caused local extinctions of *P. argus*, from which it rarely recovers. Further, the exposure of large areas of bare ground as a result of fires may lead to an increased erosion risk, particularly with forecasted heavy rain in the winter (Alcamo *et al.*, 2007).

Heavy rain may be a major mortality factor by dislodging and drowning eggs and larvae (Dennis, 1993). For example, Dennis and Bardell (1996) report that torrential rain in June 1993 had an effect on *P. argus* and more so on *H. semele* populations at a site in north Wales. They suggested that *H. semele* may have been more affected by the storm because it pupates at a shallower depth than *P. argus* and therefore received less protection from soil erosion and so experienced higher mortality. Typical depths of pupae are reported to be 70mm for *P. argus* (Bretherton *et al.*, 1989) and 10mm for *H. semele* (Bailey *et al.*, 1989). Related to this, Dennis and Bardell (1996) suggest that the fact that *H. semele* tends to occupy habitats with more bare ground than *P. argus*, and unprotected ground is subject to much greater soil erosion than areas covered with vegetation, was also a factor. Increased visitor numbers in the New Forest are expected to potentially cause trampling and widening of paths (Gallagher *et al.*, 2007), which may also cause greater erosion and damage to larval or pupal sites, particularly as *H. semele* lays eggs near bare ground.

In this respect it appears that *H. semele* may be more vulnerable to storm damage, although both populations suffered in the following year, which Dennis and Bardell (1996) suggest may have possibly been due to the weather conditions in the 1993 flight period (cooler and less sunny, which can lead to lower fecundity) or the condition of the habitats during 1993 and 1994. Additionally for *P. argus*, the storm may have caused unusually high mortality of ants on which it depends, which could have influenced survival of larvae and pupae during 1994. However, Dennis and Bardell (1996) suggest that the impact of the storm was limited, perhaps because both species were in their least vulnerable stage (the pupal stage) at the time of the storm, but greater mortality may be expected had they been in any other stage. In terms of future climate change, torrential rain would be more likely to occur in the winter, as would longer periods of flooding. At this time of the year, *P. argus* eggs are passing the winter on woody stems of their foodplants (typically less than 10-15 cm high) in sparse vegetation with patches of bare ground (Asher *et al.*, 2001). Whereas *H. semele* larvae are hibernating in tussocks surrounded by bare ground (Asher *et al.*, 2001). Therefore, both species could potentially suffer losses as a result of winter flooding.

Weather and climate are well known to influence butterfly populations (Dennis, 1993). For example, Roy *et al*. (2001) reported strong associations between weather and population fluctuations and trends in 28 of the 31 British butterfly species they studied. The main positive associations they found were with warm summer temperature (especially in June) during the current and previous year, low rainfall in the current year and high rainfall in the previous year. Roy *et al*. (2001) suggest that June temperature is likely to be particularly important because for univoltine species (as *H. semele* and *P. argus* are), June is the period of late larval and pupal development, suggesting that survival of these stages was greater when high temperatures increased development rate. Pollard (1988) also suggests that weather in the current and previous years is important for butterflies because weather in the previous year may affect oviposition and larval development or have an indirect influence through the growth of foodplants. For 31 species of butterfly in Britain, during 1976-86, Pollard (1988) found a strong association between increased numbers and warm, dry summers.

It therefore seems that warmer temperatures in themselves are likely to largely benefit most butterfly species. Being ectotherms, higher temperatures should have beneficial impacts for butterflies, such as reducing the time required to raise body temperatures to flight activity thresholds and therefore increasing flight-dependent activities: mate-location, egg-laying, nectaring, predator-evasion and dispersal (Dennis and Shreeve, 1991; Dennis, 1993). In addition, egg production may be increased and the development of early stages enhanced by reducing the time exposed to predators, and faster development rates may lead to additional broods of species with restricted seasonal voltinism and to larger population sizes (Dennis, 1993). However, the overall effect of weather on population trends will be complex and difficult to predict, particularly, for example, if warm dry summer weather tends to increase abundance in the current year but reduce it in the following year (Pollard, 1988).

The length of the flight season may also influence vulnerability to perturbations, inasmuch as those which emerge over a longer period have a greater probability of surviving periodic catastrophes such as drought or intense episodes of predation, collecting, grazing or mowing (Dennis, 1993). Indeed, a short flight period was one of the traits more likely to be possessed by threatened butterfly species in a study in Finland (Kotiaho *et al.*, 2005). The flight period of *H. semele* is approximately 2 months (mid-July to mid-September) and for *P. argus*, just under that (mid to late June, disappearing by mid-August) (Oates *et al.*, 2000). Although this is not as long a flight period as some species, it is still a fairly substantial amount of time.

The flight period of some butterflies may extend as a result of climate change, which could make them less vulnerable (depending on other factors as discussed below). Sparks *et al*. (1999) report on an analysis of the Butterfly Monitoring Scheme data over a 23 year period which shows that virtually all species have responded to warmer springs and the majority showing statistically significant trends towards earlier appearance, demonstrating how responsive butterflies are to temperature. Roy and Sparks (2000) also report on changes in phenology in 35 British butterfly species from 1976-98 at over 100 sites throughout the British Isles. They found that first appearance of most British butterflies has advanced over the last two decades and that there is a strong relationship with temperature (with central England spring temperature having increased by approximately 1.5 °C and summer temperature by approximately 1 °C during that time), particularly with increase in February temperature. They suggest that this is because spring is a critical time for larval development for many species and increased temperature over this period is likely to advance emergence.

Together with early emergence, Roy and Sparks (2000) also found there was a concurrent advancement of peak appearance and longer flight duration, and predict that in the absence of confounding factors, such as interactions with other organisms and land-use change, climate warming of the order of 1 °C could advance first and peak appearance of most butterflies by 2 to 10 days. Some of the implications of this, as noted by Roy and Sparks (2000), are that, as well as increasing the duration of each generation, earlier appearance may allow those species capable of multi-voltinism to increase the frequency with which this occurs. Further, certain species which are univoltine in Britain, but have more than one generation in warmer parts of their range, may be able to take advantage of warmer British temperatures. This was recently reported (BBC News, 2009; The National Trust, 2009) for the Duke of Burgundy butterfly (*Hamearis lucina*), which normally only has one generation a year in the UK, but in more southern areas of its range in Europe, it produces a second brood in late summer. A second generation of this endangered butterfly was recorded in Gloucestershire, the furthest north recorded so far. The flight season has been occurring increasingly earlier over the past 20 years and warm weather in late May to early July have obviously benefitted the butterflies (BBC News, 2009; The National Trust, 2009). Earlier emergence of British butterflies, and associated increases in second broods is likely to lead to increased abundance, providing that hostplants and nectar supplies, which may be increasingly affected by drought as the season progresses, are still adequate (Dennis, 1993).

Both *H. semele* and *P. argus* are univoltine (Asher *et al.*, 2001). However, *P. argus* may occasionally have partial second broods (Dennis, 1993). A second brood may therefore become more likely for *P. argus*, although this would depend on adequate quantity and quality of resources (Dennis and Shreeve, 1991; Dennis, 1993). The fact that *P. argus* emerges earlier in the year than *H. semele* may also make it more likely to advance its emergence. As both species are able to use a range of hostplants, some of which are adapted to dry conditions (see below), and flower over different periods, there should be adequate nectar supplies for them.

Successful life-cycle completion in many host-specific insect herbivores requires close synchrony with host phenology (Bale *et al.*, 2002). Therefore, lack of synchronicity, for example of insects with their foodplants, may pose a problem in some cases, particularly as insects are generally predicted to respond more rapidly to climate change than their hostplants (DEFRA, 2001). However, flowering time of British plants has also been found to have advanced by a similar order (Fitter and Fitter, 2002), although some species have advanced more than others, and some have not advanced at all or have even delayed first flowering. It is therefore difficult to predict outcomes for certain for different species at different sites, particularly when climate projections are also estimates. Nonetheless, Walther (2004) notes that the direction of change in plant and animal phenologies observed at the same location often coincides with each other. The ability of both *P. argus* and *H. semele* to utilise more than one hostplant species will likely make them less vulnerable to potential hostplant asynchrony.

Adaptation to changes in hostplant phenology may also be possible. For example, in southern latitudes, where summers are longer and drier than in the north, the flight period of *H. semele* starts earlier (García-Barros, 1988). However, summer drought has an adverse effect on the growth of grasses on which larval feeding depends and growth of the grasses is delayed as long as the drought lasts. Therefore, if adult emergence is earlier in southern latitudes where the summer dry period is longer, the insect needs to spend that dry period in a resting stage, as without a long period of pre-oviposition, larval hatching would occur too soon in the season (García-Barros, 1988). García-Barros (1988) report that in Spain, maturation of the eggs is delayed, a mechanism that can be interpreted as an adjustment of the life cycle of *H. semele* to its hostplant's phenology, allowing it to occupy wider geographical ranges.

The potential benefits of higher temperatures for butterflies will also depend on other impacts and the increased likelihood of drought in particular may have negative effects on some butterfly species, with dry summers likely to affect egg survival, hostplant growth and habitat structure (Pollard, 1988; Dennis and Shreeve, 1991; Roy and Sparks, 2000). Excessive temperatures may cause adult heat stress, reducing activity and decreasing longevity and fecundity, as well as affecting eggs and larvae (Dennis, 1993). In laboratory experiments, Karlsson and Wiklund (2005) found that *H. semele* laid the highest number of eggs at 30 °C and was also able to lay eggs at 40 °C. However at 40 °C, *H. semele* laid far fewer eggs overall (although laid the most in the first day, compared to the other temperatures, ranging from 20-40 °C) and lived approximately half as long as at any of the lower temperatures. Karlsson and Wiklund (2005) report that open landscape living species, such as *H. semele*, exhibit several life-history opportunities and adaptations to higher temperatures, such as laying more and smaller eggs than shade-dwelling species. The open landscape species tended to have higher maximum fecundity at higher temperatures, whereas woodland species had a higher maximum fecundity at lower temperatures. Therefore, although *H. semele* is able to cope with higher temperatures than its optimum, its longevity and fecundity is reduced. Nonetheless, 30 °C is still a relatively high temperature, and the decrease was not as great at 35 °C.

However, butterflies are able to regulate their body temperature to a certain extent. For example, Dreisig (1995) reports that at high temperatures, *H. semele* is able to lower its body temperature by around 2.5 °C compared to that of a non-regulating butterfly, by exhibiting thermoregulating behaviour of keeping the wing-plane parallel to the sun’s rays to minimise the projected body area, and decreasing the duration of basking bouts.

The main impacts of drought on the butterfly species are likely to be through indirect effects on their hostplants as a result of a decrease in soil moisture, which could have serious indirect effects through hostplant desiccation, leaf abscission and seedling failure, which could reduce larval growth and cause starvation (Dennis and Shreeve, 1991; Dennis, 1993). Nectar sources would diminish both by reduction in flowering and through increased nectar viscosity, potentially reducing adult longevity and fecundity in those species which are dependent on nectar for egg production (Dennis, 1993).

One of the main foodplants for *P. argus*, cross-leaved heath (*E. tetralix*) is likely to be affected by summer drought and potentially decrease as the wet heaths dry out. It occurs in moister soils than the other heathland species (Webb, 1986) and has a shallow rooting system, which makes it highly susceptible to drought (Hossell *et al.*, 2005). However, *P. argus* is also associated with heather (*C. vulgaris*) in the New Forest (Oates *et al.*, 2000), which is expected to fare much better. Heather is also better able to withstand soil drying than cross-leaved heath (Hossell *et al.*, 2005) and performs better over a wide range of soil regimes than either bell heather (*E. cinerea*) or cross-leaved heath (Webb, 1986), and occurs throughout the British Isles (National Biodiversity Network, 2009).

P. argus may also possibly be associated with bell heather (*E. cinerea*) in the New Forest (Oates *et al.*, 2000). Bell heather is generally confined to the driest parts of the heathland and is able to withstand conditions in dry, mineral soils because they can control water loss more effectively then either heather or cross-leaved heath (Webb, 1986). It is therefore likely to be largely unaffected by the forecast drier summers for the New Forest, and should remain as an alternative foodplant for *P. argus*. It also starts flowering at the same time as *E. tetralix* (Sterry, 2006) and has a similar distribution in the British Isles, occurring throughout, but local in central and eastern England (National Biodiversity Network, 2009).

H. semele is a butterfly of relatively arid places (Oates *et al.*, 2000) and it's foodplants are therefore generally species tolerant of dry habitats. *H. semele* is strongly associated with bristle bent (*Agrostis curtisii*) in the New Forest (Oates *et al.*, 2000), which is the most common grass of dry south-west heathland (Webb, 1986), but it is largely restricted to the south and south-west (National Biodiversity Network, 2009). Hossell *et al*. (2005) suggest that it is likely that increased drying and its effects on the water table may cause the boundaries of the wet heath communities and their composition to change from a large proportion of cross-leaved heath to containing a greater abundance of the species associated with drier soil conditions such as bristle bent and bell heather. It therefore appears that bristle bent (as well as species such as bell heather) may increase in the New Forest, and as the wet heaths dry out, there may be more habitat available for *H. semele* and potentially bristle bent may be able to extend its distribution northwards in the British Isles, depending on suitable habitat.

Wavy hair grass (*Deschsmpsia flexuosa*) is another species with which *H. semele* is associated in the New Forest (Oates *et al.*, 2000) and is also usually found in dry places (Hubbard and Hubbard, 1984). It is distributed throughout the UK (National Biodiversity Network, 2009). *H. semele* is also associated with purple moor-grass (*Molinia caerulea*) in the New Forest (Oates *et al.*, 2000). Purple moor-grass can occur in drier and (more often) wetter areas and has a deep root system, which may allow it to cope more effectively than cross-leaved heath with low soil moisture levels and a low water table during the summer months, which may occur as a result of climate change (Hossell *et al.*, 2005). Hossell and Rowe (2006) suggest that a shift in favour of purple moor grass would result in a more tussocky vegetation structure. Purple moor-grass is also distributed throughout the British Isles (National Biodiversity Network, 2009). Sheep's fescue (*Festuca ovina*) may also be used by *H. semele* (A. Barker, personal communication, March 6, 2009; Asher *et al.*, 2001). Sheep's fescue is very hardy and drought-resistant (Hubbard and Hubbard, 1984), so should not be negatively affected by summer drought. It is also distributed throughout the British Isles, although patchily in places (National Biodiversity Network, 2009).

The fact that both *P. argus* and *H. semele* are able to utilise more than one foodplant is an advantage, particularly for *P. argus* as one of its foodplants, cross-leaved heath, is likely to be more negatively affected by drought. In addition, as all of the *P. argus* foodplants are distributed throughout the British Isles, potentially suitable habitat may be available if the distribution of *P. argus* becomes more northerly (see below). It appears that at all of *H. semele*'s hostplants will be tolerant of summer drought, so suitable habitat is likely to remain in the New Forest for this species.

As habitats become drier and hotter and hostplants are potentially affected, another adjustment (in addition to thermoregulation) that butterflies can make is changes in use of microhabitat. Dennis and Bardell (1996) reported on the impact of drought on

P. argus and *H. semele* at a site in Wales between June and September 1995. One immediate consequence of drought conditions was the movement of individuals from their habitats to cooler microhabitats (Dennis and Bardell, 1996). Dennis and Bardell (1996) reported that more individuals of both species were found outside known habitat patches than usual and both species were also found in distinctively cooler habitats. For example, *P. argus* was found in areas where the vegetation was much denser and *H. semele* was found flying under tree cover.

Both species were commonly found over 300 m from known colonies, with a maximum distance of 1.05 km for *H. semele* and 0.97 km for *P. argus*. Further, in the year following the drought, colonies of *P. argus* vagrants were noted. However, they did not have any control data available for previous years and no systematic search for dispersing individuals was carried out. In the New Forest, it may be that *P. argus* and *H. semele* make greater use of heathland patches within Inclosures or perhaps later growth stages of heath (as opposed to pioneer), which are slightly cooler. Different areas of sites may be used at different times. For example, Dennis and Sparks (2006) reported that habitat bounds of *P. argus* at sites in North Wales appeared to change with weather conditions on scales of days and hours. Dennis and Bardell (Dennis and Bardell, 1996) also reported similar finding for *H. semele* distributions.

Thomas *et al*. (1999b) found that temperatures (in the Dorset heathlands) were 4-5 °C warmer on slopes that faced due south compared with north, and were 3-8 °C warmer under 1-3 cm tall heath compared with more than 10 cm tall heath. The combined effect is to expose a ground-dwelling ectotherm inhabiting pioneer heath on a south-facing slope to temperatures more than 10 °C warmer in spring than if it lived under mature heath with a north aspect. This therefore demonstrates how different use of habitats may provide better quality hostplants and help regulate body temperature. Thomas *et al*. (1999b) suggest that there may be scope for reducing the warming effect by maintaining more vegetation in later successional stages and that selecting conservation sites that have heterogeneous terrain – including north-facing aspects – may also enhance the persistence of biodiversity. Therefore, sites with varied microhabitats (typically the larger sites) are likely to allow longer persistence of species such as *P. argus*.

The observations of Dennis and Bardell (1996) demonstrate that both *P. argus* and *H. semele* have the ability to adjust to a slightly different use of their habitat. Similar adjustments have been reported by Thomas *et al.* (2001), who found that the silver-spotted skipper butterfly, *Hesperia comma* has been able to colonise a wider range of aspects (rather than being limited to the relatively warmer south- and south-west facing aspects), which led to an approximate doubling of habitat availability (and shorter distances between suitable habitat) resulting in an approximate trebling in expansion rate for *H. comma.* They also report that the brown argus butterfly (*Aricia agestis*), has expanded both its habitat and geographical range over the past 20 years, by expanding its use of hostplants. It previously used only one hostplant in the northern part of its distribution but as the habitats in which those species were found in the north warmed from the 1960s and 1970s to the 1980s and 1990s, it was able to use a wider range of species. Davies *et al.* (2006) also report that at sites in Surry, UK, from 1982 to 2001, warmer ambient temperatures have increased *H. comma* egg-laying rates and within 19 years the species was able to utilise a greater variety of *Festuca ovina* plants for oviposition and re-expand into unoccupied habitat patches.

Thomas *et al.* (2001) suggest that although these changes in habitat breadth could be explained completely by ecological processes, once habitat range has begun to expand for ecological reasons, phenotypes able to use either a wide range of habitats or the commonest type of habitat available in marginal areas may show relatively fast rates of range expansion. In *A. agestis*, expanding and non-expanding marginal populations differed in choice of hostplant, and thereby habitat, and the results of experiments were consistent with a genetic contribution to hostplant choice (Thomas *et al.*, 2001). Thomas *et al.* (1998) also report that *P. argus* shows evolutionary changes in life history traits (such as total body mass) in response to habitat fragmentation, which may be influenced by the effects of mate-location strategy on emigration rates in heathland patches.

P. argus consists of several different races which differ in morphology, habitat, hostplant choice, performance on different hostplant species and species of associated ant (although it is always of the *Lasius* genus) and some of these differences are maintained in captivity, which suggests evolutionary divergence

(Thomas *et al.*, 1999a). It may be that further changes take place as a result of future climate change. Bailey *et al*. (1989) suggest that there is also a great deal of ecophenotypic flexibility in *H. semele.*

It therefore appears that certain characteristics of some habitats may provide the possibility for persistence in changing habitats and survival of some species may require adjustments to the use of cooler habitat types and microhabitats, such as in longer vegetation, on northern aspects and in shaded open woodlands (Dennis, 1993). Dennis (1993) also notes the possibility of changes in hostplant use (i.e. different hostplant species, broadening of hostplant use and of hostplant quality) as butterfly species occupy habitats with different plant associations and larval development times become less limiting on ‘novel’ hostplants (Dennis, 1993).

Dennis (1993) suggests that these changes may be imposed on species by the effect of drought on hostplants either directly or through competition with more drought resistant plants, but may also be induced by the spread of hostplants and symbionts to cooler habitats and the ability of different butterfly stadia to function in cooler habitats as mean temperatures increase. Species dependent on ants may also be able to extend their use of different habitat types if the ant species they are dependent upon also extend their habitat range or if alternative ant species are acceptable (Dennis and Shreeve, 1991). In the case of *P. argus*, *Lasius alienus* (one of the mutualist ant species with which it is associated) has a largely southern UK distribution, so may potentially expand northwards, but this would also depend on other factors, such as availability of suitable habitat.

The results from Dennis and Bardell (1996) on the consequences of drought for *P. argus* and *H. semele* in Wales also demonstrate the capability of these species to readily disperse further in warmer conditions when conditions at their current location become less suitable. They suggest that hot, sunny, dry and calm conditions probably provide ideal conditions for dispersal events for the following reasons: (i) excessively high ground temperatures could render the usual habitats on southern and western slopes critical for early stages and may deter females from egg-laying, especially if the hostplants wither; (ii) loss of nectar and moisture would also induce adults to move away from these slopes; (iii) the conditions are nevertheless ideal for rapid flight out to cooler and moister habitats. They therefore expect that the conditions (during 1995) resulted in exceptionally large and distant colonisation events. *H. semele* is well known for its ability to disperse but the observations from Wales are particularly important in terms of *P. argus*, with the data indicating that it is capable of colonising vacant habitats 1 km away in a single step and possibly up to 3.8 km.

P. argus is known as an extremely sedentary butterfly, with most adults moving less than 20 metres per day and only a few travelling more than 50 m (Asher *et al.*, 2001). A small proportion of adults disperse and have been known to move up to 1.5 km between colonies, but colonisations over this distance are very rare and the maximum recorded distance is 4 km (Asher *et al.*, 2001). However, Oates *et al*. (2000) suggest that, within the New Forest, *P. argus* is reasonably mobile and so can readily colonise new habitat when it develops. *H. semele* is more mobile than *P. argus* (Dennis, 1993; see Table A-50 below) and is therefore more likely to be able to move to new habitat (providing it is available) than *P. argus*.

Thomas (2000) suggests that where organisms are unable to persist in their current location, migration rate is one trait which would be expected to have a particularly strong effect on survival, especially in fragmented, terrestrial landscapes. Thomas (2000) found that butterfly species of intermediate mobility have declined the most in fragmented British landscapes, followed by those of low mobility (*P. argus* included), whereas high mobility species are generally surviving well. They suggest that, compared to the more sedentary species, species of intermediate mobility require relatively large areas where they breed at slightly lower local densities and these species have probably fared badly through a combination of metapopulation dynamics and the mortality of migrating individuals which fail to find new habitats in fragmented landscapes.

Migrations as a result of changing climate are likely to result in changes in distributions of many species. There is good long-term population data and many studies which provide evidence that insects, in particular butterflies, are already shifting their distributions. For example, in an analysis of distributional changes of a sample of 35 non-migratory European butterflies (whose northern boundaries were in northern Europe and whose southern boundaries were in southern Europe or northern Africa), Parmesan *et al*. (1999) report that 63% have ranges that have shifted to the north by 35-240 km during the twentieth century, 29% were stable at both boundaries, 6% shifted southwards and 3% have extended at both boundaries. Two of the species included in some of the analyses carried out by Parmesan *et al*. (1999) were *P. argus* and *H. semele*. The range of *P. argus* was reported to be stable but the range of *H. semele* had extended northwards.

Parmesan *et al*. (1999) report that consistency across taxa and continents indicates that butterfly species in the northern hemisphere are shifting generally northwards in response to a common environmental change (Parmesan *et al*., 1999). Their results indicate that future climate warming could become a major force in shifting species' distributions, but it remains to be seen how many species will be able to extend their northern range margins substantially across the highly fragmented landscapes of northern Europe; something that could prove difficult for all but the most efficient colonisers (Parmesan *et al*., 1999).

As *P. argus* is currently a more southerly distributed species, it might be expected that it could expand its distribution northwards as the UK climate warms. Indeed, the MONARCH climate space models (Walmsley and Harley, 2007) showed substantial gains in potential suitable climate space for *P. argus* in all three time periods (2020s, 2050s and 2080s) with 100% overlap and no loss, with almost all of Britain and Ireland potentially becoming potential suitable climate space for this species. However, as discussed for *P. globulifera*, the MONARCH climate space models do not take account of other factors influencing species distributions and it is unlikely that this full distribution would be realised, particularly as a consequence of fragmented suitable habitat and limited dispersal ability.

The issue of available habitat was highlighted in a study by Warren *et al*. (2001), who examined recent changes in British butterfly distributions by evaluating changes in the distribution sizes and abundances of 46 non-migratory species of butterflies that approach their northern climatic range margins in Britain, where changes in climate and habitat are opposing forces. They expected that the butterflies might have responded positively to climate warming (spring-summer warming of approximately 1 to 1.5 °C) over the 30 years leading up to the study, yet three-quarters of them had declined, suggesting that the negative responses to habitat loss had outweighed the positive responses to climate warming. *P. argus* was chosen as a thermally-limited species to illustrate the patterns typical of most other species. It has low mobility and has declined in area of occupancy by 28% over the last 30 years. Warren *et al*. (2001) report that it is restricted within apparently suitable climatic areas (it has far more southerly distribution than *H. semele* (National Biodiversity Network, 2009)), and is limited by habitat and dispersal more than by climate: most declining species show comparable patterns. This may provide some explanation of why Parmesan *et al*. (1999) reported that its range has not expanded northwards as many other species (including *H. semele*) have done.

Warren *et al*. (2001) found that changes in population abundances closely matched changes in distributions. In addition, they reported that species' characteristics were also a factor; half of the species that were mobile and habitat generalists increased their distribution sites over this period (consistent with a climate explanation), whereas the other generalists and 89% of the habitat specialists declined in distribution size (consistent with habitat limitation). It was the habitat specialists and the sedentary species (traits which are highly correlated among butterflies) which had fared the worst in terms of declining distribution, as these traits restricted them in fragmented habitats. They also lagged behind climate (in terms of climatically suitable areas) more than generalist and mobile species. This suggests that *P. argus* may be more vulnerable. Further, the consequences of such lags for these less mobile species with restricted habitat requirements in fragmented landscapes will be of particular concern because these species are likely to be of high conservation value (Hill *et al*., 1999). Warren *et al*. (2001) suggest that the dual forces of habitat modification and climate change are likely to cause specialists to decline, leaving biological communities with reduced numbers of species and dominated by mobile and widespread habitat generalists. Therefore, species' characteristics and habitat distribution will be crucial in predicting species' responses to future climate change (Hill *et al*., 1999).

Hill *et al*. (1999) also report that even moderately mobile butterfly species, such as the speckled wood butterfly (*Pararge aegeria*), appear to be lagging behind current climates. However, *P. aegeria* has expanded its northern range margin substantially since 1940 and these changes in its distribution over the past 100 years are likely to have been due to climate change (Hill *et al*., 1999). Further, data also indicated that, in the past, the distribution of *P. aegeria* has expanded at a rate that is consistent with the likely dispersal rate, and that long-range dispersal events have not been a major influence on its expansion.

Hill *et al*. (1999) note that species are likely to respond individualistically to future climate change, and new associations among species are likely to arise as climate changes, particularly as species potentially move to newly suitable habitats. Warmer temperatures are likely to increase the number of migrant Lepidoptera into Britain, something which is already being seen (Sparks *et al.*, 2005; Sparks *et al.*, 2007). This will encourage the establishment of resident populations and reinforce UK populations from abroad (Sparks *et al*., 2005). Migrating species potentially impact on resident species, such as on their resources, and may represent a competitive threat to resident species which typically have lower mobility and are more specialised in habitat requirements than more mobile and adaptable migratory species (Sparks *et al*., 2007). Further, susceptibility to bacterial, microsporidial and viral infections may relate to 'stress' triggered by hot weather, although any relationship is also complicated by other factors (Dennis, 1993). Insect pests in particular are also likely to increase their abundance and new pests might occur (DEFRA, 2001; Sparks *et al.*, 2005), which could have negative impacts, such as on butterfly hostplants. For example, Tubbs (2001) reported that during droughts in the New Forest during 1974-1984, extensive tracts of heather were completely stripped of leaf by a population explosion of heather beetles. Events such as these could potentially have large negative impacts on butterflies, such as *P. argus*, if their hostplants are affected. This is another example of where utilisation of more than one hostplant species will be advantageous.

A table of factors influencing the vulnerability of butterfly species to environmental perturbations, which takes account of their capacity (resources) and their flexibility to withstand changes is provided by Dennis (1993). Each attribute is given a rating of between 1 and 4, most to least susceptible, respectively, to environmental perturbation (Dennis, 1993). Equal weight is given to all variables (although this may be unrealistic) and an overall vulnerability index value (see Dennis (1993) for details) is then provided for each species, with lower values indicating greater vulnerability. Values less than 2 (low values) generally indicate species experiencing contractions in range and distribution, whereas the status of species with values greater than 3 (high) remains unchanged or they are undergoing range and distribution expansions. Between 2 and 3 is intermediate (Dennis, 1993). The table is reproduced for *P. argus* and *H. semele* below:

	H. semele	***P. argus***
Capacity factors		
Range (based on the latitudinal extent of species on the British mainland)	4 < 100%	2 < 50%
Distribution (based on the proportion of 10 km squares occupied within the range of the species)	1 < 25%	1 < 25%
Hostplant type	3 Oligophagous - > 1 species per habitat	4 Polyphagous
Hostplant abundance	4 Ubiquitous and cosmopolitan	4 Ubiquitous and cosmopolitan
Vulnerability of major habitat seral stage occupied	4 Bare ground, short forbs and grasses	4 Bare ground, short forbs and grasses
Range of semi-natural habitat types occupied	2 < 9	2 < 9
Capacity factors total	18	17
Flexibility factors		
Dispersal ability	3 Open population structures with evidence of frequent movements between habitat units	1 Closed populations with little evidence of movement outside colonies
Voltinism	1 Biennial or univoltine	2 Univoltine but with occasional partial second broods
Length of flight period for the longest brood in the year (egglaying females)	3 2 – 3 months	2 1 – 2 months
Overwintering stage	2	2

	larva	larva
Flexibility factors total	9	7
Index	2.63	2.46

Table A-50. Factors influencing the vulnerability of *H. semele* and *P. argus* to environmental perturbations (from Dennis (1993)). Coding of attribute states: 1 – 4, most to least susceptible, respectively, to environmental perturbation.

Both species achieve an intermediate overall index value, although *P. argus* achieves a slightly lower overall index value compared to *H. semele*, indicating that it is slightly more vulnerable to environmental perturbation. This is mainly due to lower values in range, dispersal ability and length of flight period, as examined in the discussion above. This fits with the general conclusions from the discussion. However, it is very difficult to determine the overall outcome of the interaction of different impacts and potential benefits (such as greater activity and longer flight season) and costs (such as drought stress and potential loss of habitat).

Fox *et al*. (2007) report that UK-wide the distribution of *P. argus* has decreased by 43% in 1995-2004 compared to 1970-1982 and the population change from 1995-2004 has decreased by 72%; this is a severe decline in range and abundance. This is also true for *H. semele*, which has decreased its UK distribution by 45% in 1995-2004 compared to 1970-1982 and the population change from 1995-2004 has decreased by 41%. Fox *et al*. (2007) suggest that the destruction and deterioration of habitats remain the primary cause of butterfly declines and it is the habitat specialist butterflies that have fared the worst.

Both *H. semele* and *P. argus* are regional high priority butterfly species in Hampshire and the Isle of Wight, which encompasses the New Forest (Butterfly Conservation, 2007). Butterfly Conservation (2007) reports that in this area there has been a slight decline of *P. argus* over the last two decades, but around a 25% decline in *H. semele* sites in the whole region since the 1980-1994 recording period. However, few transects monitor this species in the region so they do not give a good guide to recent trends (Oates *et al.*, 2000). Oates *et al*. (2000) also note that in the New Forest there have been changes from year to year in response to weather conditions. The declines of *P. argus* and *H. semele* in the Hampshire region appear to be less than in the UK as a whole. However, there is still some cause for concern.

H. semele in particular appears to have suffered from a run of wet summer weather (Vaughan, 2008; Brock, 2009).

34.3. Potential effects of climate change on *Nemobius sylvestris*

It is expected that, being ectothermic, many insects will do well as temperatures warm, by feeding better, living longer and reproducing faster (Cook and Harrison, 2001). For example they are expected to pass through their larval stages faster and to become adults earlier (Hughes, 2000). This was shown by Willott and Hassall (1998), in a study of four British high-temperature thermal specialist grasshopper (Orthoptera: Acrididae) species, who found that growth and development rates increased with temperature (manipulated experimentally) for each species, as did egg pod production rate. Variation in sensitivity to temperature was found to be a good predictor of their distribution, with the most generalist species generally being the most widespread and the more specialist species restricted to warmer habitats. Warmer winters may potentially increase survival of overwintering *N. sylvestris*.

Willott and Hassall (1998) therefore suggest that it may be that some of these species could benefit from milder climates predicted for Britain. However, they also suggest a note of caution for broad-scale predictions of range expansion or populations of organisms as a result of climate change as they found that closely related taxa (the same subfamily) can have very different thermal strategies. Therefore, although *N. sylvestris* is also of the same order, the same changes may not occur, but similar benefits seem likely. There is also some evidence that increased atmospheric carbon dioxide may affect the physiology of some insect species, such as increasing rates of reproduction, but further work is needed in this area to understand if this response persists over time and how it interacts with temperature and rainfall changes (DEFRA, 2001).

As the New Forest is at the northern limit of the range of *N. sylvestris* it appears likely that it would be able to cope with warmer conditions (unless the species possesses different adaptations in different parts of its range). However, if conditions at some sites do become too hot or dry in the summer for *N. sylvestris*, it may be able to make adjustments by using cooler habitat types and microhabitats (Dennis, 1993). For *N. sylvestris* this may involve movement further into woodlands where it is slightly cooler, rather than on the much warmer edges. Another change might be a tolerance of a greater cover of ground vegetation cover, which would also make the ground temperatures slightly cooler.

One of the main impacts of climate change on insects, particularly increases in temperature, is likely to be changes in distributions. Bale *et al*. (2002) suggest that climatic warming will allow the majority of temperate insect species to extend their ranges to higher latitudes and altitudes, although this will depend on their ability to migrate to, and the availability of, suitable habitat. There is good long-term population data and many studies which provide evidence that insects, in particular butterflies (see Appendix 34.2), are already shifting their distributions. There is also evidence of other insect orders having expanded their range, such as the Odonata (dragonflies and damselflies). In a study of 37 species of non-migratory British Odonata from 1960 to 1995, Hickling *et al*. (2005) report that overall, southerly distributed British Odonata species increased in range size and expanded northwards at their range margins in Britain, likely attributable to increased temperatures.

Some Orthoptera species have demonstrated ecological and evolutionary processes taking place at their expanding range margins. Thomas *et al*. (2001) also provide details of two species of bush crickets, that exhibit adult wing polymorphisms, which have been spreading northwards and inland from distributions formerly confined to specific habitats in southern, coastal areas. The long-winged cone-head *Conocephalus discolor* has two forms: long-winged and extra-long-winged (macropterous). Many populations established in the past 20 years show higher frequencies of extra-long-winged (more dispersive) individuals than in those established greater than 20 years ago. Roesel's bush cricket *Metrioptera roeselii* has a short-winged form that cannot fly and a long-winged form that can and this species also shows increased frequencies of the more dispersive form in populations that have recently been established.

Thomas *et al*. (2001) suggest that these changes may represent plastic responses to new environments or genetic differences, although improving environmental conditions at existing margins, in this instance regional warming at cool margins, are likely to initiate range extensions purely on the basis of ecological, physiological and population-dynamic processes, requiring no evolutionary change. Once an expansion is initiated, individuals and populations that expand most rapidly are likely to be favoured, and expanding range fronts may become characterised by dispersive generalists or by specialists on habitats common in the area of expansion (Thomas *et al*., 2001).

N. sylvestris does not have wings, and Brouwers (2008), from a study of the species in the Isle of Wight, suggests that its dispersal ability is limited. It is therefore less likely to expand as fast as some winged insects, although variation in other traits, such as body mass (Thomas *et al.*, 1998) may give some individuals a greater likelihood or ability to disperse, which may potentially lead to an increased frequency of such traits. However, as noted by Thomas *et al*. (2001) many species will not be able to change their geographical ranges rapidly in response to climate change and large numbers of relatively sedentary and specialised species may fail to initiate any expansion across human-modified landscapes. Brouwers (2008) also suggests that because of its dispersal limitations, *N. sylvestris* demonstrates a time lag in occupying suitable habitat.

Brouwers (2008) observed that *N. sylvestris* adults and nymphs consistently preferred to move through leaf litter habitat rather than bare soil and grass habitats, although both adults and nymphs were shown to be able and willing to move through less favourable grassland habitat and able to cross obstacles such as watercourses, with individuals being found up to 55 metres (and one 90 metres) away from source populations. However, it is unlikely that individuals choosing to disperse away from woodland habitat through a relatively unfavourable matrix would establish new populations, unless they encountered suitable woodland habitat enabling them to reproduce (Brouwers, 2008). Brouwers (2008) suggests that woodland fragments separated by more than 50-60 m of non-woodland matrix habitat, might be considered as effectively isolated for this species. However, habitat corridors or ‘stepping stones’ of habitat between woodland fragments may be beneficial, although their functionality would be dependent on suitable conditions, including factors such as tree cover and leaf litter presence.

Despite this, it appears that *N. sylvestris* has been capable of spreading out of the New Forest, several miles northwards, with recent records from locations around Ampfield near Romsey in Hampshire (P. Budd, personal communication, March 5, 2009). *N. sylvestris* appears to be doing well in the New Forest, with most suitable woodlands appearing to hold populations (S. Douglas, personal observation). As *N. sylvestris* is at the northern limit of its range in Europe in the New Forest (Marshall and Haes, 1988), this suggests that temperature may be limiting its distribution; it therefore does not seem too surprising that it is starting to spread northwards. Indeed, in warmer weather, *N. sylvestris* was observed to move further from the woodland edge (N. Brouwers, personal communication, January 9, 2009) even occurring under bracken in heathland, over 100 metres from the nearest woodland edge (S. Douglas, personal observation). Being ectothermic, warmer weather is generally likely to increase the occurrence of movement further from source populations. Richards (1952) suggests that hard winters are probably one its ‘main enemies’, so as winters become milder, it is likely that it will be able to spread further north, providing suitable habitat is available and reachable. Strips of trees alongside railway lines (Richards, 1952), for example, may provide suitable corridors in which to disperse along.

Gabbutt (1959) suggests that the insulation provided by the leaf litter and its actual physical structure tend to mitigate climatic effects. Therefore, it may be that with milder winters, *N. sylvestris* does not require as deep a leaf litter layer, which may mean that more suitable habitat is available. However, a certain depth may be required to form a suitable habitat. Apart from requiring broadleaved leaf litter on a woodland edge, *N. sylvestris* is not particularly specialist in its habitat requirements, which means it should be less vulnerable to environmental change in this respect. The sort of habitat it requires is found across the UK and although woodland is fairly fragmented, there should be plenty of suitable habitat if it is able to disperse further from the New Forest.

Although winters are forecast to be milder they are also expected to be wetter. Excessive ground moisture may have a detrimental effect on overwintering nymphs. Gabbutt (1959) suggests that high precipitation resulting in an increase in the free water in the soil may cause mortality at the time of hatching. However, hatching usually takes place in June (Marshall and Haes, 1988), which will likely be a lot drier, although a one-off very heavy storm may be a problem.

Current monitoring trends suggest that changes in temperature and a lengthening growing season are affecting insect species’ phenology (DEFRA, 2001). For example, the majority of British butterfly species are showing statistically significant trends towards earlier appearance (Roy and Sparks, 2000). It may be that the warmer temperatures result in earlier emergence and later occurrences of *N. sylvestris* as male *N. sylvestris* only stridulate when temperatures are above about 15 °C (N. Brouwers, personal communication, January 9, 2009). As *N. sylvestris* is not dependent on flowering of a hostplant, there are unlikely to be issues of asynchrony. Although it depends on leaf litter, it occurs where there is a build up of leaf litter, rather than having to rely on the timing of autumn leaf fall in a single year.

The main habitat of *N. sylvestris* is broadleaved woodland, which is expected to experience several impacts as a result of environmental change. Changes in phenology of tree and shrub species (Menzel and Fabian, 1999) are unlikely to have much of an effect on *N. sylvestris*. The only effect may be that earlier and greater tree canopy coverage could reduce the amount of light and water reaching the ground, which may affect the ground flora and other ground dwelling organisms (Diack, 1999).

Likewise, changes in the distributions of some tree species as a result of increased temperatures and drought are also unlikely to have a large impact on *N. sylvestris*. Although oak leaves may be ones of its preferred litter types, it also occurs with numerous other deciduous species (Brown, 1978). Oak is more likely to persist in the New Forest than beech, which is more vulnerable to drought and tree-throw due to its shallow root system (Mountford *et al*., 1999). *N. sylvestris* may also feed on fungi associated with dead leaves (Richards, 1952; Marshall and Haes, 1988), although it is not clear how important this may be and what effect environmental change may have on this. Further, considering *N. sylvestris* will also likely feed on various other supplementary items such as dead insects, fallen acorns and leaf-galls, and honeysuckle foliage (Richards, 1952), food source is unlikely to be a problem.

The impact on woodland that is likely to have the greatest effect on *N. sylvestris* is the increased occurrences of the death of trees, which may be caused by factors related to climate change such as drought, storm damage, increases in woodland pests and pathogens and increased fire risk (Diack, 1999). This would potentially increase the number of gaps in woodlands (as occurred during the severe droughts and storms of the 1970s and 1980s in the New Forest (Mountford *et al.*, 1999)), creating new 'edge' for *N. sylvestris*. Having a similar effect, Brouwers (2008) reported that management activities (such as commercial tree harvesting) helped to create more open habitat for *N. sylvestris* in the Isle of Wight, with increased canopy openness and initial low levels of ground vegetation cover, providing good habitat conditions. However, opening up of the canopy may increase growth of ground vegetation, which could potentially be detrimental to *N. sylvestris*. Brouwers (2008) suggests that grazing may be a good way of suppressing vigorous growth of ground vegetation and as *N. sylvestris* is highly persistent and widespread across the New Forest where there is a high level grazing, it suggests that this is favourable. Therefore any potential large decreases in grazing due to changing regimes as a result of future environmental change (New Forest National Park Authority, 2007a; Trotter, 2007) could have a significant impact.

An increase in population of pests and pathogens, as a result of warmer winters (Hossell and Rowe, 2006), may have an even greater impact on trees that are already stressed or damaged as a result of drought or storms (Mountford *et al.*, 1999). However, there is also the potential that non-indigenous pests may pose a threat to native species such as *N. sylvestris*, through competition, predation or transmission of disease, for example. Finally, an increased risk of fire in drier woodlands, exacerbated by an increase in recreational use of forests, which the warmer summers are projected to encourage, could have very serious (but likely contained) consequences for *N. sylvestris*. However, there is also some anecdotal evidence that the increase in recreational use of forests, combined with an increase in mobile phone use has helped to reduce fire damage as fires can be reported more quickly (Diack, 1999).

34.4. Potential effects of climate change on *Poronia punctata*

As there are large gaps in knowledge about the ecology of *P. punctata* it is somewhat difficult to try and predict the potential impacts of climate change. Further, information on the potential impacts of climate change on fungi, particularly macro-fungi, is limited, despite the fact that fungi provide vital ecosystem services through decomposition, nutrient cycling and soil aggregation (Gange *et al.*, 2007). In addition, *P. punctata* has a specific niche, only occurring on dung.

It appears that *P. punctata* spores must be first ingested by ponies, rather than being able to disperse from dungpile to dungpile. Although *P. punctata* spores dropped onto sterilised dung in the laboratory grow and fruit freely, spores dropped onto non-sterilised dung do not (Whalley and Dickson, 1986). It therefore appears that in the wild, the spores require digestion, as perhaps the digestion of spores provides similar conditions to the sterilisation by providing a substrate free of competitors; it may be that the spores of other fungi species are not as tolerant of the digestion process as *P. punctata*. Interestingly though, *P. punctata* (which is a slow-growing and later-appearing colonist of cattle dung) has been found to contribute to a competitive hierarchy in cattle dung (in Colorado), in that it is antagonistic to all earlier-sporulating and co-occurring fungal species (Wicklow and Hirschfield, 1979). However, in Britain and Europe it only occurs on horse/pony dung (Whalley and Dickson, 1986).

It is not clear why *P. punctata* occurs on different substrates in different parts of the world. One possibility may be that cattle digestion is more thorough than pony digestion, which destroys *P. punctata* (G. Dickson, personal communication, July 21, 2009), although this would not explain its occurrence on cattle dung in America, unless the species possesses different adaptations in different locations. Alternatively, it may be something to do with wormers given to cattle (A. Lucas, personal communication, March 5, 2009). What is clear is that in the New Forest *P. punctata* is completely reliant on the presence of ponies feeding on natural vegetation and any change to this could have a large impact.

P. punctata fruiting bodies are most often observed from September to February (Poland, 2004) and most frequently recorded from the New Forest in October

(Dickson and Leonard, 1996), although they can occur at any time of year (while this could reflect a recording bias, it does not appear to occur as frequently in the summer months (S. Douglas, personal observation)). It is not clear then, what happens from when the spores are released from the fruiting bodies during autumn and winter to the following autumn when the majority of the fruiting occurs. During the autumn and winter those spores will land on vegetation that will likely be ingested relatively quickly and excreted within approximately 30 to 40 hours (Cox, 1999). Some may fruit within a few weeks (frequently about one month (Whalley and Dickson, 1986)) if the dung conditions are suitable, but towards the end of that main fruiting season, the evidence appears to suggest that some spores remain in the dung until the following autumn when conditions become more suitable again. Some occasional fruiting may occur between March and August if the right conditions occur (possibly in slightly damper, cooler locations), but on the whole it may be too dry or hot (G. Dickson, personal communication, July 21, 2009).

However, this process would partly depend on how long the dung lasts. There will likely be a range in this value, depending on site conditions (such as wet opposed to dry sites, or sheltered compared to exposed sites) as well as other factors such as weather conditions and disturbance, such as by birds raking through the dung, but it seems reasonable that it could survive several months at some sites. (Edwards and Hollis (1982) suggest that pony and cattle dung tended to persist for about 3 months in the New Forest, although, as noted above, this would depend on the conditions). It would also depend on how long *P. punctata* can persist in the dung. The Coprophilous (i.e. living or growing on dung) Xylariaceae family of fungi, of which *P. punctata* is a member, are considered to have adaptations that allow them to inhabit dry sites (Rogers, 1979). They can undergo severe desiccation, then release ascospores after a short period of wetting (Rogers, 1979) (which could potentially be one of the triggers for fruiting). It therefore seems possible that *P. punctata* is able to persist vegetatively in the dung over the summer and then fruit when it becomes wetter in the autumn. This may also add some explanation to the fact that *P. punctata* fruiting bodies are more likely to be found on intact boluses (Whalley and Dickson, 1986; Cox and Pickess, 1999; Poland, 2004), as this may provide a buffering microclimate, to reduce the impact of very dry conditions, for example. Cox and Pickess (1999) also note that in dung that has been extensively fragmented any fungal mycelium would be more likely to desiccate before stromata could be formed.

It seems likely that not all the dung piles may persist and neither may a lot of the spores, so the population may decline to a certain extent during the summer as a result of these losses but build up again relatively quickly in the autumn. However, it is not clear what level of innoculum is required to be ingested by ponies for fruiting bodies to occur on a dung pile. Remaining in dungpiles over the summer seems a likely explanation for maintaining a continuity of innoculum. An alternative might be that some spores remain on the vegetation and are not ingested until several months later. However, as Dickson (1997) notes, the grass on which they fall in the autumn will wither (or be grazed) during the year and, by the time the ponies are eating in the same area next year, fresh grass, uncontaminated by spores, is presumably their diet. Further, the spores are unlikely to be able to survive for long on grass and probably need to be ingested fairly quickly (A. Lucas, personal communication, March 5, 2009).

Assuming that *P. punctata* remains in a vegetative phase in dung piles over the spring and summer, changes in climate could have an impact on this species. Summers are expected to be hotter and drier, which could dry out the dung piles too much for *P. punctata* to persist, even with possible adaptations that may allow it to inhabit dry sites (Rogers, 1979). This may particularly be the case in the drier sites, which are expected to increase as wet heaths dry out (Cook and Harrison, 2001). However, as noted above, it has been noted that in the Xylariaceae, perithecia embedded in a stroma can undergo severe desiccation, then release ascospores after a short period of wetting (Rogers, 1979). Whalley and Dickson (1986) note that *P. punctata* is apparently fastidious concerning conditions in which it fruits, not liking drought or really wet weather. Wetter winter conditions may also be a potential disadvantage as *P. punctata* is rarely found in very wet or waterlogged habitats or on saturated dung, suggesting that excessive moisture may inhibit growth of *P. punctata* by some means (Poland, 2004). Although it appears that *P. punctata* requires a level of drainage impedance, with some moisture, with it possibly not developing on very arid sites (Poland, 2004), too much rain and a greater increase in winter flooding could reduce fruiting if the conditions do not become suitable.

Another problem could be the predicted greater frequency of torrential downpours, which Cox *et al*. (2005) observed to erode and wash away substantial portions of dung deposits. This would reduce the number of spores/mycelium developing into fruiting bodies. Cox *et al*. (2005) also reported a decline in the *P. punctata* population in Dorset after unfavourable weather conditions of a dry autumn followed by weeks of heavy rain, including torrential downpours. However, the population subsequently made a good recovery in the next season. It therefore appears that the population is able to recover fairly quickly, so if the population does decline during the summer (as discussed above), it is able to build up again relatively quickly in the autumn.

Warmer winters are unlikely to have a large effect on *P. punctata*, as Cox *et al*. (2005) observed no visible signs of frost sensitivity. However, it is not clear what triggers fruiting in *P. punctata* and whether warmer conditions might be more favourable. It is likely that the impacts of other aspects of climate change, such as drier summers and wetter winters, will be more significant. However, if one of the reasons that *P. punctata* does not occur in woodland is because of the cooler, damper microclimate conditions (A. Newton, personal communication, December 8, 2008), it seems strange that it does not occur in this habitat more often in the summer when the conditions may be more suitable for *P. punctata* than on the heaths. This could just reflect less recording, or it may be because the ponies do not spend as much time in woodland (A. Lucas, personal communication, March 5, 2009; A. Newton, personal communication, December 8, 2008; G. Dickson, personal communication, July 21, 2009), particularly in the summer. For example, Pratt *et al*. (1986) found that New forest ponies use grassland and heathland the most in summer, but less in winter when there is an increased use of gorse-brake and woodland, which provides cover and shelter, although less food. However, this begs the question of why *P. punctata* spores do not remain in the dung dropped in woodland and fruit as it becomes warmer and drier in the woodland in the summer. Perhaps it is that the woodlands do not become warm and dry enough, but this seems unlikely, particularly in some of the more open woodlands, and because woodland are likely to be warmer and drier in summer than the heaths in winter.

Potentially, as the woodlands dry out in the summer as a result of climate change, they may become more suitable for *P. punctata*. An increase in the numbers of gaps in woodland as a result of increased occurrences of deaths of trees from drought and storms (and disease) (Diack, 1999), may also create warmer and drier conditions which may potentially increase the likelihood of *P. punctata* developing on dung deposited in these woodland sites.

An alternative possibility is that fruiting indicates that the conditions are becoming unsuitable, and the dung resources are being exhausted, as has been reported for other fungi. Therefore, mycelium may be present in the dung throughout the summer, but fruiting only commences when there is limited cellulose left to digest.

One of the few studies on the impact of climate change on macrofungi reports that the first fruiting date averaged across 315 autumnal fruiting species in southern England from 1950 to 2005 has become significantly earlier, whereas average last fruiting date has become significantly later (Gange *et al.*, 2007). The overall fruiting period has more than doubled from an average of 33.2 days in the 1950s to 74.8 days in the current decade. For the species that showed significantly earlier first fruiting dates (85 in total), the average advancement was 8.6 days per decade, whereas for species showing significantly later fruiting dates (105), the delay was 7.5 days per decade, both of which are greater than equivalent spring data previously reported for higher organisms (Root *et al.*, 2003; Gange *et al.*, 2007) and for British plants (4.5 days earlier in 1991-2000 compared with 1954-1990 in south-central England) (Fitter and Fitter, 2002).

Gange *et al*. (2007) report that the alteration in fungal fruiting mirrors changes in British temperatures; over the past 56 years, August temperatures have increased, as has October rainfall. They suggest that the increase in late summer temperatures and autumnal rains has caused early season species to fruit earlier and late season species to continue fruiting later. Of the species showing an advanced fruiting date, 91% had a significant relation between first fruiting date and August temperature, whereas 88% of the species showing later last dates could be explained by positive relations between August temperature and October rainfall (Gange *et al.*, 2007). Furthermore, Gange *et al*. (2007) report that climate warming seems to have caused significant numbers of species to begin fruiting in spring as well as autumn. They suggest that given that active mycelia growth is required before sporophore production, this is strong evidence that the mycelium of certain species must be active in late winter and early spring, as well as late summer and autumn, suggesting, for example, increases in decay rates in forests.

Interestingly, contrasting results were found by Kauserud *et al*. (2008) who report that autumnal fruiting date of mushrooms in Norway has delayed by an average of 12.9 days since 1980, compared to 1940 – 1980. The changes were found to differ strongly between species and groups of species, with early-fruiting species experiencing a stronger delay than the later fruiters, resulting in a more compressed fruiting season. Incorporating monthly precipitation and temperature variables into the analyses provided indications that increasing temperatures during autumn and winter months bring about significant delay of fruiting both in the same year and in the subsequent year. Kauserud *et al*. (2008) were unable to provide an explanation for the difference in their findings to those of Gange *et al*. (2007), except that cues might relate to autumnal events that occur later than before, whereas constraints on resource acquisition and achieving ‘fruiting potential’ might be fulfilled earlier when the climate is milder. They suggest that their results indicate that important environmental cues related to autumnal events, such as a drop in temperature, might have been delayed because of global warming.

Temperature and moisture are exogenous key factors known to influence the production of fruit bodies (Eveling *et al.*, 1990; Kauserud *et al.*, 2008). Eveling *et al*. (1990) found that over a 14 year period in a coniferous forest in Northern Ireland the largest fungi sporocarp counts occurred in the autumns following the warmest four summers. Sporocarp counts were also highly correlated with the means of average daily temperature for the period 2-4 months prior to recording dates over a 10-yr-period and with rainfall for the period 3-5 months prior to recording; however, rainfall appeared to have little influence on the maximum counts obtained for each year (Eveling *et al.*, 1990). This would suggest that forecast warmer summers may be beneficial, although the accompanying drier conditions may not be as suitable for *P. punctata*.

P. punctata is unlikely to expand its distribution as a result of climate change as it is limited by the presence of ponies feeding on natural vegetation, and the number of locations where this occurs is limited and patchily distributed. Even where seemingly suitable conditions do occur, such as on Dartmoor and Exmoor, where there are also free-roaming ponies, *P. punctata* does not occur (Poland, 2004). It is not that *P. punctata* is restricted to the dung of New Forest ponies either, as it briefly occurred on the dung of Przewalski's horses at a site in Farnborough in Hampshire (Poland, 2004). The fact that occasional records occur elsewhere (likely the result of New Forest ponies being transported to new sites with the fungus in their guts) but do not persist for long suggests that the habitat requirements of *P. punctata* are very specific (Webster, 1999). A substantial decrease in the numbers of free-roaming ponies in the New Forest as a result of future environmental change as a result of changes in agricultural practices and pastoralism (New Forest National Park Authority, 2007a; Trotter, 2007) could have a significant negative impact on *P. punctata*. Newton (2009b) questions whether fluctuations in pony numbers in the New Forest in the past have been associated with the frequency of *P. punctata* occurrence and suggests that it may be the case that *P. punctata* has undergone a recent increase in abundance in the Forest, and may even still be increasing, as a result of an increase in pony numbers.

Appendix 35. Changes made to the BBN states to assess the potential impact of climate change

Tables A-51 to A-58 below show the input variables for each BBN model with changes made to the states. The changes were based on the information obtained from the literature review summarised in Tables 42 to 49 in section 5.3.2 (and section 5.3.1, Chapter 5). The suggested changes are based on numerous assumptions and interpretations. For example, ground moisture was increased to one state drier. This reflects the likely change in summer (when most of the species are flowering or active) but does not account for the likely wetter winter conditions.

Variable/node	**States**	**Change**
Habitat type	Unimproved acid grassland with 0% *Calluna* (GL11) Dry heathland/acid grassland mosaic (HL3) Dry heathland (HL1) Wet heath (HL2) Valley mire (AQ1) Unimproved neutral grassland (GL12) Semi-improved neutral grassland (GL13) Improved grassland (GL3) Woodland – Broadleaved (W1, W2), Coniferous (W5, W6), Mixed (W7, W8) Other	All habitat types were kept the same except for loss of wet heath (HL2) and valley mire (AQ1). Where these two habitats occurred, they were replaced with 50% HL1 and 50% HL3 (as it is not known which cover of *Calluna* would be more likely).
Suitable grassland	Present Absent	No change.
Bare ground	Present Absent	No change.
Ground vegetation height	< 4 cm = Average height of sward is less than 4 cm 4 – 8 cm = Average height of sward is between 4 cm and 8 cm > 8 cm = Average height of sward is greater than 8 cm Grass absent = No grassland present within the site	No change.
Tree canopy cover	Open Partially open	No change.

	Closed	
Soil type	64301 – Deep sandy to clay 64303 – Deep loam to clay 71107 – Seasonally wet loam to clayey over shale 84102 – Seasonally wet deep loam Other	No change.
Soil/ground moisture	Dry Moist/Damp Wet	Increased state to one drier (e.g. moist/damp became dry).
Winter flooding	Yes No	All sites became ‘yes’ (although it is recognised that not every site will flood in the winter, depending on topology etc.).
Application of herbicide	Recent Not recent	No change.

Table A-51. Input variables for the *C. nobile* BBN model with notes on changes made to the states to model the impact of climate change. The variable highlighted in grey is one which was not instantiated in running the BBN, but could be if a different version of the model was run (e.g. the *C. nobile* – 2 model).

Variable/node	**States**	**Change**
Tree canopy cover	Open Partially open Closed	No change.
Bare ground	Present Absent	No change.
Wetland habitat	Present Absent	‘Present’ became 50% ‘present’, 50% absent. ‘Absent’ stayed as ‘absent’.
Ground moisture	Dry Moist/Damp Wet	Increased state to one drier (e.g. moist/damp became dry).
Habitat type	Heath (HL1 and HL2) Dry heath/acid grassland mosaic (<25% *Calluna*)/unimproved acid grassland with 0% *Calluna* (HL3, GL11) Mire (in heathland situations) (AQ1) Unimproved/semi-improved neutral grassland (GL12, GL13) Woodland (W1, W2, W5, W6, W7, W8) Other	All habitat types were kept the same except for mire (AQ1) replaced with heath.
Soil type	64301 – Deep sandy to clay 64303 – Deep loam to clay 71107 – Seasonally wet loam to clayey over shale 84102 – Seasonally wet deep loam Other	No change.

Table A-52. Input variables for the *G. constrictum* BBN model with notes on changes made to the states to model the impact of climate change.

Variable/node	**States**	**Change**
Habitat type	Dry heathland (HL1) Dry heathland/acid grassland mosaic (HL3) Wet heath (HL2) Mire (AQ1) Continuous bracken (GL8) Unimproved acid grassland with 0% *Calluna* (GL11) Unimproved neutral grassland (GL12) Semi-improved neutral grassland (GL13) Woodland – Broadleaved (W1, W2), Coniferous (W5, W6), Mixed (W7, W8) Other	All habitat types were kept the same except for loss of wet heath (HL2) and mire (AQ1). Where these two habitats occurred, they were replaced with 50% HL1 and 50% HL3 (as it is not known which cover of *Calluna* would be more likely). (Bracken was kept the same at this stage).
Distance to woodland	< 100 m 100 – 1000 m > 1000 m	No change.
Tree canopy cover	Open Partially open Closed	No change.
Bracken density	Vigorous Less vigorous Sparse Absent	No change.
Bracken litter depth	Absent < 5 cm 5 – 10 cm > 10 cm	No change.
Soil type	64301 – Deep sandy to clay 64303 – Deep loam to clay 71107 – Seasonally wet loam to clayey over shale 84102 – Seasonally wet deep loam Other	No change.
Ground moisture	Dry Moist/Damp Wet	Increased state to one drier (e.g. moist/damp becames dry).

Table A-53. Input variables for the *G. illyricus* BBN model with notes on changes made to the states to model the impact of climate change.

Variable/node	**States**	**Change**
Water acidity	High acidity = pH 1 – 3 Moderate acidity = pH 4 – 5 Low acidity/Circum-neutral = pH 6 – 7 Alkaline = pH 8 – 9 Absent	No change (although it has been suggested that there may be an increase in pH in water bodies, the extent of this is not clear).
Speed of water flow	Absent Still/Slow Steady Fast	To take account of decrease in flow in summer: If ‘still/slow’ and water feature was ‘deep’ (as indicated by ‘water depth’) then kept as ‘still/slow’. If ‘still/slow’ and ‘shallow’ then changed to 50% ‘still/slow’ and 50% ‘absent’. If ‘steady’ then changed to 50% ‘steady’ and 50% ‘still/slow’. If ‘fast’ then changed to 50% ‘fast’ and 50% ‘steady’. (Although winter flow may increase, this may be more temporary (or ‘flashy’) so this was not included. Further, the still ‘flow’ of a water features such as a pond would not be able to increase.)
Water depth	Absent Shallow Deep	Water depth likely to decrease in summer and increase in winter. So used likelihood function to assign 50% increased depth (e.g. ‘shallow’ to ‘deep’). If ‘shallow’ then no change (although the wetland feature may dry up completely and be ‘absent’, this would be taken account of by the

		‘presence of wetland habitat’ node).
Water depth fluctuation	Yes No Absent	Where wetland feature present, all ‘yes’.
Disturbance/grazing	Present Absent	No change.
Bare/open substrate	Present Absent	No change.
Bank/shore substrate	Fine Coarse Absent	No change.
Bank/shore angle	Near flat/Gentle Sloped Steep/near vertical/undercut Absent	No change.
Presence of wetland habitat	Present Absent	‘Present’ became 50% ‘present’, 50% absent. ‘Absent’ stayed as ‘absent’.
Tree canopy cover	Open Partially open Closed	No change.
Ground moisture	Dry Moist/Damp Wet	Increased state to one drier (e.g. moist/damp became dry).
Soil type	64301 – Deep sandy to clay 64303 – Deep loam to clay 71107 – Seasonally wet loam to clayey over shale 84102 – Seasonally wet deep loam Other	No change.

Table A-54. Input variables for the *P. globulifera* BBN model with notes on changes made to the states to model the impact of climate change. Variables highlighted in grey are those which were not instantiated in running the BBN, but could be if a different version of the model was run (e.g. the *P. globulifera* – 2 model).

Variable/node	**States**	**Change**
Habitat type	Dry/humid heath (HL1) Dry heath/acid grassland mosaic (HL3) Wet heath (HL2) Mire (AQ1) Unimproved acid grassland with 0% *Calluna* (GL11) Woodland – Broadleaved (W1, W2), Coniferous (W5, W6), Mixed (W7, W8) Other	All habitat types kept the same except for loss of wet heath (HL2) and mire (AQ1). Where these two habitats occurred, they were replaced with 50% HL1 and 50% HL3 (as it is not known which cover of *Calluna* would be more likely).
Presence of suitable grass species	Present Absent	No change
Tussocks	Present Absent	No change
Pioneer heath	Present Only later phases Absent	No change (assumption that management maintains similar conditions).
Bare ground	Present Absent	No change (although there may be a slight increase, it is unlikely to be the difference between a change in states).
Tree canopy cover	Open Partially open Closed	No change.
Ground moisture	Dry Moist/Damp Wet	Increased state to one drier (e.g. moist/damp became dry).

Table A-55. Input variables for the *H. semele* BBN model with notes on changes made to the states to model the impact of climate change. The variable highlighted in grey is one which was not instantiated in running the BBN, but could be if a different version of the model was run (e.g. the *H. semele* – 2 model).

Variable/node	**States**	**Change**
Habitat type	Dry/humid heath (HL1) Dry heath/acid grassland mosaic (HL3) Wet heath (HL2) Mire (AQ1) Woodland – Broadleaved (W1, W2), Coniferous (W5, W6), Mixed (W7, W8) Other	All habitat types kept the same except for loss of wet heath (HL2) and mire (AQ1). Where these two habitats occurred, they were replaced with 50% HL1 and 50% HL3 (as it is not known which cover of *Calluna* would be more likely).
Suitable grazing/ management (burning)	Suitable Unsuitable	No change (assumption that management maintains similar conditions).
Suitable ericaceous species in pioneer phase	Suitable Unsuitable	No change (assumption that management maintains similar conditions. Although cross-leaved heath may decline, *P. argus* can also use heather (and bell heather)).
Bare ground	Present Absent	No change (although there may be a slight increase, it is unlikely to be the difference between a change in states).
Heathland patch size	Larger patch Smaller patch Absent	No change (although patch size may decrease if the heathland becomes too dry or the cover of acid grassland becomes much higher).

Table A-56. Input variables for the *P. argus* BBN model with notes on changes made to the states to model the impact of climate change.

Variable/node	**States**	**Change**
Woodland type	Broadleaved Mixed Coniferous Absent	No change.
Woodland cover	Low Medium High Absent	No change.
Leaf litter depth	< 1 cm/Absent 1 – 5 cm > 5 cm	No change.
Leaf litter cover	Absent Low Medium High	No change.
Ground vegetation cover	Low High	No change.
Ground moisture	Dry Moist/Damp Wet	Increased state to one drier (e.g. moist/damp became dry).
Amount of woodland edge	< 50 m 50 – 100 m > 100 m Absent	No change (although edge may increase in some woodlands, it is unlikely to be by as much as a difference in states).

Table A-57. Input variables for the *N. sylvestris* BBN model with notes on changes made to the states to model the impact of climate change.

Variable/node	**States**	**Change**
Habitat type	Dry heathland (HL1) Dry heathland/acid grassland mosaic (HL3) Unimproved acid grassland with 0% *Calluna* (GL11) Wet heath (HL2) Mire (AQ1) Unimproved neutral grassland (GL12) Semi-improved neutral grassland (GL13) Improved grassland (GL3) Continuous bracken (GL8) Broadleaved woodland (W1, W2) Coniferous woodland (W5, W6) Mixed woodland (W7, W8) Other	All habitat types kept the same except for loss of wet heath (HL2) and mire (AQ1). Where these two habitats occurred, they were replaced with 50% HL1 and 50% HL3 (as it is not known which cover of *Calluna* would be more likely).
Soil type	64301 – Deep sandy to clay 64303 – Deep loam to clay 71107 – Seasonally wet loam to clayey over shale 84102 – Seasonally wet deep loam Other	No change.
Tree canopy cover	Open Partially open Closed	No change (although there may be some opening up in the canopy, overall it is unlikely to as much much of a difference to warrant a change in state).
Ground vegetation height	< 15 cm 15 – 30 cm > 30 cm	No change.
Ground moisture	Dry Moist/Damp Wet	Increased state to one drier (e.g. moist/damp became dry). AND (also ran with): Decreased state to one wetter.
Presence of (dung from) ponies feeding on natural vegetation	Present Absent	No change (although numbers of ponies may change an assumption is made that sufficient numbers will remain).

Table A-58. Input variables for the *P. punctata* BBN model with notes on changes made to the states to model the impact of climate change.

Appendix 36. Results of testing BBNs with fieldwork data with states changed to account for the potential impact of climate change

	Presence sites	**Absence sites**
1	70.18	38.89
2	87.39	49.56
3	86.42	25.67
4	43.74	16.81
5	33.71	0.84
6	81.57	5.72
7	17.55	3.60
8	78.18	16.12
9	18.18	10.56
10	70.79	10.71
11	92.72	1.74
12	7.08	49.56
13	90.30	2.58
14	59.26	2.07
15	39.78	30.16
16	26.80	38.89
17	24.83	0.45
18	97.84	25.50
19	90.30	8.16
20	15.84	12.75
Average	56.62	17.52
Range	7.08 – 97.84	0.45 – 49.56

Table A-59. Habitat suitability values for the fieldwork data applied to the *C. nobile-1* BBN.

	Presence sites	**Absence sites**
1	27.84	20.14
2	21.00	0.00
3	40.38	0.00
4	15.00	0.00
5	3.80	0.00
6	15.00	0.00
7	7.13	0.00
8	19.50	25.50
9	29.30	24.00
10	10.60	0.83
11	15.00	19.50
12	12.75	0.00
13	19.50	0.00
14	19.00	2.32
15	19.00	0.00
16	12.03	0.00
17	9.50	0.00
18	21.20	2.29
19	15.00	0.00
20	13.21	7.08
Average	17.29	5.08
Range	3.80 – 40.38	0 – 25.50

Table A-60. Habitat suitability values for the field work data applied to the *G. constrictum* BBN.

	Presence sites	**Absence sites**
1	3.92	12.11
2	23.35	1.57
3	35.14	23.77
4	60.33	1.76
5	5.00	3.78
6	55.16	32.15
7	43.27	0.00
8	20.75	0.47
9	4.04	0.56
10	9.27	0.00
11	11.38	16.91
12	13.69	0.67
13	13.29	2.74
14	34.62	42.00
15	51.12	1.12
16	45.29	2.42
17	38.09	1.86
18	47.51	29.86
19	35.74	0.88
20	32.53	34.05
Average	29.17	10.43
Range	3.92 – 60.33	0 – 42.00

Table A-61. Habitat suitability values for the fieldwork data applied to the *G. illyricus* BBN.

	Presence sites	**Absence sites**
1	100.00	100.00
2	100.00	0.00
3	70.00	100.00
4	70.00	16.64
5	100.00	4.73
6	100.00	0.00
7	36.60	0.00
8	100.00	0.00
9	40.00	65.00
10	70.00	0.00
11	40.00	21.60
12	40.00	54.01
13	90.80	40.00
14	64.06	0.00
15	70.00	0.00
16	36.20	56.20
17	100.00	2.69
18	40.00	0.00
19	36.60	11.12
20	100.00	0.00
Average	70.21	23.60
Range	36.20 - 100	0 - 100

Table A-62. Habitat suitability values for the field-work data applied to the *H. semele-1* BBN.

	Presence sites	Absence sites
1	51.94	22.36
2	58.10	5.70
3	18.31	3.31
4	70.15	2.28
5	84.60	23.39
6	83.37	43.04
7	83.37	0.00
8	83.37	0.00
9	83.37	0.00
10	38.32	0.00
11	39.25	2.28
12	83.37	0.00
13	83.37	28.48
14	2.28	2.28
15	18.31	0.00
16	18.31	0.00
17	83.37	0.00
18	41.01	2.28
19	87.42	0.00
20	36.65	2.08
Average	57.41	6.87
Range	2.28 – 87.42	0 – 43.04

Table A-63. Habitat suitability values for the fieldwork data applied to the *N. sylvestris* BBN.

	Presence sites	Absence sites
1	42.03	0.00
2	63.73	0.00
3	46.40	0.00
4	46.40	0.00
5	65.88	42.03
6	92.00	43.16
7	87.90	46.40
8	43.62	46.40
9	70.76	46.40
10	48.80	0.00
11	92.00	44.60
12	69.20	0.00
13	92.00	49.74
14	69.20	0.00
15	92.00	0.00
16	76.50	5.68
17	92.00	0.00
18	51.08	0.00
19	26.48	76.64
20	35.60	22.79
Average	65.18	21.19
Range	26.48 – 92.24	0 – 76.64

Table A-64. Habitat suitability values for the fieldwork data applied to the *P.argus-2* BBN.

	Presence sites	Absence sites
1	15.35	15.35
2	15.35	0.00
3	15.35	1.72
4	29.21	0.00
5	15.35	0.00
6	18.04	0.00
7	26.71	0.00
8	2.85	0.00
9	15.35	0.00
10	18.04	0.00
11	18.04	0.44
12	15.35	15.35
13	16.32	0.00
14	18.04	0.00
15	18.46	0.27
16	18.62	1.42
17	18.04	0.99
18	10.92	0.00
19	18.04	0.00
20	15.35	0.00
Average	16.94	1.78
Range	2.85 – 29.21	0 – 15.35

Table A-65. Habitat suitability values for the fieldwork data applied to the *P. globulifera-1* BBN.

	Presence sites	Absence sites
1	82.30	58.10
2	89.58	40.63
3	89.58	57.05
4	97.71	78.03
5	87.80	48.36
6	81.83	43.90
7	39.84	70.22
8	81.44	69.37
9	89.88	56.38
10	82.04	73.60
11	85.16	81.44
12	98.81	68.21
13	84.16	80.85
14	97.71	38.47
15	81.44	72.72
16	76.06	73.97
17	86.99	70.57
18	81.44	85.79
19	97.71	87.27
20	57.67	25.89
Average	83.46	64.04
Range	39.84 – 98.81	25.89 – 87.27

Table A-66. Habitat suitability values for the fieldwork data applied to the *P. punctata-dry* BBN, with assumption that ground moisture becomes drier.

	Presence sites	**Absence sites**
1	58.85	3.01
2	64.09	1.80
3	64.09	2.94
4	69.94	4.40
5	33.51	18.15
6	58.52	31.24
7	14.83	50.17
8	58.24	3.80
9	64.30	40.22
10	31.27	52.60
11	32.48	58.24
12	70.72	48.73
13	32.09	30.80
14	69.94	27.35
15	58.24	27.64
16	54.37	28.12
17	62.22	3.88
18	58.24	38.26
19	69.94	4.18
20	41.15	1.64
Average	53.35	23.86
Range	14.83 – 70.72	1.64 – 58.24

Table A-67. Habitat suitability values for the fieldwork data applied to the *P. punctata-dry* BBN, with assumption that ground moisture becomes wetter.

Explanations of the different models can be found in section 5.4 (Chapter 5).

Appendix 37. Comparison of predicted values for GLM and BBN models

	Presence sites		Absence sites	
Site number	**GLM**	**BBN**	**GLM**	**BBN**
1	51.57	65.16	37.66	23.71
2	40.22	89.58	9.36	15.55
3	15.16	89.58	70.99	25.24
4	51.23	97.71	74.46	26.33
5	81.71	68.42	26.54	41.41
6	99.40	81.83	29.01	43.90
7	0.27	34.08	1.15	70.22
8	83.18	81.44	78.97	26.74
9	86.27	89.88	73.89	56.38
10	54.48	59.90	50.17	73.60
11	15.36	58.43	65.35	81.44
12	86.33	98.81	26.05	59.76
13	91.39	54.59	97.98	65.46
14	85.32	97.71	47.29	38.47
15	99.40	81.44	80.26	52.37
16	95.87	76.06	7.48	59.00
17	97.48	86.99	59.46	24.19
18	96.44	81.44	94.48	58.84
19	95.26	97.71	24.36	45.82
20	62.78	57.67	2.61	9.12

Table A-68. Comparison of predicted habitat suitability values for GLM and BBN models for *P. punctata* (using the ‘dry’ BBN model) for the BBN presence and absence testing fieldwork sites for that species. Note that the GLM values were multiplied by 100 so that they were of the same order as the BBN values. It should be noted that several of the presence sites were sites that were used to train the GLM model, so the model would tend to perform particularly well in predicting values for those sites.

Table A-68 above shows that the habitat suitability (HS) values for the presence sites for both the GLM and BBN models for *P. punctata* were generally high and both models achieved a high sensitivity (true positive) rate (see section 3.3.3 (Chapter 3) and section 4.3.11.2 (Chapter 4)). The BBN HS values were all above 54, except for one (site 7) at 34. (This site also achieved a very low value for the GLM model). There were three sites that were predicted particularly low values by the GLM model (sites, 3, 7 and 11), whereas the BBN HS values were much higher for these sites, demonstrating the benefit of taking account of the finer scale habitat variables. There were some sites where the GLM HS values were higher than the BBN values, but the BBN values were still high.

It can be seen from Table A-68 above that some of the values for the absence sites were quite high for both models. As noted in section 3.4 (Chapter 3) and 6.1.1 (Chapter 6) this may be because many sites are likely to be suitable for *P. punctata*, making it difficult for a model to discriminate between more and less suitable sites. However, the BBN models provide a more accurate estimation due to them taking account of finer-scale factors. There were several particularly high HS values predicted by the GLM model (e.g. sites 3, 4, 8, 9, 13, 15 and 18) but the BBN model results showed that the sites were not actually that suitable (although some of the values are still quite high). There were also some high HS values predicted by the BBN model (sites 7, 10 and 11), which the GLM model predicted as lower (although still quite high for sites 10 and 11). As the BBN model predicted high HS for these sites it indicates that they were seemingly suitable but *P. punctata* did not occur there, just because species do not always occur at every suitable site (see section 1.2.3, Chapter 1).