Numerical Analyses of Distribution Patterns in the British and European Floras

Thomas James Ronald Finnie

A dissertation submitted in partial fulfilment

of the requirements for the degree of

Doctor of Philosophy

of the

University of London

and for the

Diploma of Imperial College.

Department of Biology

Imperial College, London

June 1, 2007

Declaration

Some of the material included in this thesis was produced in collaboration with other workers:

Chapter Three Was originally produced as a manuscript for journal submission in collaboration with C. D. Preston, M. O. Hill and M. J. Crawley. Collaborators provided the detailed botanical descriptions of the groups. The author's contribution was in the development of the methods used, carrying out the statistical analyses and in writing the first draft of the manuscript.

Chapter Five The theory behind the second measure of nesting, NEST2, was created in collaboration with M. O. Hill. The author then developed this into a working solution. All other measures and statistics were created and implemented by the author.

All other work contained in this thesis is my own.

Tom Finnie

June 1, 2007

T. J. R. Finnie

June 1, 2007

Prof. M. J. Crawley FRS

Abstract

This work is a study in plant distribution, phytogeography. It covers the identification, classification and interpretation of landscape-scale patterns within the British and European floras.

I show the implementation and characteristics of a new mathematical method for species group definition. This method provides an objective, multistage procedure which extracts both well-known and previously undetected species groups. Application of this method to the complete native British flora found fresh phytogeographical groups and confirmed familiar groups such as calcicolous, coastal and montane species. Many of these groups are shown to be formed by broad-scale environmental variables. The new classification of the British flora is detailed.

Classification of the European flora revealed wider phytogeographic patterns. Differences between species richness and group densities highlighted regional variations of alpha and beta diversity across regions, as well as indicating distributional differences between national and continental scales.

Finally, I investigate the utilisation of species distribution patterns within mathematical ecology. Nestedness of bryophyte species is considered with particular reference to climatic influence. The results show the relative geographical correspondence of the species in each group, indicate those species that do not fit the general pattern and give the overall geographical similarity of each class. I also examine the effect of species patterns on species-area relationships. Different results may be obtained from the choice of the initial site or by the scale used. Increasing scales caused a decrease in the power law exponent, z, and low species density in the initial study area was strongly linked to high z values.

Species distributions are spatially uneven and repeat across many taxa, they allow classifications of distributions and influence outcomes of theoretical and practical studies. My new classifications provide a reference point for future studies and, combined with further mathematical research, will offer insights into the relationships between species.

Acknowledgements

No scientific work stands in isolation or exists without the aid of others. To this end I would like to thank the following for their assistance during the course of this project:

My supervisors, Mick Crawley and Chris Preston, for their advice, imagination and perseverance: the best ideas come from the wisest heads.

Mark Hill for the many suggestions involving the mathematics of this thesis and for corrections of historical, bryological and numerical fact.

The many staff at C.E.H. Monks Wood, including Henry Arnold for help in the production of the maps in chapter 3 whose knowledge and assistance with BRC data holdings is unrivalled; David Roy, a man whose ability to furnish one with esoteric data-sets should not be underestimated; and Pete Rothery for ensuring that my statistical analyses were on track. A mention should also go to the C.E.H. library staff, Pam Moorhouse and Helen Suddaby for tracking down some of the more elusive references.

NERC for their sponsorship of this project.

Finally I must thank my parents Janice and Mike without whose love and support none of this would have been possible.

†

Contents

List of Figures

List of Tables

Without mathematics the sciences cannot be understood, nor made clear, nor taught, nor learned...

Roger Bacon 1214-1292

Chapter 1

Introduction

1.1 Objectives

This work is a study in plant distribution, phytogeography. Investigating the patterns that can be found in the British and European flora, I look at the factors that may cause these patterns and study the effect that plant distribution has on other population processes. I aim to show that the distributions of species provide more information than just the locations in which species live.

Critical to this study is the availability of systematically recorded and comprehensive databases (in particular the *New Atlas of the British and Irish Flora* (Preston et al., 2002)). These allow more complex and detailed analyses to be performed than were ever before possible or justifiable. I focus on the study of repeating patterns within species distribution but also consider the nesting of species distributions, how species-area relationships behave at various scales and what the role of species distribution is in these.

1.2 Relevance

At the most fundamental level, the distribution pattern for a plant species demonstrates the places a plant can grow when subjected to competition with other species. In ecological terms, these patterns are the species' realised niches. Many ecological processes are strongly influenced by competition and the niche occupied by a species. Accordingly, insight into the working of these processes can be gained by the study of patterns of species distribution.

Maps of species distribution typically show repeated patterns. These repeating patterns reveal the large-scale assemblages of species. In this context they are phytogeographical elements or groups. For the most part these groups are controlled by factors such as temperature and geology that have a similar effect across large areas of countryside. However, broad-scale factors alone are not the only arbiters of the final communities. Other factors working directly on the plants are also important but may, ordinarily, become hidden in studies at the landscape scale. Rather than using environmental, historical or geographical surrogates, defining these landscape-scale communities mathematically enables us to identify species groupings by species distributions.

Community assemblage is not the only ecological process where the distribution patterns of species act as either indicator or causal agent. With a group of species, no matter how it is defined, measuring the degree of nesting of the distribution patterns can help quantify the strength of the geographical associations within that group. Results from the nestedness calculations serve to indicate those species that behave in fundamentally different ways to the other group members. Distribution patterns play the opposite, causal, role when considering species-area relationships. Here the distribution of species and groups can lead to wildly differing conclusions depending on such factors as the scale of the study or the area chosen.

In this work I present the first phytogeographical classification of the British flora based solely on the distributional data for the species at a 10km square resolution. The new method of analysis demonstrates robustness and applicability to other species groups and over a range of scales. I also describe a series of numerical techniques to test group robustness through nestedness. Finally, I show that the patterns that define phytogeographical classification have profound implications for other types of ecological study.

1.3 Background

1.3.1 History of British Phytogeography

Beginnings in ancient history

Knowledge of where food and predators are to be found is vital to the survival of many animals. While such knowledge was undoubtedly exchanged amongst early human ancestors, its earliest recorded example in relation to plants is attributed to Aristotle's pupil Theophrastus (370-285 BC). His *Enquiry into plants* covers plant morphology, behaviour, distinctive characteristics, reproduction, life history and, in his fourth book, the places where plants grow. He discusses this at country and regional scale as well as at finer scales such as where on a hillside a plant is to be found (Theophrastus). After Theophrastus' broad study much of the botanical literature of the next 1500 years follows Dioscorides' first century *De materia medica* being simply concerned with medicinal plants. A very few medieval works give directions to the locations where plants can be found but most leave this subject untouched. References to localities are more frequent in the first modern botanical works, published from the mid 16^{th} century onwards, and in county floras, which have been published since 1660. However, it was many years before any synthesis of these data was attempted.

Enlightenment and the Victorian era

True biogeography had to wait for the rise of the scientific method. The first biogeographical map is considered (Ebach and Goujet, 2006) to have been published by Lamarck and Candolle (1805). In this same year Humboldt was instrumental in launching the subject into the public consciousness with his *Essai sur la Géographie des Plantes* (Humboldt, 1805). The groundswell of scientific learning and the rise of the amateur naturalist during the 19^{th} century brought the subject to prominence because then, as now, the subject as a whole and species recording in particular was highly reliant on amateur efforts.

Table 1.1: Previous classifications of the British flora showing author and group names. None of the classifications show one to one mapping.

Watson	**Mathews**	**Preston-Hill**
British	Endemic	Endemic
English	Mediterranean	Mediterranean-Atlantic
Scottish	Oceanic Southern	Southern-Temperate
Highland	Continental Southern	
Atlantic	Oceanic West European	
(Western Britain and Southwest England)		
Germanic (Southeast England)	Continental	
	Continental Northern	Boreo-Temperate
	Northern Montane	Boreal-Montane
	Oceanic Northern	
	North American	
	Arctic-Subarctic	Arctic-montane
	Arctic-Alpine	Boreo-arctic montane
	Alpine	
	Wide	Wide boreal
	Eurasian	Wide-Temperate
	European	Temperate

Watson, in 1847, produced the first complete work on the groups of species found in the British flora and sought to define where a species could be located (Watson, 1847). The work divided Britain up into areas called vice counties, a system that is still used, in part, today. The presence of species in each of the vice counties allowed Watson to build up a series of "geographical types" (phytogeographical groups) for Britain. Watson's work contains two main parts, one of collating species distributions, the other of assigning the species in them to phytogeographical elements. Watson relied heavily on correspondence for his records. He maintained by post a network of fellow naturalists scattered throughout the country and so may have created the first modern biological recording scheme. Within his work the flora of

Britain is considered in isolation with no, or very little, reference to the wider European or worldwide distribution of a species. The treatise relied on Watson's knowledge of the species involved and his judgement as to which element each species belonged. It recognised six main types of distribution and several intermediate types. The main groups for the three major past classifications are presented in table 1.1.

His work was a great achievement. However, following the publication of the final volume of *Cybele Britannica* in 1853, interest in the flora seems to switch from the geographic to the taxonomic despite Watson's later updated volume *Topographical Botany* (1883). This shift in emphasis may well be ascribable to Watson himself as he is obviously wrestling with the problem of species concepts: in 1857, just after the final volume is published, he writes to Darwin "In writing the final volume of my *Cybele Britannica*, I find myself unable to carry out the ideas or inquiries originally intended. And why? Mainly, because the limits of species are so uncertain in nature" (Burkhardt and Smith, 1990).

Early twentieth century

It was not until the early 20^{th} century that this subject was revisited. In the intervening period many of the taxonomical problems that had affected Watson were resolved by the activities of Botanical Exchange Clubs and the publication of a large number of county floras (Matthews, 1937). The question at the heart of the subject expanded from what groups existed in the flora to the way the British flora formed and, in particular, how did it reform after the last ice age (Moss, 1914; Reid, 1911). Partial systems for species associations were proposed by several authors, but none were generally adopted (Moss, 1914; Salisbury, 1932; Stapf, 1914) until Matthews published a series of papers on the subject (Matthews, 1923, 1924, 1926). These culminated with his presidential address (Matthews, 1937) which sought to define groups of British plant distribution with the plants' wider European distributions. This was later published in revised form as a book (Matthews, 1955).

Matthews, as might be expected from the then newly-elected president of the British Ecological Society, was quite insistent that problems within plant distribution (in particular the construction of the flora's current composition) be wrested from the plant geographers and brought under the umbrella of ecology. The consequence of this act of moving the field from the purely descriptive toward the explanatory underlies not only this current work but a great deal of what is today considered ecology.

Into the computer age

When Matthews' 1955 book was published there was a revolution underway in the manner in which species recording functioned in Britain. This took the form of the project to produce an atlas of the British flora (see below). The publication of dot map distributions for the majority of Britain's native species in the atlas (Perring and Walters, 1962) would, in time, lead to a change in the way in which analysis of plant species distribution was carried out. The vast number of records involved in the production of the atlas necessitated the use of mechanical devices in their collation.

Although computers were used for data collation and other studies in the field of plant ecology (Williams and Lambert, 1960, 1961) their use in phytogeographical analysis took rather longer to become popular. The dot map provides a wealth of information to the researcher but for many years the information available far outstripped the computational capacity needed to study them in detail.

Some of the earliest uses of computational methods occur in the analysis of animal distribution rather than that of plant distribution (Fisher, 1968; Hagmeier and Stults, 1964; Holloway, 1969; Holloway and Jardin, 1968). A limited number of papers had been published in the late 1960s describing use of computers to study plant distribution (Proctor, 1967; Sneath, 1967) but in 1972, at a conference on taxonomy, phytogeography and evolution, Jardine who was an early pioneer of the use of computer methods stated "Perhaps one reason why computers have been so rarely used by phytogeographers is uncertainty about the nature of the numerical methods which could be profitably used" (Jardine, 1972).

There was obvious interest in the subject during this period with Birks perhaps being the most prolific author. His 1973 paper on *A numerical analysis of the past and present flora of the British Isles* harked back to the earlier era by attempting to define the origin of the British flora from phytogeographical and historical data. The technology only allowed the country to be divided into ten units. He concluded "despite the many limitations of the primary floristic data, the numerical methods have considerable potential for handling and synthesising large amounts of historical biogeographical data" (Birks and Deacon, 1973). He went on to produce a numerical analysis of the European pteridophytes (Birks, 1976), the technology having sufficiently advanced to allow 144 species in 65 areas to be directly analysed. By 1993 sufficient computing power was available to permit an analysis of the 65 European Salix species with 484 grid squares (Myklestad and Birks, 1993).

As shown through the increasing amount of data used by Birks, but true for all workers, the restrictions of the computers and the data available at the time presented difficulties for those working with numerical analyses in phytogeography. By necessity these studies had to limit the volume of data to be processed. Strategies adopted to limit the data included the use of relatively small study areas, restricted numbers of species, few divisions, or any combination of the above. In many of the early studies the chosen method was that of cluster analysis, it having been specifically developed for this task. The greatest difficulty with its use is the quantity of data and hence computer storage required in its performance. The entire raw dataset and a distance matrix of half the dataset's size need to be stored in active memory. Searching for greater resolution, later studies turned to some form of correspondence analysis (reciprocal averaging). The advantage of this method is the much smaller memory footprint obtained through the use of climatic surrogates. Furthermore, because the species data are not used directly in the analysis their quality becomes a more minor issue. This very indirectness is also this method's weakness: by using surrogates inaccuracies are gathered and only predictable factors (even if in novel combinations) are observed.

Carey et al. in 1995 produced a paper on the biogeographical zonation of Scotland. Based on environmental data this attempted to represent the species distributions of various groups including plants. In 1997 Preston and Hill, working from a similar premise but limiting the species involved to just plants, published a paper analogous to an updated version of Matthews' work. At this time the data on the distribution patterns restricted them to using climatic data and envelope (broad niche) data rather than the species geographic distributions. In defence of this they argued that the extra information provided by full distributional data would not provide additional useful insights. By separating the analysis in two and separately calculating north-south and east-west bands this classification produced more possible phytogeographical groups than ever before. In practice the most important were the north-south groups and it is these that are shown in table 1.1.

Today technology has sufficiently advanced to allow entire datasets both to be stored and processed on a standard desktop system. Use of climatic data allows good approximations to be made to the phytogeographical element distribution and is still of particular import when distributional data is suspect or absent, as it was for Preston and Hill. But by using full distributional data we are able to find both patterns that are and those that are not caused by large scale environmental variables. This direct use of plant distributions means that all conditions that affect plants are considered in determining phytogeographical groups. The direct use of distributional data, as set out in the following chapters, allows us to distinguish more nuanced phytogeographical patterns.

1.3.2 Distribution mapping

Beginnings and the first atlas

Critical to any computational method is the data with which it is supplied. The data used for all of the work carried out in this thesis is in the form of a dot map. This is a map of the study area, generally divided by a regular grid, on which the presence of the species of interest in a grid square is marked by a dot. Squares where the species is absent are left unmarked. The first dot map of a British plant appeared in Good's 1936 study of *Himantoglossum hircinum* (Perring and Walters, 1962). Prior to this national descriptions of species distributions were rather more prosaic.

The widespread use of dot maps did not occur until after the Second World War. The publication of the *New Naturalist* and *Biological Flora* series promoted their inclusion in taxonomic and ecological literature as standard practice. This rise to prominence can be seen in the 1950 conference of the Botanical Society of the British Isles (BSBI) asking the question "which plants could be reliably mapped?". The rare plants of Britain, the meeting concluded, could be mapped with a fair degree of certainty. For the more common species very little beyond their existence in certain vice counties was known. Clapham, in the final paper of that meeting, proposed that the BSBI should "take steps to ensure that before long we have a set of distribution maps for British species... the maps be comprehensive and accurate... [and] that the unit area should be the 10km grid square". The proposal was "carried with acclamation"(Perring and Walters, 1962).

The selection of the 10km square (i.e. 10km by 10km) as the primary recording unit was facilitated by the inclusion of a 10km pitch national grid in the new sixth edition of the Ordnance Survey maps (1945-1947). The national maps now printed with these squares and the Ordnance Survey, publishing its 1:25,000 maps on sheets that exactly corresponded to the 10km squares, provided a ready-made system for botanical recording. This selection of 10km squares as the primary recording unit remains with us today. The majority of British biological recording is still being performed on this grid, or a version of it. Beyond the standard 10km square the most common variant is the 10km square subdivided into 2km squares (tetrads) for greater precision for local projects.

Immediately after the 1950 conference the BSBI set up a maps committee to investigate the practicalities of carrying out such a project. This committee concluded that the proposition was practical and approached various organisations including the Nuffield Foundation and the Nature Conservancy for funding. The former gave a grant of £10,000 and the latter a £4,000 grant and a punched card system for mechanised map production on the understanding that the machinery and the data would go on to become the basis of a permanent recording system. With funding secure the recording scheme started in 1954. The collection of records for the Atlas ran until 1960 and the Atlas itself was published in 1962.

The *Atlas of the British Flora* (Perring and Walters, 1962) set the standard for modern biological recording. Here there was no use of the historic county boundaries: recording was undertaken on the Ordnance Survey grid. All native plant species were recorded as present or absent in each of the squares in the grid. The focus of the work was the accurate and complete mapping of the species. Particular care was taken to obtain even coverage of the British Isles. The analysis of the data was left to later papers and so we see a break between data recording and analysis within plant phytogeography.

The legacy of the 1962 atlas was fourfold: it proved that an amateur network could successfully conduct a wide-scale scientific survey, it stimulated interest in biological recording amongst the general public, created a network of local experts to whom they could turn for help and advice and, finally, the core data from the recording scheme became the basis of the Biological Records Centre (BRC) at the Nature Conservancy's Monks Wood Experimental Station. Perring himself went from Senior Worker/Director of the mapping scheme to being the head of the new BRC.

A European aside

The authors of *Atlas of the British Flora* were also instrumental in launching the European equivalent. Walters was involved in the work that formed the taxonomic backbone, *Flora Europaea* (Tutin et al., 1964, 1968-1980), while the inspiration for the *Atlas Florae Europaeae* came from "a map presented by

Dr. Perring at the Tenth International Botanical Congress in Edinburgh, 1964 which showed the 50km square distribution of *Silene acaulis*" (Jalas and Suominen, 1972). Perring also acted as secretary to the nascent committee before responsibility for the secretariat was handed to Helsinki in 1965. The first volume of *Atlas Florae Europaeae* was published in 1972 (Jalas and Suominen, 1972), ten years after the British atlas. Due to the vast amount of work involved and the political difficulties for field recorders in parts of Europe the project has proceeded rather slowly, releasing a volume every three to five years. With the publication of the thirteenth volume in 2004 the project has currently covered around 20% of the total European flora. Despite its incomplete nature, for phytogeographical purposes the data are useful. Omission of the species yet to be mapped are countered by the taxonomic order in which the volumes have been published. The plants so far covered represent all the major habitats found within Europe.

The new atlas

The strong recording ethos and support structures left by the 1962 atlas in the British biological community caused an explosion in biological recording. Many of Watson's vice counties now had teams of people recording their flora and a central person, in the form of a vice county recorder, to turn to for help and advice and to act as a repository and intermediary for records. In this positive atmosphere a large number of county floras were published. Preston et al. (2002) list 62 floras that map plants with a grid square system and a further 36 that use more traditional checklist-type techniques. These 98 works in the 40 years subsequent to the publication of the *Atlas of the British Flora* contrast with a mere 32 in the preceding 50 years.

The national situation was also positive. Around the nucleus of the atlas records the central national repository of the BRC was formed. With advances in technology it moved from its initial punched card system to a computerised digital database in the 1970s. The vast majority of the field records from the Atlas were digitised between 1970 and 1971 (Harding and Sheail, 1992). Together with keeping abreast of technology BRC continued to publish a series of updates and amendments to the Atlas. These included *Critical Supplement to the Atlas of the British Flora* (Perring, 1968) for species difficult to identify and excluded from the atlas, *Atlas of the Ferns of the British Isles* (Jermy, 1978) due to changes in pteridophyte taxonomy the 1962 maps became quickly outdated and *British Red Data Books 1 Vascular Plants* (Perring and Farrell, 1977) from resurvey work on rare plants. The atlas enjoyed two further editions in 1976 and 1982 and although the rare species had been resurveyed and their distribution maps updated the more common species remained unchanged except for the correction of a few minor errors. By the time that the third edition had been published it was becoming apparent that a replacement was needed.

The formal proposal of Wells at the 1983 BSBI recorders conference, that a project to produce a new atlas be begun, proved somewhat divisive. While many felt that the project was worthwhile there were also senior members of BSBI who doubted that there was either the volunteers available or sufficient change in the flora to justify a complete resurvey. As a compromise a partial survey was proposed: one in nine of the British and Irish 10km squares (11%) were to be resurveyed during the 1987-1988 field season. At the same time, a detailed survey of three smaller 2km squares within each chosen 10km square was also performed for later comparison. The "unqualified success" (Rich and Woodruff, 1990) of the survey led to a proposal that the BSBI council should undertake "a comprehensive survey of the British and Irish floras, in order to produce a replacement for the 1962 atlas" (Rich and Woodruff, 1990). In 1992 the council accepted this proposal and began detailed work towards the eventual production of a new atlas. Being a long term project, funding proved initially difficult but by 1995 with a grant from the Department of the Environment, Transport and the Regions the project began in earnest (Preston et al., 2002).

Following its start in 1995 there had been detailed work including the production of field cards, planning of special field meetings and a series of articles in various journals to bring the project to the attention of botanists and naturalists. Recording was begun during the 1996 field season and carried on until the end of 1999. Vice county recorders were responsible for surveying their own area although they were supported by workshops, meetings and the appointment of the Atlas Organiser, T.D. Dines. During this time interim maps and articles were drawn up to minimise omissions and under-recording. Data from the vice county recorders was then sent to BRC to be collated, checked, added to a database and mapped.

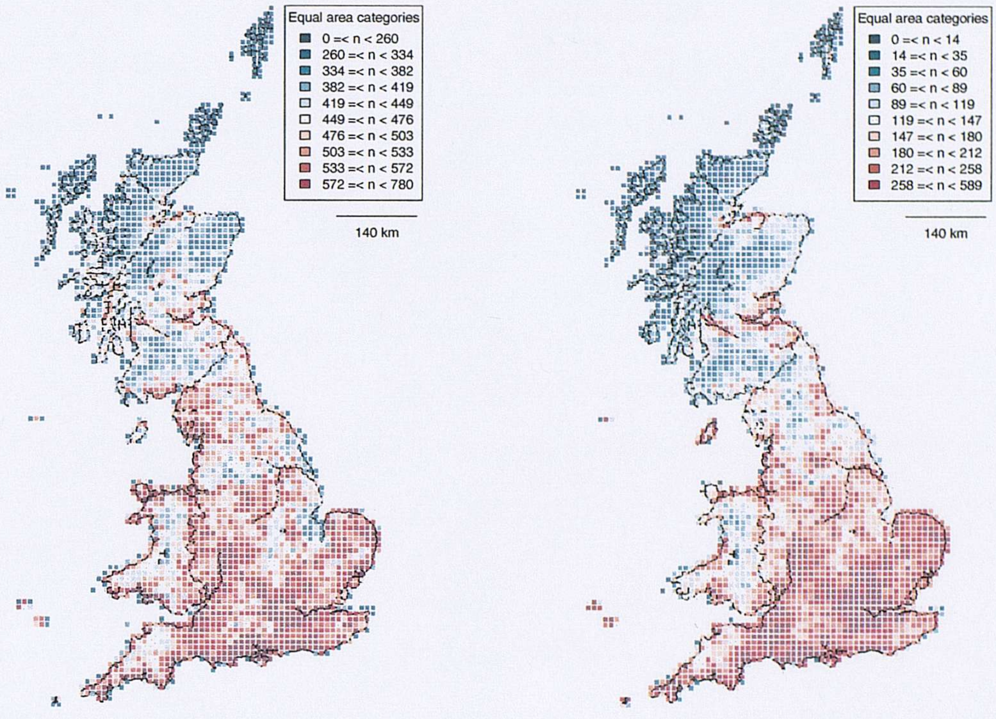
The *New Atlas of the British and Irish Flora* (Preston et al.) was published in 2002. It contains maps and descriptions of 2412 plant species with a further 940 available on the included CD. Along with the paper volume the project created a new vascular plant database. This carefully checked and verified, nine-million-record-strong database is an exceptional, unparalleled resource to the modern scientist and represents the current state of the art in large-scale species recording.

1.4 Species distribution miscellanea

1.4.1 Hotspots and alien species

Coincidence maps of the species from the *New Atlas of the British and Irish Flora* data set are presented in figure 1.1. For both the native and alien species these maps highlight those areas of high species richness. Due to the recording methods used the native map can be considered a true reflection of actual biodiversity. However, to be considered to be a hotspot in the sense of Myers (1989) these areas would also have to contain threatened species. Comparison of figure 1.1a and figure 1.2 shows that, while there is some correspondence between the places where rare species occur and the areas of highest floral diversity, they are not inextricably linked.

The degree to which alien species were recorded and whether a species was considered alien in a



(a) Native species

(b) Alien species

Figure 1.1: Density maps of native and alien species in Britain from the *New Atlas of the British and Irish Flora*. Data collection for native plant coverage was carefully regulated to ensure even and accurate recording; both presences and absences are highly reliable. Records of alien species were dependent on individual county recorders although the species still show an urban bias.

particular location was left to individual botanists as they conducted their surveys for the new atlas. For this reason, the data for alien species is rather more unevenly recorded than that for the native species (Preston et al., 2002). Despite the uneven recording the effect of anthropogenic factors can be seen in figure 1.1b: the distribution of alien species is highly urban.

The notable degree of correspondence between 'native' and 'alien' species in the two density maps of Fig. 1.1 is consistent with recent studies that report positive relationships between the richness of exotic and native species at regional (Deutschewitz et al., 2003) and at neighbourhood (Gilbert and Lechowicz, 2005) scale. Thompson et al. (2001) urge that 'the processes that facilitate invasion by exotic plant species and colonizations by native species are fundamentally the same', citing the experimental studies of invasion by Crawley et al. (1999) and Robinson et al. (1995) in support. Davis et al.'s 2000 paper argues that these same ecological processes may explain the positive correlation often found between native and alien species richness. Brown and Peet's 2003 study notes that at the scale of an individual plant competitive interactions lead to a 'negative correlation between diversity and invasibility' but also observes that otherwise Elton's 1958 prediction that more diverse communities should be more difficult

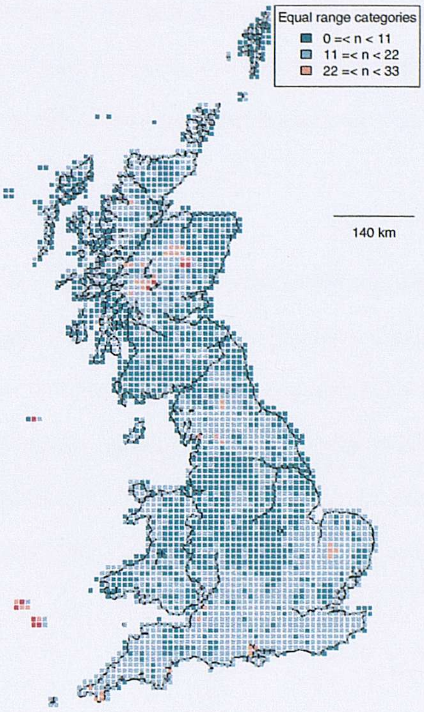


Figure 1.2: Density of rare species (native species that occur in 50 or fewer squares). Note that the concentrations of rare species do not exclusively coincide with the concentrations of native species.

to invade 'was not born out'. All the studies in this thesis fall above the neighbourhood scale, considering national or international regions.

In this thesis I do not intend to cover British biodiversity hotspots or alien species further, except where they have a direct effect on the results. Hotspots have been well covered in the literature since Myers' original paper (including but not limited to Griffiths et al. (1999); Myers et al. (2000); Prendergast et al. (1993); Reid (1998)) and vascular plant diversity is covered in chapter six of the *New Atlas of the British and Irish Flora*. The patchiness of the alien data negates its usefulness for many of the analyses I perform. These analyses have only been carried out using records of native species.

1.4.2 Tracking change in distribution

The two British atlas data-sets, completed thirty years apart, make it natural to consider looking for patterns of change in plant distributions. Indeed, changes in species range between the *Atlas of the British Flora* and the New Atlas survey are considered in chapter 7 of the New Atlas. Nor is the study in the New Atlas unique: the partial survey that led to the New Atlas, *The BSBI monitoring scheme 1987-1988* (Rich and Woodruff, 1990) has been repeated and used as the basis of *Change in the British Flora 1987-2004* (Braithwaite et al., 2006). With the rise in the political importance of the climate change 'debate' other prominent work also uses atlas data to measure changes in species distributions (Thomas et al., 2004; Wilson et al., 2004). However the lack of recorder effort figures, for the *Atlas of the British Flora* in particular, is a major impediment to the detailed study of change. Without this information only broad, aggregated studies of change can be made, as was the case for Preston et al. (2003). A decision was made early in the course of this project, therefore, not to study change in species distribution.

1.5 Technical notes

In so far as they do not conflict with the University of London thesis regulations I have set out this thesis and references in the style of Ecology. All numerical work has been carried out using the R statistical package (R Development Core Team, 2005). This package has also been used to produce the majority of the figures including most of the maps. Additional figures were produced and manipulated using OmniGraffle, Inkscape, the Gimp and Excel. Maps in chapter 3 were produced using the Dmap package. References were stored in Endnote and converted to BIBTEX. The text was written in Word and typeset using LATEX.

1.6 Thesis outline

Each chapter of this thesis has been written to be self-contained whilst advancing the overall argument. This scheme allows the major work to occupy chapters two to four with the satellite work appearing in chapters five and six. A discussion of the work as a whole appears in chapter seven with brief conclusions in chapter eight. Short descriptions of each chapter are provided below.

Chapter Two describes a new mathematical method for defining groups of species distributions. It provides an objective, multistage procedure which can extract both well-known species groups and groups previously undetected. It give examples of each from sample data.

Chapter Three analyses the complete species data for plants of the British Isles using the new method. Fresh phytogeographical groups are found and familiar groups are confirmed. A new classification of the British flora is detailed.

Chapter Four demonstrates the method applied to the flora of Europe and constructs a phytogeographical classification for the whole continent. Differences between species densities and groups densities highlight regional differences in the alpha and beta diversities across different regions.

Chapter Five considers the degree of nestedness to be found in the British bryophytes. This phylum contains a pair of classes; one traditionally considered to be nested (liverworts) the other not (mosses). The results of the methods generated show the relative geographical similarity of the species in each group, indicate those species that do not fit the general pattern and give the overall geographical similarity of each class.

Chapter Six examines the effect of species patterns on species-area relationships. It is shown here that different results may be obtained from such studies by choice of initial site or the scale used. Increasing scale caused a decrease in the power law exponent (and hence causes a reduction in the expected number of species when making predictions), whereas the effect of initial study area was strongly linked to its species density.

Chapter Seven draws together the main scientific points of the previous chapters in a general discussion and briefly covers more general use of species distribution patterns in ecological studies and other work in progress.

Chapter Eight gives a short summary of the work covered and presents the key conclusions.

Chapter 2

A new method of species classification by cluster analysis

2.1 Summary

Traditional methods of distribution analysis have often failed to illuminate dispersed or non-obvious patterns within species ranges. Cluster analysis has the potential to highlight such patterns but the shortcomings of individual techniques have so far precluded their widespread use. The novel combination presented here remedies these problems.

The method is demonstrated using the dataset from a new atlas of the British and Irish flora. I include only native records of vascular plants from mainland Britain, the Channel Islands and the Isle of Man.

Here two standard clustering techniques are used in an innovative combination. Initially I employ hierarchical clustering with complete linkage agglomeration and the asymmetric binary distance metric. This finds core groups. A second step is added so that all species within a group have similar range sizes. This step employs k-means type clustering using the cluster centres found in the first step as a starting point. Allocation of species to these centres is achieved by cosine angle measurement between the cluster centre and species vector.

Use of the method produces distinct, objective groups. These groups display clear separation between the areas members of the group occupy and those they do not.

The combination of these existing techniques produces discrete groups in an objective and repeatable manner. The resulting classification contains new groups of biological interest as well as the traditional phytogeographical groupings for the British flora. It is also applicable to other datasets.

2.2 Introduction

When distribution maps are studied, it becomes obvious that some patterns are repeated across species and taxa (Jardine, 1972). These patterns in species occurrence have long been of interest to scientists, from Victorian naturalists such as Watson (1847) and such twentieth-century workers as Matthews (1937, 1955) and Birks and Deacon (1973) to contemporary researchers (Preston and Hill, 1997). The classification of these patterns is one of the fundamental activities of descriptive biogeography. As well as being an end in itself, a classification can be put to use in tasks such as siting nature reserves to maximise conserved biodiversity (Marquez et al., 2001) or in the evaluation of the biological resources of an area (Carey et al., 1995). The classification can also be used as part of a wider enquiry, for example using its repeated patterns in the study of ecological driving factors or broad niche description. Historical or experimental studies of a species can now also be taken out of isolation and compared or contrasted with phytogeographically similar or dissimilar species (Birks, 1976).

The visual grouping of species into geographical patterns is fraught with difficulties, many of which are allied to human inconsistency and subjectivity. Chief amongst these are the finding of non-existent patterns, biasing of patterns from prior knowledge of the species' ecology, taxonomy or history and the difficulty in perceiving patterns with very disjoint outlines (Birks, 1976; Jardine, 1972). Use of a computational method to separate out patterns of species occurrence can eliminate some of the difficulties and add a level of objectivity to the classification method. Within any classification there will always be an arbitrary element as species distributions are intergrading, many being partially or completely nested within one another.

In this chapter I propose a new method of classification of species that generates groups based on the characteristics of their geographical distribution alone. Individual species can then be assigned to these groups with a high degree of confidence. This method is a novel combination of two well-known cluster analysis techniques. Output from this method can then be used to identify the main factors that directly influence the presence of species. This chapter will focus on the method as applied to a mapped species distribution to produce groups of species with similar ranges.

2.2.1 Objectives of classification

A classification method must be aligned with the final use to which the classification is put. My aim, in this study, is to classify all British native plant species by their distributions and to uncover previously unrecognised patterns. From a methodological standpoint this means that the classification system has to be all-inclusive, requires only the distribution data supplied and has to be robust. Inclusivity in this context means that the system has to produce a result for all the species that are included within the analysis. Because some species will fit the overall patterns less well than others it is desirable for the system to produce a measure of how well each species fits the group to which it has been allocated.

The requirement for the system to use only distributional data is so that all the potential influences (environmental, physical and biological) on plant distribution are included within the analysis and that no *a priori* assumptions are made as to the importance of these various factors. Robustness means that the system should, given similar starting conditions, finish with broadly similar groups. From a practical standpoint, the results should make biological sense. This is perhaps the most difficult of the limits to assess objectively, especially as the input data contain no explicit biological information. Results from a classification may be reasonable from a distributional perspective but not from an ecological one. However, the advantage is that any biological information extracted from the final result will have come from the implicit biological information contained within the distributions of the species concerned and not have been pre-determined.

2.2.2 Previous methods

With multiple sites (squares in this case) and many species, the data that makes up presence/absence data-sets are inherently multivariate. Any method that analyses this type of data must take this into account and will therefore be multivariate too. Existing statistical methods for the analysis of distributions can be split into two broad types: those that predict and those that classify. Predictive methods attempt to generalise the distribution pattern for a species over a large area from its known distribution and environmental information within a smaller area. Classification methods take known species distributions and attempt to group them together by known characteristics. These characteristics may be contained within the distributions themselves or be provided as extra information to the system. Methods such as generalised linear modelling (GLM), generalised additive modelling (GAM), and the various methods in the correspondence analysis family (CA, CCA, DCA, etc.) are of the predictive type and are not useful in the production of classifications (Braak, 1986; Cox, 1968; Guisan et al., 2002; Hastie and Tibshirani, 1986; Hill, 1974; Hill and Gauch, 1980; McCullagh and Nelder, 1989; Nelder and Wedderburn, 1972). Three of the more important methods of the second type, those that are applicable to classification, are outlined below.

Twinspan

This uses reciprocal averaging techniques to produce an ordination of species along a gradient (Hill, 1979). The gradient is then repeatedly sub-divided to produce the final groups. When the data do truly fall along some kind of gradient this method produces good results. However, as Belbin and McDonald (1993) note, the assumption of a dominant gradient obscures any secondary or greater order gradients that may exist within the data-set and the subdivision along the main axis is perhaps somewhat arbitrary. Of these two, the first objection makes this method unsuitable for the task here. The very high dimensionality of the data (one dimension for each square) means that multiple gradients are to be expected.

Bayesian classification

The use of naive Bayesian classifiers has become quite common in the field of machine learning especially since the misclassification rate was shown to be optimal not only under conditions where the explanatory variables are independent but also under conditions where 'strong attribute dependences are present' (Domingos and Pazzani, 1997). Despite their common and mainstream use in document classification their uptake and application to ecological problems has been rather slow. Although this system performs classification it is unsuitable for use here as not only are the classes predetermined but the system also needs to be trained with known values for each class before it is able to classify unknown individuals. In this sense the method is part way between true classification and the predictive methods (Phillips et al., 2006).

Cluster analysis

Although there are innumerable variants, cluster analysis techniques are of two broad types: hierarchical methods that assume that there is underlying structure to the the data and attempt to express this in the form of a relatedness tree; and bucket sorting, or partition-type clustering algorithms that attempt to assign individuals to a limited number of predetermined groups. Both work with mathematical measures of the individuals involved. Hierarchical techniques measure either the difference or similarity of the individuals to one another and use this to construct a tree. Partition clustering uses a measure of group fit and a reassignment mechanism to assign individuals to groups, to calculate a group identity and then to repeat the re-assignments until a point where all species have reached their optimal group or no further progress is made.

Hierarchical clustering is weakest when assigning distant individuals to specific groups. It also performs less well if there is is no underlying structure to the data. Partition clustering methods require that the number of classes be determined before the analyses are performed. Although the group count can be determined from the outset by a number of external methods none of these solve the greater problem of local minima. This is where the partition clustering stops prematurely because although there may be a grouping that produces a better overall fit for the data no rearrangement of the species to the current set of groups would allow this solution to be reached without first degrading the classification: a procedure that is not permitted within the mechanisms of partition clustering.

2.3 Method

2.3.1 Method choice

Of the available methods the cluster analysis techniques are the most suited to the purposes of this work. They have the advantage of directly working with the clusters themselves and demonstrate the patterns of distribution within the flora. However, both of the two main classes of clustering techniques have shortcomings that must be addressed before the criteria above can be met. Hierarchical clustering, as used by Birks (1976), would seem to be the most appropriate as the data are likely to be structured by species' physical and environmental requirements. Indeed, a simple hierarchical clustering of the data meets all but the inclusivity requirements above. However, it fails on inclusivity as many of the rarer plant species are placed within unique clusters and treated as distributionally unrelated to any other species. This inclusivity is guaranteed by a reassignment and partition clustering second stage.

An overview of my full method is presented as a flow diagram in figure 2.1.

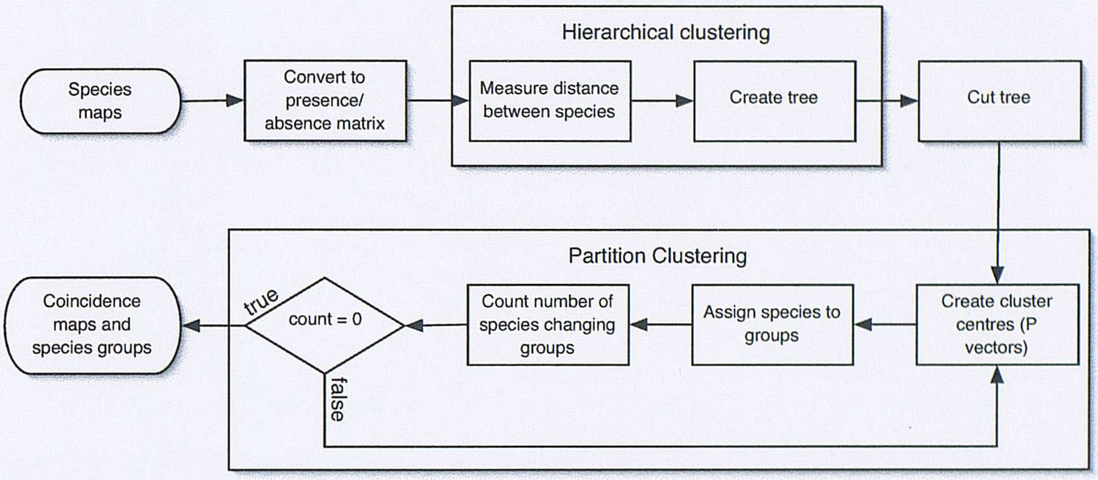


Figure 2.1: Flow diagram outline of the two-part clustering method. The method starts with individual species maps from atlas data (top left) and works through each, square-boxed, procedure in turn. In the penultimate step it checks that no species moves groups, if any do the partition clustering stage is repeated if not the method finishes with coincidence maps and group listings.

2.3.2 Metric choice

For hierarchical clustering I chose to use complete linkage agglomeration. In this type of agglomeration a species has to be similar to the most distant existing member of a group before the species can join it. This means that it becomes systematically more difficult for species to join a group (Legendre and Legendre, 1998). The clusters produced tend to be spherical with complete linkage. In classification this is desirable as it produces clearly delineated clusters with sharp discontinuities. These form the basis for the groups.

The choice of partition clustering countered the tendency of hierarchical clustering to place geographically restricted species into more widespread groups and provided a more informative classification by ensuring that loosely associated species did not get lumped into ubiquitous groups. It is important to ensure that the ubiquitous groups are truly made from species that exist everywhere rather than be formed from a loose patchwork of species occurring throughout the study area, creating a 'dustbin' group. It was inevitable that some species would move between groups because a different metric was used in this step. The species that moved groups were those which occupied the edges of the groups

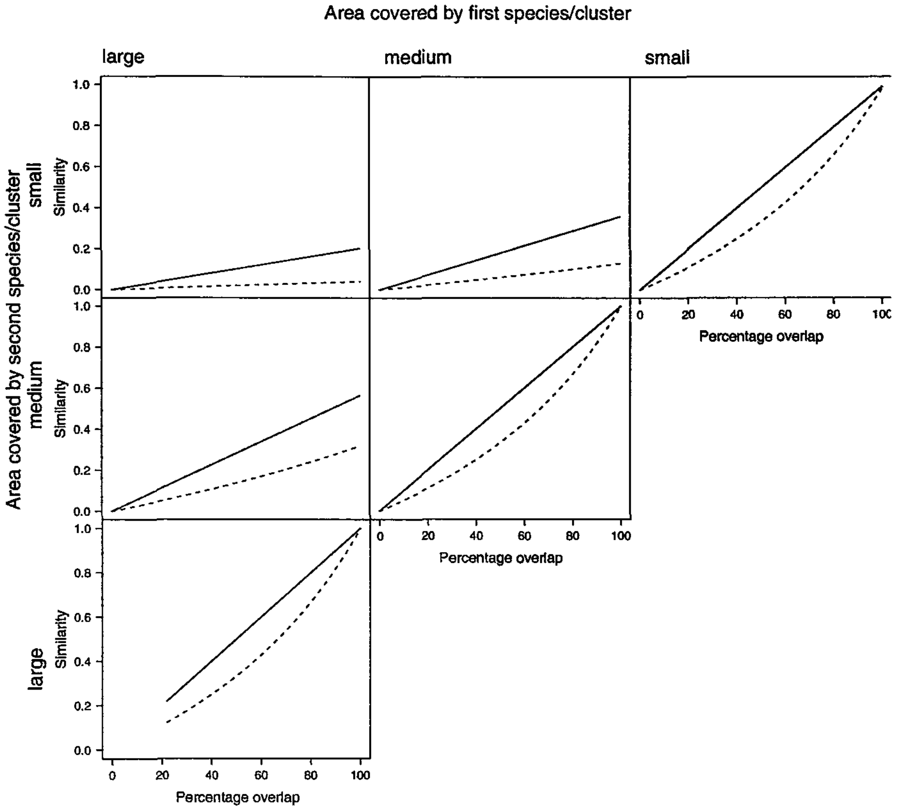


Figure 2.2: Graph series showing the values of the two metrics used versus the overlap between species or groups of varying size. The solid line is the response of the asymmetric binary distance and the dashed line that of the cosine metric. Note that there is no difference, mathematically, between a species and a group vector. The size of the species/group distributions were defined as: large; species representing the 75th centile (1620 10 km squares) of square occupancy, medium; the median square occupancy (513.5 10km squares) and small; 25th centile (67.25 10km squares) occupancy.

formed by hierarchical clustering.

The metrics chosen for the cluster analysis in this study had to fulfil different objectives. The metric used in hierarchical clustering was needed to place species into relatively few groups to limit the final size of the classification. This objective was achieved by the hierarchical clustering metric (the asymmetric binary distance) producing a stronger response to group-species overlap than the cosine metric used in the later partition clustering. By comparison, the cosine metric favours group-species interactions of equal size and so has the dual function of creating truly ubiquitous, ubiquitous groups and also of creating strongly defined groups of more limited range. The responses of the selected metrics to percentage overlap are illustrated in figure 2.2, the solid line representing the asymmetric binary distance and the dashed line representing the cosine metric. The use of the asymmetric binary distance metric always produces a stronger response, tending to lump species into groups, whereas the use of the cosine metric only produces a strong response to equal sized groups.

2.3.3 Detailed description

The study area was divided into equal-area squares in a grid pattern. The species distributions were then represented as a presence-absence matrix. A species distribution is a row within this matrix consisting of 1 or 0 indicating presence or absence respectively. An individual row is a species vector. Each element within this vector represents one unique square within the study area. The order of the squares remains constant throughout the analysis.

I selected the Hclust algorithm (Hartigan, 1975) as implemented within the amap package of R (Ihaka and Gentleman, 1996) for hierarchical clustering. Hclust is a tree-building algorithm. Each species starts as an individual "node" unconnected to any other. The algorithm measures the distance between each species and every other. I used the asymmetric binary distance metric (equivalent to 1 minus Jaccard's coefficient of community (Jaccard, 1908)) to measure distance between species, the formula for which is given in Equation 2.1. This measurement is similar to any other measure of distance in that smaller values mean that distributions are closer together. The only difference here is that instead of being in the familiar two or three dimensions this measurement has one dimension for every square in the data-set. When all the distances between species have been measured the two closest species are joined to form a new node. The process is repeated until just a single node remains.

Measuring the distance between two species is similar to measuring the distance between two trees in a park. Measuring the distance between a single species alone and a node containing more than one species is akin to measuring the distance between one tree and a clump of trees in this park; measuring the distance between two multi-species nodes is equivalent to measuring the distance between two clumps of trees. The question arises: where do you measure to in a clump of trees? The nearest tree, the average position for trees in the clump or the furthest tree? Rather counter-intuitively the most appropriate joining rule, so that the final groups are compact and properly include outlying species, is that which only allows species to join an existing group when the smallest distance recorded is between that species and the most distant member of that group. This furthest-distance joining rule is known as complete linkage agglomeration. In the park analogy, to form a 'spinney' group you would first require, for example, a 'north spinney' and a 'south spinney' group. These two subgroups would contain, between them, not only trees in the main spinney but also outlying trees (outliers are included in these subgroups because they are closer to a tree at the far edge of the subgroup than trees at the subgroup edges are to one another). Exactly how far out the furthest outliers are allowed to lie before the main group forms will depend on the distance metric used. The chief advantage of this method over other joining rules such as nearest neighbour or centroid is that as the algorithm proceeds it becomes more difficult for two groups to merge (their furthest members become progressively further away). With a nearest neighbour type rule the cost of joining remains constant and can lead to chaining. Chaining is where otherwise distinct groups are merged because their outer edges are close together and such chaining is undesirable if we are looking for distinct "clumps" of species as we are when seeking phytogeographical groups.

The final output from Hclust is a linkage chart starting with individual species and working down to the complete flora. This is known as a dendrogram or tree diagram. Every join between two species, a species and a node or two nodes has an associated height calculated from the distance between the nodes merged. This is equivalent to one minus the statistical probability of the nodes being equivalent. In line with standard statistical practice, I considered that merging nodes with less than 5% probability of being equivalent was unjustified and so the dendrogram was cut at the 95% dissimilarity level to produce the initial groups. Cutting the dendrogram at a higher dissimilarity level would have had the effect of producing larger, more widespread groups; cutting at a lower level would have had the reverse effect with a greater number of more localised groups.

Cluster centres for the second part of the two-part analysis were created by retaining groups containing ten or more species (0.71% or more of the flora). Species not in these initial groups were assigned to one of them using the cosine distance metric described in the partitioning algorithm below. I chose a minimum group size of ten species after evaluating all possible group sizes. Ten gives the minimum final ingroup deviation while maximizing the value of the minimum group size. That is, a minimum group size of ten gives a final deviation that is within 0.1% of the minimum possible in-group deviation and is at least an order of magnitude smaller than choosing minimum group sizes of 11 or greater.

To carry out partition clustering every species in turn is compared with each cluster centre and then assigned to the group it most closely resembles. To do this there are two requirements: a representation of each group and a method of comparing a species to this representation. This group representation is a vector with a location for every square in the study area. Each element within this vector is calculated as the number of species occurring in the square divided by the total number of species within the group. It is called a probability vector because each element gives the probability of finding a group member in any given square. To compare a species to a group I used the cosine angle between the species and the probability vector of the group (see Equation 2.2).

In this case a probability vector for each of the initial cluster centres was created. Then for each species in turn a value of the cosine metric was calculated. Species were assigned to the group that gave the greatest cosine angle value (this is a measure of similarity rather than distance). When all the species had a group assignment, the probability vectors for these new groups were calculated and the process repeated until no species moved group between interations.

Output from this technique can be taken either as a list of species assigned to each group or as the final probability vectors. The degree of fit of a given species to a group can be seen as the final cosine metric value. Let

$$A = \text{presence absence matrix}$$
$$n = \text{number of species}$$
$$m = \text{number of squares}$$
$$p_{ik} = \text{proportion of species in cluster } k \text{ that occur in square } i$$

Then

$$A = [a_{ij}] \quad (i = 1 \ldots m; j = 1 \ldots n)$$

Where a_{ij} is 1 or 0 depending on species presence or absence respectively.

The asymmetric binary measure is only used to obtain the distance between two species (j and j'). These species are always binary.

$$\text{Asymmetric binary distance}_{(j,j')} = \frac{\sum\limits_{\text{all squares}} \text{only one species present}}{\sum\limits_{\text{all squares}} \text{one or more species present}}$$

$$= \frac{\sum\limits_{i=1}^{m} a_{ij} + a_{ij'} - 2a_{ij}a_{ij'}}{\sum\limits_{i=1}^{m} a_{ij} + a_{ij'} - a_{ij}a_{ij'}} \qquad \text{(Equation 2.1)}$$

$$\text{Cosine angle}_{(j,k)} = \frac{\sum\limits_{i=1}^{m} a_{ij}p_{ik}}{\sqrt{\sum\limits_{i=1}^{m} a_{ij}^2 \sum\limits_{i=1}^{m} p_{ik}^2}} \qquad \text{(Equation 2.2)}$$

The data used to demonstrate this classification were from the database created for the ***New Atlas of the British and Irish Flora*** (Preston et al., 2002). Only native records mapped in the New Atlas were included, except for those species where native and alien records are not distinguished (e.g. *Taxus baccata*). For these, all records were included. The mapping units were those used by the ***New Atlas*** project and are the 10-km grid squares of the Ordnance Survey grid.

2.4 Results

When the method was applied to the British Flora data set 21 groups of plant distribution were defined. The group with the greatest area occupied 2646 of the 2833 British 10-km grid squares while the smallest group by area occupied only 173 squares. The mean occupancy was 1589 squares. The largest group contained 225 species (16.4% of the native flora) with the smallest group containing 17 species (1.2% of the flora). The mean group size was 65.4 species. Full details for these results with more detailed interpretation are provided in chapter 3.

2.5 Discussion

2.5.1 Choice of numerical methods

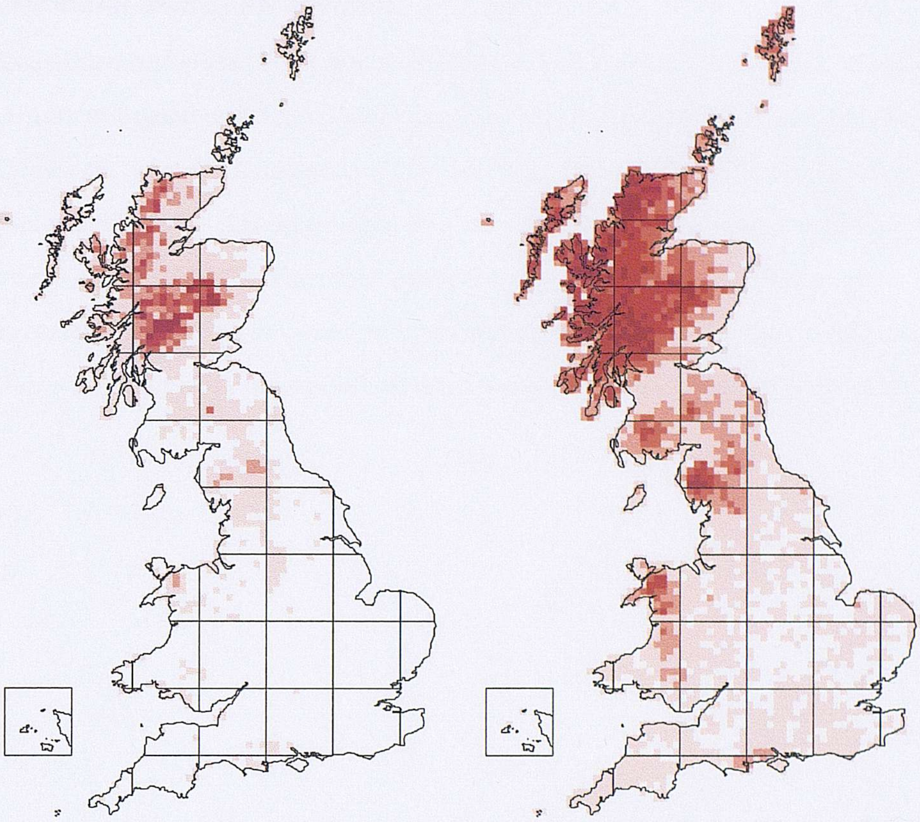
The aim of a successful classification is to take the unique distributions of all the individual native species and group them in a robust and meaningful manner. Ideally this should leave a small number of clearly defined groups from which information can be extracted. Cluster analysis in theory has the ability to produce these desirable characteristics. In practice, one type of cluster analysis alone is not sufficient. Hierarchical clustering is unable both to provide desirable clear-cut groups and to deal with species whose distributions do not fit well to a widespread pattern. Partition clustering, with this type of data, is unable to determine a suitable number of groups into which the data can be divided. The employment of the two separate analyses together using the advantages of each (hierarchical clustering creating clearly defined groups, partition clustering allowing manipulation of the groups) makes it possible to classify the more difficult species.

Measured against the original criteria outlined in the introduction, the system I use meets all of the methodological goals and for the most part meets the practical one too. The classification provided results for all included species, is robust (additions or deletions of sites or species do not cause vastly different classifications to be produced) and is produced entirely from the distributional data. The practical biological interpretation of the resultant groups is generally straightforward (see figure 3.2). Where it is not, such as the joint heath and broadland *Carex elata* group (figure 3.2r), this is generally caused by the inseparability of ecologically distinct species occurring within the same 10km square. Improvements to the biological accuracy of the groupings could be obtained by providing the method with explicit information at the outset but this would negate the original goal of finding patterns in the distributions alone. Moreover, the availability of the confidence of the fit of each species to the group to which it is assigned ameliorates this biological uncertainty while maintaining the original aims.

It should be noted that only recently has generally-available computer hardware been able to cope with agglomerative hierarchical clustering of data-sets containing several thousand individuals although at the current rate of increase in computational power this is rapidly becoming a trivial constraint. While the metrics chosen for this method produced a useful and usable classification, with increasing computing power it should soon be possible to assess many metrics against objective criteria such as group deviance or the global sum of squares to find those that perform most reliably.

2.5.2 Scale dependency

Results produced depend on the scale at which the data are recorded. It should not be assumed that given national species data the method is able to separate species that live, say, in a ditch from those that live on the banks or hedgerows surrounding the ditch. The more squares that the analysis has at its disposal the better the separation of different groups, exemplified in the splitting of species living on the tops of



(a) Restricted group (b) Widespread group

Figure 2.3: Maps showing separation of species into a restricted mountain top group and a widespread montane group.

the Scottish mountains from those living in the general montane area (figure 2.3). These groups also show how the groups nest within one another and overlap. The main limitation of this method lies in its requirements for the data. These are that the recorded data should cover all of the species equally within a square. Variations in recorder effort, if kept constant within the square, have little effect on the final group lists but would adversely affect the mappings of the groups. While it is possible to run the analysis on an incomplete data-set, caution must be advised as an excess of missing data can cause the hierarchical clustering step to fail by producing many unlinked groups. This could also be a problem if patterns are sought in sparsely distributed species with insufficient overlap between species ranges. One further caution would be that this is a pattern-seeking technique and as such provides no explanation for the patterns produced. To seek explanations for the pattern statistical techniques such as linear regression or correspondence analysis would be appropriate.

2.5.3 Example data

Three important findings from the example data are: firstly, the restriction of the ubiquitous group to truly widespread species; secondly, the method is able to separate groups that occur in very similar areas; and thirdly, with equal validity, groups that were expected occur alongside groups that are unexpected.

2.5.4 Potential uses of the method

This technique has been applied to geographical data at various scales with a good deal of success. Apart from the British data recorded in this chapter I have used both European 50km square data from the *Atlas Florae Europaeae* (Jalas and Suominen, 1972) to generate a European classification (see chapter 4) and county level tetrad (2km by 2km square) data for Cornwall, Dorset, Somerset and Norfolk.

The method as described in this chapter provides a way of objectively classifying species data without *a priori* assumptions. It has been specifically formulated to produce classifications from geographical data and to combat some of the more problematic aspects of using cluster analysis with this kind of data.

Chapter 3

Large-scale spatial patterns in the British flora

3.1 Summary

Classification of the large-scale distribution of vascular plant species in Great Britain has historically been a subjective and sometimes partial process. With the publication of *The New Atlas of British and Irish flora* the opportunity arose to apply, for the first time to such data, modelling algorithms which give, with a known degree of confidence, distinctive groups of species. This allows an estimate of the relative importance of climate, topography and geology in determining spatial patterns. Described here is the application of the statistical method, based on the sequential use of hierarchical and partition clustering, detailed in chapter 2. The method is robust, being independent of environmental factors and area measurement and dependent only on each species presence or absence within a map grid cell. Important new distributional groups within the British flora are shown and give a base-line for future studies.

3.2 Background

There has not been an analysis of the groupings of species for any national flora at the ten kilometre square grid scale. The question is whether there are distinctive patterns in the spatial distribution of vascular plant species within Great Britain and, if so, how many patterns there are, and what factors determine the location of their boundaries? Theory does not predict the number of expected groups to be observed. At the largest scale there is just one group, because all the species are British natives. At the smallest scale there are 1400 different distributions, because no two species, even the most frequent, have identical distributions. The principal ecological factors affecting plant distribution were proposed by the nineteenth century plant geographers (Brown and Lomolino, 1998), and included temperature, rainfall, geology, soil pH, altitude, aspect, exposure, snow-lie and maritime influence. Watson's classic categorization of British native plant species (Watson, 1847) identified 6 basic "types" of distribution (shown in table 3.1). Subsequently, Matthews (Matthews, 1937, 1955) identified 17 categories that he called "elements", placing British plant species in a European setting. Later, Preston and Hill produced a classification based on a combination of biogeographic zones and climatic data (Preston and Hill, 1997).

Table 3.1: Watson's classification, the most comparable with this work, is shown with the percentage of the flora in each group.

Group name	
British	44%
English	30%
Scottish	8%
Highland	5%
Atlantic (western Britain and southwest England)	8%
Germanic (south-east England)	5%

The early attempts at classification by large-scale distribution were subjective and sometimes partial, not least because the data were so sparse. The opportunity to re-evaluate and modernise the classification came in 2002 with the publication of the *New Atlas of the British and Irish Flora* (Preston et al., 2002) based on intensive field work by members of the Botanical Society of the British Isles. Because of its rigorously even coverage, and more critical refereeing than the original Atlas (Perring and Walters, 1962), there is a much greater confidence that the absences in species' distributions really are absences, rather than errors of omission resulting from under-recording. Because the Irish flora is still seriously under-recorded, the analysis has been confined to Great Britain.

The aim was to use data on species' presence or absence in all of Britain's 2851 10x10km squares to determine whether there are characteristic groups of species with broadly similar large-scale distributions. The use of "group" is in the sense of a "floristic element" (Cain, 1947) to describe a group of species with similar current geographic distributions within Britain, not *sensu* Dansereau (1957) who assumed a common origin and shared migratory history for the plants and perhaps even a common evolutionary association (Davis, 1987). Definition of such groups will enable us to address many of the fundamental questions of biodiversity: what aspects of their ecology do the species in each group have in common and what distinguishes them from species of similar growth form in other groups? Will species in the same group respond to climate change in similar ways? Do the widespread species show higher levels of phenotypic plasticity and/or greater genetic variability than shown by the species with more restricted geographic distributions?

3.3 Method

A presence/absence matrix was constructed for all 1400 native plant species and all 2851 10km squares comprising Great Britain, then two computationally intensive statistical methods were used in series.

First, hierarchical clustering was employed to uncover group structure in the presence/absence matrix (Hartigan, 1975), an approach pioneered by Proctor (1967) working on British liverworts. Archaeophyte and alien species, whose large-scale distributions are strongly influenced by people (Preston et al., 2004), were excluded from all stages of the analysis. This first-pass procedure identified 21 top-level groups. Each species was taken in turn and subjected to a separate partition cluster analysis. This exercise was repeated until none of the species changed groups between iterations (details in chapter 2). On completion of this analysis, final groups were named after the species which showed the highest correlation between its distribution and that of the group as a whole. Thus, the allocation of species to groups was based entirely on their spatial distributions within Great Britain, without regard to any environmental, geographic or ecological explanatory variables.

Explanatory variables were assembled for climate, geology, altitude and soil, averaged across each 10km square. For each group in turn, the proportion of the species found in each 10km square in Britain was modelled against the assembled environmental variables. Modelling was done via logistic regression; all squares were included even if they contained no species belonging to a group. No interaction terms or other non-linearities were included. Continuous explanatory variables entered into the maximal model were January temperature, July temperature, annual precipitation, minimum altitude, maximum altitude and proportional surface gradient (log(maximum altitude/minimum altitude)) this is shown as steepness in the tables, with two-level categorical explanatory variables for presence/absence within a 10km square of peat, chalk, limestone and coastline. Some of these variables are mapped in chapter 5 figure 5.5. Standard methods of model simplification (e.g. AIC) are ineffective as they treat each square as an independent data-point, this makes almost all explanatory variables significant. To find the most important variables in these full models they were simplified to give a reduced model by removing as many explanatory variables as possible while keeping the R^2 within 5% of that of the full model. This provided the most likely ecological factors controlling these distributions without involving complex (and not necessarily justifiable) spatial-autocorrelation analyses. Latitude and longitude were not included in the statistical models, but they are correlated with many of the explanatory variables (table 3.2).

Table 3.2: Correlations (Spearman's rank, n = 2851) between the explanatory variables used in modelling the distribution of groups, and the correlations between the explanatory variables and latitude and longitude (northing and easting). Note that because of the roughly triangular shape of Great Britain, latitude and longitude are significantly correlated in this data set, not orthogonal as might be expected.

	Jan Temp	**July Temp**	**Rain**	**Peat**	**Chalk**	**Limestone**	**Altitude Min**	**Altitude Max**	**Steepness**	**Sea**	**Easting**
July Temp	0.586										
Rain	-0.268	-0.719									
Peat	-0.407	-0.688	0.647								
Chalk	0.158	0.409	-0.332	-0.281							
Limestone	0.015	0.362	-0.333	-0.275	0.471						
Altitude Min	-0.562	-0.342	0.326	0.256	-0.107	-0.015					
Altitude Max	-0.655	-0.637	0.704	0.522	-0.253	-0.153	0.636				
Steepness	-0.584	-0.632	0.715	0.519	-0.262	-0.171	0.445	0.964			
Sea	0.481	-0.133	0.121	0.046	-0.075	-0.193	-0.494	-0.326	-0.248		
Easting	0.014	0.632	-0.802	-0.495	0.456	0.411	-0.131	-0.474	-0.504	-0.324	
Northing	-0.511	-0.862	0.409	0.544	-0.359	-0.306	0.058	0.308	0.336	0.242	-0.427

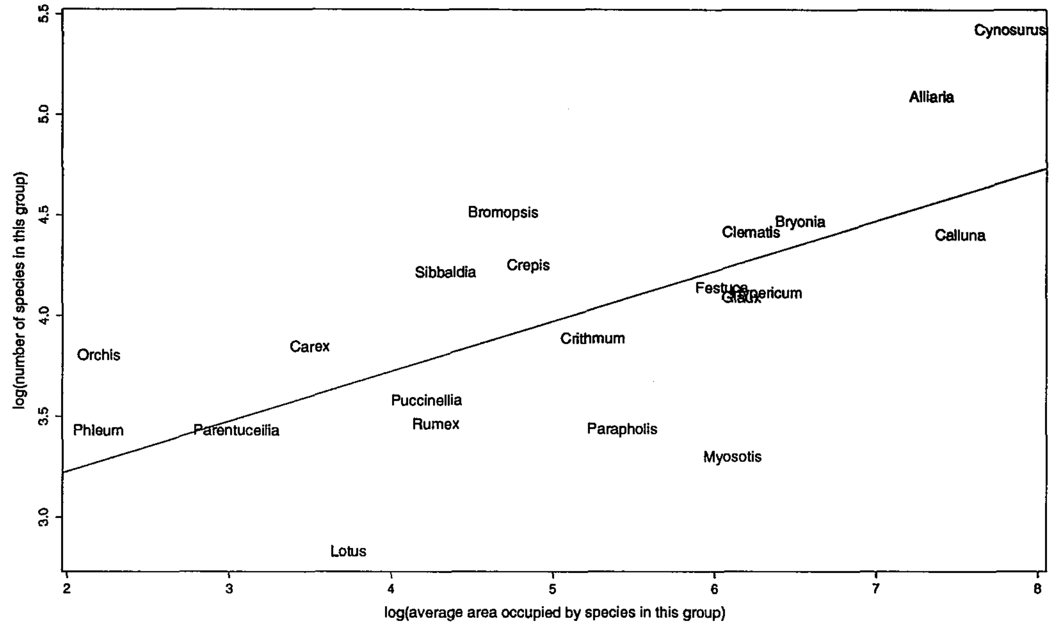
3.4 Results

Of the 21 groups recognised by this procedure, 15 were non-coastal and 6 were coastal. The attributes of the groups and the corresponding values of the explanatory variables are listed in table 3.3; their distribution maps are shown in figure 3.2 (lists of the component species are in electronic appendix 1, see CDrom). The number of species in each group is related to the mean area occupied by members of the group (figure 3.1a). The slope for this area per species within a group is close to that of the standard species area relations with a z value equal to 0.239 (Crawley and Harral, 2001; Rosenzweig, 1995). The most widespread was the *Cynosurus cristatus* group, comprising species with more-or-less ubiquitous distributions within Great Britain, while the *Glaux maritima* group comprises the species that occur all around the British coastline. These two groups comprised only 20% of native species, in marked contrast to Watson's estimate of 44% of species in his "British" group (Watson, 1847) demonstrating the greater specificity of the method described here.

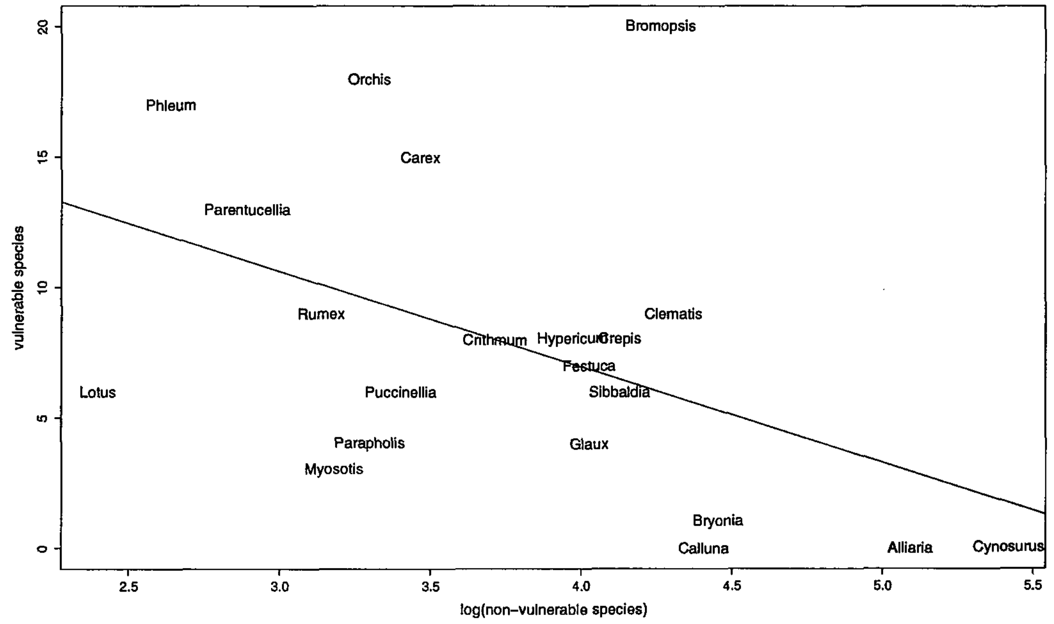
Four groups had predominantly upland distributions (higher than 200m mean altitude; table 3.3). The *Crepis paludosa* group is centred on the English Pennines and covers the Welsh Hills and most of upland Scotland, while the *Festuca vivipara* group picks out the highest ground in Western Britain (Snowdonia in Wales, the Lake District in England, and the Western Highlands of Scotland). The western peaks and plateaux of Scotland support the *Sibbaldia procumbens* group, while the *Phleum alpinum* group is confined to the very highest altitudes in Central and Eastern Scotland (figure 3.2l). Scottish alpine species are often found at lower altitudes in the west than in the east, and in these cases, altitude is not a good explanatory variable for predicting species' spatial distributions (see the strong correlation between temperature and altitude in table 3.2). Intermediate altitudes (100m-200m) are characteristic of three of the groups: the extensive *Calluna vulgaris* group (82 species), the eastern Scottish *Rumex longifolius* group (32 species) and the *Hypericum humifusum* group (61 species) in Wales and southern England. The remaining groups are essentially lowland (lower than 100m mean altitude).

The *Alliaria petiolata* group is all but ubiquitous in British lowlands, though species richness is greatest in central and south-eastern England. Many lowland woodland and meadow species are included in this group, and mean plant height is greater in this group than in any other. There are two large groups with broadly overlapping distributions in lowland south and central England: *Clematis vitalba* group (83 species), *Bryonia dioica* group (97 species). These groups are distinguished on soil nitrogen and soil moisture. The drier, N-poor locations are characteristic of the *Clematis vitalba* group. The moister, more N-rich locations of the *Bryonia dioica* group include many wetland and alluvial species. The *Parentucellia viscosa* group is extreme south-western but predominantly non-maritime, and has species of lowland heaths like *Agrostis curtisii, Erica ciliaris* and *E. vagans*. The driest, most continental parts of England support *Myosotis ramosissima* group and the *Carex elata* group centred on the East Anglian

Fenland.



(a) Species-area relations for the 21 groups of large-scale spatial distribution, showing log number of species in a group as a function of the log of the average area occupied by members of the group (ln(average number of occupied 10km squares); see table 3.2). Positive outliers (groups with more species than predicted on the basis of their geographic extent) are *Orchis laxiflora, Sibbaldia procumbens, Bromopsis erecta, Alliaria petiolata* and *Cynosurus cristatus* groups, while negative outliers are *Lotus subbiflorus, Rumex longifolius, Parapholis strigosa* and *Myosotis ramosissima.*



(b) The number of IUCN-classified vulnerable and endangered species in a group as a function of the log of the number of non-vulnerable species in the group. Positive outliers indicate greater than expected absolute numbers of endangered species, and are seen in the *Bromopsis erecta, Orchis laxiflora, Carex elata* and *Phleum alpinum* groups, while negative outliers are *Lotus subbiflorus, Parapholis strigosa, Myosotis ramosissima, Calluna vulgaris* and *Bryonia dioica* groups (but note that the proportion of rare species is high in *Lotus subbiflorus* and *Parentucellia viscosa* groups).

Figure 3.1: Area and species-richness effects.

Table 3.3: Statistics for the groups.

Group Name	**Members**	**Percent of flora**	**Threatened species**	**Species Area (10km squares)**			**Median altitude(m)**	**January Temperature (^{o}c)**	**July Temperature (^{o}c)**	**Annual Precipitation (mm)**	**Ellenberg values**					**Environmental factors in reduced model**	**R^2 Full**	**R^2 Reduced**
				(min)	**median**	**(max)**					**L (Light)**	**F (Moisture)**	**R (Acidity)**	**N (Nitrogen)**	**S (Salt tolerance)**			
Clematis vitalba	83	5.9	9	(8)	506	(1281)	96	3.5	15.7	765	6.9	5.3	4.8	6.6	0.1	+JulT +JanT +MaxAlt	0.906	0.885
Bryonia dioica	87	6.2	1	(35)	685	(1416)	113	3.4	14.9	876	7.2	8.9	5.9	6.9	0.3	-JanT +JulT -MinAlt	0.836	0.806
Myosotis ramosissima	27	1.9	3	(4)	450	(1219)	106	3.4	15.4	805	7.6	4.1	3.1	5.6	0	+JulT -Rainfall +Steepness	0.513	0.489
Calluna vulgaris	82	5.9	0	(436)	1847.5	(2498)	119	3.4	14.6	908	7.1	7.2	2.5	4.2	0.1	+JanT -JulT +Steepness	0.601	0.585
Hypericum humifusum	61	4.4	8	(1)	559	(1744)	118	3.4	14.6	901	6.8	6.6	3.6	5.1	0.1	+JanT +JulT +Steepness	0.531	0.446
Bromopsis erecta	91	6.5	20	(1)	109	(1002)	106	3.4	15.4	801	6.6	4.4	3.5	7.7	0	+JulT +Lime +Steepness	0.763	0.703
Alliaria petiolata	163	11.6	0	(7)	1549	(2189)	119	3.4	14.6	901	6.5	6	5.3	6.8	0.1	+JulT +Steepness	0.867	0.819
Rumex longifolius	32	2.3	9	(1)	72	(601)	132	3.2	13.9	940	6.7	8.2	3.7	5.5	0.2	-JanT -MaxAlt +Steepness	0.416	0.277
Crepis paludosa	70	5	8	(1)	127.5	(1271)	137	3.3	14.2	961	6.7	5.9	3.6	6.1	0	+JanT -Chalk +Steepness	0.632	0.572
Festuca vivipara	63	4.5	7	(1)	422	(1100)	139	3.3	14	1025	7.5	7.2	2.2	4.5	0	-JulT +Rainfall	0.762	0.726
Sibbaldia procumbens	68	4.9	6	(1)	76.5	(394)	238	2.6	12.8	1242	7.5	6	2.4	5.7	0	-JulT +Steepness	0.823	0.791

Table 3.3 continued.

Group Name	Members	Percent of flora	Threatened species	Species Area (10km squares)			Median altitude(m)	January Temperature (^{o}c)	July Temperature (^{o}c)	Annual Precipitation (mm)	Ellenberg values					Environmental factors in reduced model	R^2 Full	R^2 Reduced
				(min)	median	(max)					L (Light)	F (Moisture)	R (Acidity)	N (Nitrogen)	S (Salt tolerance)			
Phleum alpinum	31	2.2	17	(1)	9	(100)	417	0.8	11.7	1438	7.7	6	2.4	5.2	0	-JulT +MaxAlt	0.687	0.664
Lotus subbiflorus	17	1.2	6	(1)	42	(100)	70	5	16.1	915	7.8	4.8	3.8	5.8	0.6	+JanT +JulT -MaxAlt	0.807	0.791
Parentucellia viscosa	31	2.2	13	(1)	21	(207)	99	4.5	15.7	967	7.1	5.7	3.2	4.6	0.3	+JanT +MinAlt -Sea	0.768	0.716
Crithmum maritimum	49	3.5	8	(1)	190	(857)	83	3.7	15.5	808	8.1	4.6	4.2	7	1.5	+JanT -MinAlt	0.682	0.622
Puccinellia fasciculata	36	2.6	6	(5)	68	(213)	52	4	16.1	723	8.4	6	4.9	7.1	2.9	+JulT -MinAlt +Sea	0.722	0.703
Parapholis strigosa	31	2.2	4	(1)	228	(554)	68	3.8	15.5	822	8.3	7.7	5.4	7.4	5	+JanT -MinAlt	0.598	0.547
Carex elata	47	3.4	15	(1)	33	(294)	74	3.5	15.8	708	7.3	6.7	3.8	6.7	0.1	+Chalk -MaxAlt -Sea	0.447	0.39
Orchis laxiflora	45	3.2	18	(1)	9	(62)	79	4.4	15.9	869	8.3	4.8	2.5	5.9	0.6	+JanT +Lime +MaxAlt	0.315	0.271
Glaux maritima	60	4.3	4	(1)	476.5	(1461)	111	3.4	14.6	911	8.2	6.3	4.2	6.5	2.2	+JulT -MinAlt +Sea	0.607	0.558
Cynosurus cristatus	226	16.1	0	(623)	2509	(2821)	119	3.4	14.6	904	6.6	5.9	4.8	5.9	0.1	+JulT +Steepness	0.614	0.593

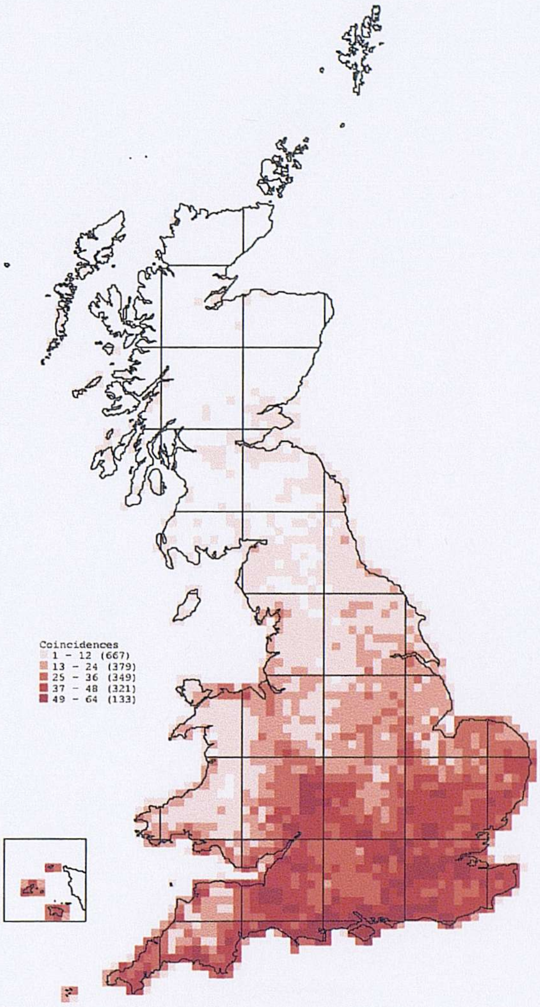
The *Orchis laxiflora* group is confined to the mildest climates of the extreme south and west, centred on the Channel Islands and the Isles of Scilly, while *Lotus subbiflorus* group is found on the coasts of Cornwall and Devon as well as on these islands. These are typically coastal species from sandy or rocky, nutrient-poor soils in regions with high January temperatures. The *Crithmum maritimum* group of coastal species is centred on Wales and the south-west, while the more widespread *Parapholis strigosa* group encompasses species of the English and Welsh coasts. Finally, the *Puccinellia fasciculata* group is restricted to the coasts of Southeast England. Some rare coastal species with distinctive but idiosyncratic distributions do not fall clearly within any of the non-ubiquitous classes. Scotland has no group of coastal species of its own, so Scottish rarities like *Primula scotica* (confined to the north coast of mainland Scotland) and more widespread Scottish coastal species such as *Mertensia maritima* are both subsumed within the *Glaux maritima* group of ubiquitous coastal species.

Few groups were determined directly by underlying geology. The clearest correlation is shown by the *Bromopsis erecta* group which is confined to the chalklands of south-east England. Other lime-loving species (calcicoles) with wider distributions fall in the *Cynosurus cristatus* group (e.g. *Koeleria cristatus* which is chalk-loving in the south, but predominantly coastal in the north), whereas calcicoles with very restricted distributions are placed within the nearest local group (e.g. the rare orchid *Ophrys fuciflora* is placed in the predominantly coastal *Puccinellia fasciculata* group in extreme south-eastern England). Two groups were associated with acidic substrates. The *Calluna vulgaris* group is found throughout Britain, and comprises species of wet heaths on peat soils, and dry heaths on freely draining acid sands. Acid-loving (calcifuge) lowland heath species that are rare or absent in upland Scotland fall into the *Hypericum humifusum* group.

Ellenberg's scores (Hill et al., 2004) averaged over the constituent species are shown in table 3.3. The mean fertility indicator (n) was higher than average for species in the *Bryonia dioica* and *Cynosurus cristatus* groups, and lower than average in *Calluna vulgaris, Myosotis ramosissima, Orchis laxiflora, Bromopsis erecta* and *Hypericum humifusum* groups. Soil moisture was high for *Bryonia dioica* and *Rumex longifolius* groups, and lower than average for *Myosotis ramosissima, Bromopsis erecta, Lotus subbiflorus, Orchis laxiflora* and *Crithmum maritimum* groups. Soil pH was high for *Bromposis erecta* and *Parapholis strigosa* groups, and low in *Festuca vivipara, Calluna vulgaris, Hypericum humifusum* and *Parentucellia viscosa* groups. In some coastal groups the availability of suitable substrate plays a major role, as in the *Puccinellia fasciculata* and *Parapholis strigosa* groups of salt marsh plants. Different coastal groups show clear differences in salinity tolerance: high in *Parapholis strigosa* and *Puccinellia fasciculata* groups, where many of the species are regularly inundated in sea water, much lower in *Crithmum maritimum* and *Orchis laxiflora* groups where none of the species typically suffer immersion although some are occasionally salt-sprayed.

Table 3.4: Wider geographical element membership when compared to British groups found.

Group	**Major biome category**									
	Arctic-montane	**Boreal-arctic montane**	**Wide-boreal**	**Boreo-montane**	**Boreo-temperate**	**Wide-temperate**	**Temperate**	**Southern-temperate**	**Mediterranean-Atlantic**	**Unknown**
Clematis vitalba (a)							42	21	20	
Bryonia dioica (b)			1		16	2	49	16	3	
Myosotis ramosissima (c)					3		16	6		2
Calluna vulgaris (d)		2	2	6	39	1	28	3		1
Hypericum humifusum (e)	1			1	8		32	14	5	
Bromopsis erecta (f)			1	1	5		66	10	8	
Alliaria petiolata (g)					19	3	99	36	4	2
Rumex longifolius (h)	1	1		20	6		2	1	1	
Crepis paludosa (i)	3	4	2	25	24		11		1	
Festuca vivipara (j)	13	6		28	11		5			
Sibbaldia procumbens (k)	37	15		14	1					1
Phleum alpinum (l)	24	3		4						
Lotus subbiflorus (m)							3	4	10	
Parentucellia viscosa (n)		1					14	10	6	
Crithmum maritimum (o)						1	12	15	21	
Puccinellia fasciculata (p)		1			1	1	7	13	13	
Parapholis strigosa (q)				1	3	2	9	13	2	1
Carex elata (r)		1	1		11		21	8	4	1
Orchis laxiflora (s)					1	1	11	16	16	
Glaux maritima (t)	1	4	5	13	8	3	13	13		
Cynosurus cristatus (u)			7		71	22	92	34		



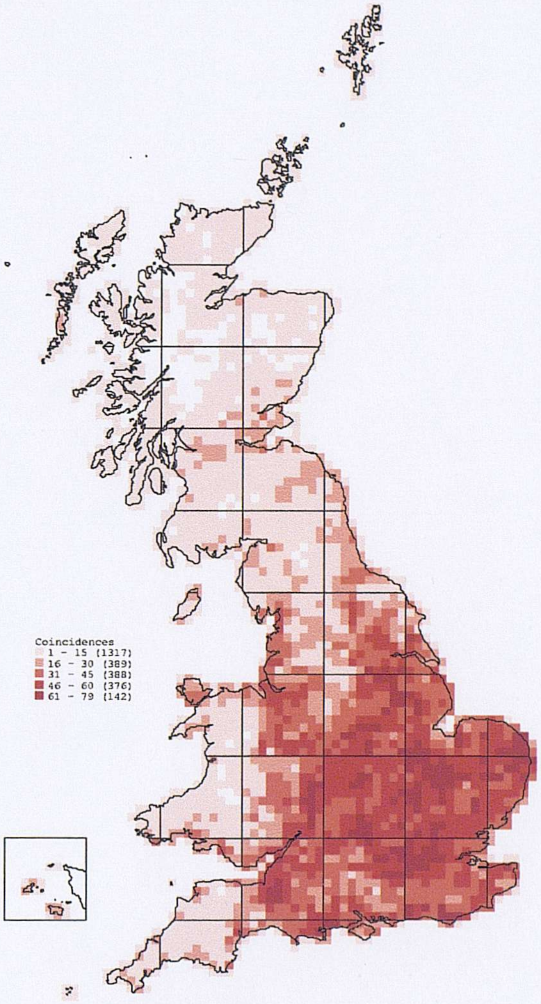
(a) Clematis vitalba

This group of species is concentrated in S.E. England, particularly south of a diagonal line between the Severn and the Wash. In S.W. England the group is widespread but there are 'cold-spots' in the interior, moorland regions; in Wales the group extends around the coastal fringe.

The main habitats of this group of species are grasslands and woodlands, though there are some species of heathland (e.g. *Chamaemelum nobile*), wetland habitats (e.g. *Leersia oryzoides*, *Wolffia arrhiza*) and waterside mud (*Damasonium alisma*, *Pulicaria vulgaris*). There is a relatively high proportion of annual species. The mean Ellenberg F value is relatively low, and the R and N values are high.

Half the members of this group are from the Temperate elements, with the remainder equally divided between the Southern-temperate and Mediterranean-Atlantic elements.

Figure 3.2: Maps showing the distributions of the 21 groups within Britain. Darker colours indicate increasing proportion of the species in a group present in a particular 10km square. Absence of all members of a group in a 10km square is indicated by white space.



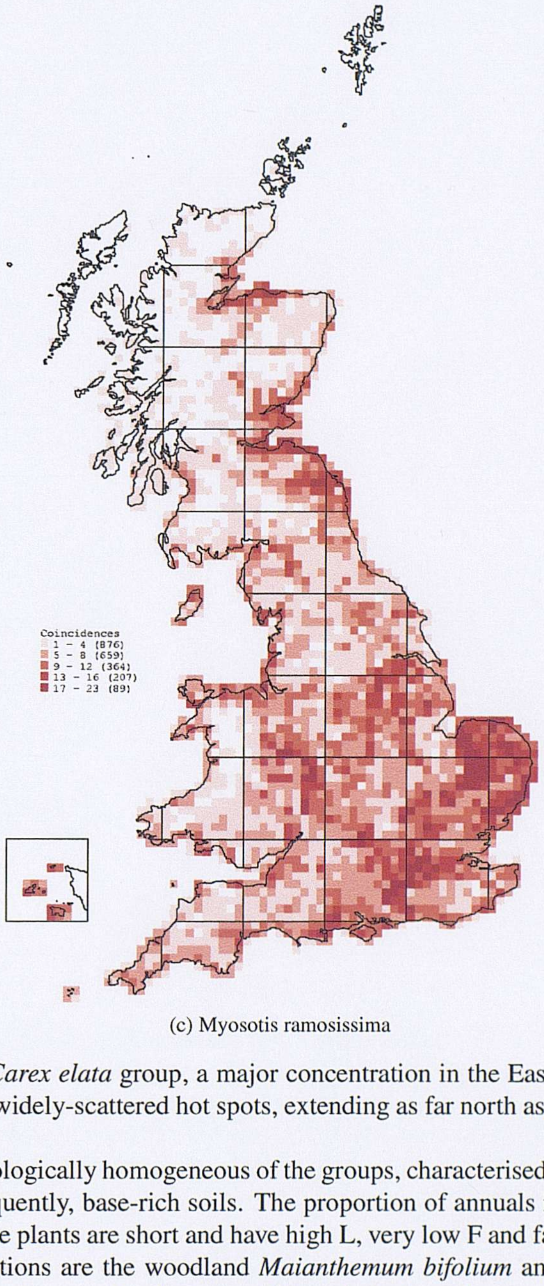
(b) Bryonia dioica

In general the distribution of this group is similar to that of the *Clematis vitalba* group. The two differ in a number of details: the species in the *Bryonia dioica* group are much less frequent in the Channel Islands and S.W. England, the greatest concentrations are in wetland areas such as Broadland, the Cambridgeshire Fenland and the Somerset Levels, and the members of the group extend (at low frequency) much further north into Scotland.

Most species in this group (>70%) are plants of seasonally or permanently flooded habitats or are floating or submerged aquatics. The proportion of perennials is very high, and includes a fairly high proportion of rhizomatous perennials. The proportion of plants which spread by irregular fragmentation (a characteristic of many aquatics) is higher than in any other group. The aquatic nature of the group is also reflected in the highest of all mean Ellenberg F values; N values are also high. There are, however, some plants of terrestrial habitats, including woodland, hedgerows and grassland.

This is another predominantly Temperate group, with sizeable minorities of Boreo-temperate and Southern-temperate members.

Figure 3.2 continued



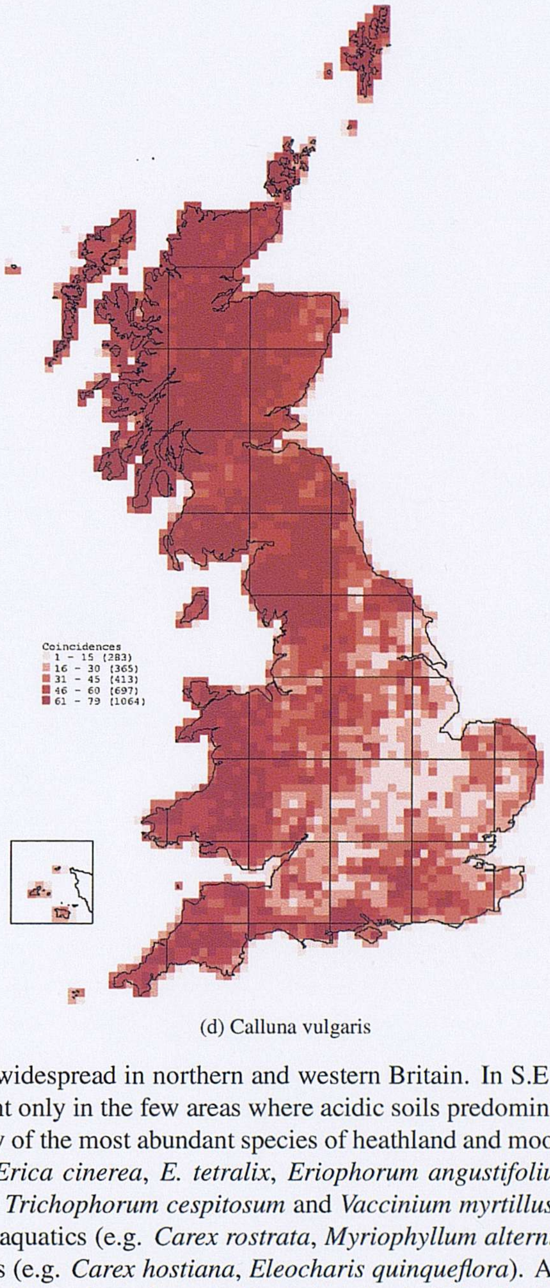
(c) Myosotis ramosissima

This group has, like the *Carex elata* group, a major concentration in the East Anglian Breckland, but it shows a number of other widely-scattered hot spots, extending as far north as the coasts of S.E. Scotland and Morayshire.

This is one of the most ecologically homogeneous of the groups, characterised by species of open habitats on dry, sandy or, less frequently, base-rich soils. The proportion of annuals in the group is greater than that in any other group; the plants are short and have high L, very low F and fairly low N mean Ellenberg values. Ecological exceptions are the woodland *Maianthemum bifolium* and three wetland or aquatic plants, *Epipactis palustris*, *Ranunculus fluitans* and *Scrophularia umbrosa*.

This is yet another group dominated by Temperate species, with smaller Southern-temperate and Boreo-temperate minorities. Mediterranean-Atlantic species are completely absent.

Figure 3.2 continued



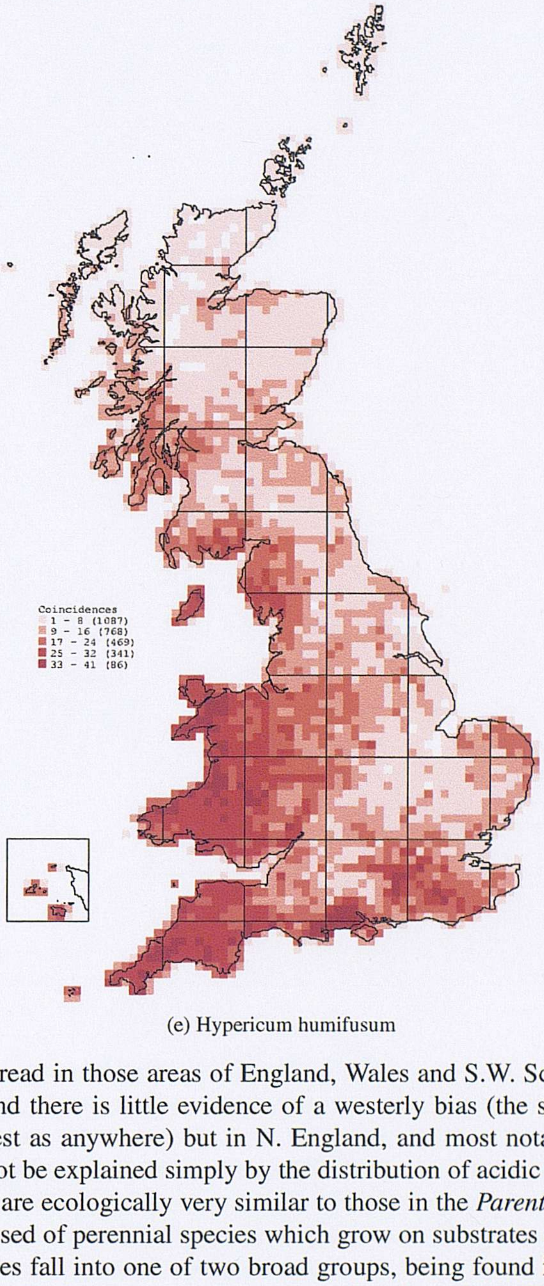
(d) Calluna vulgaris

This group of species is widespread in northern and western Britain. In S.E. England they are sparsely distributed, being frequent only in the few areas where acidic soils predominate.

This group includes many of the most abundant species of heathland and moorland vegetation, including *Calluna vulgaris* itself, *Erica cinerea*, *E. tetralix*, *Eriophorum angustifolium*, *E. vaginatum*, *Molinia caerulea*, *Nardus stricta*, *Trichophorum cespitosum* and *Vaccinium myrtillus*. Most are plants of acidic habitats, including some aquatics (e.g. *Carex rostrata*, *Myriophyllum alterniflorum*) although some are found in base-rich flushes (e.g. *Carex hostiana*, *Eleocharis quinqueflora*). A few are plants of nutrient-poor, calcareous habitats (*Botrychium lunaria*, *Sagina nodosa*). Like all the northern groups, the vast majority of species are perennials. The mean Ellenberg L and F values are unremarkable but the R value is very low and the N value is also low.

The proportion of Boreo-temperate species in this group is higher than that in any other group; the majority of the remaining species have Temperate distributions.

Figure 3.2 continued



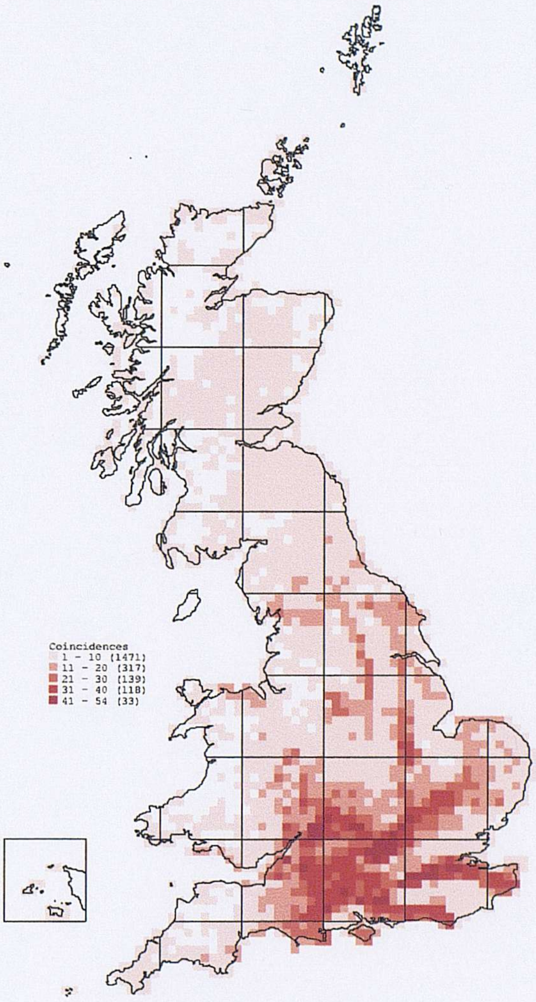
(e) Hypericum humifusum

These species are widespread in those areas of England, Wales and S.W. Scotland with predominantly acidic soils. In S. England there is little evidence of a westerly bias (the species in this group are as frequent in the New Forest as anywhere) but in N. England, and most notably in Scotland, there is a westerly bias which cannot be explained simply by the distribution of acidic soils.

The species in this group are ecologically very similar to those in the *Parentucellia viscosa* group; both are predominantly composed of perennial species which grow on substrates of low pH and low nutrient status. Many of the species fall into one of two broad groups, being found in a range of wetland habitats (e.g. *Eriophorum gracile*, *Luronium natans*) or in rocky places (e.g. *Hymenophyllum tunbrigense, Umbilicus rupestris*). *Ceterach officinarum* and the five *Sorbus* species are notable exceptions to the predominantly calcifuge character of the group.

This is a predominantly Temperate group, with a smaller group of Southern-temperate species and minor Boreo-temperate and Mediterranean-Atlantic components.

Figure 3.2 continued



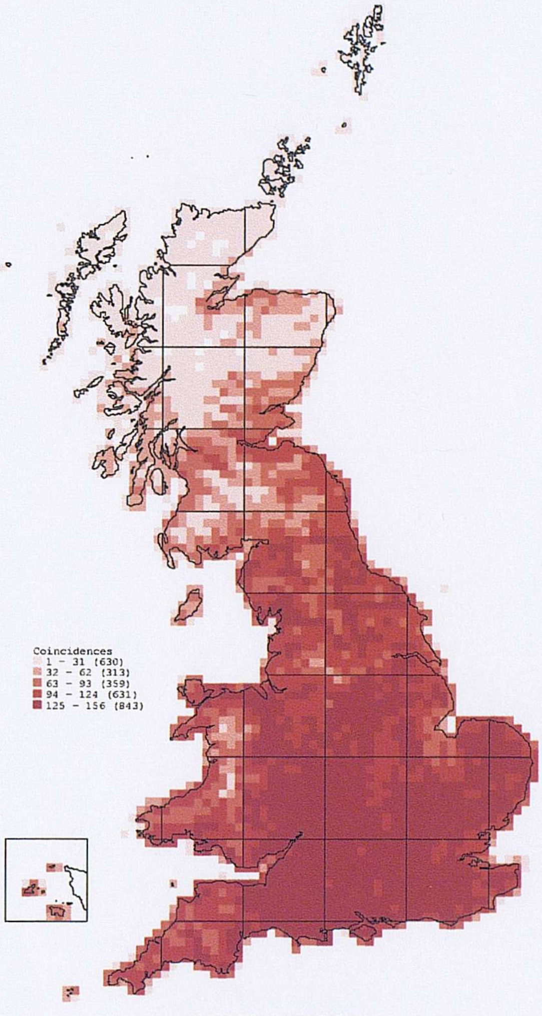
(f) Bromopsis erecta

The distribution of this group picks out the familiar pattern of calcareous rock (chalk, oolitic and magnesian limestone) in England.

Bromopsis erecta is an appropriate label for this group, for most species in it are plants of disturbed base-rich soils, calcareous rocks, grassland, scrub and, to a lesser extent, woodland. There are also a few species characteristic of moister, base-rich grassland on the flood plains of the Thames and other base-rich rivers and even a single aquatic (*Potamogeton nodosus*). The mean Ellenberg F and N values are low, and the R value high.

The proportion of Temperate species in this group is higher than that in any other group, reducing the size of the Boreo-temperate, Southern-temperate and Mediterranean-Atlantic elements to a small minority.

Figure 3.2 continued



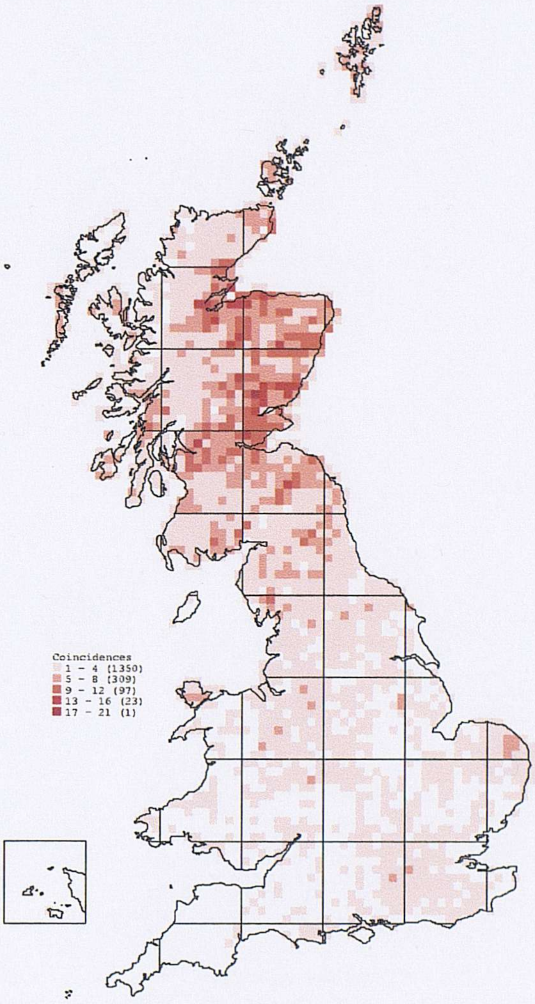
(g) Alliaria petiolata

This group is well-represented throughout lowland England and Wales, thinning out in S.W. England and Wales, especially in upland areas. Other 'cold spots' include the Pennines and the area around the Wash. In Scotland the group is concentrated in the south-east.

This is the second largest of the groups. The 163 species inevitably include plants from a wide range of habitats, including woodlands, hedgerows, grasslands and wetlands. They are predominantly perennials, relatively tall in stature, and with relatively high mean Ellenberg values for R and N.

Despite the differences in British range, the phytogeographical affinities of this group are very similar to those of the *Myosotis ramosissima* group, with the Temperate species dominating the members of the Boreo-temperate and Southern-temperate elements.

Figure 3.2 continued



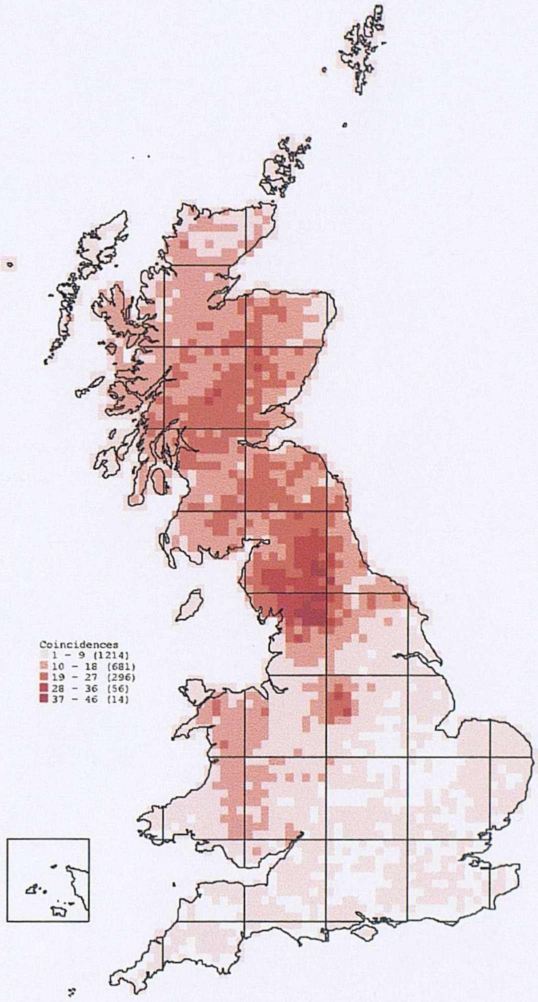
(h) Rumex longifolius

This is a group of plants concentrated in lowland, eastern Scotland.

The majority of species are aquatic plants, found in water or in open lakeside habitats, marshes and swamps. The 10-km square with the greatest concentration of species, NO14, includes a series of species-rich lochs near Blairgowrie. The group also includes species characteristic of Boreal woodland (*Goodyera repens*, *Linnaea borealis*, *Moneses uniflora*) and a few plants of moorland or other types of woodland. *Rumex longifolius* itself is perhaps most frequent as a roadside weed. The perennial species include a large proportion of rhizomatous plants. The mean Ellenberg F value is high (though the spread is large).

This group has a higher proportion of Boreal-montane species than any other; the Boreo-temperate elements are the only substantial minority. The only species of Mediterranean affinities, *Potentilla rupestris*, has a Mediterranean-montane distribution.

Figure 3.2 continued



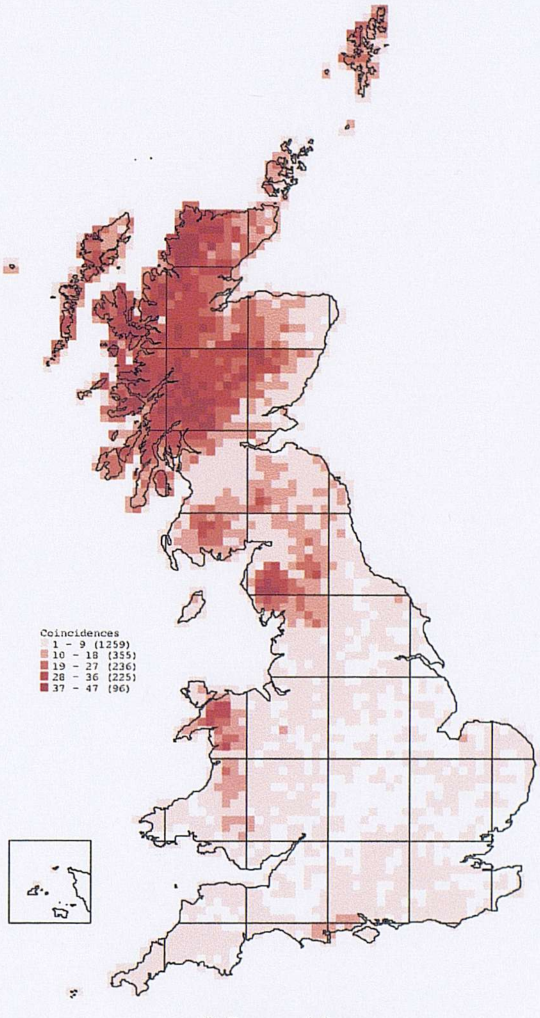
(i) Crepis paludosa

The greatest concentration of these species lies in the Carboniferous limestone massif of N. England, but the species also occur at reduced frequency in upland Wales and in Scotland, although they are notably scarce around the coastal fringes.

Most of these species are plants of open calcareous rocks and scree (including limestone pavement), short dry or moist calcareous turf, hay meadows and open woodland. Two species are often found on sites with high concentrations of heavy metals (*Minuartia verna*, *Thlaspi caerulescens*). In contrast to these species, a minority are calcifuge plants of moorland or bogs (*Andromeda polifolia*, *Carex curta*, *C. magellanica*, *Lyopodium clavatum*, *Vacciniun oxycoccos*, *V. vitis-idaea*) or acidic rocks (*Cryptogramma crispa*). Most of the calcifuge species in this group, despite their habitat requirements, are rare in N. and W. Scotland. The Ellenberg values show little departure from the norm, perhaps because of the contrasting ecology of the two ecological subgroups comprising this group.

Most species in this group have Boreal-montane or Boreo-temperate distributions, with a minority from the Arctic-montane, Boreo-arctic Montane, Wide-boreal and Temperate elements, and one Mediterranean-montane species (*Dryopteris submontana*).

Figure 3.2 continued



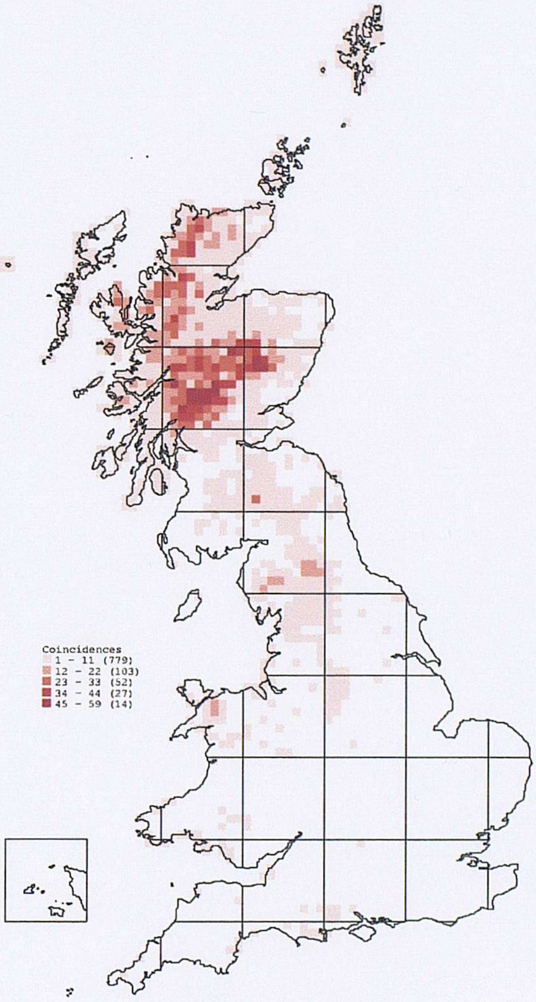
(j) Festuca vivipara

A group of northern and western species, concentrated in N. Wales, the Lake District and northern and western Scotland. These species are concentrated in the same areas as the *Calluna vulgaris* group, but their distributions are much more restricted.

This group including plants which grow in a wide range of habitats, including standing waters, swamps, moorlands, rock outcrops and montane communities. Most are calcifuge but there are some calcicoles (e.g. *Asplenium viride*, *Eriophorum latifolium*). They tend to be short and have a high mean Ellenberg L value, a relatively high F value, a very low R value and a low N value.

This is another group of species with predominantly Boreal-montane distributions, although there are also Arctic-montane, Boreo-arctic montane, Boreo-temperate and even Temperate species included in it.

Figure 3.2 continued



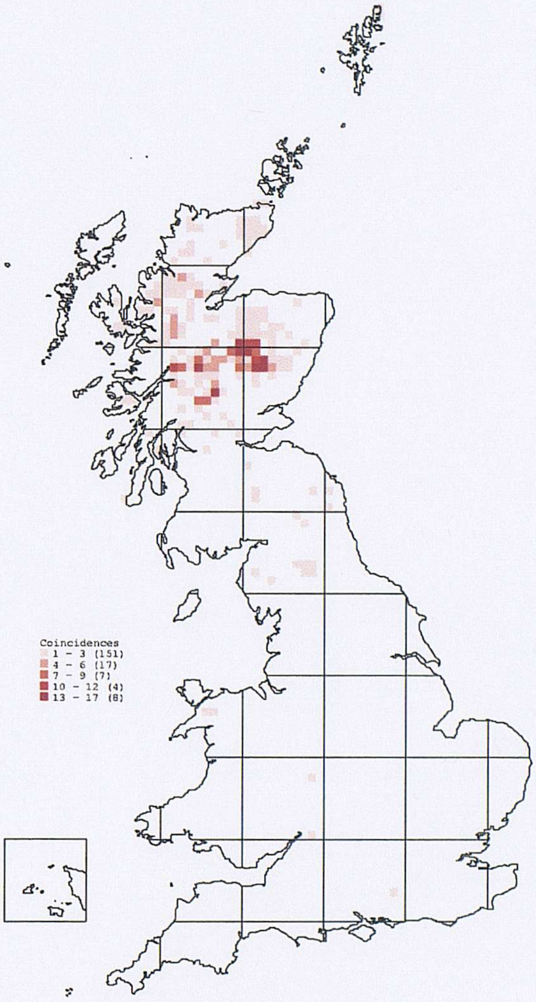
(k) Sibbaldia procumbens

The first of two montane groups, with a median altitude of 238 m. These species have three main concentrations, in the Breadalbane mountains, the Cairngorms and Caenlochan, but lesser concentrations elsewhere in the N.W. Highlands, in N. England and N. Wales.

The great majority of species are plants of montane habitats, including exposed heathland and grassland, flushes, stony habitats, rock outcrops and areas of late snow-lie. There are also species of moorland (e.g. *Rubus chamaemorus*) and even the woodland dominant *Pinus sylvestris*. Some species are calcicoles, others calcifuges and others indifferent to soil reaction. The group is characterised by plants of low stature, high mean Ellenberg L values and low F and N values; the mean R value is middling.

Over half the species in this group have Arctic-montane distributions; the remainder are fairly evenly split between the Boreo-arctic Montane and Boreal-montane elements.

Figure 3.2 continued



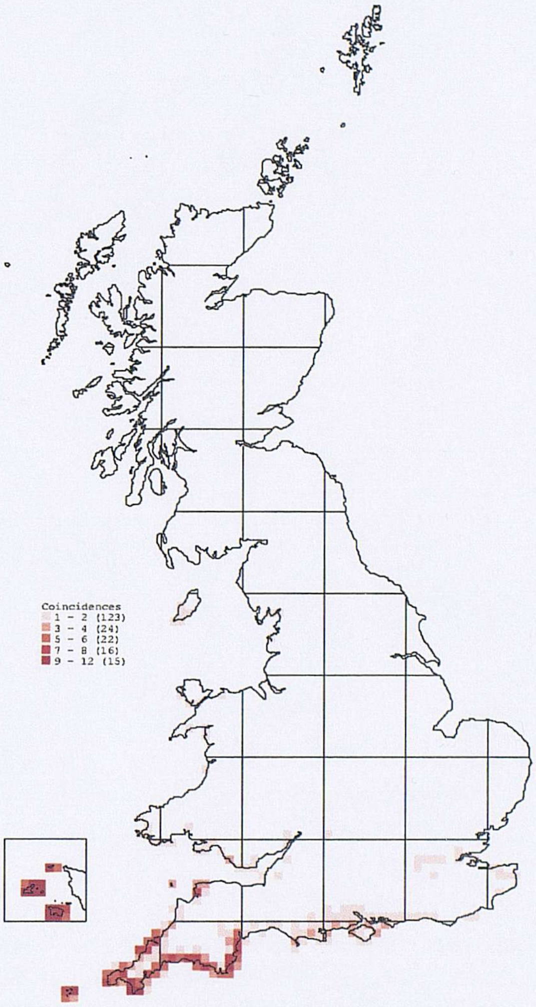
(l) Phleum alpinum

This is the second montane group, comprising species with much more restricted distributions than those of the *Sibbaldia procumbens* group. They are almost absent from England and Wales, and in Scotland there are two major concentrations of species. These are in the Eastern Highlands, in the same areas as the two easterly hotspots of the *Sibbaldia procumbens* group, the Cairngorms and Caenlochan. There are scattered hotspots of this group to the south-west, but very few in N.W. Scotland.

The plants in this group grow in similar montane habitats to those of the *Sibbaldia procumbens* group. They also have very similar ecological attributes, the only real difference being a slightly lower mean R value.

Phytogeographically, this is the most specialised of all groups. Over 75% have Arctic-montane distributions, with smaller Boreo-arctic Montane and Boreal-montane components.

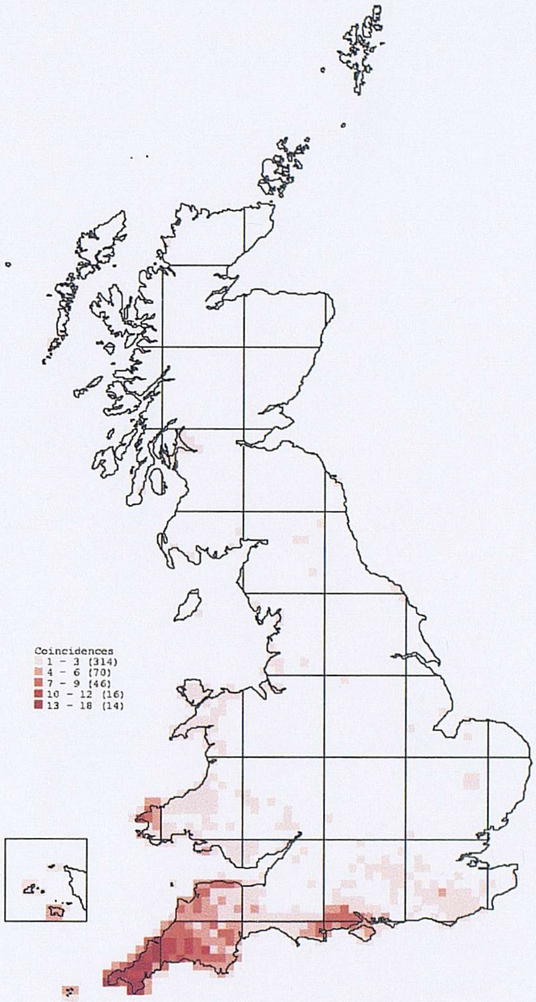
Figure 3.2 continued



(m) Lotus subbiflorus

This group is concentrated in the Channel Islands, Isles of Scilly, Cornwall and S. Devon, extending eastwards to Kent and northwards along the west coast to the Isle of Man and Westmorland.
These species grow in a range of coastal habitats, including shingle and sand-dunes, cliffs and (*Arum italicum*) coastal woodland. They occupy a variety of microhabitats from summer-droughted sites to areas with freshwater seepage; one species, *Cyperus longus*, is found by streams near the sea. The group contains an almost equal number of annuals and perennials and lacks rhizomatous perennials; the mean Ellenberg L value is high.
Uniquely, over half the species in this group have broadly Mediterranean-Atlantic distributions; the remainder are composed of Southern-temperate and Temperate species.

Figure 3.2 continued



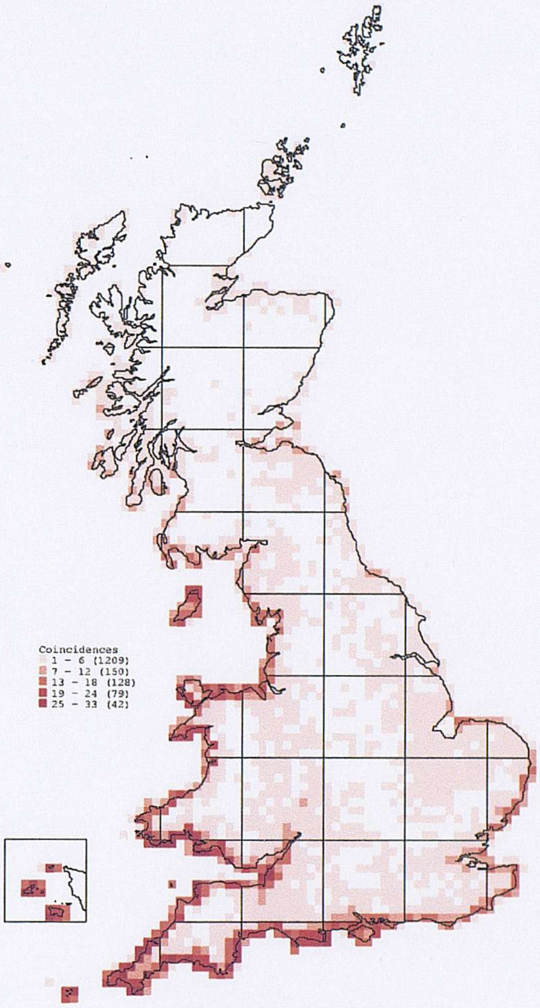
(n) Parentucellia viscosa

These species are concentrated in S.W England, with subsidiary hot-spots in Dorset/S. Hampshire and S.W. Wales. The group differs from the *Lotus subbiflorus* group, which is also predominantly south-western, in being poorly represented in the Channel Islands, Isles of Scilly and the coastal fringe of S.W. Britain. One of the few areas where both these groups are strongly represented is the Lizard peninsula in W. Cornwall.

This relatively small group is ecologically diverse. It includes a number of heathland plants, both community dominants (*Agrostis curtisii*, *Erica ciliaris*, *E. vagans*) and rare species which occupy more specific niches within heathland habitat complexes (e.g. *Cicendia filiformis*, *Euphrasia vigursii*). Other species are found in coastal habitats (from fore-dunes to coastal woodland), moist acidic habitats, inland field margins, hedgebanks and woodland. Almost 80% of the species in this group are perennials, although few of these have far-creeping rhizomes. The Ellenberg L values are lower than those of the coastal groups and of course salinity values are usually zero. The *Parentucellia viscosa* group has a notably low Ellenberg mean R value, and a lower mean N value than many southern groups.

The largest contribution to this group comes from the Temperate element, with a substantial minority of Southern-temperate species and a smaller group of Mediterranean-Atlantic plants. Seven of the 31 species are endemic to Britain (6) or the British Isles (1), and several others have restricted distributions in the Atlantic zone of Europe.

Figure 3.2 continued

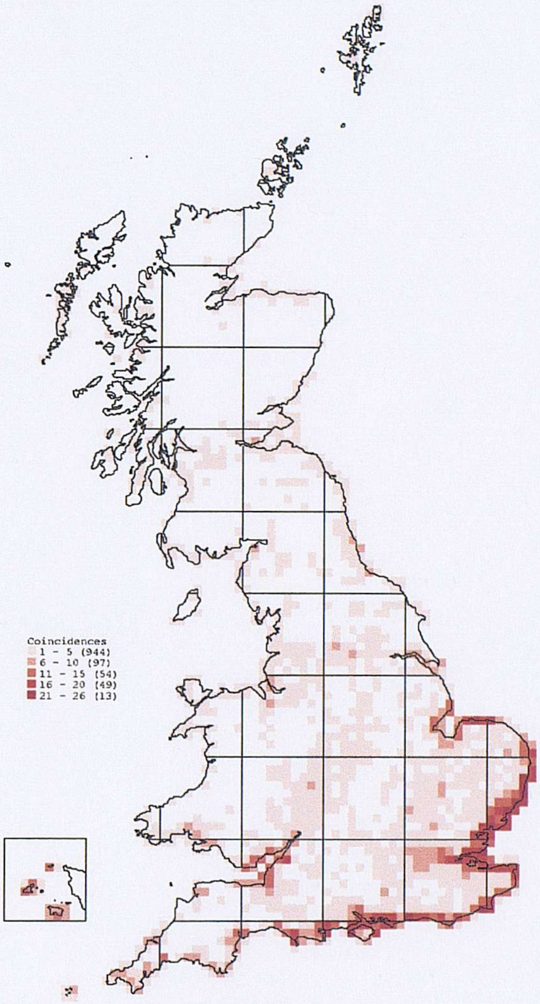


(o) Crithmum maritimum

Members of this group are widely distributed round the coasts of England, Wales and S.W. Scotland; the greatest concentrations are in the Channel Islands, S. and S.W. England and Wales.

Most species are plants of coastal habitats, including open communities on shingle, cliffs and sand dunes and, to a lesser extent, scrub and woodland (e.g. *Lithospermum purpureocaeruleum*, *Orobanche hederae*, *Rubia peregrina*). Some require moist or flushed conditions (e.g. *Carex punctata*, *Samolus valerandi*). Also included in this group are the less maritime species concentrated on open Carboniferous limestone sites in the Mendips (*Helianthemum apenninum*, *Koeleria vallesiana*, *Sorbus eminens*, *Trinia glauca*) and the weedy western *Fumaria* species, *F. bastardii*, *F. capreolata* and *F. purpurea*. The proportion of annuals in this group, although substantial, is less than in the *Orchis laxiflora* and *Lotus subbiflorus* groups. The mean Ellenberg L, F and N values are similar to those of the *L. subbiflorus* group but the *Crithmum maritimum* group has a higher R value, reflecting the occurrence of many species on limestone. The phytogeographical composition of this group is almost identical to that of the *Orchis laxiflora* group.

Figure 3.2 continued



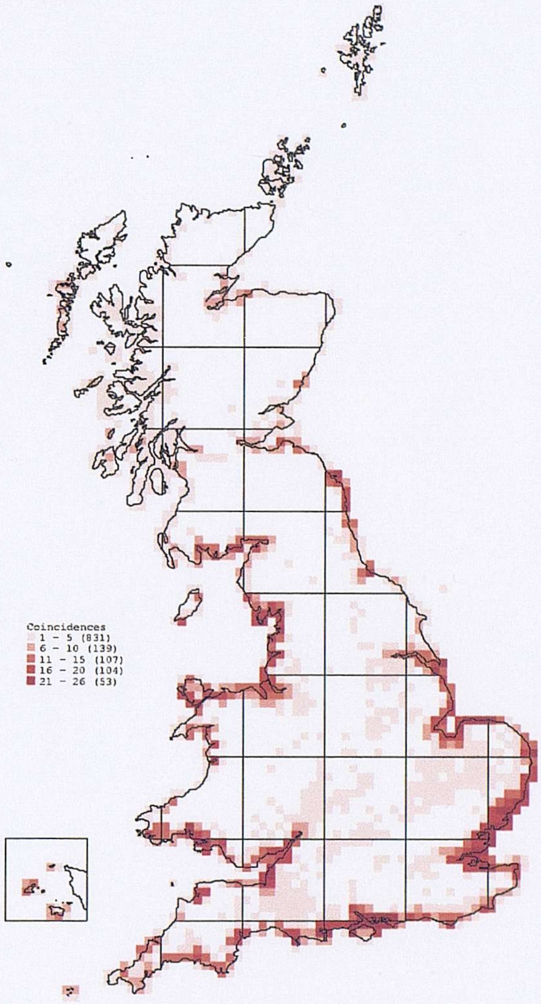
(p) Puccinellia fasciculata

This is the most south-eastern of the coastal groups, with concentrations limited to S.E. England from the Solent to the Wash; the group is especially well represented in Kent and East Anglia.

The habitats of these species are similar to those of the *Parapholis strigosa* group, including brackish water, shingle, sand dunes, parched grassland, ditches in grazing marshes and sea walls. There are two plants of chalk cliffs (*Orobanche fuciflora*, *Orobanche artemisiae-campestris*). None of the more successful colonists of salted roadsides are included in this group. The life-form spectrum of this group is similar to that of the *Parapholis strigosa* group, although there are fewer perennials with far-creeping rhizomes. Mean Ellenberg F, R, S and N values are slightly lower than those of the *P. strigosa* group.

This group, like the restricted southern and western coastal groups but unlike the more wide-ranging *Glaux maritima* and *Parapholis strigosa* coastal groups, has a high representation of Mediterranean-Atlantic species. The total representation of Southern-temperate and Mediterranean-Atlantic species is as high as that of the *Orchis laxiflora* and *Crithmum maritimum* group, and exceeded only by the *Lotus subbiflorus* group. Most of the remaining species have Temperate ranges.

Figure 3.2 continued



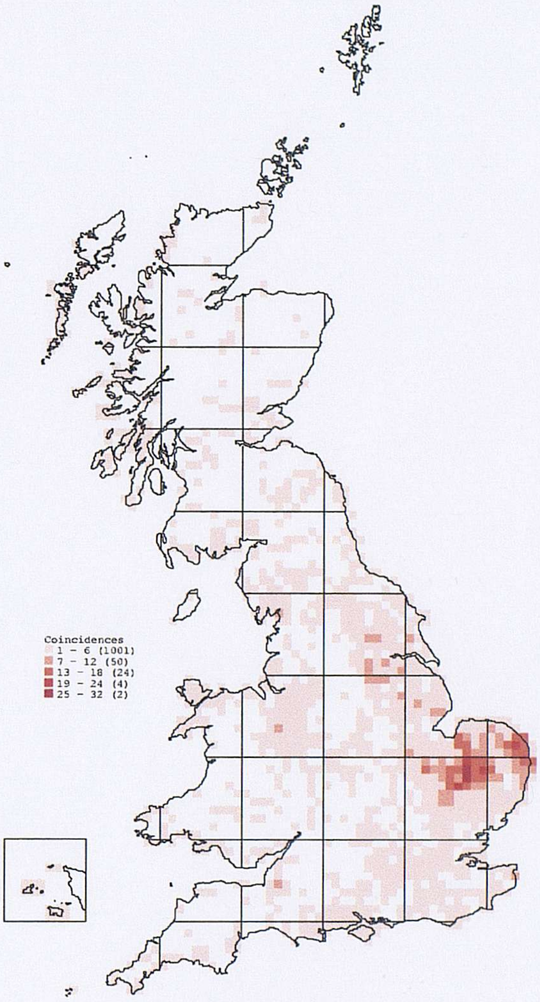
(q) Parapholis strigosa

This is in some ways a mirror image of the *Glaux maritima* group, as the species are concentrated in coastal squares in S.E. England, from Hampshire to Lincolnshire. The two groups overlap in the Bristol Channel, Wales and N. England but members of the *Parapholis strigosa* group are rare in Scotland, where the *Glaux* group is so well represented.

The habitats of the species in this group are those of the southern coastline: brackish water, sand dunes, salt marshes and coastal ditches. Two species, *Atriplex littoralis* and *Puccinellia distans*, have successfully colonised salted roadsides in recent decades. The only non-coastal species are three rarities, *Carex flava*, *Scorzonera humilis* and *Sorbus domestica*. The life-form spectrum of this group is very similar to that of the *Glaux maritima* group. The *Parapholis strigosa* group has a high mean Ellenberg L value; the F, R, N and S values are also high.

Unlike the *Glaux maritima* group, this group is dominated by Temperate and Southern-temperate species; the more southern Mediterranean-Atlantic group and the more northerly Wide-temperate, Boreo-temperate and Boreal-montane groups are represented by a few species.

Figure 3.2 continued



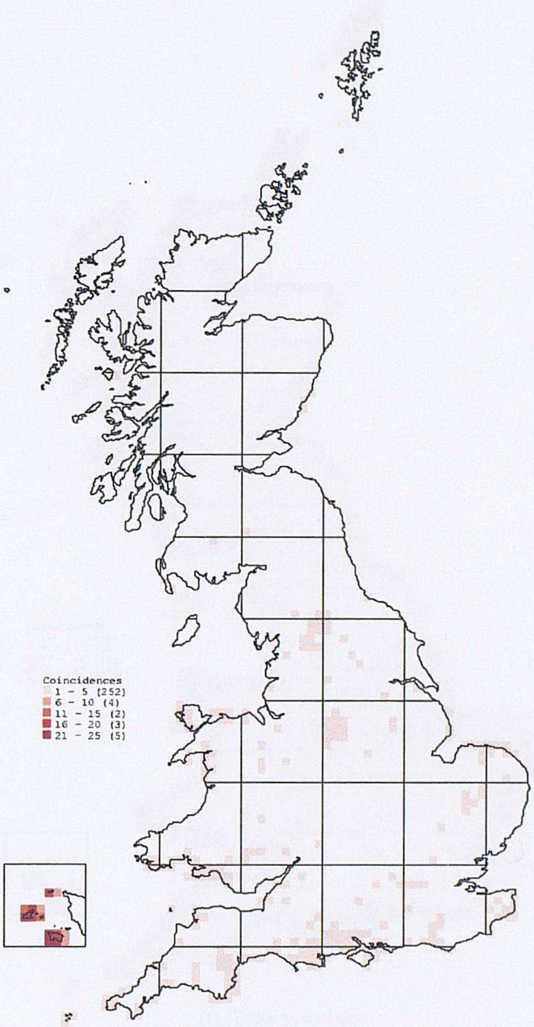
(r) Carex elata

This is an East Anglian group, with two concentrations of species, one centered on Breckland (which has, or formerly had, some rich wetlands as well as the more familiar dry, open habitats) and the other centered on Broadland. There are minor hotspots scattered elsewhere, notably in Lincolnshire and Yorkshire.

The species in the *Carex elata* group fall into two groups. One consists of plants of dry, calcareous or acidic grassland or disturbed soils. The other is a group of aquatic plants, swamps and other wetland habitats, mainly but not exclusively associated with calcareous waters. Two woodland species (*Primula elatior*, *Pulmonaria obscura*) are also included in the group. The plants in the group are mainly perennial; the wide differences in the ecological requirements of the component species is reflected in the spread of Ellenberg F and N values.

This is a predominantly Temperate group. The smaller Southern-temperate and Mediterranean-Atlantic components are outnumbered by the members of the Boreo-temperate element, which is better represented in the *Carex elata* group than any other group which is concentrated in England.

Figure 3.2 continued



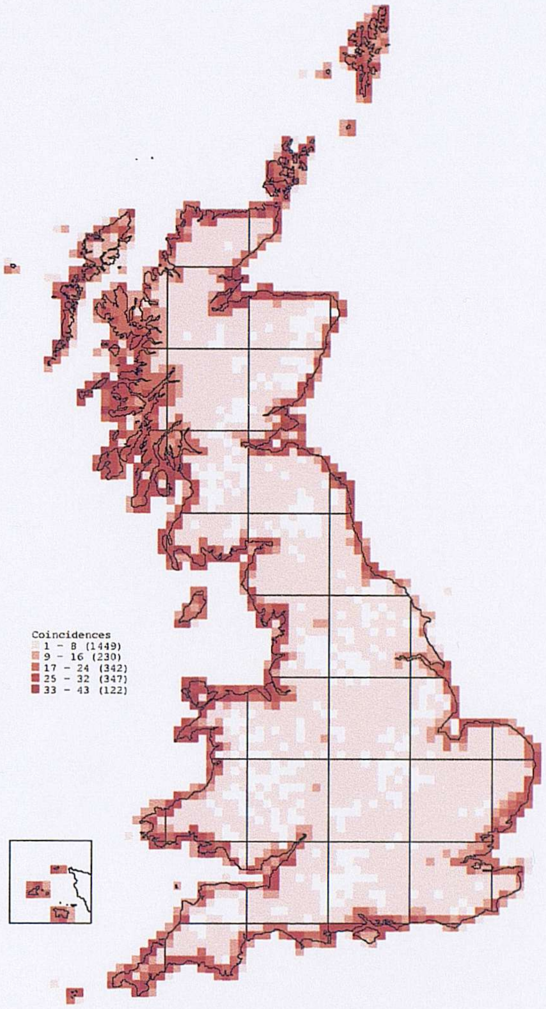
(s) Orchis laxiflora

This group is centered on the Channel Islands; all 43 species occur there and within the study area no fewer than 13 are confined to them as natives, with a further 3 found only there and in the Isles of Scilly. Unlike the *Lotus subbiflorus* group, these species are poorly represented in mainland Cornwall (except on the Lizard peninsula) but there are minor concentrations in S. and S.E. England, notably in the New Forest (where *Galium constrictum* and *Ludwigia palustris* have their main British populations), and in coastal Wales.

Most species grow in coastal habitats, including sand dunes, dune slacks and seasonally droughted sites on cliffs; some also occur inland in similar sandy or rocky habitats. The group also includes a minority of plants of other seasonally or permanently moist habitats, including *Orchis laxiflora* itself, *Galium constrictum*, *Ludwigia palustris*, *Ranunculus ophioglossifolius*, *R. paludosus* and *Spiranthes aestivalis*. The ecological profile of these species is similar to that of the *Lotus subbiflorus* group; note also the low stature of the plants, and their low mean Ellenberg N and F values.

This is one of seven groups with at least half the members with Southern-temperate or Mediterranean-Atlantic distributions; most of the others are Temperate species with only single representatives of the Boreo-temperate elements (*Pilosella peleteriana*) and Wide-temperate elements (*Spergula arvensis*).

Figure 3.2 continued



(t) Glaux maritima

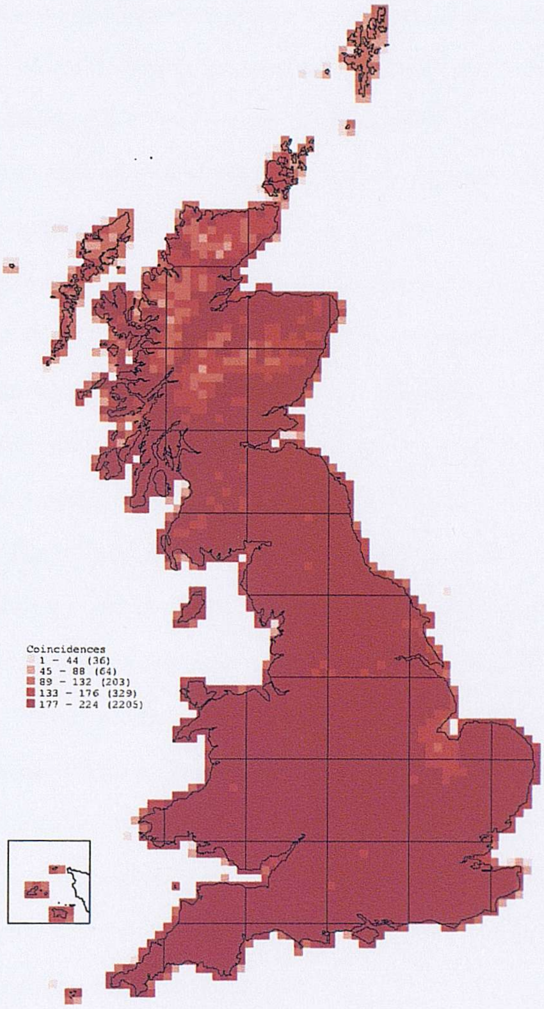
This is a group of widespread, coastal species, occurring all round the British coast and in the Channel Islands but perhaps more frequent in Scotland, N. England and N. Wales than elsewhere.

Almost all the species in this relatively large group are plants of coastal habitats, including sea water (*Zostera marina*), strand-lines, dunes, salt marshes, rocky cliffs and cliff-top grassland. The group includes *Cochlearia danica* and *Spergularia marina*, two of the most successful colonists of salted roadsides in recent decades. In addition to the coastal species, there is a number of rare or scarce species which also happen to grow in sites close to the sea but not in coastal habitats, especially in northern and western Scotland (*Calamagrostis scotica*, *Cerastium nigrescens*, *Dactylorhiza majalis*, *Geranium sanguineum*, *Orobanche alba*, *Schoenus nigricans*, *Spiranthes romanzoffiana*).

Most members of this group are perennials, although there is a substantial minority of annuals; unlike the three preceding southern and south-western coastal groups, the perennials contain an appreciable proportion of plants with far-creeping rhizomes. Mean Ellenberg L and N values are similar to those in the *Crithmum maritimum* group, but F and S values are higher.

In sharp contrast to the other coastal groups, Mediterranean-Atlantic species are absent from this group and even Temperate and Southern-temperate plants account for less than half the species. Over half the species occur in the Boreal zone, in a range of elements from the Boreo-arctic Montane to the Wide-temperate.

Figure 3.2 continued



(u) Cynosurus cristatus

This group is widespread and well-represented throughout most of Britain, thinning out only in the species-poor area around the Wash and in the Highlands and Islands of Scotland.

This is the largest group, and like the *Alliaria petiolata* group it includes plants of a wide range of habitats. The mean Ellenberg values are similar to those of the *A. petiolata* group, although the mean R value is lower.

The Temperate species are the largest of the phytogeographical elements in this group. The Boreo-temperate species contribute a substantial proportion of the remainder, and are almost twice as numerous as the Southern-temperate representatives. The proportion of species belonging to the Wide-temperate element, plants which are particularly widespread in the northern hemisphere, is higher in this than in any other group.

Figure 3.2 continued

Threatened species (those classified by IUCN criteria as Vulnerable, Endangered, Critically Endangered or Extinct) are non-randomly distributed over the 21 groups (figure 3.1b). Naturally, there is a negative correlation between the area occupied by a group and the number of vulnerable and endangered species it contains, since vulnerability is defined on the basis of restricted distribution. As expected, therefore, the two most widespread groups contain no vulnerable or endangered species at all (*Cynosurus cristatus* and *Alliaria petiolata* groups). Clearly, however, some groups have high absolute numbers of rare species (*Bromopsis erecta, Orchis laxiflora, Phleum alpinum* and *Carex elata* groups; figure 3.1b) and others have high relative numbers of rarities (*Parentucellia viscosa, Lotus subbiflorus* and *Rumex longifolius* groups; table 3.3).

Concentrations of plant rarities can be associated with unusual geology (e.g. serpentine rocks close to sea level on the Lizard Peninsula of Cornwall, or calcareous mica-schists at high altitude on Ben Lawers in the Breadalbane region of Perthshire). The Lizard rarities are divided between the *Orchis laxiflora, Parentucellia viscosa* and *Lotus subbiflorus* groups, whereas the majority of the Ben Lawers rarities are in the *Sibbaldia procumbens* group (31 species) and the remaining 6 species are in the *Phleum alpinum* group. The New Forest rarities are mostly in the *Parentucellia viscosa* group, but the Norfolk rarities (a mix of Breckland (dry) and fen (wetland) species) have their own *Carex elata* group.

Membership of wider European geographic elements (Preston and Hill, 1997) were only loosely correlated with the distributional groups. The high altitude *Phleum alpinum, Sibbaldia procumbens, Crepis paludosa* and *Festuca vivipara* groups showed the closest match, with good representation of Arctic-montane, Boreal and Boreal montane elements. Likewise, the extreme south-western *Lotus subbiflorus, Crithmum maritium, Puccinellia maritima,* and *Orchis laxiflora* groups showed disproportionate representation of Mediterranean-Atlantic species. None of the other groups, however, showed greater than 50% overlap with any of the elements (table 3.4).

3.5 Discussion

It is important to emphasise that geographic rather than ecological groupings have been defined, and that ecological factors were not used in constructing the groups. Presumably, the groups arise from broad similarities in the way that the fundamental niches of the associated species interact with climate and soil to determine persistence (Crawley, 1997*a*; Hutchinson, 1957). The extent to which group membership is influenced by competition between species, and by interactions with mutualists and natural enemies cannot be resolved with data at this scale (Gaston and Blackburn, 2000). The groups should be thought of as convenient reference points that describe easily-recognisable clusters within a multi-dimensional distributional continuum (Birks and Deacon, 1973). They should not be regarded as corresponding to real discontinuities; rather they are the colours in a distributional spectrum. With this classification the species fall into two sets, those that belong strongly to a given group and those with a more loosely-fitting association as demonstrated by their second- and third-best fit values on the CD as supplementary table 1. Further study will show whether this partitioning remains when the entire European range for these species is taken into account or if this is an effect either of scale or of constraining the analysis to Britain.

Group membership affords a datum for testing a broad range of hypotheses about biogeography. For example, do species become less abundant towards the edge of their range? Do species form self-replacing populations in the core of their distribution but are casual (dependent on immigration) at the edge of their range?

There is no doubt that spatial patterns were dynamic in the past (Godwin, 1975) and are likely to be so in the future. Plants sharing the same contemporary geographic range may have evolved sympatrically or allopatrically and so it is not inevitable that plants sharing the same geographic range today will share ranges in future: historically, different species moved at different rates (Gleason, 1926) and hence would have been included in different spatial groupings at different times (Davis, 1987). The method is applicable to the classification of the distribution of any uniformly recorded binary data set, be it fine-scale local plant distributions or broad-scale continental biological, ecological or geographical patterns. This novel classification of the British flora provides a detailed reference point from which future changes in distribution can be measured and offers an insight into the current relationships between the species of the twenty-one floristic groups identified.

Chapter 4

Patterns in European phytogeographic diversity

4.1 Summary

A two-stage cluster analysis is performed on the data published in the *Atlas Florae Europeae* to produce a new and extensive phytogeographical classification for the European continent. The classification shows 18 groupings. When species density (from the raw data) is compared with the phytogeographical groups it is found that while the two are linked there are interesting areas of difference. These poorly correlated areas highlight regions of either enhanced or reduced secondary (beta) diversity. Areas of low diversity are especially apparent in the European interior and those of high diversity border the Mediterranean Sea. These are new groupings and the relative differences in diversity give fresh insights into the continental-scale plant communities and will provide a tool to aid the study and preservation of Europe's biodiversity.

4.2 Introduction

4.2.1 Objectives

I set out to find which areas of Europe have the greatest phytogeographical diversity, that is, the greatest secondary diversity. To answer this question a modern, detailed, large-scale, high-resolution phytogeographical classification for Europe was created.

4.2.2 Background

The objective of descriptive biogeography has always been to investigate the patterns of occurrence found in the biota of a study area (Marquez et al., 2001). Of particular significance are those patterns that are repeated across taxa (Jardine, 1972). This leads to the desire to group taxa according to these patterns thus producing phytogeographical distributions.

The beginning of the use of computational methods to derive phytogeographical classifications occurs with the more general availability of computing power to the scientific researcher in the early 1970s. Jardine's review (1972) seems to have sparked renewed interest in the subject and brought to wider attention the possibility of using computers in this way. Previous workers such as Watson (1847) and Matthews (1937) compiled their classifications by hand from detailed knowledge and their own data.

Computational methods help resolve some of the difficulties in compiling such classifications. They allow greater amounts of data to be processed so permit greater resolution and numbers of species to be considered. They also have the effect of reducing unintentional human bias. Indeed, while humans are excellent at determining patterns that have strong edges they are bad at discerning disjointed patterns and have a tendency to be biased by prior knowledge of the ecology, taxonomy or history of a species being considered (Jardine, 1972; Birks, 1976).

Simultaneous to the rise in computing power, the formation of the Atlas Flora Europeae committee allowed the creation of a continent-wide project to map all the vascular plant species in a consistent manner. The slow progress of this project and the great demands placed upon computer systems by multivariate methods has limited the number of studies on continent-scale patterns. These difficulties have lead to previous workers either restricting their studies geographically (Birks, 1973; Birks and Deacon, 1973; Marquez et al., 2001) or taxonomically (Birks, 1976; Myklestad and Birks, 1993). While these studies allow a picture to be formed for a country or a group they do not allow any overview of the patterns of plant distribution in Europe to be formed. The work here should go some way to remedying this deficiency.

4.3 Methods

4.3.1 Data

Data for this analysis were based on the first twelve books (*Psilotaceae* to *Platanaceae*) of the *Atlas Florae Europeae* (AFE) (Jalas and Suominen, 1972-1999) and comprises around a fifth of all taxa within the European vascular flora. The AFE provides the only published, large-scale and consistently-mapped dataset for plants in Europe. The data takes the form of presence or absence for each species recorded for every 50 x 50km square in the AFE grid. All native records were used, including extinctions but excluding doubtful and alien records (as defined within the AFE). Known extinctions were included because these very few records are actually of presence: to record an extinction not only does a species have to be recorded as present it also, at a later stage, has to have been recorded as absent. Not all squares in Europe will have had sufficient effort expended upon them that an extinction could have been recorded. To prevent inconsistencies caused by unrecorded extinctions these were included as presences. All hybrids were excluded. Records for subspecies were merged into a consolidated distribution for the species. Where species were difficult to distinguish and an aggregate is recorded by the AFE, all species that fell into this aggregate were consolidated and only the aggregate distribution was used.

4.3.2 Creating the phytogeographical groups

A novel combination of well-proven cluster analysis techniques was employed. This method has been described and discussed in chapter 2. Briefly, the data were converted into a presence/absence matrix for each species recorded in each grid square in the AFE. The primary classification groups were created using a hierarchical clustering algorithm to generate cluster centres. This was followed by an error-correcting partitioning algorithm as a second step.

4.3.3 Analysis of the groups

The correlation between the species richness of a square and phytogeographical richness was calculated using a standard regression technique between the species and group counts for each square. A group was considered present within a square if one or more of its constituent species occurred in that square.

4.4 Results

4.4.1 European groups

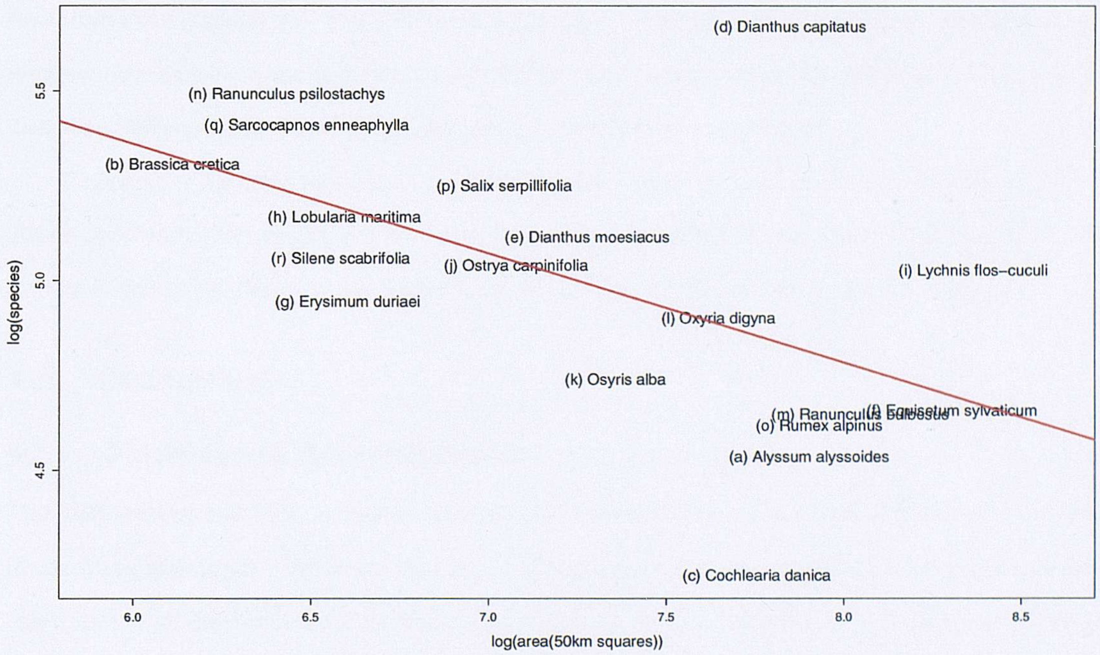


Figure 4.1: The relationship between group range and species richness. On the continental scale widespread species are infrequent so groups with wide ranges have many fewer species than those with more restricted ranges. This is shown as a decline in group membership as range increases in this figure. The trend-line, shown in red, has a gradient of -0.29.

The 18 distinct groups that result from the cluster analysis technique are illustrated in figure 4.2. These groups have been named after the species that conforms most closely to the overall group distribution. There is a declining relationship between the number of species contained in a group and the group's range (figure 4.1). This shows that the relationship is not a standard species area relationship as the behaviour exhibited is the complete opposite of that which would be expected. This is an effect of groups being non-exclusive in an area and of many of the species rich groups being geographically restricted.

4.4.2 Species richness and group density

Figure 4.3 shows the species richness and the group density in the AFE 50km squares. The species richness map, despite the taxonomic simplification performed, still follows the overall species density map as published by the Finnish Museum of Natural History. However, the density map produced from the new groups published here exhibits a more southerly bias and a relative reduction in the number of groups in the central European region.

4.4.3 Regression analysis

Unsurprisingly, group density shows a strong correlation with species richness (figure 4.4). The regression analysis gives an intercept of 3.10 ($p < 2 \times 10^{-16}$) and a gradient of 0.03 ($p < 2 \times 10^{-16}$) (R^2=0.63). This is shown as the black line in figure 4.4. Further simplification of this model (by, for example, removing the intercept term) is not statistically justified ($p < 2 \times 10^{-16}$). Although the data may appear to be following a power law-type relationship ($\text{groups} = \text{species}^a$) fitting a quadratic model shows no improvement on the simple linear model (R^2=0.67). Use of a polynomial smoother (LOWESS) with this data also indicates a linear non-polynomial model, the blue line in figure 4.4.

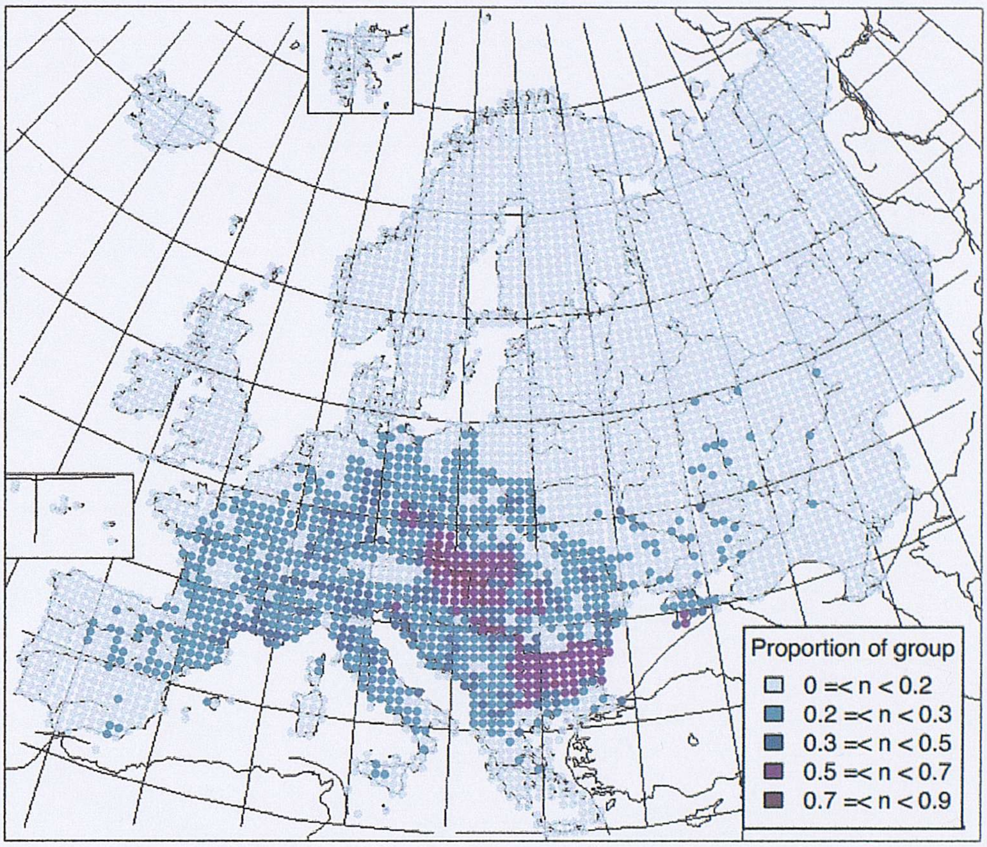
Figure 4.5 is a graphic representation of the residual values, the number of species by which a square differs from that predicted, from the regression shown in figure 4.4. Those squares that have fewer than expected species are shown in red whilst those that have a greater number are shown in green.

4.5 Discussion

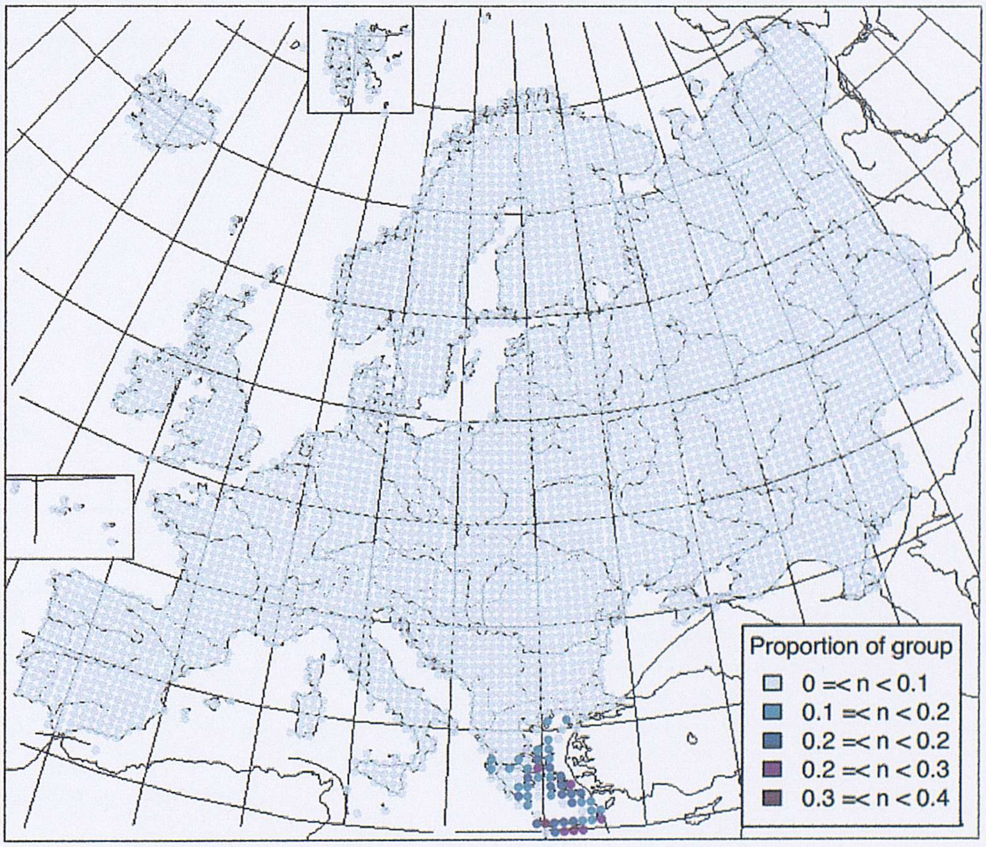
4.5.1 Mechanics of the classification

The AFE species used only included one fifth of all European plant taxa (2808 of a total of 11557 taxa in the Flora Europeae). However, because of the manner in which the AFE has been published, by plant family, all environments have been represented by the taxa included in this analysis. When the Atlas is complete it is not anticipated that the overall emphasis of the groupings will change significantly although it is inevitable that the boundaries of groups will be redefined.

It is to be expected that some species will have been better recorded than others. Overall however, the classification will remain unaffected so long as within a single square there is approximately equal treatment of species across all groups. The failure mode for the classification is relatively benign. Gross under-recording, such as the omission of data for entire countries, would cause groups that only occur in these areas to be omitted. This omission of a group does not affect the number or, in large part, the makeup of the other groups in the classification. Poor recording in the east of Europe is insufficient for

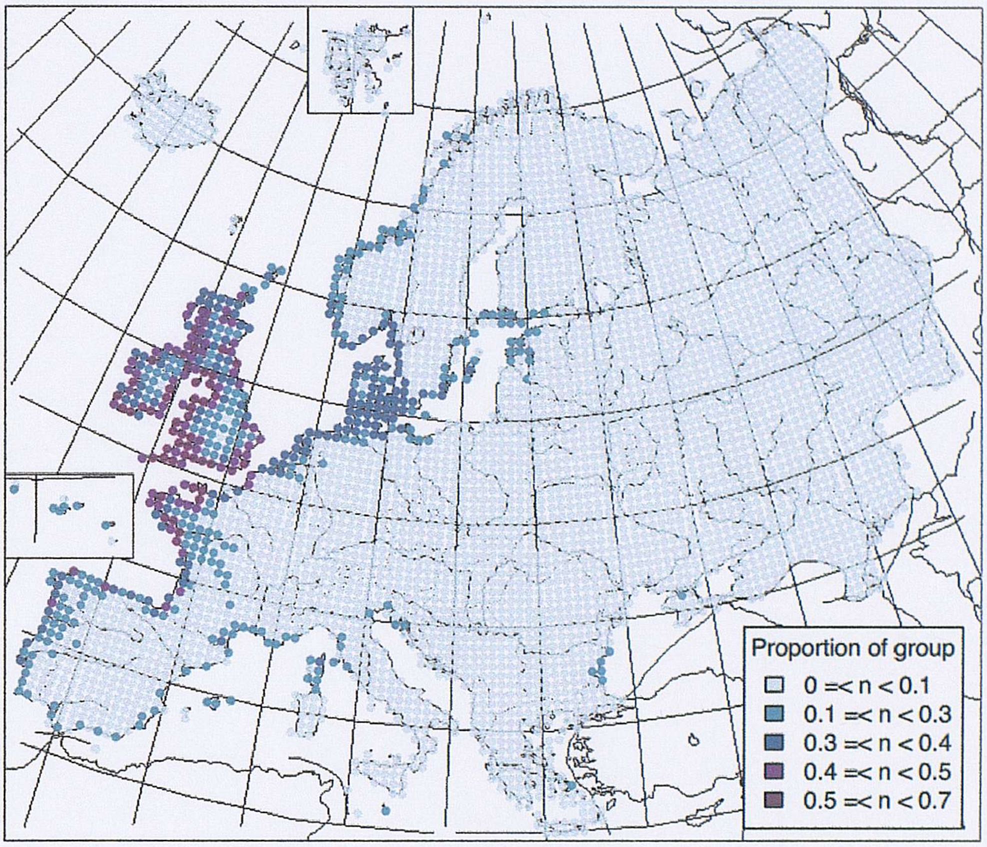


(a) Alyssum alyssoides

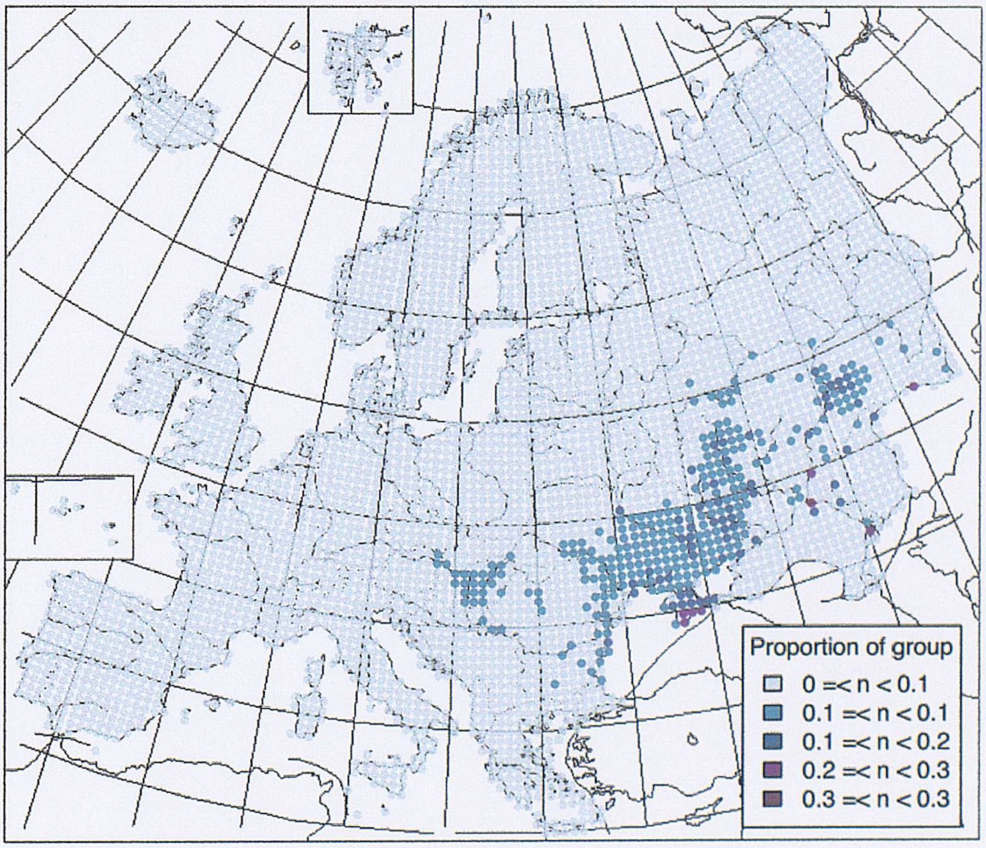


(b) Brassica cretica

Figure 4.2: Maps showing the distributions of the 18 groups within Europe. Darker colours indicate increasing proportion of the species in a group present in a particular 50km square.

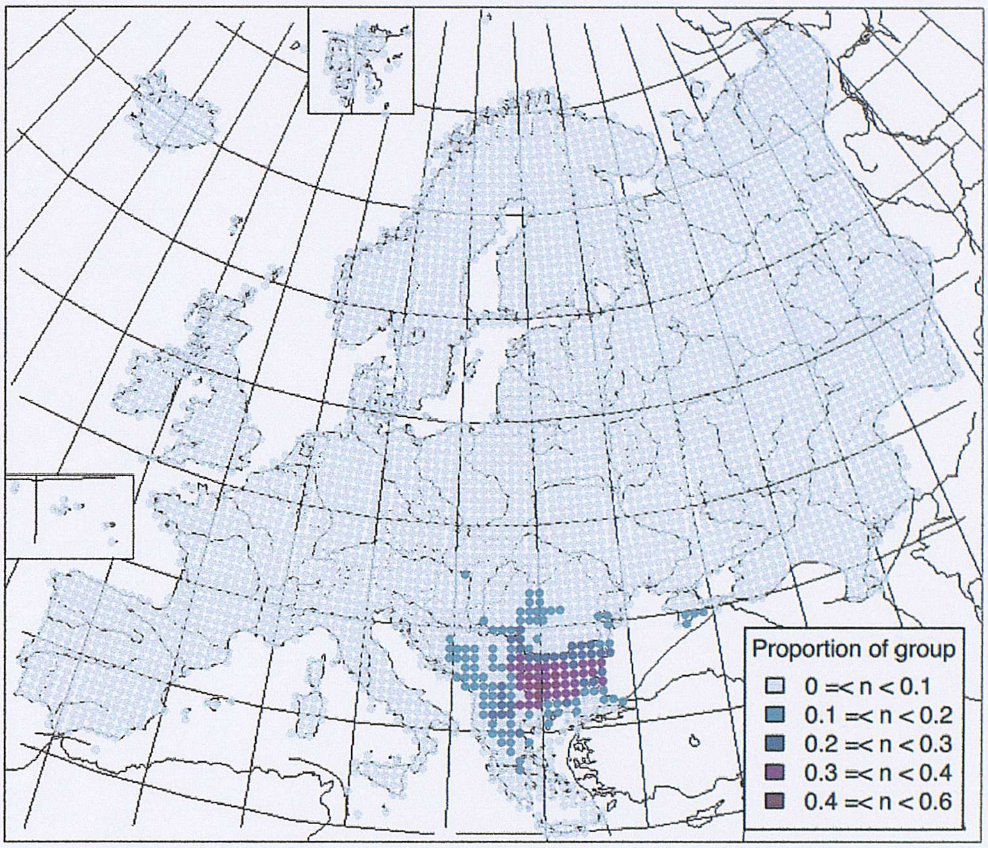


(c) Cochlearia danica.png

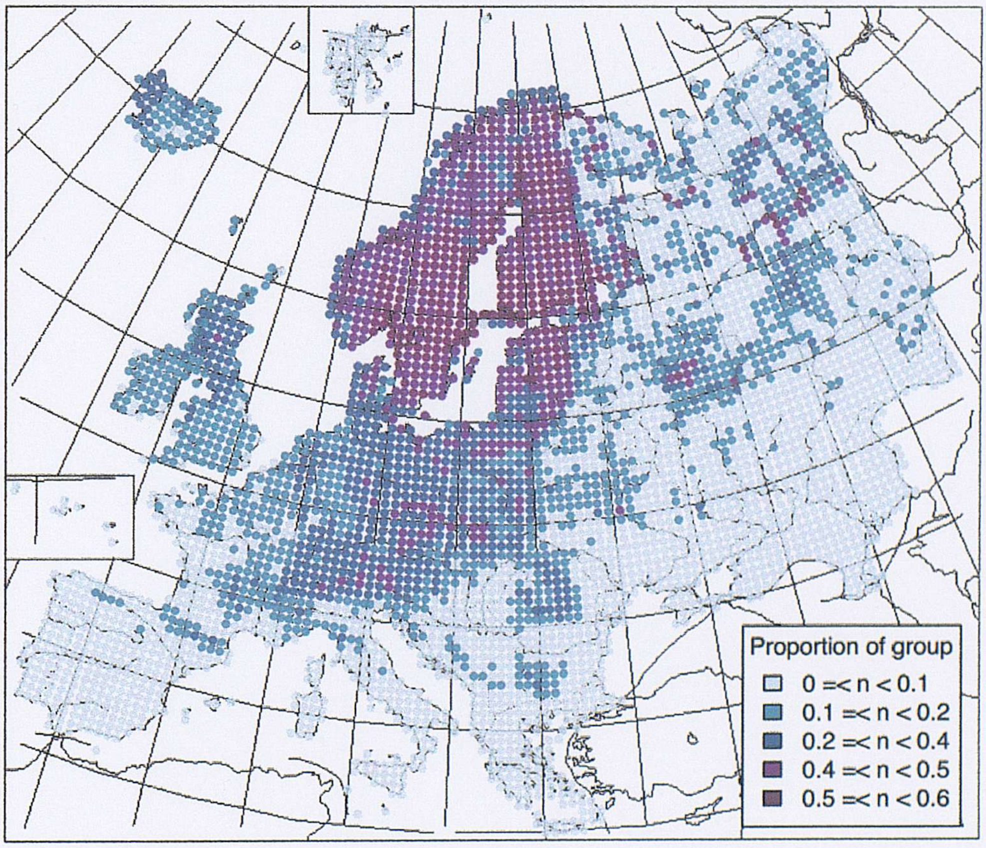


(d) Dianthus capitatus.png

Figure 4.2 (continued)

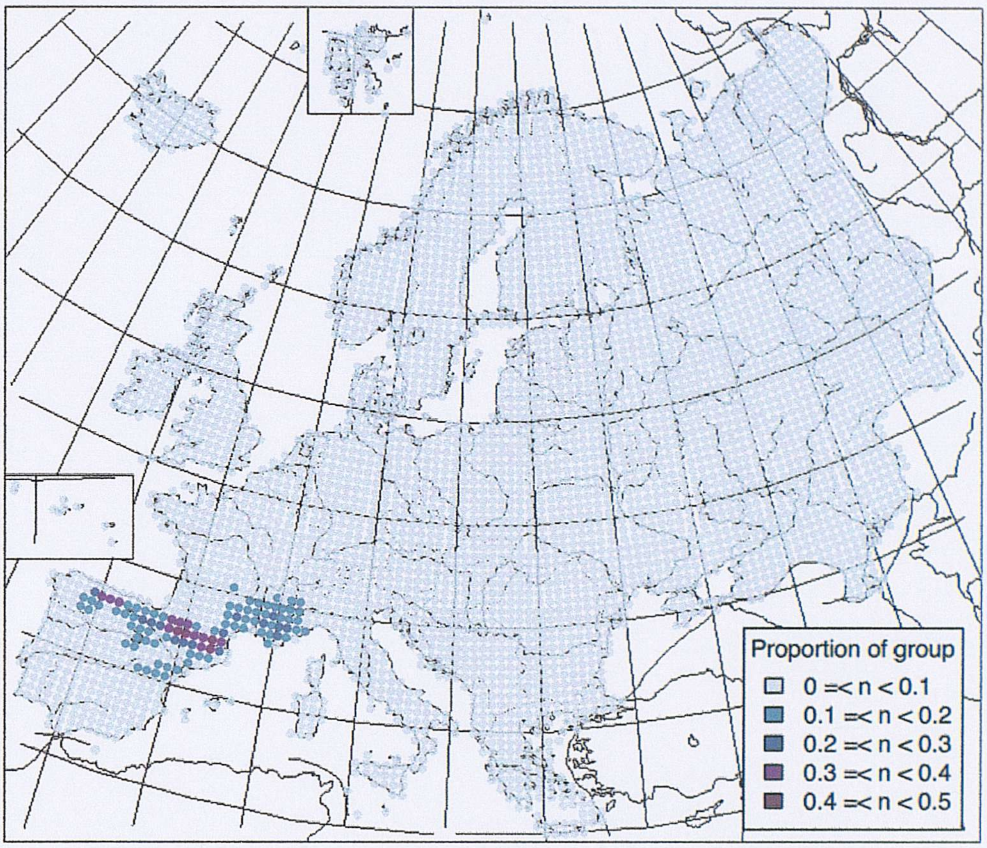


(e) Dianthus moesiacus

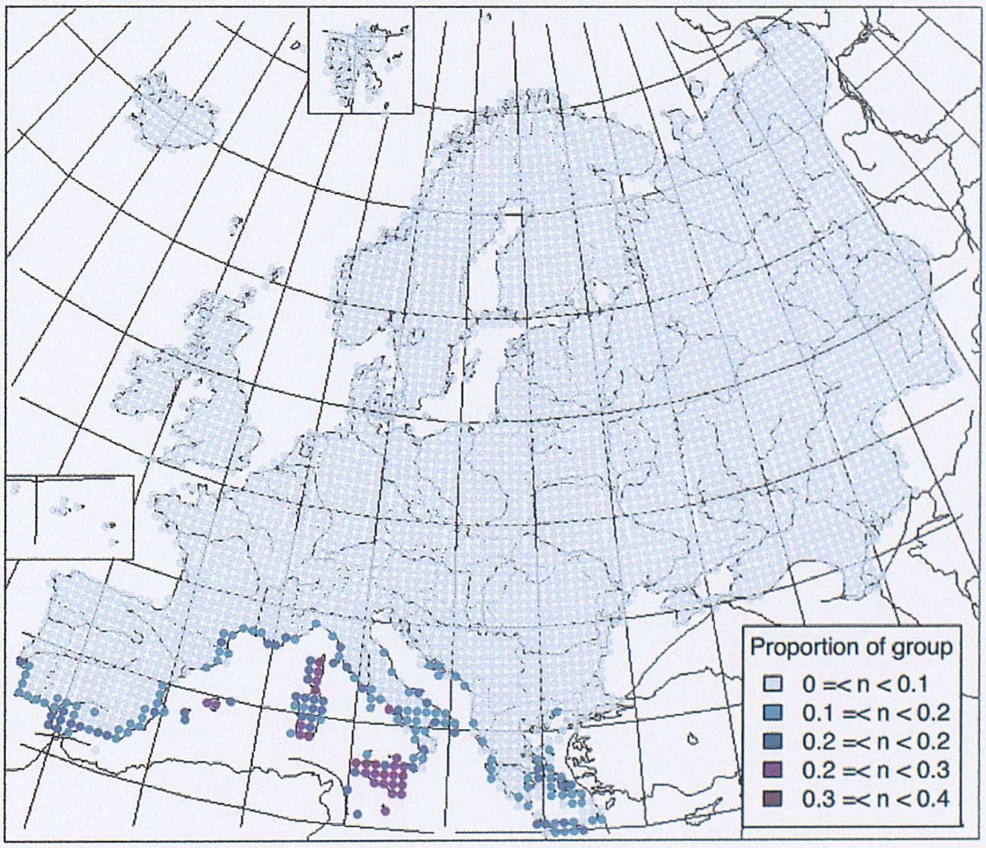


(f) Equisetum sylvaticum

Figure 4.2 (continued)

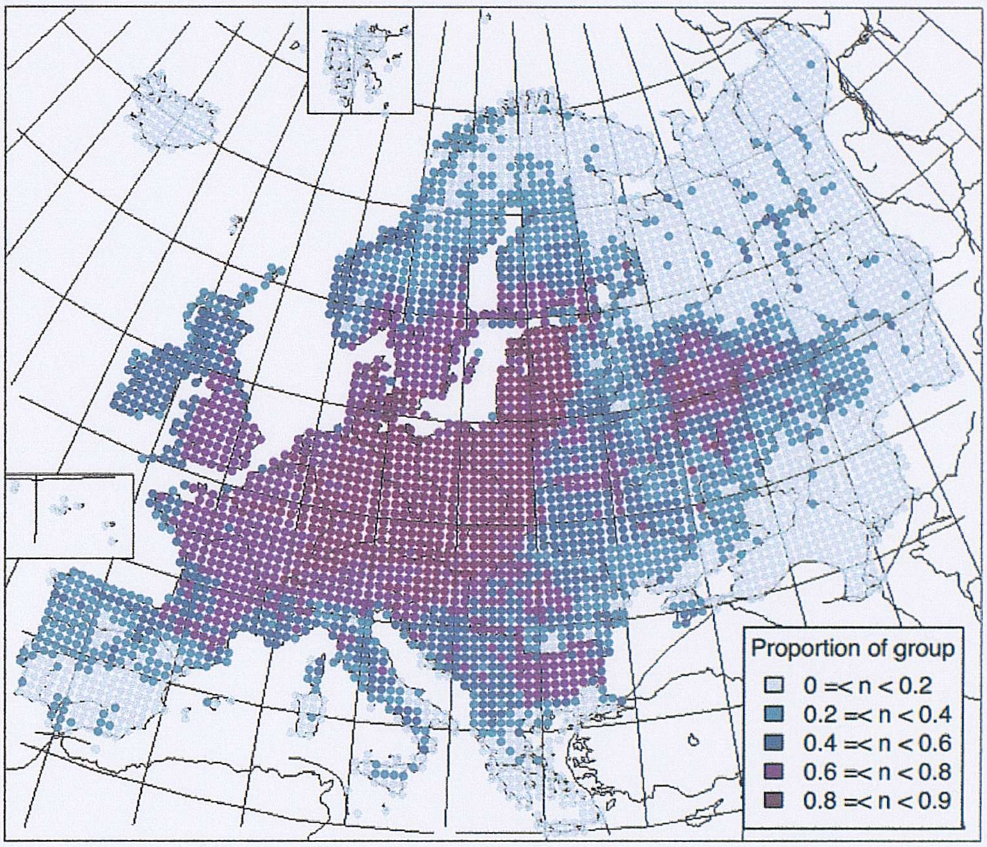


(g) Erysimum duriaei

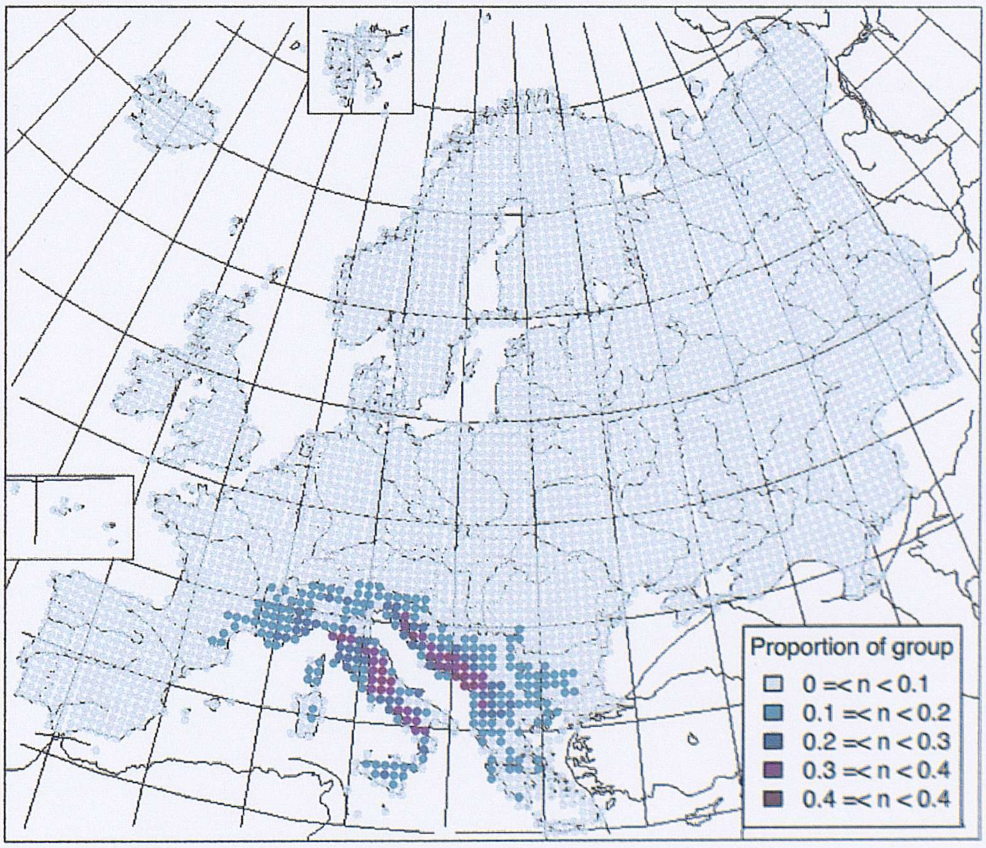


(h) Lobularia maritima

Figure 4.2 (continued)

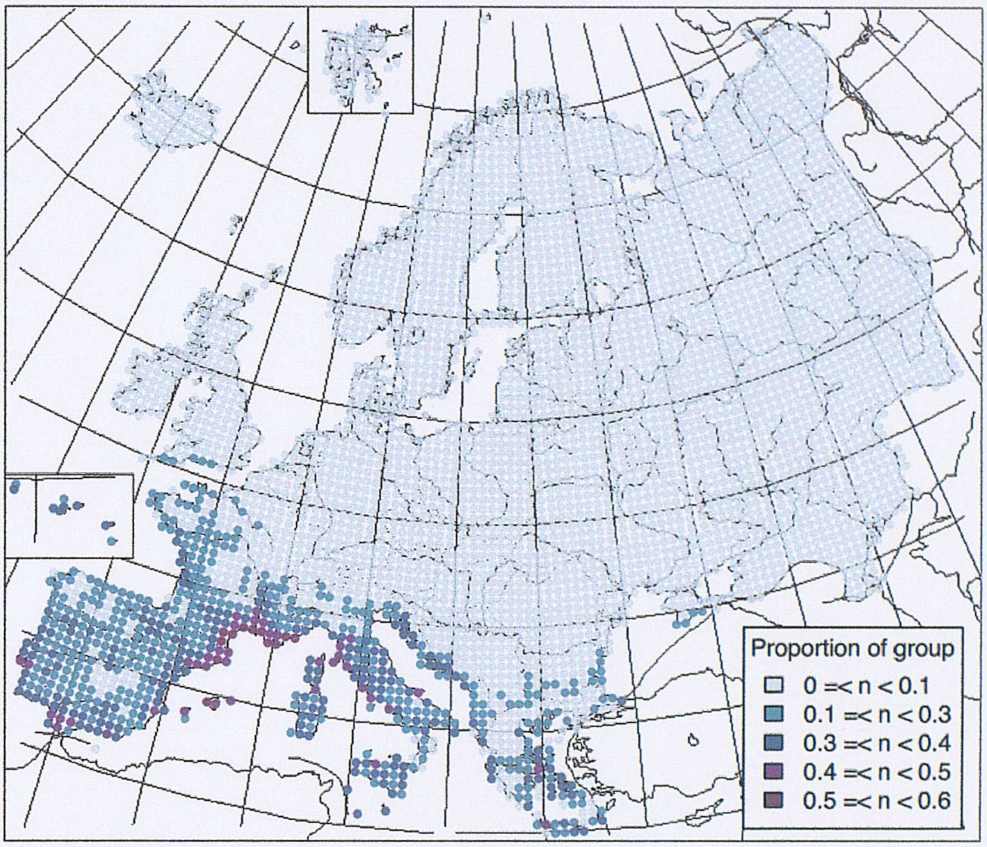


(i) Lychnis flos-cuculi

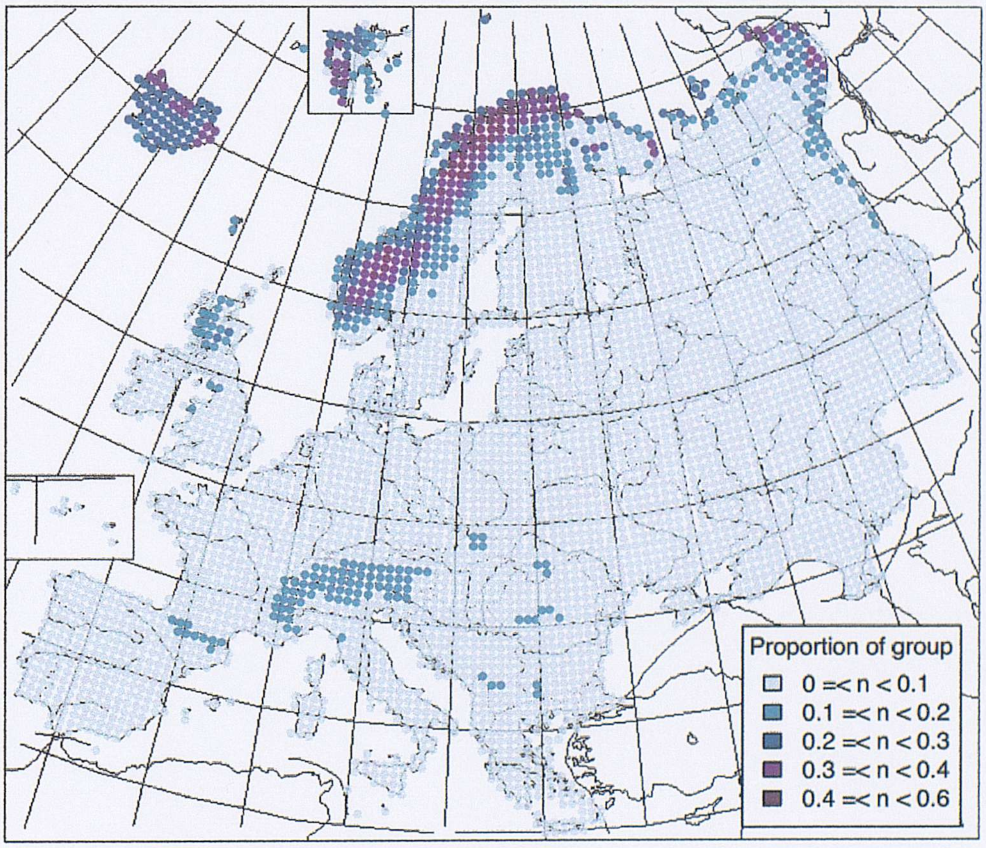


(j) Ostrya carpinifolia

Figure 4.2 (continued)

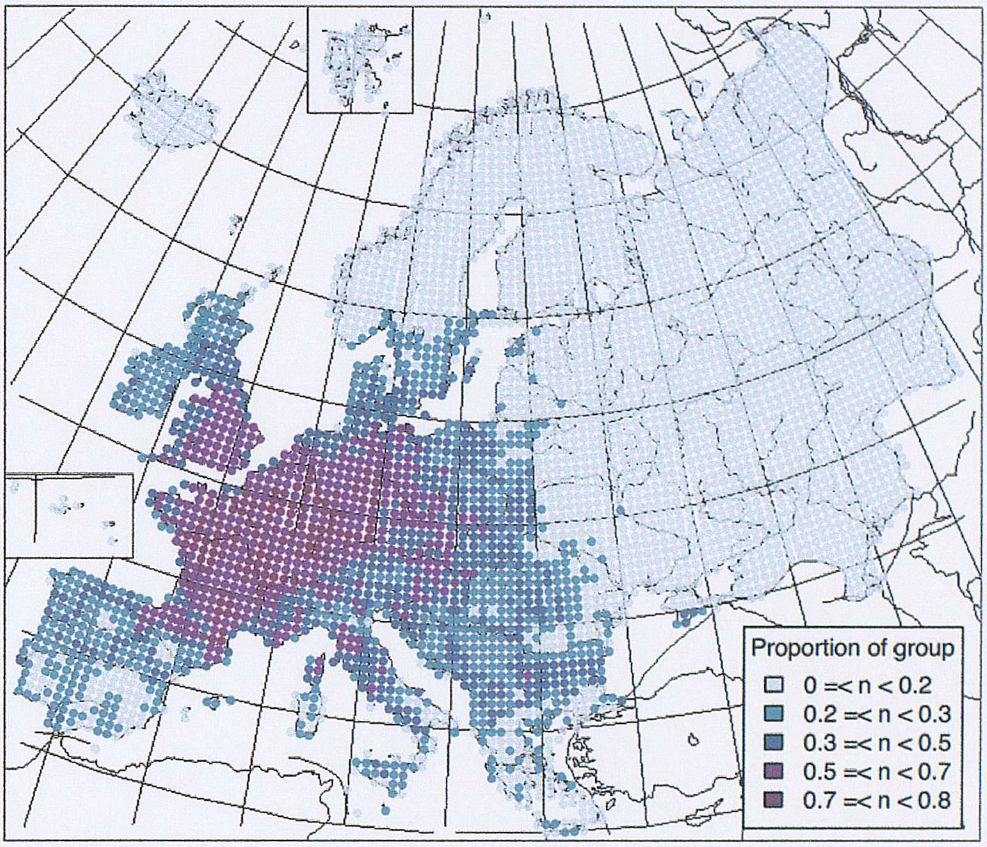


(k) Osyris alba

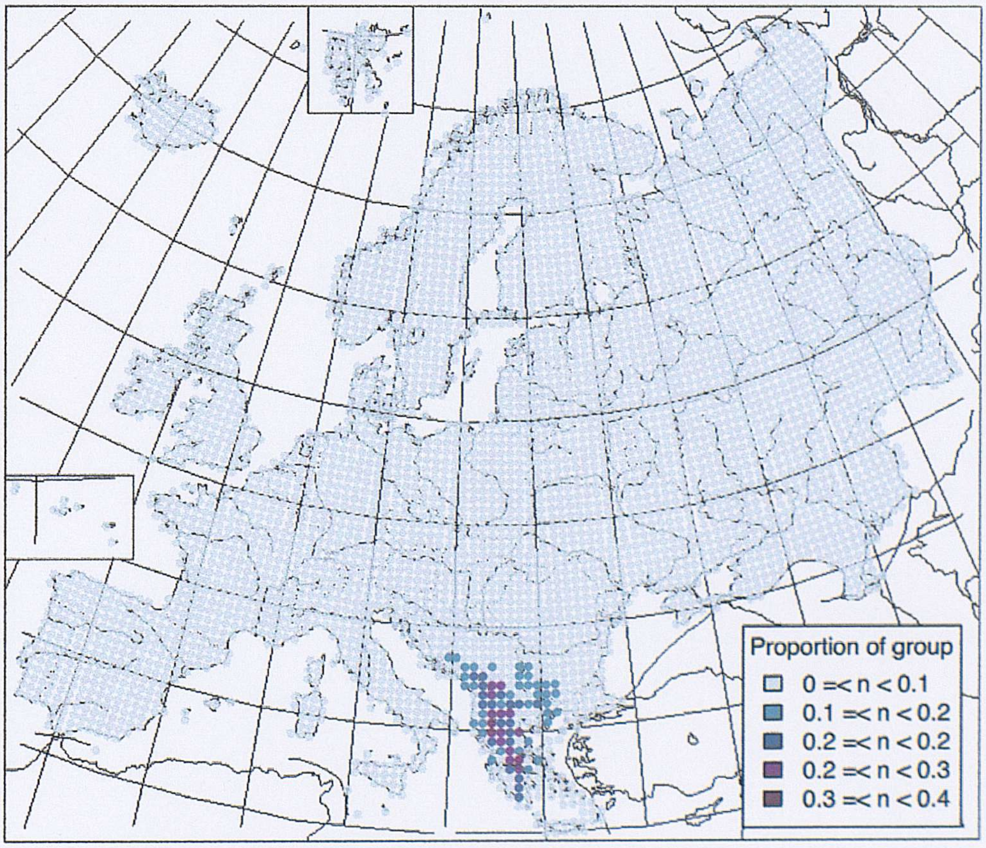


(l) Oxyria digyna

Figure 4.2 (continued)

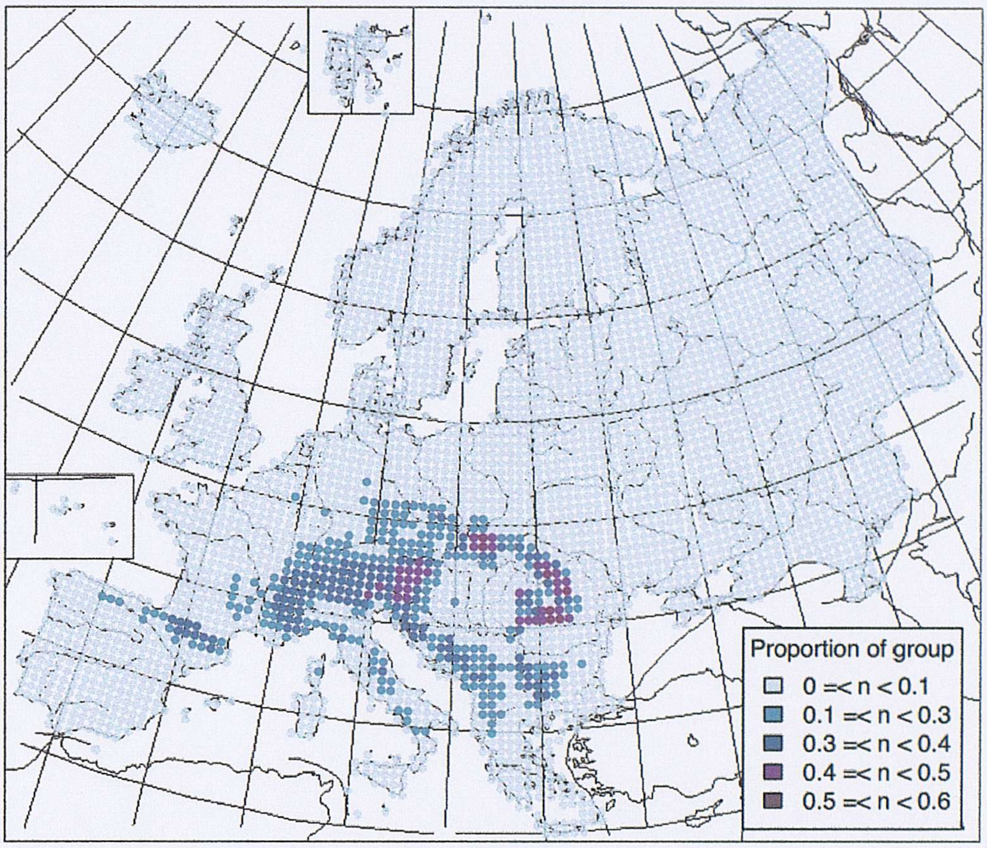


(m) Ranunculus bulbosus

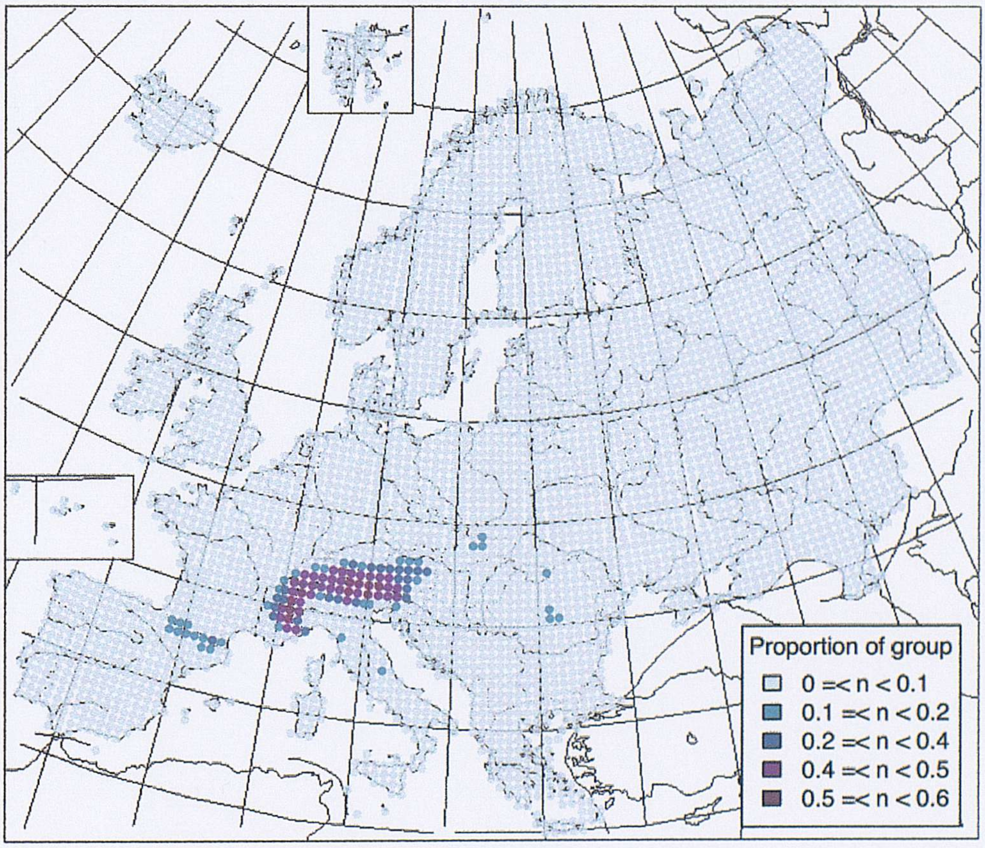


(n) Ranunculus psilostachys

Figure 4.2 (continued)

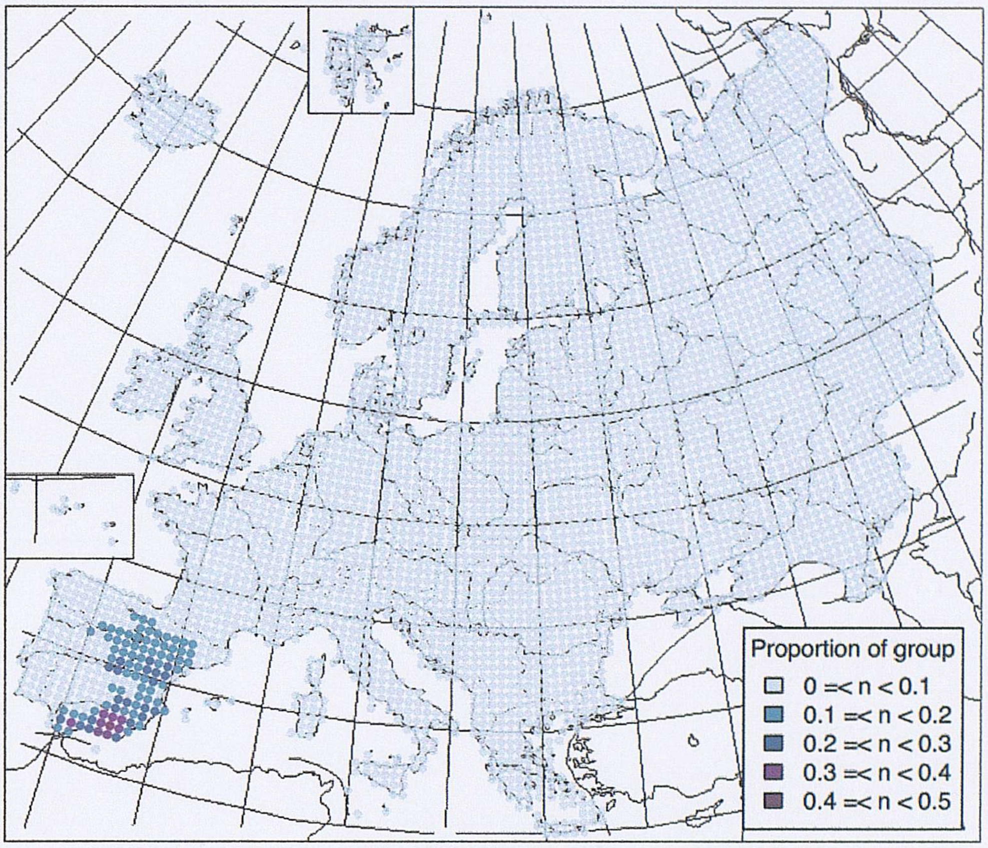


(o) Rumex alpinus

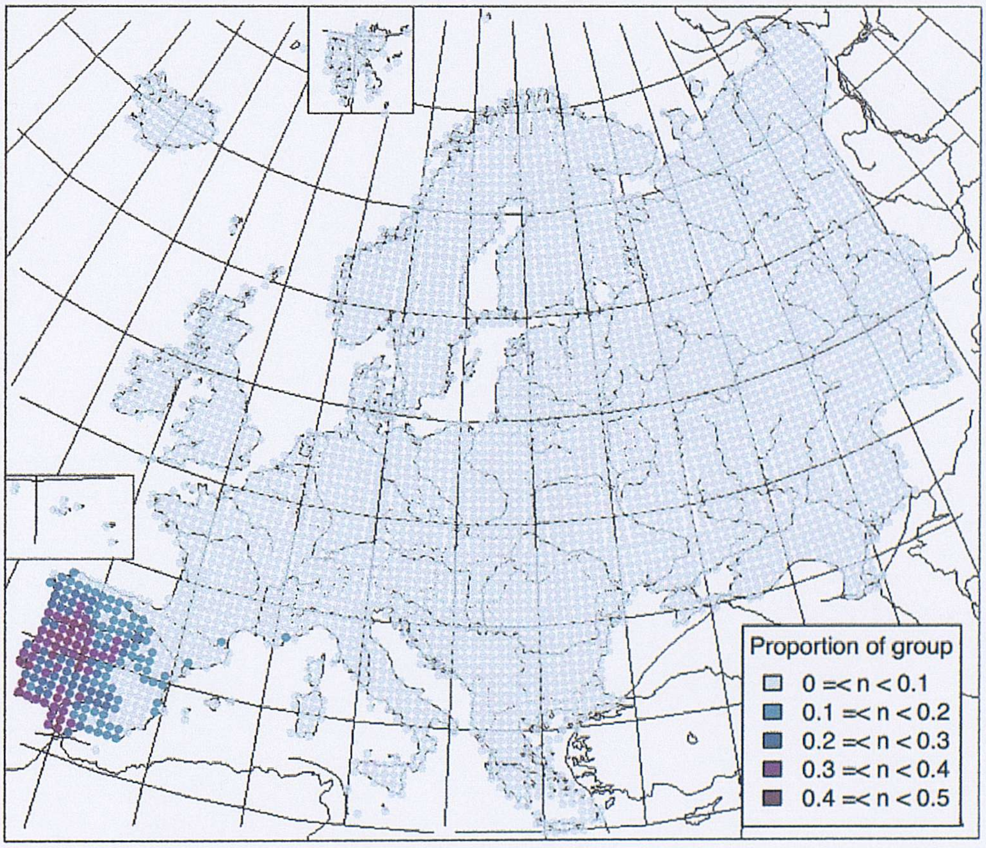


(p) Salix serpillifolia

Figure 4.2 (continued)

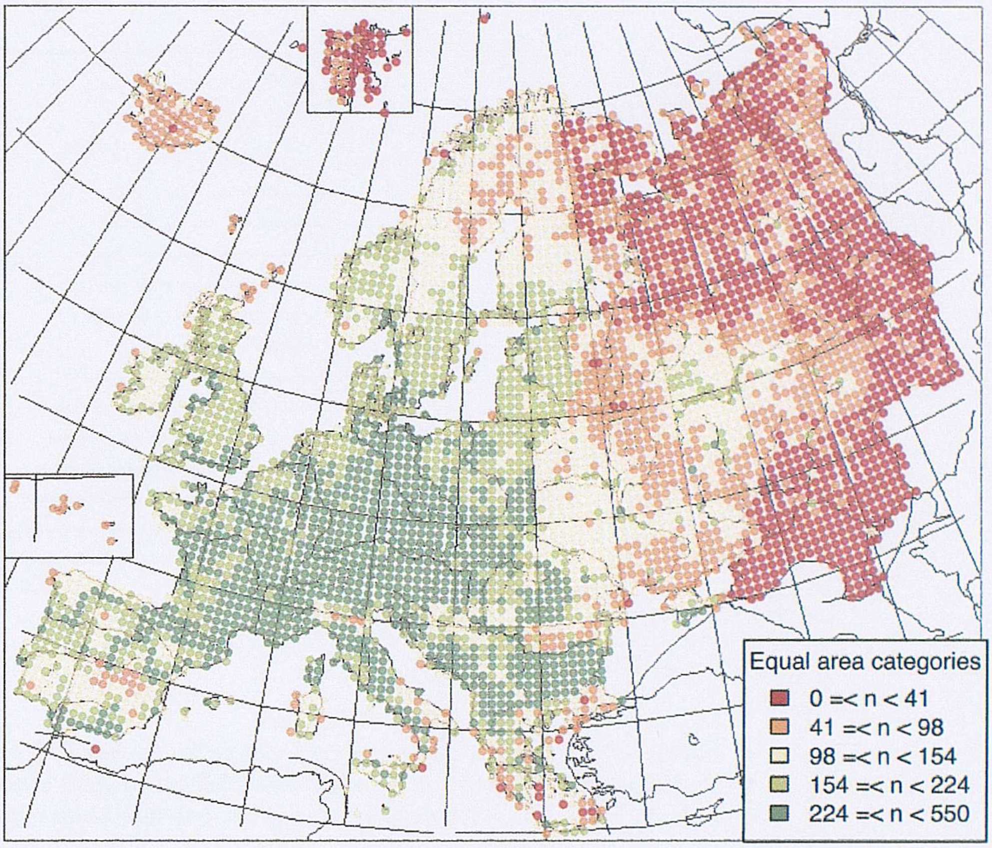


(q) Sarcocapnos enneaphylla

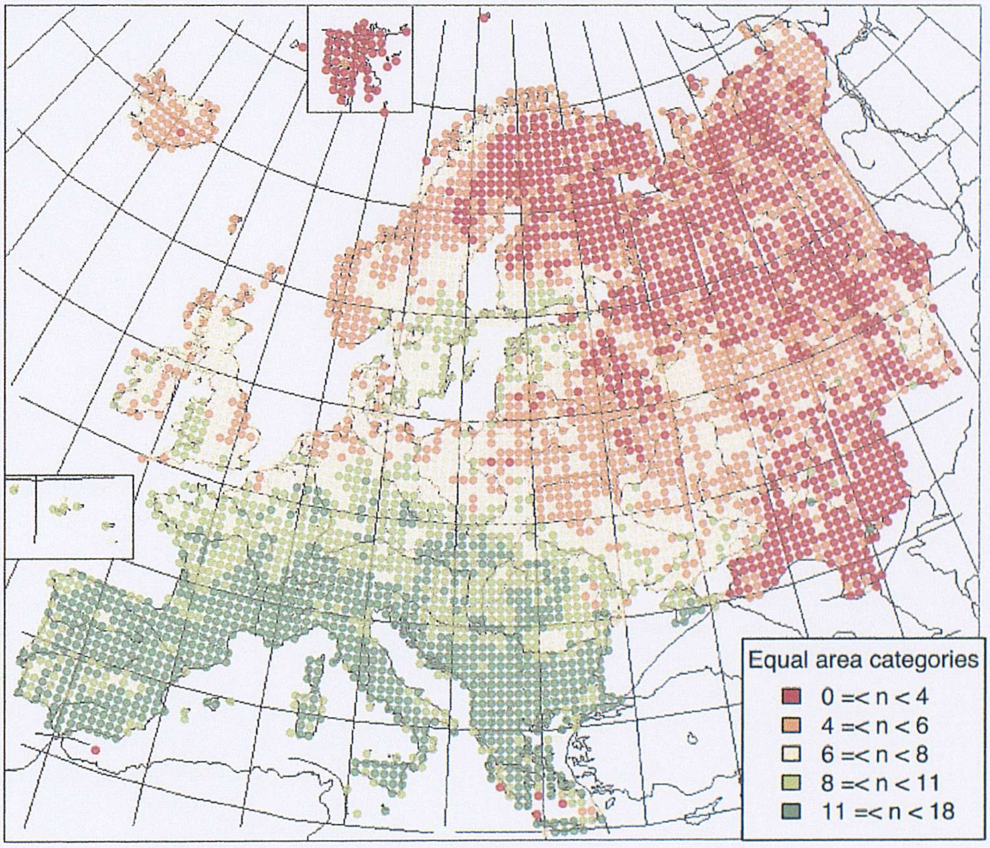


(r) Silene scabrifolia

Figure 4.2 (continued)



(a) Species richness in Europe



(b) Group density for Europe

Figure 4.3: Species Richness and Group density for Europe.

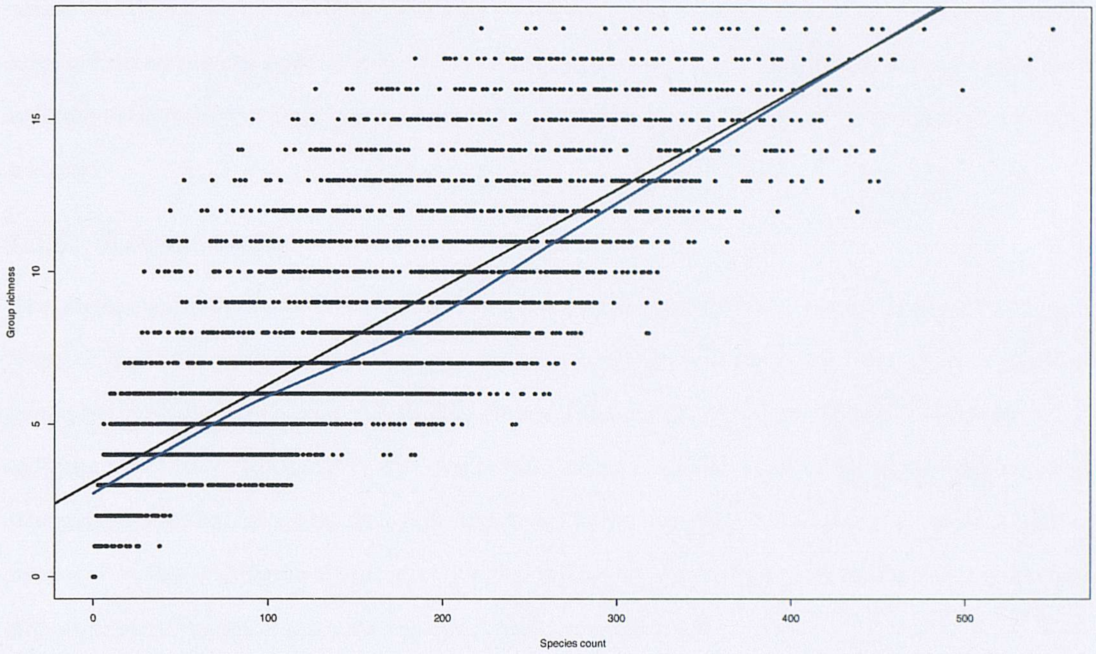


Figure 4.4: Regression of species richness versus group count. Solid black line is the best-fit model. The blue line is fitted by the LOWESS smoother to show any local trends within the data. This polynomial technique agrees well with the simple linear model.

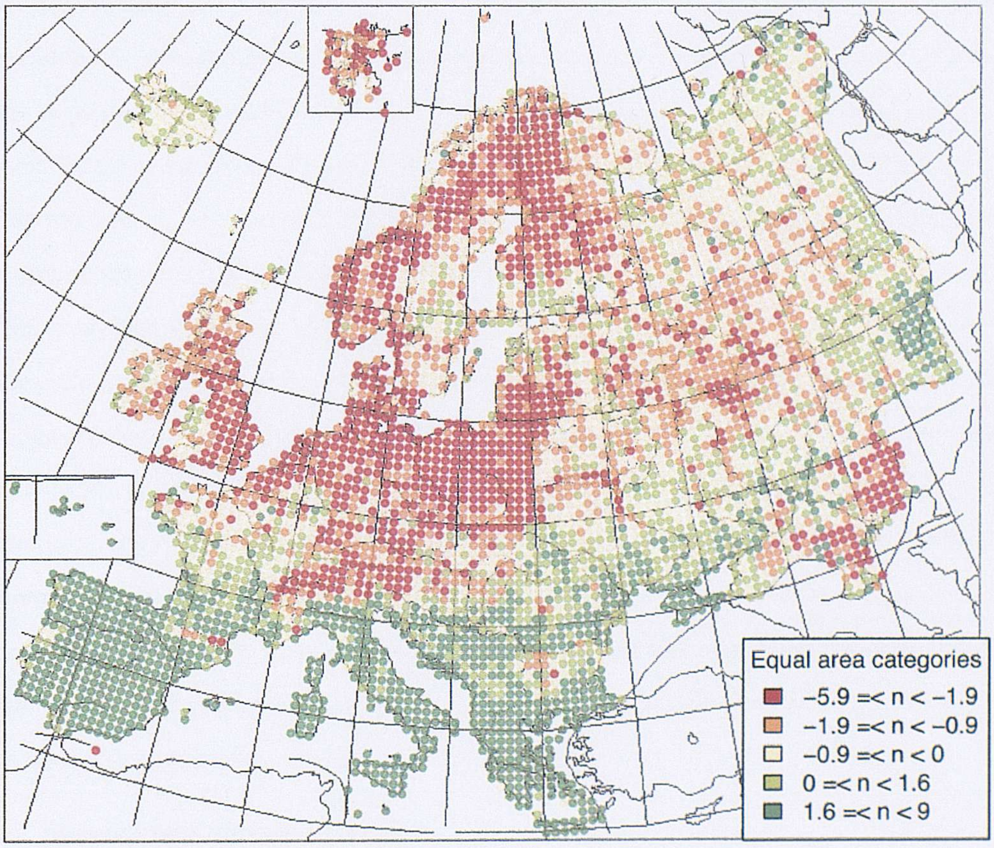


Figure 4.5: Difference between the number of groups expected from the species number and the actual group count.

group loss to occur due to the large number of squares present. The averaging effect of group formation means that variations in recording effort between squares and between species are mitigated. Similarly, edge and island squares will have no adverse effect on the analysis so long as a species in each square is as likely to be recorded as any other. This is because each square is considered individually within the analysis.

4.5.2 Group results

The phytogeographical analysis produced 18 distinct groups. Each of these is mapped in figure 4.2. There is no single Mediterranean group. The Mediterranean Sea area is split into at least six distinct groups not including the Iberian and Balkan groups which could also be considered Mediterranean. This splitting up of the Mediterranean area into various smaller regions is an effect of the fractured nature of the landscape with the flora changing very rapidly over very short distances (i.e. there is high beta diversity). What is particularly interesting is that this is not a series of repeating floras but a procession of completely different floras replacing each other along the coast.

The three, widespread, central to northern European groups (figures 4.2f, 4.2i & 4.2m) are caused by the inverse of the same beta diversity effect, very little change in the overall flora over great distances. However, between them these three groups represent only 13% of the species studied. Despite the largest *Lychnis flos-cuculi* (figure 4.2i) having species representatives in 97% of the land area, the number of species from this generalist group is few when compared with the local specialist species. The reduced number of groups in central Europe is an effect of the large areas of environmentally homogeneous landscape found in the central European region. With little difference in topology or environmental conditions to cause the flora to diversify, the species within these groups have become widespread with fewer species overall and perhaps more in-species variation.

Major geological features including the three main mountain ranges in Europe, the Alps (figure 4.2p), the Pyrenees (figure 4.2g) and the Norwegian Alps (figure 4.2l), together with the Atlantic coast (figure 4.2c), mountainous regions in the Balkans (figures 4.2n & 4.2e) and the Black Sea (figure 4.2d) each seem to have developed a distinct flora of their own. In general these distinct, restricted, area floras are more speciate than the widespread groups. This finding is in stark contrast to the result in the preceding chapter where in excess of 50% of the flora is found in the ubiquitous *Cynosurus cristatus*(u) group. The important factor here is that these groups are sufficiently distinct not to merge with one another and form a more widespread group. In this context the Mediterranean Sea is such a large feature that its flora is broken down into more localised groups.

4.5.3 Species and Group diversity

There is an undoubted link between the number of species in a square and the number of phytogeographic groups found (figure 4.4). Those areas where species richness incorrectly predicts the number of groups are geographically associated. Hotspots of species diversity are revealed to be a localised phenomenon. An area that would be considered a hotspot in Britain (such as Ben Lawers, seen as a single dark green square in figure 4.3a) contains what would be considered an average number of species on the mainland of Europe. Although overall species diversity is a good predictor for phytogeographical richness, figure 4.5 demonstrates that in central Europe the phytogeographical richness is lower than would be expected from the number of species records and that in the Mediterranean area that the opposite is true.

Over-prediction of groups is particularly acute where strong geographical features cause strongly-defined local floras to be produced, such as is the case for mountain ranges. It also occurs where there is little change in the flora over large areas, for example the band running from Britain through Holland, Northern Germany, Denmark and Northern Poland. In both cases species occurring in these areas join phytogeographical groups whose shape is strongly predicted by the geography of the area. Under-prediction occurs where the shape of species ranges is weakly defined by geography and by more transient factors such as rainfall, wildfires and even perhaps temperature as is the case in the Mediterranean region.

4.6 Conclusion

At this scale of 50km squares some detail is lost when this classification is compared to finer scale classifications (chapter 3) but the overall differences between groups demonstrated by this technique give a new insight into the broad continental-scale patterns. In partial answer to the question raised in chapter 3 regarding the strength of partitioning, at the continental scale species are shown to be more localised than at the country scale. Species become more strongly associated with other species from their region and form strong local groups. A very few join more widespread groups due in most part to the vastly greater range of climates within Europe.

The construction of this classification should help in answering questions about the patterns found within the European flora. With suitable data, an environmental analysis such as that performed with the groups found in chapter 3 should provide answers to questions of the relative importance of historical and evolutionary factors compared with those of the environment. Regions that have approximately equivalent environments but are dominated by different groups should provide particular insight. The use of groups in these studies, rather than individual species, is preferable as group statistics mitigate the peculiarities of distribution any single species exhibits.

It will be interesting to compare these findings with those of a similar classification performed on species of North America and other continents. Do species behave in the same way, with predictable landscapes such as the central prairies, and sharp geographical features such as the Florida Everglades producing few groups; with more broken, unpredictable landscapes like those found in the mountainous, fire-prone regions of California producing a high numbers of groups? More locally, with the burgeoning of genetic studies, comparison between these geographic European group patterns and those found genetically could lead to insight into historic processes such as post-glacial re-colonisation and the process of community formation on the broad scale.

Chapter 5

Nestedness of bryophyte species

5.1 Summary

Nesting, the study of how species ranges overlap to produce patterns of matrix fill, is an essential component of assemblage structure and helps in the understanding of the underlying mechanisms involved (Greve and Chown, 2006). In this chapter I look at the patterns of British bryophytes and determine the degree of nesting displayed by both the individual species and the two taxonomic groups, liverworts and mosses. Nesting is measured using three different methods based on standard statistical practices.

Both taxonomic groups display nesting. Liverworts are seen as more nested by measures that penalise phytogeographic inconsistency in the group. Both groups were considered equally nested when this group diversity was allowed.

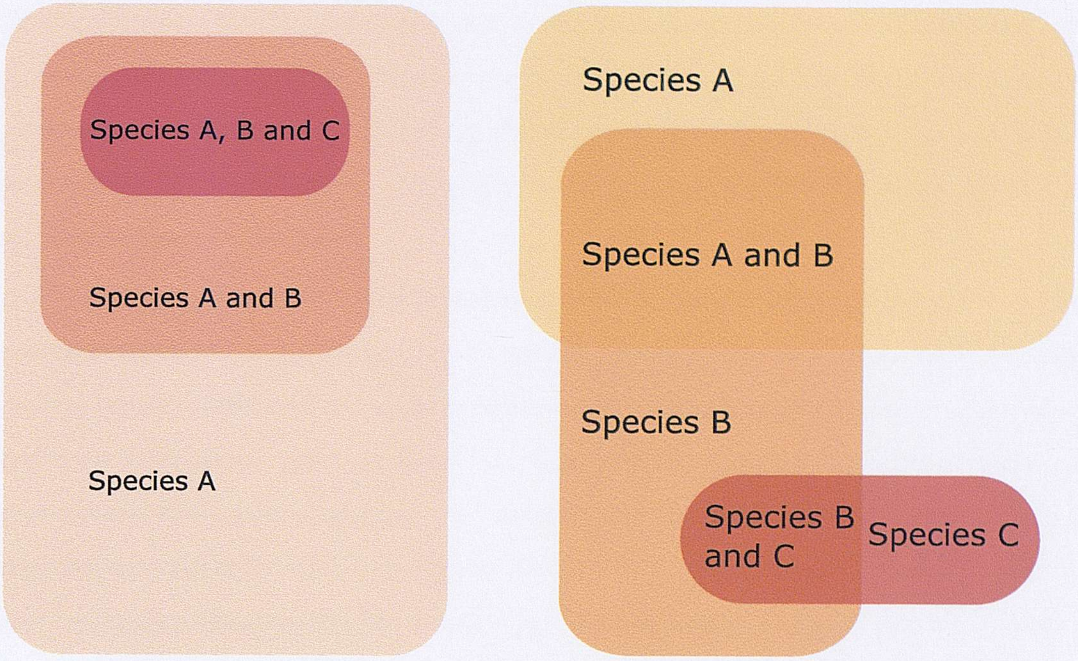
The pattern of nestedness in liverworts appears to be controlled by winter temperature and rainfall; that for mosses is more cryptic and is likely to be controlled by small-scale local factors.

5.2 Aim

I set out to find if patterns of distribution in the British bryophyte flora show nestedness; that is, if the pattern of distribution of a species with restricted range is a subset of the pattern of a more widespread species. A casual inspection of the distribution maps of British liverworts suggests that the ranges of many are nested. I test whether this is so and also investigate whether the mosses also display such nesting, albeit of a less obvious form. Species that are weakly nested may be phytogeographically or ecologically dissimilar to the other group members. I intend to investigate this through the use of phytogeographical elements.

5.3 Background

The study of nestedness began with the observation that the assemblages of species on ocean archipelagos often had a defined extinction order (Brown, 1986). This results in communities that are subsets of one another (Patterson and Atmar, 1986), although extinction is not the only mechanism proposed



(a) Nested distribution patterns (b) Unnested distribution patterns

Figure 5.1: Diagrammatic representation of distributionally nested and unnested groups. These are not to scale. While (a) shows the only way these species can be completely nested the pattern shown in (b) is only one in a whole range of ways that the species could be unnested.

that could result in nested communities (Patterson, 1990; Simberloff and Martin, 1991; Cook and Quinn, 1995; Loo et al., 2002). These results have led to a search for a mathematical method of expressing the nestedness of species compositions for a group of sites. There has been much discussion in the literature of both the manner in which the nesting of the species occurrences in a series of sites can be measured (Atmar and Patterson, 1993; Brualdi and Sanderson, 1999; Worthen, 1996) and, when a measure is found, what null model should be applied to allow a valid statistical test to be performed (Sanderson et al., 1998; Jonsson, 2001; Fischer and Lindenmayer, 2002). Nestedness as a subject shows ever increasing popularity. A repeat of Jonsson's 2001 Science Citation Index search for "nested subsets", "nested biota" or "nestedness" which found only 83 papers in the ten years to 2000 now returns 144 papers published between 2002 and today, an expansion that is not solely accounted for by the expansion of the life sciences. These previous studies follow the original Atmar and Patterson (1993) method in measuring the nestedness of the species assemblage of islands (or areas). When dealing with a large number of sites, such as is found in data from an atlas, it is more rigorous to consider species rather than sites as the primary unit. So in this study I look at the nestedness of species by the sites (grid squares) in which they are found rather than the nesting of sites by species: in effect rotating the species-sites matrix by 90 degrees. The rotating of the species-sites matrix places the emphasis on species rather than sites and de-emphasises the importance of any one site. The early studies used the site (island) as their primary unit of study. Here I use species as they are of greater biological relevance than the arbitrarilydefined grid squares. This is important because a common criticism of previous studies has been that any nesting observed is an effect of the difference in the areas of each site. The change does mean that I am assuming that individuals of a species are comparable across the species entire range (sites are the same for all species but the genetic variation within a species may mean that it is not the same for all sites, here it is treated as if it were).

It may be easier to consider species nesting by distribution pictorially. Two extremes of nestedness are shown in figure 5.1. It is interesting to note that one extreme is absolute. To be completely nested a group of species' distributions has to be arranged such that each species range is entirely enclosed by species with larger ranges. The distributions behave like a series of Russian dolls. The other extreme can be reached in any number of ways. At a national scale, the most unnested a group could be is that no species occupies the same square as any other, a completely disaggregated pattern. Practically, we are bounded by land area and each species' range. There is simply insufficient land area in Britain for complete disaggregation of the species in groups as large as the liverworts and mosses, meaning that some overlap in inevitable. At sufficiently small spatial scales most bryophytes are disaggregated although some are almost obligate epiphytes on other bryophytes (e.g. *Odontoschisma sphagni*).

Nesting of species distribution patterns suggests that there is a degree of commonality between the factors that control the distribution of species in a group. In previous chapters I have considered groups created from distribution patterns. It is likely that any of the groups already seen would show phytogeographical nesting. In this chapter I am considering groups that are formed taxonomically. It is plausible that the taxonomic features that define these groups could influence the pattern of these species' distributions. If these taxonomic groups are nested then we are some way to showing a link.

5.4 Method

5.4.1 Data

Data used in this chapter came from the three volumes of the *Atlas of the bryophytes of Britain and Ireland* (Hill et al., 1991-1994). Although less well recorded than the vascular plants these species were chosen as they form two distinct taxonomic groups which are expected to show different degrees of nesting. Data were restricted to the most taxonomically robust, traditional, species concepts. This data set is recorded as a presence or absence in all the 10km squares of the British Ordnance Survey grid system for all native bryophytes. All squares that contained at least one species were included in the analysis. For the purposes of this analysis the four native hornwort species were included with the liverworts. The data were taken from the master table used to construct the maps for the published atlas and converted into a presence-absence matrix covering all squares and all species, species as rows and squares as columns. No records added after the publication of the atlases were included.

5.4.2 Method selection

The determining criterion for any measures of nestedness used in this study was that they were capable of performing the required calculations in a reasonable time-scale with current, generally-available, computer hardware. The datasets I worked with in this chapter are comparatively large compared to those used in previous studies: 280 and 695 species for the liverworts and mosses respectively in 2851 squares compared to Sanderson *et al*'s 1998 Vanuatu data with 60 rows and 30 columns or the 28 rows by 26 columns in Brualdi and Sanderson (1999). This was the greatest limitation and precluded the possibility of using certain pre-existing methods, in particular those that use boot-strap type techniques to form their null hypotheses. Increasing matrix size is onerous because a linear increase in the size of the species presence-absence matrix causes an exponential increase in the difficulty of finding null-model matrices or computational solutions to matrix sum type equations.

Additionally, the data-sets chosen for use here are not as complete as those used for the vascular plants. The data suffer from geographic variation in the recorder effort used for each square. Any method proposed should therefore consider the in-square variation rather than between-square variation. It should perhaps also have a mechanism to avoid reliance on the most poorly recorded squares and species.

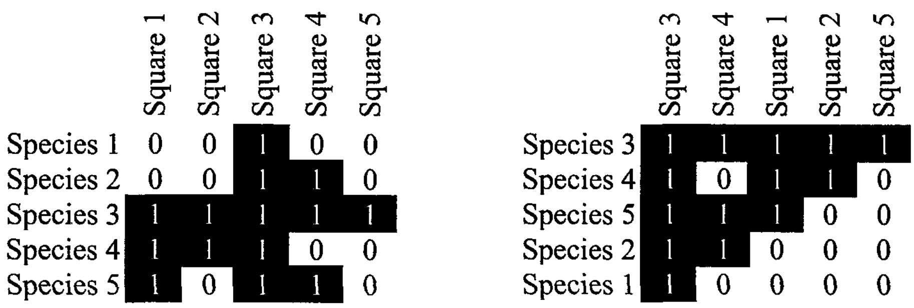
These two constraints, then, form the primary criteria for any measures of nestedness: the measures of nesting have to be capable of coping with the very large size of the datasets involved and they should limit the impact of 'rogue' data on the final result. Further requirements such as the ability to provide a score for the degree of nesting of a species within a group, the ease of the calculations and the stability of the measure (in particular when additional records are added to the data-set) should also be considered when assessing the measures.

5.4.3 Measuring Nesting

Underlying all the analyses is a sorted matrix. This means that the presence-absence matrix is sorted so that the species are ordered from most common to least common and the squares are ordered from most occupied to least occupied (see figure 5.2). Before sorting, the positions of the rows and columns carry no information. The information recorded in the matrix, the presence or absence of a species in a square, remains unaffected by the sorting as sorting simply transposes entire rows or columns.

D⋆

To consider the nestedness of an individual species I study its discrepancy within the sorted matrix. A discrepancy exists when a species occurs in a less-populated square while not occurring in a more-populated square. It is in effect a hole in the sorted matrix. The greater the number of discrepancies the less well-nested a species is. This approach is similar to the work of Sanderson et al. (1998) although due to the size of the data-set I am unable to use even their fast null matrix generating technique (Brualdi



	Square 1	Square 2	Square 3	Square 4	Square 5
Species 1	0	0	1	0	0
Species 2	0	0	1	1	0
Species 3	1	1	1	1	1
Species 4	1	1	1	0	0
Species 5	1	0	1	1	0

(a) Example of an unsorted matrix direct from the database

	Square 3	Square 4	Square 1	Square 2	Square 5
Species 3	1	1	1	1	1
Species 4	1	0	1	1	0
Species 5	1	1	1	0	0
Species 2	1	1	0	0	0
Species 1	1	0	0	0	0

(b) The same matrix after sorting

Figure 5.2: Demonstration of the effect of sorting on a matrix. Presence of a species is indicated by a solid black square. Initially no information is contained in the order of the rows and columns. After sorting, relative size of the species ranges and square biodiversity is implicit in the row and column orders. The matrix displayed here demonstrates almost perfect packing with only one discrepancy (species 4 in square 4), and hence indicates that this group of species is nested. Sorted matrices from groups with weaker nesting would contain more gaps in the black area in the top left corner.

and Sanderson, 1999). I use, in preference, standard statistical techniques and in particular weighted averaging.

We now calculate the number of discrepancies. Let

$$A = [a_{ij}] \qquad\qquad (i = 1 \ldots m; j = 1 \ldots n)$$

be a presence-absence matrix specifying the occurrence of n species in m squares where a_{ij} is the presence or absence of species j in square i. Then, Q_j, the number of squares occupied by species j, is defined as

$$Q_j = \sum_{i=1}^{m} a_{ij} = a_{.j}$$

Mathematically, discrepancy for species j is the count of its absences from the first Q_j columns of the matrix. Define D_j to be the discrepancy for species j

$$D_j = \sum_{i=1}^{Q_j} a_{ij}$$

The more squares the species occupies the more likely it is to display a discrepancy simply by chance up until 50% of the available squares are occupied. After this point the species' distribution is constrained by the finite nature of the study area and the chance of a discrepancy falls until it reaches zero when a species occupies all squares. In this study the value of discrepancy is bounded by the finite land area of Britain. A completely ubiquitous species will have no discrepancy but it is unclear whether such a species is nested within the group. A standardised discrepancy allows comparisons to be made between species that occupy different numbers of squares and across study areas of different sizes. This measure is standardised by comparing the observed value against what would be expected in the random case.

Specifically, the standardised value for species j is

$$D\star_j = \frac{D_j^{\text{Rand}} - D_j}{D_j^{\text{max}} - D_j^{\text{min}}}$$

where D_j^{Rand}, D_j^{max} and D_j^{min} are the random expected, maximum and minimum values of D_j respectively. Specifically

$$\begin{aligned} D_j^{\text{Rand}} &= Q_j(1 - \frac{Q_j}{n}) \\ D_j^{\text{max}} &= \min(Q_j, n - Q_j) \\ D_j^{\text{min}} &= 0 \end{aligned}$$

The overall standard discrepancy for the group is then simply

$$D\star = \text{average}(D\star_j)$$

and can be used to define a measure of departure from nestedness for the whole matrix (group).

NEST1

With $D\star$, those species that occupy fewer squares gain a disproportionate sway over the final value of this metric. Restricted species are a particular problem. Consider a species occupying a single square. The standardised discrepancy score of this species is highly susceptible to being one or zero depending on whether the species occurs in exactly the most occupied square or not. A better approach is to consider how well packed the sorted matrix is as a whole. To this end two further methods of evaluating the overall nestedness of the group are considered.

The first of these is simply calculating the degree of the packing in the sorted matrix, where the observed centre of gravity of occupied squares is compared with that expected for random occurrences. Packing measures the number of discrepancies in the matrix. A dense (highly packed) matrix has few discrepancies. The degree of packing in the matrix can only be improved by moving a species between squares, this is the gedanken equivalent to moving the entire population of a 10km square to a different square, so that the nestedness of the group is improved. Packing the matrix in the orthogonal direction would be equivalent to converting one species to another and thus makes no sense. This packing score is calculated by taking an average of the column numbers within the matrix and weighting this by the number of occupied squares in that column. A better packed matrix will show a smaller value for this packing score. To allow cross-group comparisons to be made the packing score is standardised by scaling it between the weighted average of a fully-packed matrix, that is one in which the species display perfect nesting, and a value of half the number of squares demonstrating a completely random matrix. This standardised packing score is the first measure of group nesting (NEST1). Specifically

$$\text{NEST1} = 1 - \frac{R-C}{R-B}$$

where R, C and B are the random expected, actual (observed) and best possible centres of gravity that could be achieved for this combination of species frequencies. R, C and B are calculated as

$$\begin{aligned} R &= \text{Centre of gravity of randomly distributed group} \\ &= \frac{(m+1)}{2} \\ C &= \text{Actual centre of gravity of matrix} \\ &= \frac{\sum_{j=1}^{n} \sum_{i=1}^{m} i a_{ij}}{\sum_{i=1}^{m} i} \end{aligned}$$

$$\begin{aligned} B &= \text{Best possible centre of gravity} \\ &= \frac{\sum_{j}^{n} Q_j (Q_j + 1)}{2 \sum_{i=1}^{m} i} \end{aligned}$$

NEST2

As stated above, the simple sorting of the squares by occupancy can lead to rare species disproportionately skewing a measure of discrepancy simply by not occurring within the most occupied squares. In the NEST1 method this effect was counteracted by considering not just if a species fitted the pattern but how different it was from the group pattern. NESTf2 is similar in approach, in that the distance from the ideal pattern is considered, but into this method is factored the practical consideration that a square will never be completely sampled. Each species is considered in turn excluding the most common. For each species two figures are calculated. The first is the number of differences between the species and the next most abundant species. The second figure is the number of differences between the species and that species to which it is most similar. The ratio of the two counts is the second measure used.

The following counts are the distance between two species (j and j'). These species are always binary. The matrix is sorted such that $j = 1$ is the most common species and $j = n$ is the least common. For species j the number of squares where it does not match species j' is

$$D_{j;j'} = \sum_i a_{ij}(1 - a_{ij'})$$

By definition, in a sorted matrix

$$j - 1 = \text{index of the next most common species to } j$$

If we measure the similarity of two species j and j' by the Jaccard coefficient

$$S_{jj'} = \frac{\sum_{i=1}^{n} a_{ij}a_{ij'}}{\sum_{i=1}^{n} a_{ij}a_{ij'} + D_{j;j'} + D_{j';j}}$$

and define $j*$ to be the index of the most similar species to j when similarity is measured by the Jaccard coefficient, then by summation across all species we get

$$\mathrm{NEST2} = \frac{\sum_{j=1}^{m} D_{j;j-1}}{\sum_{j=1}^{m} D_{j;j*}}$$

5.4.4 Nesting by European element

Hill and Preston in 1998 produced a phytogeographical element analysis for all the bryophyte species then recorded in Britain. In their work they assigned each taxon to a phytogeographic element that represented the taxon's northern hemisphere range. The northern hemisphere was split into seven bands east to west and nine bands from north to south. An element number was made up of these two parts. The first number gives how northerly the distribution, the smaller the more northerly (1 Arctic-montane, 2-Boreo-arctic montane, 3 Wide-boreal, 4 Boreal-montane, 5 Boreo-temperate, 6 Wide-temperate, 7 Temperate, 8 Southern-temperate and 9 Mediterranean). The second gives how easterly an element is: the smaller this time, the more westerly the group (0 Hyperoceanic, 1 Oceanic, 2 Suboceanic, 3 European, 4 Eurosiberian, 5 Eurasian and 6 Circumpolar). Combined, these numbers give 63 possible phytogeographical elements.

By considering the range of nesting scores displayed by species in the Hill and Preston's elements it was possible to see how the wider geographic range of a species affected how well that species fitted the general British pattern for the group. The range of nesting scores for each of these elements was analysed by ANOVA. Not all possible elements contain species, in fact, only 39 elements are occupied by the mosses and 31 by the liverworts.

Model simplification was used with the ANOVA analysis so that where the distribution of nesting scores for two elements was not statistically different the elements were merged. These new groups of elements were also considered for amalgamation and the process repeated until no further mergers were possible. The groups left at the end of this process were, as far as nesting is concerned, statistically distinct.

5.4.5 Environmental models

The two classes of bryophytes were subdivided into a nested and an unnested group. The unnested group was defined as the species with a nesting score in the lowest 20% of values. This was equivalent to 0.9% nested for liverwort species (56 species) and 2.6% nested for moss species (139 species). The distributions of the four groups of bryophytes were then modelled using a GLM with binomial errors in the same manner as the groups in chapter 3. The same environmental data as chapter 3 were also used: that is, continuous explanatory variables for January temperature, July temperature, annual precipitation (rainfall), minimum altitude, maximum altitude and proportional surface gradient (log(maximum altitude/minimum altitude)). Gradient is again shown as steepness in the tables. Two-level categorical explanatory variables were used for presence-absence within a 10km square of peat, chalk, limestone and coastline.

5.5 Results

Table 5.1: Summary statistics for nesting analysis. Individual nesting measures are not comparable but all scale, for nesting, between -1 and 1 (negative scores in the species nesting scores indicate dissociation with the nested species).

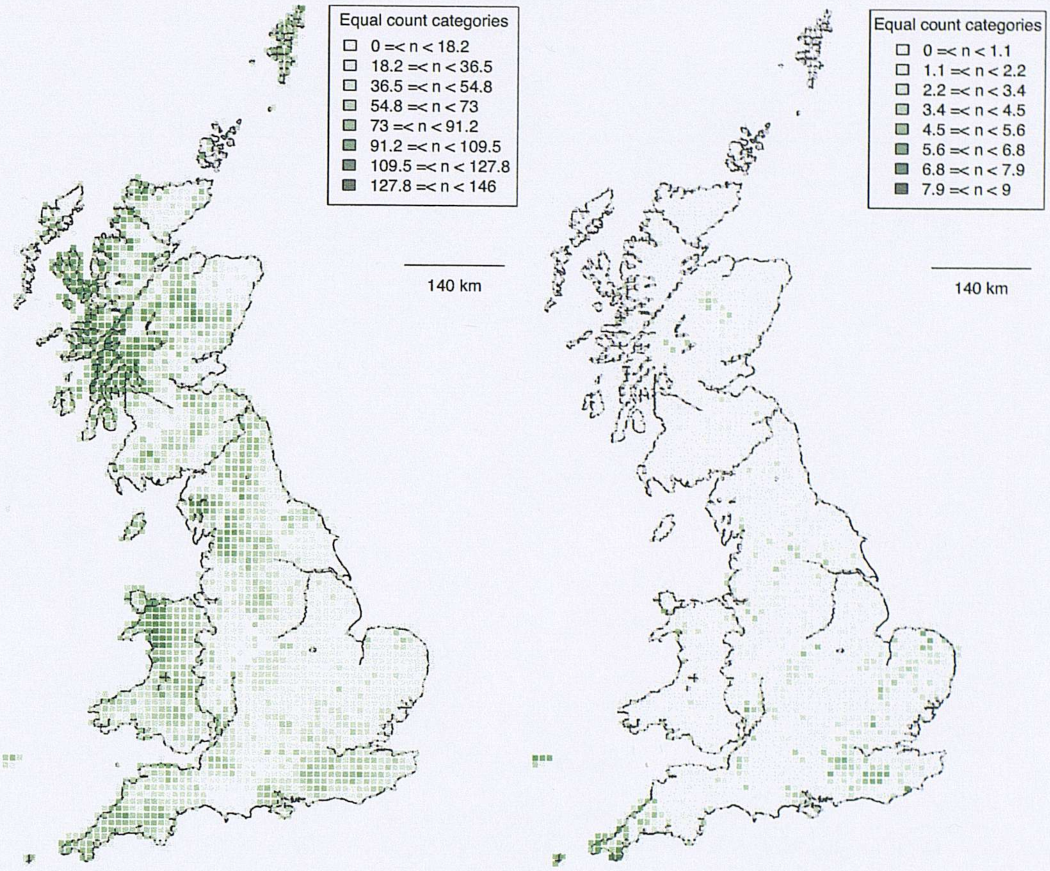
Group	Species count	Species nesting scores				Group nesting	
		Min	Max	Mean($D\star$)	SD	Nest1	Nest2
Liverworts	280	-0.07	0.58	0.23	0.17	0.67	0.676
Mosses	695	-0.03	0.59	0.18	0.14	0.59	0.682

The results of applying these methods to the bryophyte data-set are shown in table 5.1. Of the nesting measures used two, $D\star$ and NEST1, show the liverworts to be more nested than the mosses. NEST2 shows the mosses marginally more nested than the liverworts. Individual species nesting scores range from -0.07 to 0.59. This result indicates that no bryophyte species is entirely nested within the group and that some species show the opposite of nesting, a dissociation from the rest of the group. Coincidence mapping for both classes are shown in figure 5.3.

Table 5.2: Environment models for the nested and unnested bryophyte groups. The nested and unnested groups of mosses and liverworts were modelled in the same way as the groups in chapter 3. All figures to 2DP. Factors with a '+' showed a positive correlation with the group distribution, those with a '−' a negative correlation. Factors are listed in order of their significance.

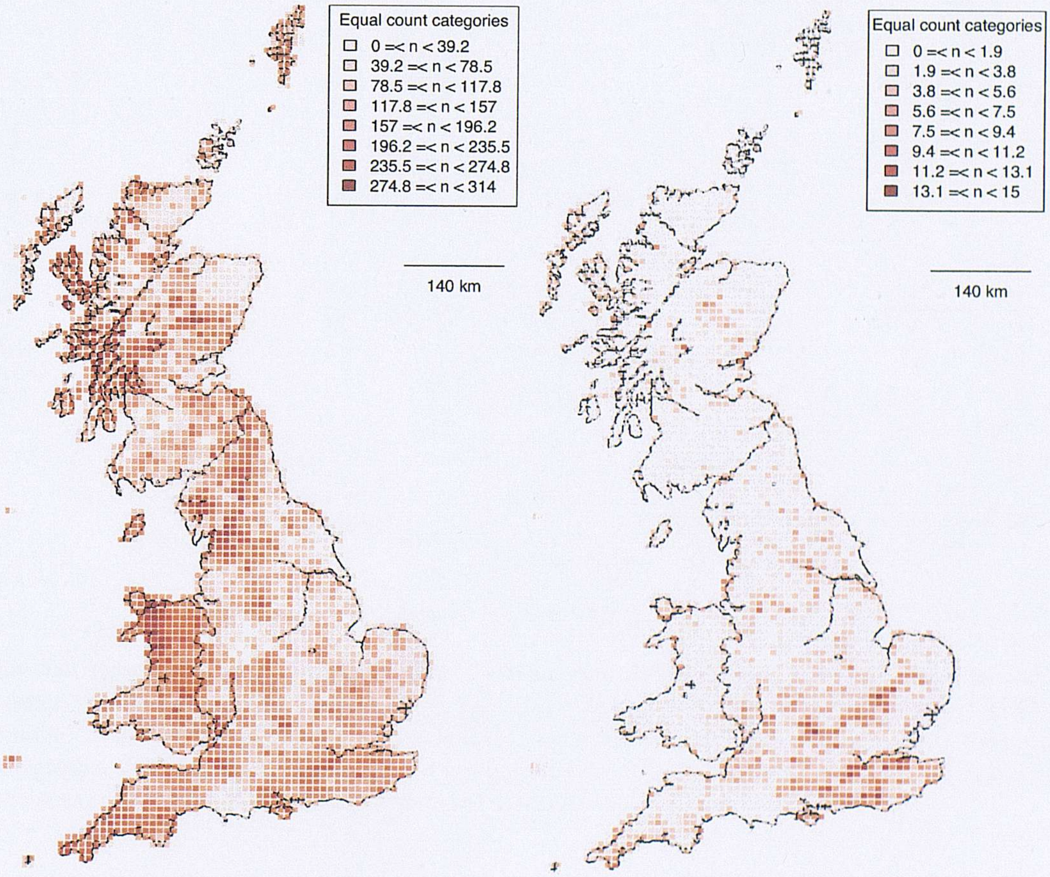
Group	Environmental factors in reduced model	Full model R^2	Reduced model R^2
Nested mosses	+MaxAlt -JanT +JulT +Steepness	0.28	0.26
Unnested mosses	+JulT	0.24	0.17
Nested liverworts	-JanT +Rainfall +Steepness +MaxAlt	0.45	0.44
Unnested liverworts	+JulT	0.18	0.14

The distribution of nesting scores by phytogeographical element is shown in figure 5.4. The final merged groups are shown as box and whisker diagrams. In this figure the overall mean nesting score and one standard deviation in both directions are plotted as lines. Due to the low numbers of species in each element only coarse trends can be considered. As might be expected those species with a more southerly European distribution, as shown by a higher leading digit in their element number, are also those that



(a) Nested liverworts

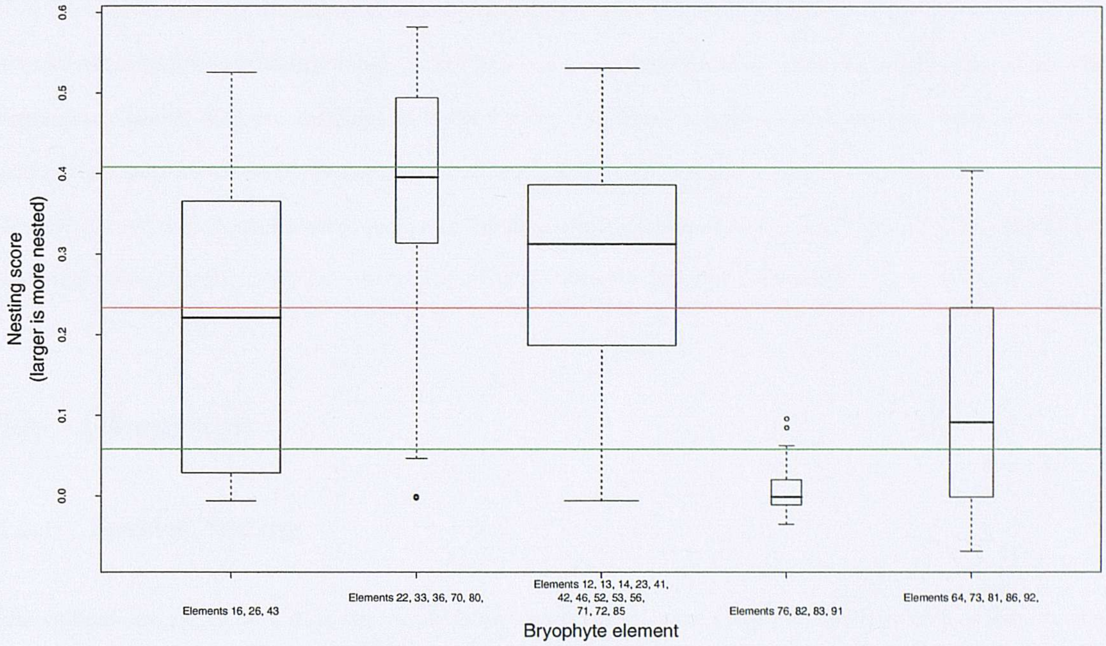
(b) Unnested liverworts



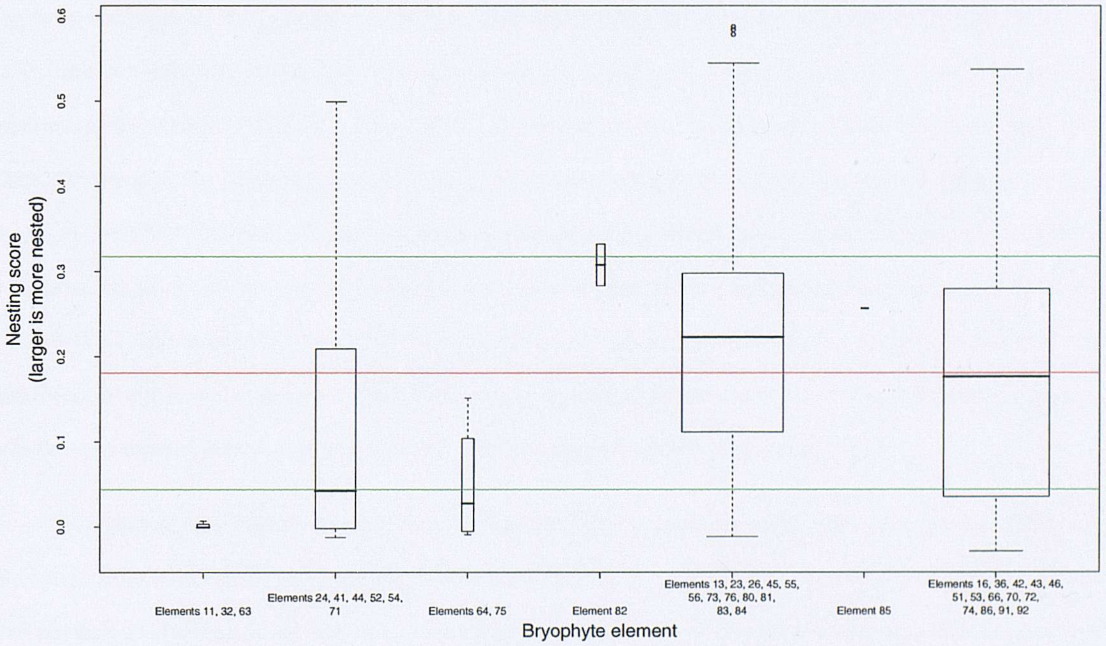
(c) Nested mosses

(d) Unnested mosses

Figure 5.3: Coincidence maps of nested and unnested liverwort and moss species. In both classes species that show nesting display a more north-westerly bias.



(a) Liverwort nesting by element



(b) Moss nesting by element

Figure 5.4: Nesting score by phytogeographic element. Red lines indicate the mean nesting scores for the entire group. Green lines are 1 SD from the mean. Groups of elements were defined by ANOVA and model simplification. Box areas indicate the relative number of species contained within the group. Those groups above the red line show more nesting than average; those below show less. Descriptions of the meaning of element numbers can be found in section 5.4.4.

are least nested. Typically the British distributions of these species do not fit the general north-westerly pattern of British bryophytes.

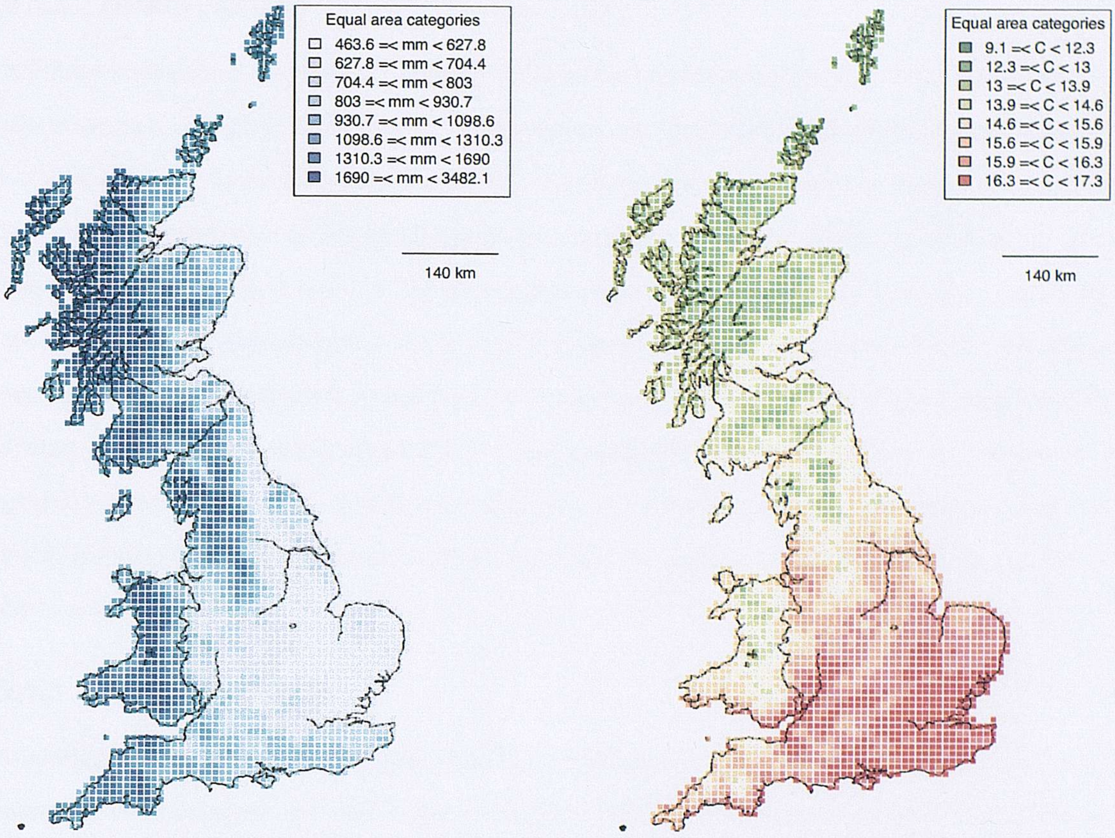
An examination of the environmental modelling results (table 5.2) shows that for all but the nested liverworts these broad environmental factors provide a poor explanation of the distribution patterns. The explanatory power of these variables is worse for the unnested than the nested groups. For the unnested groups the majority of the explanatory power for the full model is accounted for by summer temperature. Despite the apparently calcareous pattern in the unnested distributions (figures 5.3b and 5.3d) chalk does not appear as an explanatory factor for either the unnested mosses or liverworts.

5.6 Discussion

5.6.1 Species Nesting

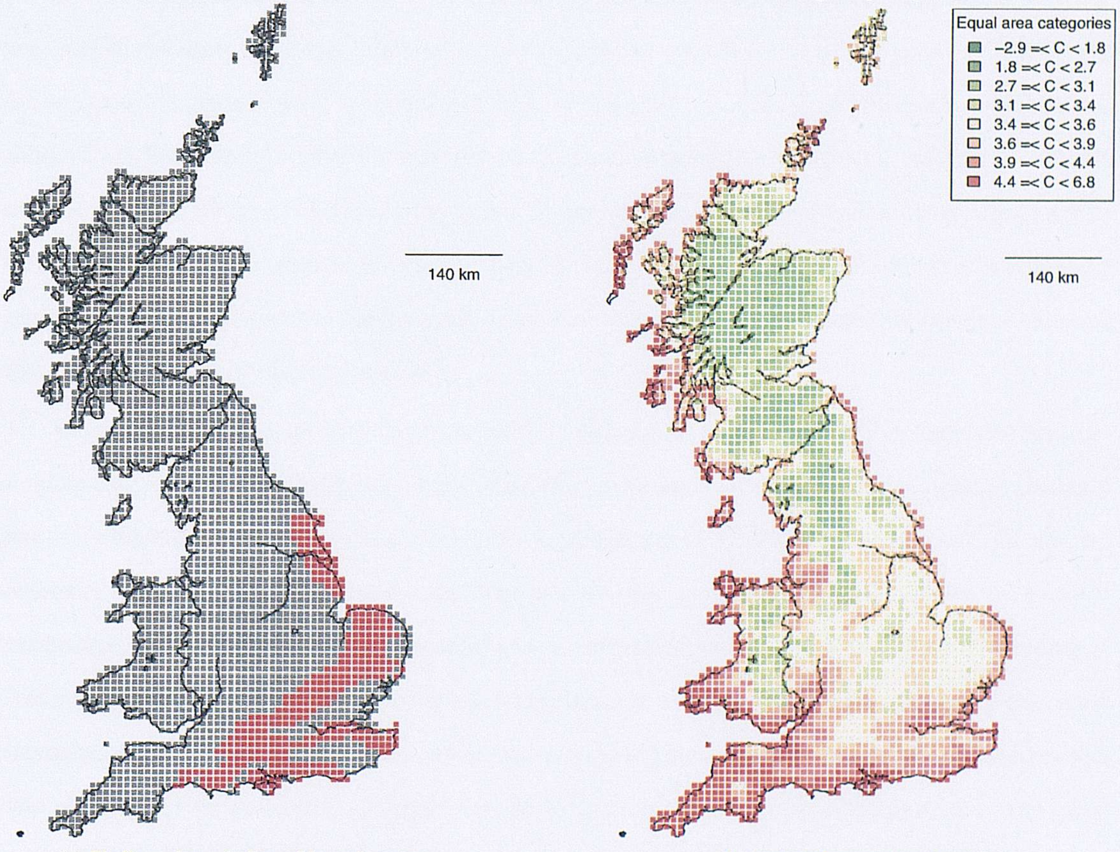
The species nesting scores indicate that in both bryophyte classes there are species that conform to the general nesting pattern and those that do not (see electronic appendix for complete species listings). In the liverwort species the general pattern, focused on the Western Isles, spreads out of the north and west of Britain with the progressive loss of species to the south and east. This is the classical pattern of species-richness in liverworts (Hill et al., 1991-1994). Those species that are not nested have two concentrations. One in Cornwall, the Lizard peninsula in particular, and a second set of aggregations in Kent and the East Anglian counties. The lack of chalk geology as an explanatory factor is somewhat unexpected because at a local scale many of the species defined as unnested appear to have calcareous associations (Hill et al., 1991-1994). The probable reason for the lack of chalk association at a national scale is that overall the unnested group is well dispersed with the few aggregations on the chalk insufficient for the group as a whole to be considered strongly associated with the chalk variable (see figure 5.5c).

The greatest aggregation in the liverwort unnested group contains only 16% of the group's members in comparison to an aggregation of 65% of species for the nested group. The picture of nesting for the mosses is less obvious. Unnested species are even more dispersed, the greatest aggregation of unnested species only accounts for 10% of the unnested group. The nested group is also more dispersed with 57% of species present in the the places of greatest aggregation. Geographically the nested group is much less constrained though it still favours the north and west. Where the unnested species show aggregation there is an obvious south-eastern bias and again what may be considered to be a calcareous signal. However, aggregations in both the unnested moss and liverwort groups are low. With this low aggregation and overall south-eastern bias the only environmental variable to show any association with these groups is summer temperature.



(a) Average annual rainfall

(b) Average July temperature



(c) Chalk geology; indicated by red squares.

(d) Average January temperature

Figure 5.5: The most likely environmental factors to influence bryophyte distribution in Britain. These were taken from the data set used for the environmental analysis in chapter 3.

5.6.2 Group Nesting

All three methods indicate that both bryophyte classes are nested (using $D\star$, $P < 2.2 \times 10^{-16}$) and for two of the measures (D/*star* and NEST1) the liverworts are significantly more nested that the mosses ($P = 8.474 \times 10^{-06}$). In the third method (NEST2) no difference was found between the two classes. This is because this method does not penalise phytogeographically dissimilar groups as much as the other two methods. When sorted by range, the sum of all the discrepancies for species within the group is discounted by the minimum possible discrepancy score obtained from the difference between each species and its nearest phytogeographical neighbour, an arrangement that is favourable to phytogeographically diverse groups. The mosses display more phytogeographic diversity than the liverworts and contain a greater number of subgroups. So this reduced penalty in the group's nesting score, despite being fractured into two or more phytogeographical groups, makes the degree of nesting comparable with that of the liverworts.

5.6.3 Interpretation

Requiring water during the sexual phase, having no true vascular system or cuticle to obtain and retain moisture, the bryophytes and their entire ecology are dominated by water (Schofield, 1985; Watson, 1972). That the patterns of distribution seem to be more strongly influenced by factors such as temperature may at first seem puzzling. There is however a negative correlation in Britain between rainfall and January temperature (see table 3.2 and figure 5.5). At this scale it is difficult to separate January temperature effects from rainfall. That the majority of these plants with poor desiccation tolerance are found in the cooler, wetter areas of the country is not surprising; the unexpected finding here is that despite the physiological and ecological forces acting on these plants, some species not only fail to follow the general distribution pattern but are inversely correlated with it (figures 5.3b and 5.3d and see electronic appendix for listings of unnested species).

For liverwort species the pattern of species nesting scores is clear: those that are unnested are from temperate to southern continental elements, those that are strongly nested are hyperoceanic or northerly and the remaining species fall into a marginally nested group in the middle (figure 5.4a). Of the two classes of bryophyte, the liverworts are the ones that are most water dependent. Typically, the rhizoids (anchorage structures) are single cells in comparison to the multi-cellular structures found in the mosses. This, combined with many species having a single-cell-thick thallus (leaf analogue) and poor desiccation tolerance, means that these species are almost entirely bound to the rainfall pattern. Those species that flout the pattern have developed strategies that allow them to tolerate low rainfall but by the same token are unable either to compete or survive in wetter areas. The single most important factor in determining vascular plant distribution, July temperature (figure 5.5b and chapter 3), is a much poorer descriptor of liverwort distribution than January temperature or rainfall (figure 5.5a). This is reflected in the much greater explanatory power of the nested liverwort model (table 5.2).

The more complex patterns in the mosses comes from greater group size coupled with the phytogeographical and ecological inconsistency in this group. Defined taxonomically this group contains many species as susceptible to drought as the liverworts and so we see a pattern similar to that of the liverworts within the moss distribution. However, many mosses have evolved structures and physiological mechanisms that allow them to tolerate low water conditions. Some have also developed ruderal strategies enabling them to quickly exploit gaps in vegetation and bare ground, in arable fields for example. These additional traits are expressed in this class's more widespread pattern. With the additional range, other ecological factors become more important. A recent study by Hylander and Dynesius (2006) at the much smaller scale of stream-side forests found bryophytes to be nested, this being due to the availability and colonisation of micro-habitats rather than to broad factors such as pH. Similarly, in this study the reduced importance of January temperature (seen in the reduced explanatory power of the model) for the mosses leads to little reduction in the nestedness of the group despite none of the other standard, large-scale ecological factors supplying any additional explanation. Local conditions perhaps via dispersal ability, extinction or colonisation as in the cases of Greve et al. (2005), Patterson (1987) or Loo et al. (2002) are more important than broad country-wide ecological factors.

5.6.4 Assessment of the methods

Table 5.3: Summary of the nesting measures compared to the criteria.

Criteria	D⋆	NEST1	NEST2
Computational requirements	yes	yes	yes
Robustness	yes	yes	excessively
Species score	yes	no	no
Ease of calculation	moderate	good	low
Stability	poor	good	poor

All the methods proposed here adequately met the computational limit requirements set out above (no method took more than 20 minutes to run with the larger moss data-set). This is in contrast to any previously available method. However, performance against the other requirements varied. Both D⋆ and NEST2 are reliant on matrix order in that they compare each species in turn to the one directly above. The use of species means that these measures are robust to geographic variation in recording at the expense of stability. Both are vulnerable to species which are not, distributionally, the same as the rest of the group. D⋆ will penalise two species scores for the addition of one 'odd' species while NEST2 can compensate almost completely for two or more 'odd' species. Because NEST2's final score does not reflect the presence of an ill-fitting species in a group it negates much of the utility of this measure. Stability is also impaired as the addition of 'odd' species can lead to large changes in the final score (in particular, moving from a single ‘odd’ species to two or more). NEST1 being a weighted rank average is buffered against the effects of rogue species while still producing an appropriate response. The lack of a nesting score for each species is a slight hindrance but this is more than made up for by the stability and ease with which this measure can be calculated. Beyond this, because of NEST1’s similarities with standard non-parametric methods, confidence intervals and p values are also fairly simple to calculate. For future studies $D\star$ is perhaps useful when the individual nestedness of a species within a group is required, NEST2 is best avoided and NEST1 should be the method of choice for calculating the degree of nesting within large presence-absence matrices.

5.7 Conclusion

Nesting is strongly displayed at a national scale in both groups of British bryophytes. The liverworts seem to be controlled by single, broad factors across their entire range. While the mosses also show nesting the factors that control this are less clear. They are affected by rainfall but are influenced to a much greater degree by other factors such as geology, suitable micro-sites and temperature. For unnested bryophyte species the only influential broad-scale factor is summer temperature; local factors are probably key in defining these species ranges.

Chapter 6

Scale Dependence in z at intermediate and large spatial scales

6.1 Summary

I investigate the effect of scale on species-area relationships by the use of samples created from the British flora data-set. These samples are taken both via an agglomeration method and in a random manner.

The results show declining trend between area and species count area on a log-log plot. This trend indicates that the power law constant, z, varies with scale. Examination of spatial trends in z shows that changing the initial location of a study will change the resultant value of z.

6.2 Introduction

Species richness, the number of species present in an area, is fundamental to many questions in ecology and conservation. It underlies questions as diverse as niche parameters, predator-prey interactions and siting and maintenance of conservation areas. At the smallest temporal and spatial scales the species number is controlled by basic population processes such as births, deaths, dispersal rates and interactions between individuals and their competitors, predators and mutualists (Rosenzweig, 1995); whereas at larger scales processes such as speciation and extinction take precedence (Pacala, 1997). Overarching all these processes is the effect of the size of the area studied.

Because of its importance the relationship between the number of species present and the size of the study area is perhaps one of the most studied questions in ecology (for a review see Rosenzweig 1995). These relationships have the potential both to answer some of the most profound questions about ecological processes and, if correctly quantified, to allow small-scale, inexpensive investigations to substitute for those on a broader scale.

It is of course obvious that as you increase the area of study you will encounter more species.

Scientifically, the exact form of the relationship between species and area has been the subject of much discussion from the earliest work right up to the present day (Arrhenius, 1921; Adler et al., 2005; Fisher et al., 1943; Gleason, 1922, 1925; Plotkin et al., 2000*a*,*b*; Preston, 1962*a*; Skellam, 1952; Tjorve, 2003). Throughout this time various associations have been proposed. By convention, this is most commonly modelled as a power law, $S = cA^z$ (Arrhenius, 1921). It has been found to best represent field data (Connor and McCoy, 1979) and has been described as one of the few laws of ecology (May, 1975).

There is not a single pattern of species area curves. Williams (1943) identified 3 patterns and this was later expanded by splitting the second and third categories apart, as shown below (Preston, 1962*a*,*b*). Canonically, from Rosenzweig (1995), these categories are:

1. Species area curves amongst tiny pieces of single biotas.
2. Species area curves amongst larger pieces of single biotas.
3. Species area curves amongst islands of one archipelago.
4. Species area curves amongst areas that have separate evolutionary histories.

With the exception of number three, these categories may also be viewed as being points along a continuum of spatial scale. As different factors - including not only the biological, physical and ecological but also sampling effects - have differing effects at different scales (Turner and Tjorve, 2005) this suggests that it may be better to view these as points along a continuum of changing ecological driving factors rather than distinct inseparable categories. Using Rosenzweig's categories the current study is of the second type: a species area curve amongst larger pieces of a single biota.

The effect of area is mediated through the absolute number of species present and the heterogeneity of the species' distributions (Fisher et al., 1943; Preston, 1962*a*; Plotkin et al., 2000*b*). Of these factors the first is expressed in the power law model as the c term. The latter, as well as the interactions of broad ecological processes, is encompassed in the z term. It is for this reason that quantifying the z term is of such importance and although in this study I attempt to quantify z by scale it has also been shown to vary with effects such as latitude, sampling scheme and size of organism (Drakare et al., 2006).

The value of z has, traditionally, been taken to be around 0.25 although reported values vary (Connor and McCoy, 1979; May, 1975; Sugihara, 1980). The largest values of z are found in studies of small to intermediate scale (Crawley and Harral, 2001). With these it is greater on islands (Rosenzweig, 1995) and lower at smaller plant-sized scales (Condit et al., 1996; Hopkins, 1955; Kilburn, 1966). The problem no longer is one of community ecology at the very smallest scale the but one of packing. Because of this, z approaches zero (Williamson, 2003). These studies show that the z term is susceptible to change. In a recent review Turner and Tjorve (2005) consider that the historical disregarding of scale within species-area relations has been a major impediment to the field. Crawley and Harral (2001) studied, in detail, scales of between 0.01m^2 and 100km^2 and found that z followed a quadratic relationship with scale. As a simple demonstration that z cannot be constant we can take the number of native plant species and the area of Britain (1400 and 285100km^2 respectively, from chapter 3) and calculating the value of c ($c = \frac{1400}{(2851 \times 10^2)^{0.25}} = 60.5$) would then allow us to calculate the expected number of species in a ten-kilometre square $S = 60.5 \times 100^{0.25} = 191.3$. The median number of species in such a square in Britain is 449 species and the lower quartile is 359. An actual value as small as 191 species only occurs in about 4% of cases. Crawley and Harral (2001) demonstrate that the mismatch between prediction and reality is worse when the difference in scales is greater and that prediction from a 1m^2 to global scales either vastly overestimates the total global flora or gives a poor representation of species richness at intermediate scales. Changing of any constant is not sufficient and simply alters the scale at which the misfit occurs. So, an Arrhenius model that correctly predicts the species number in a 10km square from the total British species would fail to accurately predict the number of species within a 100km square.

6.3 Aim

In this study I will look at how the relationship between species and area varies for plant species in the British flora. I do this at scales larger, for plants in Britain, than those that have previously been considered in detail. Specifically, I consider z at scales ranging between 100km^2 and 285100km^2. I look at how including a spatially-explicit component causes this relationship to change and how the initial location of a study can affect its results.

6.4 Method

6.4.1 Data

The data for this study were drawn from the *New atlas of the British and Irish flora* (Preston et al., 2002). I included only those species recorded as native and then only those squares in which they were considered to be native. Hybrids were excluded. Where a set of species was particularly difficult to identify then it was included as their aggregate (e.g. *Rubus fruticosus*). This dataset was recorded on the 10 by 10km square grid system developed by the Ordnance Survey. Each of the 1400 species was recorded as a presence or absence in each of the 2851 squares of Britain. These squares formed the basis for the samples used in this chapter.

6.4.2 Sampling

To allow the effect of habitat to be studied, the samples were taken in two ways: by spatially-explicit agglomeration and in a randomised manner. In each case the sample originated from just a single square. In the agglomeration method each square within the data set was used in turn as a starting point. With the randomised method the starting square was simply the first to be drawn at random from all the possible

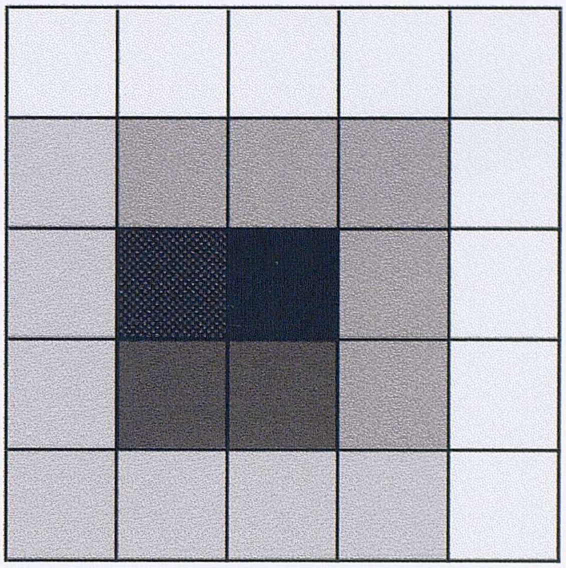


Figure 6.1: Order in which squares are chosen to expand a sample. The sample starts with the black central square and sequentially adds squares in order of increasing lightness.

squares.

Selection through agglomeration

Once an initial square had been chosen (a kernel) the sample was expanded in a set sequence. Because the smallest samples are of particular interest, for each kernel a two-square sample was taken consisting of the kernel and the square immediately to its west. Squares were then added to the south. Further squares were added to the north and east to surround the existing sample. This was then repeated in the south and west. This sequence of L-shaped additions was repeated until all squares in the dataset were included (see figure 6.1). On each expansion the number of squares and the number of unique species present were recorded. Information to allow separation of squares which were only partially land was not available so where these contain species they have been counted as being whole. When a square was selected that contained no land at all (i.e. the sample had expanded beyond the coast of Britain) that square was ignored and was not included in the area of the sample. This is a nested sampling scheme because within a sample all smaller areas are contained within the larger areas.

Samples selected via random square choice

When taking samples in a random manner it was envisaged that all the squares from the dataset had been placed in a common pool. The initial square was drawn at random from this pool and, to increase the size of the sample, further squares were also drawn at random from this pool. Once taken from the pool squares were not replaced. Samples were increased in size from the initial single square to include all 2851 squares in steps equivalent to the theoretical increases in square count for square samples in the agglomeration method (so sizes of 1, 4, 9, 25, 36, etc. were used). A total of 5690 random samples were created in this way. This was almost double the number created via the agglomeration method. This

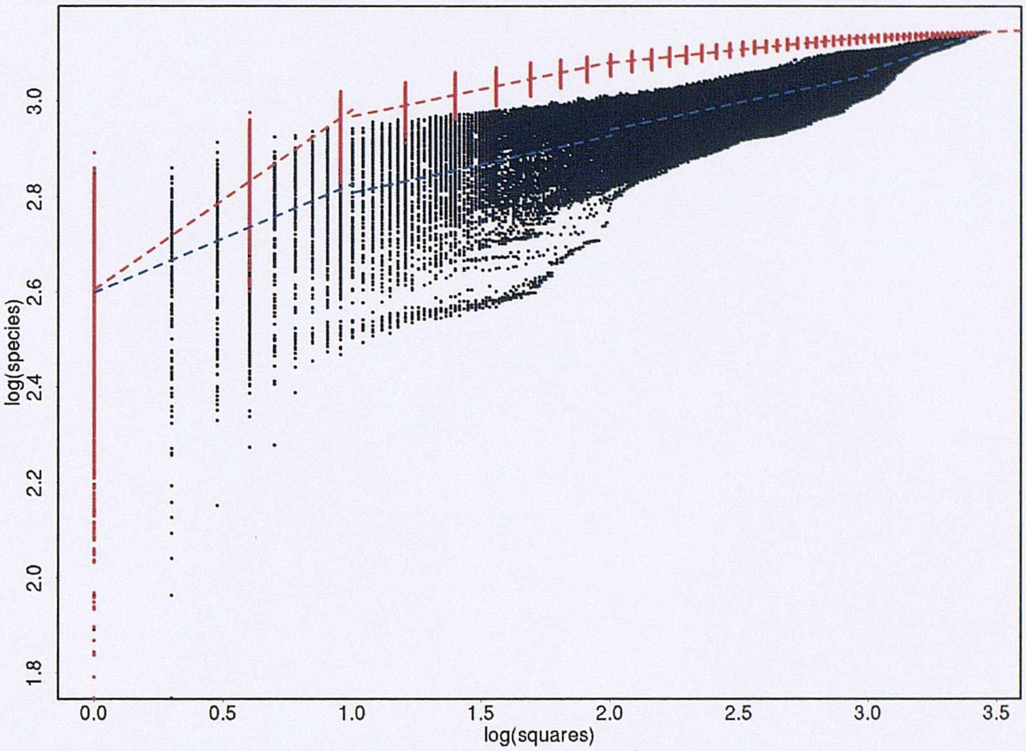


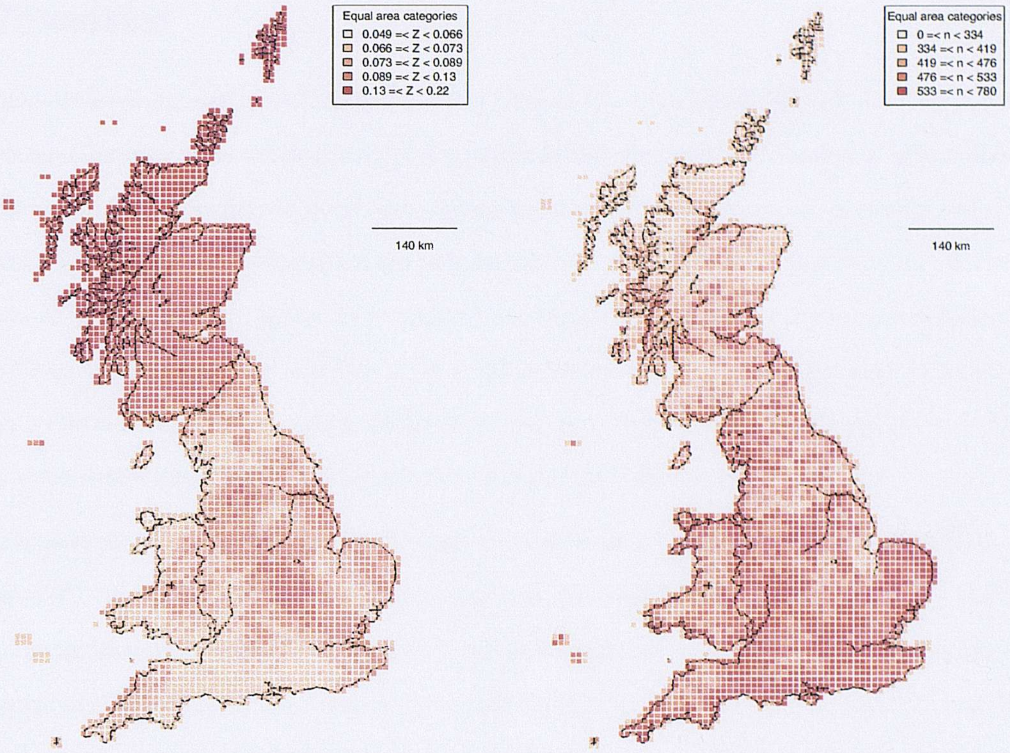
Figure 6.2: Plot of number of squares vs the number of species for the agglomeration method (black) and the random method (red). Piecewise regression results are shown as dashed lines, red for the random, blue for the agglomeration.

method of sampling is equivalent to Rosenzweig's 1995 scattered sub-plots.

6.4.3 Analysis

For each expansion of a sample the number of species and the number of squares along with the starting point or random run number were recorded. To obtain the value of z these square and species counts were log transformed and the gradient calculated. z is the gradient of any line plotted on log transformed species and area axes.

To look at the effect of scale, regression was also carried out in a piece-wise manner on samples with squares in the range: 1 to 9, 10 to 99, 100 to 999 and greater than 1000 squares. Individual z values for every increase in sample size were also calculated. This calculation is simply the gradient of the line, on the log-log plot, between individual points as a sample increases in size. The distribution of these z values allowed the behaviour of individual samples to be considered at varying scales, in aggregate and individually. For each kernel in the agglomeration dataset the median of the individual z values during sampling were calculated. These median values were then plotted on a map to show the geographical distribution of z values (figure 6.4).



(a) Median z by kernel for square in Britain.

(b) Species density for each square in Britain.

Figure 6.4: Spatial distribution of median z values across all scales for a square plotted with species density of the British flora so a comparison may be made. This is a spatially-explicit version of the median line in figure 6.3 at the median value of scale (10^4 km^2). Areas of low species density have the greatest median z values. This is an effect of their requiring the greatest increase in the number of species to gain the entire flora of Britain. Central England's high z values are partially controlled by the species richness effect but also in the non-uniform manner that these samples accumulate species.

z values for the piece-wise regressions are shown in Table 6.1. The P values for all of these regression models are smaller than 2 x 10^{-16} although care should be taken with this value because it is calculated from repeated measures.

These changes in the value of z as scale is increased are even more noticeable when the individual z values for each change are considered. Figure 6.3 clearly demonstrates these changes with a decline in z with increase at small scales, a constant region between 10^3 km^2 and $10^{4.5}$ km^2 and then finally an increase in z at the very largest scales.

Median z values for a set aggregation across scales are strongly affected by the species count of the starting kernel. Figure 6.4 shows these median z values for each kernel plotted alongside species density for the British flora. For a samples with squares randomly selected, the initial square's species count has no effect on the median value of the z for that sample. This is due to the rapid dilution of any local effects by national squares. With random samples made up from four squares, three quarters of the squares chosen have no relationship with the first square. This dilution further increases as the samples get larger.

6.6 Discussion

The method of sampling used, in effect the sampling pattern, has a profound effect on the final shapes of the species-area curves and the distribution of z values at any one scale. The randomly accumulated samples demonstrate a smooth asymptotic curve on log-log axes with greater variance in species count at smaller values of area. The agglomeration samples demonstrate a more stepwise pattern in the species-area curve. At scales between $\log^{-1}(1.5) \times 100$ km^2 and $\log^{-1}(3) \times 100$ km^2 there is clear separation between the curves. Study of the distribution of z values for these curves shows that random samples have high z values at small scales rapidly falling to almost zero. Agglomeration samples can have very low z values at almost any scale. It is these two factors that create these distinctive curves.

At virtually all of the scales used in this study the values of z found were significantly less than the figure of 0.25 that is typically reported in the literature. Only random aggregation of squares shows a z value higher than the standard figure and this only at the smallest scales (100 - 900 km^2 via piece-wise regression). The agglomeration data-set at no point shows typical values greater than 0.25 and the greatest values of z are at the smallest scales. At small scales the value of z declines as scale is increased. Values of z are smaller than those found from previous work at these scales (Crawley and Harral, 2001). At medium scales z is stable with a value around 0.1 while at the largest scales z again increases. This result is similar to that reported by Preston (1960) for land birds in Pennsylvania although the flattening in this study occurs at a larger scale. Although also consistent with triphasic SAR (Rosenzweig, 1995) the slight upturn in the trend at the very largest scale should be considered with caution as the system is constrained to having all species present at the greatest scale.

The low values of z at larger scales are not altogether surprising. Lower values of z correspond to less variance in assemblage (Southwood et al. 2006). Although the flora of Britain varies from arctic-alpine species through to Mediterranean species (Preston and Hill 1997), when compared to the standard single-site field study where the typical value is observed, local differences in species assemblages are far greater than those across the whole country. That is, β diversity is greater than δ diversity within the British flora. This is typical for such temperate systems (Crawley, 1997*b*). We could see this were we to survey a hedgerow in southern Britain. It would be unlikely to supply many different species than a hedgerow in Scotland, one in France or indeed any hedgerow in northern Europe. A move into field, ditch or woodland would cause, relatively, a much greater change in the species found.

Working at these smaller scales, 0.01m^2 to 100km^2, Crawley and Harral (2001) found z values in the range 0.530 to 0.088. Species-area relationships at these scales are governed by ecological processes that control plant-plant interactions, species assemblage rules and habitat boundaries. The largest repeated unit in that study is the smallest in this. At the scales I consider here each square will contain many such habitats and so mask the effect of small-scale ecological processes. Controlling factors for species-area relations in this study are landscape-scale processes such as those that define those patterns found in chapter 3.

The abundance of species over various spatial scales has been argued to follow a fractal relationship. A fractal relationship would indicate an accelerating increase in species number with area on the log-log plot (Lennon et al., 2002). The results presented here show a completely opposite trend with a deceleration in the rate of species accumulation (see figure 6.2). This result is not altogether unexpected. Many factors that could affect the distribution of plants demonstrate some degree of fractal nature, for example topography, river systems and geology (Mandelbrot, 1982; Rodríguez-Iturbe and Rinaldo, 1997). However, those factors that are not fractal in nature, temperature and rainfall (Lennon et al., 2000), are precisely those that are most important in defining the distribution of British plants (see chapter 3).

With agglomeration sampling the species richness of the initial square is strongly related to the median z value of the entire sample. Low initial species richness leads to greater z values overall. This occurs not only because these samples have to accumulate more species but also because habitats with similar numbers of species are aggregated. Species-poor squares are likely to have species-poor neighbours. As the sample is aggregated the pattern of species richness leads to a stair-step curve with little change in species across several increases in area, giving low z, and then a jump in species count as a new habitat is encountered, leading to high z values. This low β diversity leading to greater z values, along with the spatial aggregation of these z values can be seen in figure 6.4.

The value of z is depressed throughout this study because of the similarity of species lists in widely separated squares (see the large numbers of species in the ubiquitous group in chapter 3). This is due to the correspondence of habitats across the study area and will likely not hold true across scales or for all areas of the world. Areas with particularly high δ diversity, such as South Africa, should display greater values of z at these scales as should a similar study at a European scale using 50km squares. However, the low values of z across parts of the temperate zone may well go some way to explain the miscalculation in the global floral diversity outlined in the introduction.

6.7 Conclusion

At intermediate and large scales, 100km^2 to $285{,}100\text{km}^2$, z is affected by two main factors. The first is that z has a declining trend with respect to initial area size. Large-scale studies will return z values lower than similar smaller-scale studies. The second factor is the location of the study area. Should a species-poor site be chosen as the starting point, z across the whole study will be greater than had a more species-rich site been chosen and so adversely affect any conclusions drawn.

Chapter 7

Discussion

7.1 Introduction

In this chapter I draw together the key findings of the preceding chapters and examine the overriding principles that control distribution. Species distributions have been shown to form repeating patterns. By grouping these patterns I am able to show the relative importance of various ecological factors to the individual plants involved and how, at larger scales, species behave. The non-random distribution of plant species has also been shown to produce interesting geographical effects and to be highly influential in the modelling of plant ecology. Before discussing the final findings of this thesis I address more general matters that have not been specifically covered in the preceding chapters as they have a bearing on how atlas data is used, analysed and interpreted.

7.2 Atlas data

7.2.1 The nature of atlas data

Atlas data, as with all scientific recording, is not absolute. The recorded distributions of species within the study are reflections of the absolute species ranges filtered through the mechanisms of recording and verification. The more effort that is expended on ensuring even coverage, accurate species identification and the verification of records, the closer the distributions recorded in an atlas conform to reality.

For each species, recording starts with a blank map and places dots where species have been reported to be present. For atlases the records of presence are based on fact. New records for vice-counties and squares undergo a high degree of scrutiny before acceptance. It is where a species has not been recorded as present (i.e. an absence) that is of particular difficulty in numerical analysis. An absence leads a dual existence: it may be that the species truly does not exist in that square or that the species simply has not been recorded there. Despite its difficulties this is manifestly better than attempting to prove that each species does not exist in each square!

When no limit is placed upon the effort expended in surveying a square (typically measured as

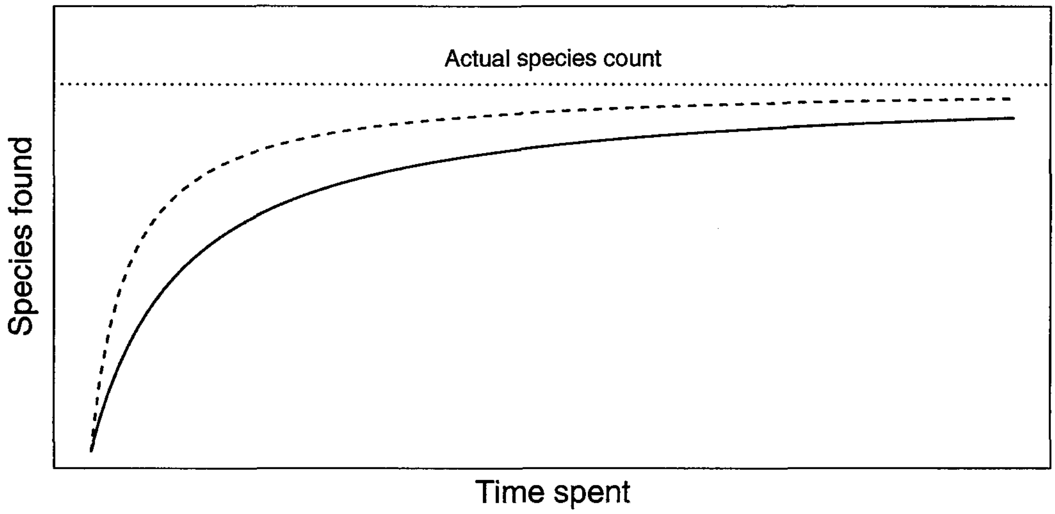


Figure 7.1: The relationship between recorder effort (measured as time spent) and species count. The actual number of species present is indicated by the dotted line, simple recording as the solid line and simple recording with checking mechanisms as the dashed line. With infinite effort both lines converge on the actual species count but with realistic time-scales any technique will under-record the number of species to a degree.

recording time), the overall species count will follow the law of diminishing returns when plotted against the effort. That is, initial effort will yield many species relatively quickly but, as the search goes on, fewer new species will be found (figure 7.1). This is a standard result of probability theory. The chance of recording a particular species when it exists in a square is highly dependent on recording effort. These however, are not the only variables. Factors such as how dense the population or how cryptic the habit of a species is or the square's location within a species' range, affecting a recorder's expectations, will also play a part in determining if a species is found.

Knowing that we can never, at least in realistic time-scales, obtain a list of all the species in a square through free recording, what processes can be employed to ensure that survey returns are as comprehensive as possible? Standard checking mechanisms indeed do much to negate the law of diminishing returns. Expressly seeking missing species ensures that they are more likely to be found as is having multiple recorders with differing biases towards the taxa they find. Both are effective techniques as they increase the number of species found for a given amount of effort. These strategies push the curve higher from the solid to the dashed line in figure 7.1. However, no technique or combination of techniques will ensure location of all species.

Numerically, solutions to the problem of absence data are limited. Without knowing the effort that has been expended in the recording of a square we are unable to estimate the number of species unrecorded. Where possible, we are able to limit the effect of poor absence data by assuming that squares that are poorly recorded are poorly recorded for all species equally. As in chapter 5, by treating each square separately and considering all species recorded and unrecorded within it, 'missing' presences become less important than recorded presences for all but those species with the most restricted ranges. Diligent recording and verification, as in the *New Atlas of the British and Irish Flora* (the New Atlas) (Preston et al., 2002) leads to a high degree of confidence in the absence data and it is from this practical perspective that these problems are best tackled.

7.2.2 Future atlases

The work involved in the production of the *New Atlas of the British and Irish Flora* was impressive. Over 1600 recorders between them submitted in excess of 4.8 million new records to the project. However, the final maps are plotted on what is, at 10km by 10km, a very coarse grid to an individual plant. Many county floras are produced on much finer grids, perhaps 2km by 2km (tetrad) or even 1km by 1km. Should a future national atlas project consider mapping at one of these finer scales? From a practical perspective, even moving to a tetrad-grid system would require a 25 fold increase in effort to maintain the standard of recording. The benefits of such a move are decidedly unclear. Many of the insights that could be gained through the shift to a finer grid, such as how species groupings behave at different scales, can be as easily obtained through studies of the county floras themselves. National coverage at a fine scale is not necessary although fine-scale mapping of regions of particular interest during the course of such a survey may help in questions of scale dependence. Of more importance from an analytical viewpoint would be the inclusion of some measure of recorder effort. Rich et al. (1996) conducted a highly structured survey of the Ashdown Forest in east Sussex. In this survey each recorder visited every area and logged recording time. For a survey such as that for the New Atlas this would be impractical (Preston et al., 2002). However, in many cases the inclusion of time spent in a square combined with a uniquely identifiable recorder number with a record would be sufficient to build a picture of the effort expended.

7.3 Major findings

7.3.1 Defining geographical groupings

The method defined in chapter 2 provides a way of objectively classifying species data without *a priori* assumptions. The groupings it defines are geographic rather than ecological. By not using explicit ecological factors in group definition and using only geographic data, any and all of the ecological factors that affect the distribution of plants are included implicitly. We do not have to start with an initial 'guess' at which factors may be important. The fact that the new method is able to identify, say, a group of ecologically distinct, sandy-soil loving species (see figure 3.3c on page 50) emphasises the importance of approaching phytogeographical classification without preconceptions. The presence of this group would have gone unnoticed if broad ecological factors had been used instead of geography.

The groups produced in chapters 3 and 4 provide a strong basis for testing biogeographical theories. Individually, the classifications have uses in defining broad-scale niches and providing informative data on the relative importance of various ecological factors which determine plant distribution. Together they show how species distribution behaves at the 10 km^2 and the 50 km^2 scale. The fact that at the world scale species are not ubiquitous is recognised in the biome model of species distribution.

We should not expect communities and ranges to remain the same. Plants that share the same geographical range today are unlikely to have had the same distribution throughout time (Davis, 1987). This is especially true when we consider how species originate and associate to form communities. Only species which evolved together and spread together will have maintained consistent ranges, a trick that only obligate symbionts are likely to perform. The groups, as aggregations of species, can provide robust descriptions of how plant communities are changing. For many species future changes between the groups found are possible. The buffering effect against individual species' peculiarity allows us to be more certain of how changes are occurring, particularly as range boundaries change and species migrate.

The classifications also provide information for practical uses. By determining floristic groupings it is now possible to determine areas of importance to conservation. Groups, in this case, provide a stronger inference that an area is important than the behaviour of single indicator species. These areas will not always be the same as those identified through hotspot analysis (Prendergast et al., 1993). The areas of most importance under hotspot analysis tend to be the areas of intersection between groups rather than a group's core range. These classifications also allow an experimental study of a taxon to be compared or contrasted with that of other group members or with members of other groups entirely. Communities of species can be examined through the use of the species lists found in the electronic appendix.

7.3.2 Effect of geographical groupings

Spatial aggregation in species distributions allows species to be placed into phytogeographical groups. This aggregation has also been seen to show effects in other population processes and affect the modelling of species patterns (chapters 5 and 6). As these patterns are externally driven by ecological factors, mostly large scale but occasionally very localised, these wider population processes can be linked back to the individual ecologies of the species concerned.

Highly nested species groups have been seen to be controlled by single broad factors across their entire range (liverworts) but equally the distribution of these nested groups can seem unconnected to broad ecological factors perhaps being driven to this pattern by cryptic factors (mosses). Those species that display only weak nesting do not form a coherent phytogeographical group. This might be expected from a series of species that share only a lack of commonality with a distribution different from their own. The overall pattern of unnested species (those that do not behave as typical bryophytes) is loosely related to summer temperature. This factor naturally correlates with increasing species diversity amongst vascular plants. Perhaps these atypical bryophyte species, having escaped from the overriding effect of water, are now being influenced by the factors that control vascular plant distribution.

On average, at national spatial scales, homogeneity in species distribution causes depression in the modelling parameter used to simulate spatial patterns and heterogeneity of species distributions within species-area relationships, z. At more local scales this same 'patchiness' causes the value of z to vary considerably. This variation leads to two main conclusions. The first is that the traditional models scale but not as predictably as first thought. Care should be observed when findings from local studies are used to describe national trends. Leading on from this, the second conclusion is that locations for field studies should be chosen so that they neither overestimate such parameters by choosing sites with low initial species diversity nor underestimate the parameters by selecting sites that are very diverse as a starting point.

Of all the methods used in this thesis only the final species-area relationship method is explicitly spatial. The other methods instead rely on the species data to supply the spatial information. As in the case of using implicit rather than explicit ecological factors in the methods chosen, by not including explicit spatial information I remove preconceptions of how species distributions should behave. Should a species distribution naturally be patchy these methods will allow this to be accurately reflected. Although I do expect species distributions to be spatially auto-correlated (i.e. an occupied square on a distribution map is more likely to be surrounded by other occupied squares) this does not present a difficulty to these other methods. Whereas in other analyses this spatial auto-correlation would need to be accounted for and removed, here it is essential to mapping groups of distributions.

7.4 Future research directions

Beyond the obvious application of the methods developed here to other atlas-based data-sets, several other research directions have presented themselves during the course of the study. I cover them briefly below.

7.4.1 Null models

Underlying much of science, and ecology in particular, are rigorous statistical tests. It is, for example, perfectly possible to measure the difference between the means of two sets of numbers and find if they are statistically distinct. To do this we build what is known as a null model, an expectation of how the two data-sets would behave if they were not different. From this model we are able to tell how different the means must be, given the variation in the data-sets that produced them, in order for us consider them to be different. For means, as for many statistical parameters there are formulae-based analytical methods to calculate the null models. When performing a t-test, as in this example, we don't need to carry out simulations of hundreds of data-sets to find if our two sets of numbers are different. Although methods exist to create null models both for phytogeographic patterns (Roxburgh and Chesson, 1998) and for patterns of nesting (Brualdi and Sanderson, 1999; Sanderson et al., 1998) all are still reliant on simulation and so require a great deal of computing power to generate. Indeed for data-sets of the size used in this project many of these methods are impractical. Future research effort should be directed toward the mathematical quantification of distribution patterns allowing simulations to be replaced with calculated probabilistic models.

7.4.2 Scale

We have seen in the difference in the results between chapters 3 and 4 that scale has an profound effect. At national scale many plants have ubiquitous distributions. This is not true when either continental scale patterns or when scales less than $1\mathrm{m}^2$ are considered (Crawley et al., 2005). We have also seen how scale affects the modelling of species-area relationships. Further study, including investigations at the local scale, should attempt to investigate and quantify the relationship between range size, study area size, the groups found and their distributions.

7.5 Conclusion

The common theme running throughout this work is that heterogeneity in species distribution patterns is engendered by a multitude of ecological factors. Because patterns of species distribution are spatially uneven and repeat across many taxa they permit classifications of distributions to be created and influence the outcome of theoretical and practical studies. The novel classifications presented in this thesis provide a detailed reference point for future studies and combined with the later work in this thesis offer an insight into the current relationships between the species.

Chapter 8

Conclusions

- Repeated patterns of species distribution can be objectively enumerated using cluster analysis to form groups that have botanical relevance. A two-stage analysis allowed the efficient grouping of species that are difficult to place.
- Classification of the British flora by plant distribution rather than surrogates allowed otherwise cryptic patterns to be observed and provides a strong baseline for future biogeographical study.
- The classification of the European flora reveals that groups at continental scale are relatively localised when compared to those at a national scale. Group count is correlated with species richness but with important exceptions for major geographical features.
- Both liverwort and moss species show a nested pattern to their distributions. For liverworts this correlates with the broad climatic factors of winter temperature and rainfall. The nested pattern of moss species show only weak correlation with ecological factors, in particular with altitude and winter temperature.
- When studying species-area relationships both the scale and the starting point of the study will profoundly influence the values of the parameters found. Species-poor areas give higher overall z values than species-rich areas. Studies at larger scale will give smaller z values than their smaller-scale counterparts.
- The distribution of species is non-random. Collation and interpretation of this information can provide insights beyond the locations of the species studied.

Bibliography

Adler, P. B., E. P. White, W. K. Lauenroth, D. M. Kaufman, A. Rassweiler, and J. A. Rusak. 2005. Evidence for a general species-time-area relationship. Ecology **86**:2032–2039.

Arrhenius, Ö. 1921. Species and Area. Journal of Ecology **9**:95–99.

Atmar, W., and B. D. Patterson. 1993. The Measure Of Order And Disorder In The Distribution Of Species In Fragmented Habitat. Oecologia **96**:373–382.

Belbin, L., and C. McDonald. 1993. Comparing three classification strategies for use in ecology. Journal of Vegetation Science **4**:341–348.

Birks, H. J. B. 1973. Past and Present Vegetation of the Isle of Skye. Cambridge University Press, Cambridge.

Birks, H. J. B. 1976. The distribution of European pteridophytes: a numerical analysis. New Phytologist **77**:257–287.

Birks, H. J. B., and J. Deacon. 1973. A Numerical analysis of the past and present flora of the British Isles. New Phytologist **72**:877–902.

Braak, C. J. F. t. 1986. Canonical Correspondence Analysis: A New Eigenvector Technique for Multivariate Direct Gradient Analysis. Ecology **67**:1167–1179.

Braithwaite, M. E., R. W. Ellis, and C. D. Preston. 2006. Change in the British Flora 1987-2004. Botanical Society of the British Isles, London.

Brown, J. H. 1986. Two decades of interaction between the Mac-Arthur-Wilson model and the complexities of mammalian distributions. Botanical Journal of the Linnean Society **28**:231–251.

Brown, J. H., and M. V. Lomolino. 1998. Biogeography (2nd Edition). Sinauer Associates, Sunderland, Mass.

Brown, R. L., and R. K. Peet. 2003. Diversity and invasibility of southern Appalachian plant communities. Ecology **84**:32–39.

Brualdi, R. A., and J. G. Sanderson. 1999. Nestled species subsets, gaps, and discrepancy. Oecologia **119**:256–264.

Burkhardt, F., and S. Smith, editors. 1990. The Correspondence of Charles Darwin (1856-1857). Cambridge University Press, Cambridge.

Cain, S. A. 1947. Characteristics of natural areas and factors in their development. Ecological Monographs **17**:185–200.

Carey, P. D., C. D. Preston, M. O. Hill, M. B. Usher, and S. M. Wright. 1995. An environmentally defined biogeographic zoneation of Scotland designed to reflect species distributions. Journal of Ecology **83**:833–845.

Condit, R., S. P. Hubbell, J. V. Lafrankie, R. Sukumar, N. Manokaran, R. B. Foster, and P. S. Ashton. 1996. Species-Area and Species-Individual Relationships for Tropical Trees: A Comparison of Three 50-ha Plots. The Journal of Ecology **84**:549–562.

Connor, E. F., and E. D. McCoy. 1979. The Statistics and Biology of the Species-Area Relationship. The American Naturalist **113**:791–833.

Cook, R. R., and J. F. Quinn. 1995. The Influence Of Colonization In Nested Species Subsets. Oecologia **102**:413–424.

Cox, D. R. 1968. Notes on some aspects of regression analysis (with Discussion). Journal of the Royal Statistical Society B **49**:1–39.

Crawley, M. J., editor, 1997*a*. Plant Ecology, Pages 475–531 . Blackwell Science, Oxford.

Crawley, M. J., editor. 1997*b*. Plant Ecology. Blackwell Science.

Crawley, M. J., S. L. Brown, M. S. Heard, and G. R. Edwards. 1999. Invasion-resistance in experimental grassland communities: species richness or species identity? Ecology Letters **2**:140–148.

Crawley, M. J., and J. E. Harral. 2001. Scale Dependence in Plant Biodiversity. Science **291**:864–868.

Crawley, M. J., A. E. Johnston, J. Silvertown, M. Dodd, C. de Mazancourt, M. S. Heard, D. F. Henman, and G. R. Edwards. 2005. Determinants of species richness in the park grass experiment. American Naturalist **165**:179–192.

Dansereau, P. 1957. Biogeography: An Ecological Perspective. Ronald Press Co., New York.

Davis, M., 1987. Invasions of forest communities during the Holocene: beech and hemlock in the Great Lakes Region. Pages 373–393 *in* A. Gray, M. Crawley, and P. Edwards, editors. Colonization, Succession and Stability. Blackwell Scientific Publications, Oxford.

Davis, M. A., J. P. Grime, and K. Thompson. 2000. Fluctuating resources in plant communities: a general theory of invasibility. Journal of Ecology **88**:528–534.

Deutschewitz, K., A. Lausch, I. Kuhn, and S. Klotz. 2003. Native and alien plant species richness in relation to spatial heterogeneity on a regional scale in Germany. Global Ecology and Biogeography **12**:299–311.

Domingos, P., and M. Pazzani. 1997. On the Optimality of the Simple Bayesian Classifier under Zero-One Loss. Machine Learning **2-3**:103–130.

Drakare, S., J. J. Lennon, and H. Hillebrand. 2006. The imprint of the geographical, evolutionary and ecological context on species-area relationships. Ecology Letters **9**:215–227.

Ebach, M. C., and D. F. Goujet. 2006. The first biogeographical map. Journal of Biogeography **33**:761–769.

Elton, C. S. 1958. The ecology of invasions by plants and animals. Methuen, London.

Fischer, J., and D. B. Lindenmayer. 2002. Treating the nestedness temperature calculator as a "black box" can lead to false conclusions. Oikos **99**:193–199.

Fisher, D. R. 1968. A Study of Faunal Resemblance Using Numerical Taxonomy and Factor Analysis. Systematic Zoology **17**:48.

Fisher, R. A., A. S. Corbet, and C. B. Williams. 1943. The Relation Between the Number of Species and the Number of Individuals in a Random Sample of an Animal Population. Journal of Animal Ecology **12**:42–58.

Gaston, K., and T. Blackburn. 2000. Pattern and Process in Macroecology. Blackwell Science, Oxford.

Gilbert, B., and M. J. Lechowicz. 2005. Invasibility and abiotic gradients: the positive correlation between native and exotic plant diversity. Ecology **86**:1848–1855.

Gleason, H. 1926. The individualistic concept of the plant association. Bulletin of the Torrey Botanical Club **53**:7–26.

Gleason, H. A. 1922. On the Relation Between Species and Area. Ecology **3**:158–162.

Gleason, H. A. 1925. Species and Area. Ecology **6**:66–74.

Godwin, H. 1975. The History of the British Flora. 2nd edition. Cambridge University Press, Cambridge.

Greve, M., and S. L. Chown. 2006. Endemicity biases nestedness metrics: a demonstration, explanation and solution. Ecography **29**:347–356.

Greve, M., N. J. M. Gremmen, K. J. Gaston, and S. L. Chown. 2005. Nestedness of Southern Ocean island biotas: ecological perspectives on a biogeographical conundrum. Journal of Biogeography **32**:155–168.

Griffiths, G. H., B. C. Eversham, and D. B. Roy. 1999. Integrating species and habitat data for nature conservation in Great Britain: data sources and methods. Global Ecology and Biogeography **8**:329–345.

Guisan, A., T. C. Edwards, Jr., and T. Hastie. 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecological Modelling **157**:89–100.

Hagmeier, E. M., and C. D. Stults. 1964. A Numerical Analysis of the Distributional Patterns of North American Mammals. Systematic Zoology **13**:125.

Harding, P. T., and J. Sheail, 1992. The Biological Records Centre: a pioneer in data gathering and retrieval. Pages 5–19 *in* P. Harding, editor. Biological recording of changes in British wildlife. HMSO, London.

Hartigan, J. A. 1975. Clustering Algorithms. Wiley, New York.

Hastie, T. J., and R. J. Tibshirani. 1986. Generalized Additive Models. Statistical Science **1**:297–318.

Hill, M. O. 1974. Correspondence Analysis: A Neglected Multivariate Method. Applied Statistics **23**:340–354.

Hill, M. O., 1979. TWINSPAN: A fortran program for arranging multivariate data in an ordered, two-way table by classification of the indivduals and attributes. Technical report, Cornell University, New York.

Hill, M. O., and H. G. Gauch. 1980. Detrended correspondence analysis: An improved ordination technique. Plant Ecology **42**:47–58.

Hill, M. O., and C. D. Preston. 1998. The geographical relationships of British and Irish bryophytes. Journal of Bryology **20**:127–226.

Hill, M. O., C. D. Preston, and D. B. Roy. 2004. PLANTATT: Attributes of British and Irish plants: Status, size, life history, geography and habitats. Centre for Ecology and Hydrology, Huntingdon.

Hill, M. O., C. D. Preston, and A. J. E. Smith, editors. 1991-1994. Atlas of the bryophytes of Britain and Ireland. Harley Books.

Holloway, J. D. 1969. A numerical investigation of the biogeography of the butterfly fauna of India, and its relation to continental drift. Biological Journal of the Linnean Society of London **1**:373–385.

Holloway, J. D., and N. Jardin. 1968. Two approaches to zoogeography: a study based on the distributions of butterflies, birds and bats in the Indo-Australian area. Proceedings of the Linnean Society of London **179**:153–188.

Hopkins, B. 1955. The Species-Area Relations of Plant Communities. The Journal of Ecology **43**:409–426.

Humboldt, A. v. 1805. Essai sur la géographie des plantes. F. Schoell, Paris.

Hutchinson, G. E., 1957. Cold Spring Harbor Symposium in Quantitative Biology. Pages 415–457 , volume 22.

Hylander, K., and M. Dynesius. 2006. Causes of the large variation in bryophyte species richness and composition among boreal streamside forests. Journal of Vegetation Science **17**:333–346.

Ihaka, R., and R. Gentleman. 1996. R: A Language for Data Analysis and Graphics. Journal of Computational and Graphical Statistics **5**:299–314.

Jaccard, P. 1908. Nouvelles recherches sur la distribution florale. Bull. soc. vaud. sci. nat. **44**:223 – 276.

Jalas, J., and J. Suominen. 1972. Atlas Florae Europaeae, Volume 1 (1972). Committee for Mapping the Flora of Europe and Societas Biologica Fennica Vanamo, Helsinki.

Jalas, J., and J. Suominen. 1972-1999. Atlas Florae Europaeae, Volume 1 (1972), 2 (1973), 3 (1976), 4 (1979), 5 (1980), 6 (1983), 7 (1986), 8 (1989), 9 (1991), 10 (1994). Committee for Mapping the Flora of Europe and Societas Biologica Fennica Vanamo, Helsinki.

Jardine, N., 1972. Computational methods in the study of plant distribution. Pages 381–393 *in* D. Valentine, editor. Taxonomy, Phytogeography and Evolution. Wiley, London.

Jermy, A. C., editor. 1978. Atlas of ferns of the British Isles. Botanical Society of the British Isles, London.

Jonsson, B. G. 2001. A null model for randomization tests of nestedness in species assemblages. Oecologia **127**:309–313.

Kilburn, P. D. 1966. Analysis of the Species-Area Relation. Ecology **47**:831–843.

Lamarck, J. B. P. A. d. M., and A. P. d. Candolle. 1805. Flore Française, ou descriptions succinctes de toutes les plantes qui croissent naturellement en France, disposées selon une nouvelle méthode d'analyse, et précédées par un exposé des principes élémentaires de la botanique. 3 edition. Desray, Paris.

Legendre, P., and L. Legendre. 1998. Numerical ecology, 2nd English edition. Elsevier, Amsterdam.

Lennon, J. J., J. J. D. Greenwood, and J. R. G. Turner. 2000. Bird diversity and environmental gradients in Britain: a test of the species-energy hypothesis. Journal of Animal Ecology **69**:581–598.

Lennon, J. J., W. E. Kunin, and S. Hartley. 2002. Fractal species distributions do not produce power-law species-area relationships. Oikos **97**:378–386.

Loo, S. E., R. Mac Nally, and G. P. Quinn. 2002. An experimental examination of colonization as a generator of biotic nestedness. Oecologia **132**:118–124.

Mandelbrot, B. B. 1982. The fractal geometry of nature. Freeman, Oxford.

Marquez, A., R. Real, and M. Vargas. 2001. Methods for comparison of biotic regionalizations: the case of pteridophytes in the Iberian peninsula. Ecography **24**:659–671.

Matthews, J. R. 1923. Plants restricted to England and Wales. Annals of Botany **37**:277–298.

Matthews, J. R. 1924. Plants restricted to England, Scotland and Wales. Annals of Botany **38**:707–721.

Matthews, J. R. 1926. Irish and Anglo-Irish plants. Annals of Botany **40**:773–797.

Matthews, J. R. 1937. Geographical relationships of the British Flora. Journal of Ecology **25**:1–90.

Matthews, J. R. 1955. Origin and Distribution of the British Flora. Hutchinson's University Library, London.

May, R. M., 1975. Patterns of species abundance and diversity. Pages 81–120 *in* M. L. Cody and J. M. Diamond, editors. Ecology and evolution of communities. Belknap Press, Harvard.

McCullagh, P., and J. A. Nelder. 1989. Generalized Linear Models. 2nd edition. Chapman and Hall, London.

Moss, C. E., 1914. Vegetation and Flora. Pages 92–112 *in* The Oxford survey of the British Empire, volume 1. The Clarendon Press, Oxford.

Myers, N. 1989. Threatened biotas: "Hotspots" in tropical forests. The Environmentalist **8**:1–20.

Myers, N., R. A. Mittermeier, C. G. Mittermeier, G. A. B. da Fonseca, and J. Kent. 2000. Biodiversity hotspots for conservation priorities. Nature **403**:853–858.

Myklestad, O., and H. J. B. Birks. 1993. A numerical analysis of the distribution patterns of Salix L. species in Europe. Journal of Biogeography **20**:1–32.

Nelder, J. A., and R. W. M. Wedderburn. 1972. Generalized linear models. Journal of the Royal Statistical Society A **135**:370–384.

Pacala, S. W., 1997. Dynamics of plant communities. Pages 532–555 *in* M. J. Crawley, editor. Plant Ecology. Blackwell, Oxford.

Patterson, B. D. 1987. The principle of nested subsets and its implications for biological conservation. Conservation Biology **1**:323–334.

Patterson, B. D. 1990. On The Temporal Development Of Nested Subset Patterns Of Species Composition. Oikos **59**:330–342.

Patterson, B. D., and W. Atmar. 1986. Nested Subsets And The Structure Of Insular Mammalian Faunas And Archipelagoes. Biological Journal Of The Linnean Society **28**:65–82.

Perring, F. H., editor. 1968. Critical supplement to the atlas of the British flora. Thomas Nelson, London.

Perring, F. H., and L. Farrell. 1977. Vascular plants. (British red data books: 1). The Society for the Promotion of Nature Conservation, Lincoln.

Perring, F. H., and S. M. Walters. 1962. Atlas of the British Flora. Thomas Nelson and Sons Ltd, London.

Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. Ecological Modelling **190**:231–259.

Plotkin, J. B., M. D. Potts, N. Leslie, N. Manokaran, J. LaFrankie, and P. S. Ashton. 2000*a*. Species-area curves, spatial aggregation, and habitat specialization in tropical forests. Journal Of Theoretical Biology **207**:81–99.

Plotkin, J. B., M. D. Potts, D. W. Yu, S. Bunyavejchewin, R. Condit, R. Foster, S. Hubbell, J. LaFrankie, N. Manokaran, L. H. Seng, R. Sukumar, M. A. Nowak, and P. S. Ashton. 2000*b*. Predicting species diversity in tropical forests. Proceedings Of The National Academy Of Sciences Of The United States Of America **97**:10850–10854.

Prendergast, J. R., R. M. Quinn, J. H. Lawton, B. C. Eversham, and D. W. Gibbons. 1993. Rare species, the coincidence of diversity hotspots and conservation strategies. Nature **365**:335–337.

Preston, C. D., and M. O. Hill. 1997. The geographical relationships of British and Irish vascular plants. Botanical Journal of the Linnean Society **124**:1–120.

Preston, C. D., D. A. Pearman, and T. D. Dines, editors. 2002. New atlas of the British and Irish Flora. Oxford University Press, Oxford.

Preston, C. D., D. A. Pearman, and A. R. Hall. 2004. Archaeophytes in Britain. Botanical Journal of the Linnean Society **145**:257–294.

Preston, C. D., M. G. Telfer, D. B. Roy, P. D. Carey, M. O. Hill, W. R. Meek, P. Rothery, S. M. Smart, G. M. Smith, K. J. Walker, and D. A. Pearman, 2003. The changing distribution of the flora of the United Kingdom. Technical report, Centre for Ecology and Hydrology.

Preston, F. W. 1960. Time and space and the variation of species. Ecology **41**:785–790.

Preston, F. W. 1962*a*. The Canonical Distribution of Commonness and Rarity: Part 1. Ecology **43**:185–215.

Preston, F. W. 1962*b*. The Canonical Distribution of Commonness and Rarity: Part 2. Ecology **43**:410–432.

Proctor, M. C. F. 1967. The Distribution of British Liverworts: A Statistical Analysis. Journal of Ecology **55**:119–135.

R Development Core Team, 2005. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Reid, C. 1911. The relations of the present plant population of the British Isles to the glacial period. Reports of the British Association pages 573–577.

Reid, W. V. 1998. Biodiversity hotspots. Trends in Ecology and Evolution **13**:275–280.

Rich, T. C. G., P. Donovan, P. Harmes, A. Knapp, M. McFarlane, C. Marrable, N. Muggeridge, R. Nicholson, M. Reader, P. Reader, E. Rich, and P. White. 1996. Flora of Ashdown Forest. Sussex Botanical Recording Society.

Rich, T. C. G., and E. R. Woodruff. 1990. The BSBI Monitoring Scheme, 1987-1988. Botanical Society of the British Isles, London.

Robinson, G. R., J. F. Quinn, and M. L. Stanton. 1995. Invasibility of experimental habitat islands in a California winter annual grassland. Ecology **75**:786–794.

Rodríguez-Iturbe, I., and A. Rinaldo. 1997. Fractal river basins : chance and self-organization. Cambridge University Press, Cambridge.

Rosenzweig, M. L. 1995. Species Diversity in Space and Time. Cambridge University Press, Cambridge.

Roxburgh, S. H., and P. Chesson. 1998. A new method for detecting species associations with spatially autocorrelated data. Ecology **79**:2180–2192.

Salisbury, E. J. 1932. The East Anglian Flora. Transactions of the Norfolk and Norwich Naturalists Society **8**:191–263.

Sanderson, J. G., M. P. Moulton, and R. G. Selfridge. 1998. Null matrices and the analysis of species co-occurrences. Oecologia **116**:275–283.

Schofield. 1985. Introduction to Bryology. Macmillan, New York.

Simberloff, D., and J. L. Martin. 1991. Nestedness Of Insular Avifaunas - Simple Summary Statistics Masking Complex Species Patterns. Ornis Fennica **68**:178–192.

Skellam, J. G. 1952. Studies in Statistical Ecology: 1 Spatial Pattern. Biometrika **39**:346–362.

Sneath, P. H. A. 1967. Conifer Distributions and Continental Drift. Nature **215**:467–470.

Stapf, O. 1914. The southern Element in the British Flora. Engler's Bot. Jahrb **50**:509–525.

Sugihara, G. 1980. Minimal Community Structure: An Explanation of Species Abundance Patterns. The American Naturalist **116**:770–787.

Theophrastus. 1916. Enquiry into Plants [With an English translation by Hort, A.]. William Heinemann, London.

Thomas, C. D., A. Cameron, R. E. Green, M. Bakkenes, L. J. Beaumont, Y. C. Collingham, B. F. N. Erasmus, M. F. de Siqueira, A. Grainger, L. Hannah, L. Hughes, B. Huntley, A. S. van Jaarsveld, G. F. Midgley, L. Miles, M. A. Ortega-Huerta, A. T. Peterson, O. L. Phillips, and S. E. Williams. 2004. Extinction risk from climate change. Nature **427**:145–148.

Thompson, K., J. G. Hodgson, J. P. Grime, and M. J. W. Burke. 2001. Plant traits and temporal scale: evidence from a 5-year invasion experiment using native species. Journal of Ecology **89**:1054–1060.

Tjorve, E. 2003. Shapes and functions of species-area curves: a review of possible models. Journal Of Biogeography **30**:827–835.

Turner, W. R., and E. Tjorve. 2005. Scale-dependence in species-area relationships. Ecography **28**:721–730.

Tutin, T. G., N. A. Burges, A. O. Chater, J. R. Edmondson, V. H. Heywood, D. M. Moore, D. H. Valentine, S. M. Walters, and D. A. Webb, editors. 1964. Flora Europaea. 1st edition. Cambridge University Press, London.

Tutin, T. G., V. H. Heywood, N. A. Burges, D. M. Moore, D. H. Valentine, S. M. Walters, and D. A. Webb, editors. 1968-1980. Flora Europaea. 1st edition. Cambridge University Press, London.

Watson, E. V. 1972. Structure and Life of Bryophytes. Hutchinson, London.

Watson, H. C. 1847. Cybele Britannica: British Plants and their Geographical Relations. Longmans, Green, Reader & Dyer, London.

Watson, H. C. 1883. Topographical Botany. 2 edition. Barnard Quartich, London.

Williams, C. B. 1943. Area and the number of species. Nature **152**:264–267.

Williams, W. T., and J. M. Lambert. 1960. Multivariate Methods in Plant Ecology. Journal of Ecology **48**:689–710.

Williams, W. T., and J. M. Lambert. 1961. Multivariate Methods in Plant Ecology. Journal of Ecology **49**:717–729.

Williamson, M. 2003. Species-area relationships at small scales in continuum vegetation. Journal of Ecology **91**:904–907.

Wilson, R. J., C. D. Thomas, R. Fox, D. B. Roy, and W. E. Kunin. 2004. Spatial patterns in species distributions reveal biodiversity change. Nature **432**:393–396.

Worthen, W. B. 1996. Community composition and nested-subset analyses: Basic descriptors for community ecology. Oikos **76**:417–426.

Electronic appendices on CD-ROM