

Habitat diversity determines ghyll woodland species diversity. Adding complexity to the species area relationship, a new predictive model?

S. Waite¹, N.G. Burnside^{2a} and A. R. Flint^{2b}.

¹Hartpury College, Hartpury House, Gloucester, GL19 3BE, UK.

e-mail: Stephen.Waite@hartpury.ac.uk

²Biogeography and Ecology Research Group (BERG) ^aSchool of the Environment & Technology; ^bSchool of Pharmacy and Biomolecular Sciences, University of Brighton, Brighton, BN2 4GJ, UK.

Abstract

Ghylls are linear valley features cut into the sandstone beds of the Weald of south-eastern England. The indigenous ghyll woodlands are highly species rich, supporting distinctive assemblages of cryptogamic plants. We assessed the ability of the species-area relationship, the Choros model and a modified version of the Choros model to describe variations in higher plant diversity. Principal component analysis (PCA) was used to produce proxy environmental variables, multi-regression analysis of which allowed factors determining diversity to be investigated. Field layer diversity varied substantially between ghylls. The data did not fit the species area relationship. A statistically weak fit was provided by the standard Choros model. The modified Choros model, incorporating a PCA generated proxy variable for habitat diversity, provided a good fit. The results suggest that ghyll field layer diversity is primarily determined by habitat diversity and to a lesser extent by ghyll isolation, length and area.

Key words: Ghyll woodlands; Species area relationship; Choros model.

Introduction

Species richness, at a range of spatial scales, is a function of area (Boecklen & Gotelli, 1984; Rosenzweig, 1995; Rosenzweig 2004; Whittaker & Fernández-Palacios, 2007). The species area relationship has frequently been described by Equation 1 (see overleaf). Several explanations for this relationship have been suggested. One formulation of this relationship, referred to as the Island Biogeographic Theory (IBT), originally developed by MacArthur and Wilson (1967) to model species richness on true islands, has been extensively applied to habitat islands. Under the IBT species richness is determined by the equilibrium between rates of species immigration and extinction, both are seen as functions of area. IBT has provided an important conceptual framework for investigations of the effects of habitat fragmentation on species richness (Powledge 2003), with the continued occupation of fragments by species and their re-colonisation, being considered in part a function of fragment size and their spatial distribution. It is clear that while the conceptual simplicity of IBT is attractive, it has a number of fundamental limitations when applied to habitat fragments (Laurance 2008; Smith 2010).

Boecklen & Gotelli (1984) state that the species-area relationship '*is at least partly epiphenomenal*', a secondary phenomenon resulting from another. MacArthur and Wilson (1967) viewed area as a 'proxy' variable for a range of factors including environmental and resource heterogeneity. Some studies have tested the fit of a modified version of the normal species area relations (Equation 2), where area is multiplied by a parameter linked to habitat diversity and resource availability (Whittaker & Fernández-Palacios, 2007, Triantis et. al. 2003, 2008).

$$\begin{aligned} \text{Equation (1)} \quad S &= cA^z & \log S &= \log c + z \log A \\ \text{Equation (2)} \quad S &= c(A \times F)^z = c K^z & \log S &= \log c + \log K \end{aligned}$$

Where S = species diversity, c and z are constants, F is a scaling factor e.g., measure of habitat diversity and A is area. K is the product of area (A) and the scaling factor (F).

The Weald of southeast England supports woodland valley systems internationally distinct in both their ecology and geomorphology. Ghyll woodlands are typically linear features occupying deep and narrow valleys cut into the sandy and silty Hastings Beds and the Weald and Wadhurst clays. The sheltered valleys occupied by ghyll woodlands buffer temperature fluctuations and maintain high humidity levels resulting in unusually oceanic micro-climatic conditions. Spatial and structural analyses suggest that ghyll woodlands are an important and significant landscape feature of high conservation value (Rose and Patmore, 1997; Burnside *et al.*, 2002, 2006). Ghyll woodlands support a unique assemblage of cryptogamic plants with both oceanic and sub-oceanic affiliations (Ratcliffe, 1968). The presence of rich liverwort and moss communities suggests that the ghyll woodlands are of considerable age and most can be regarded as 'ancient woodland' (*sensu* Rackham, 1980; Peterken, 1981; Forestry Commission, 2001). Much of south-eastern and southern England underwent preferential exploitation of woodland following a phase of human expansion in the late Neolithic/early Bronze Age (Waller and Marlow, 1994). From a land-use perspective, however, the steep slopes of the ghylls have meant that they have mostly remained uncultivated and wooded. The historical persistence of wooded areas may have resulted in many ghylls maintaining 'old forest type' epiphytic lichens and bryophytes (Rose and Patmore, 1997). Thus ghyll woodlands may be viewed as the fragmented remnants of a more extensive ancient wooded landscape.

The occurrence and distribution of species within [habitat] islands is of key concern to landscape ecologists, biogeographers and conservationists. There is a need to understand the patterns of species occurrence and richness among habitat patches within a fragmented landscape (Bruun, 2002; Krauss *et al.*, 2004; Fattorini, 2010). We assess the ability of the standard species area and modified species area models (equations 1 and 2) to describe the species richness among fragmented woodlands. In addition a range of more complex models derived from multiple regression are used to investigate the factors that determine ghyll woodland high plant diversity. The development and validation of such models can aid the identification of key environmental factors and processes determining patterns of species richness at a landscape scale. They also provide tools allowing the impacts of future habitat fragmentation to be quantified and predictions made about the consequences of landscape scale conservation and planning.

Methods

Field survey and data set

This study has used data from field surveys of ghyll woodland systems throughout East and West Sussex. Using a uniform walkover survey approach, both biotic and abiotic characteristics of ghylls were recorded. The data were used to construct a GIS database containing site-specific information on the ecology, landform and management of 48 ghyll woodlands randomly selected from 1130 ghyll woodlands. The findings presented in this paper are based on a sub-set of 35 ghylls for which a complete set of data was available. The variables included in the survey reflected those considered to be important in characterising ghyll woodlands (Rose and Patmore, 1997) and are shown in Table 1. FragStats (McGarigal and Marks, 1994) was used to measure woodland patch area, perimeter length, and inter-patch distance. Additional spatial statistics including shape indices, were calculated from the digital boundary data. Further information on ghyll woodlands and the collection and initial treatment of the data may be found in Burnside *et al.* (2006).

Table 1: Field survey information for the ghyll woodland sample in the Weald

Characteristics	Descriptors
Geology	Predominant geological bed, geological beds in chronological sequence
Geomorphology	Valley form profiles, channel dimensions, height of ghyll, degree of fall, length of ghyll
Present land use	Coppicing, thinning/selective felling, clear felling, adjacent land use
Woodland type	Broadleaved, coniferous, mixed-woodland NVC - e.g. W6, W8, W10
Vegetation structure	Canopy, Understorey, Field layer vegetation composition using DAFOR scale, presence of alien species.
Presence of important features	Geomorphology - Waterfalls, sandstone outcrops Ecology - Bryophytes, invertebrates

Analysis

Two approaches, principle component analysis (PCA) and linear regression were used to establish and model the relationship between species richness and the area, habitat diversity and other environmental characteristics of the ghyll woodlands. Multiple regression allows the generation of linear predictive models incorporating more than one variable. However its use is limited where individual explanatory variables are highly correlated with one another (multicollinearity), only uncorrelated explanatory variables should be included in the analysis. To overcome this problem a mixed strategy was adopted. Pearson correlation coefficients were calculated between all variables and used to identify sub-sets of independent explanatory variables for subsequent inclusion in multi-regression analysis. Variables most highly correlated with ghyll woodland species richness were included in linear regression models. Additional independent variables were then systematically added and the explanatory power of the new model tested. In parallel with this approach an iterative regression sub-routine which performs stepwise multi-regression was used to identify combination of variables from the complete set of independent variables that yielded the greatest explanatory power.

The second approach adopted to overcome the multicollinearity inherent in the dataset without necessitating the exclusion of variables and therefore a loss of information, used principal component analysis (PCA). Principal component analysis was performed on the correlation matrix of measured ghyll woodland habitat and environmental variables. The generated principal components scores uniquely summarise the characteristics of each data point. Component scores which are normally distributed and orthogonal (Waite, 2000) were used to provide independent proxy environmental variables that can be incorporated into multiple regression models. This approach allows the complete set of environmental variables and measured habitat characteristics to be included in the analysis. However, while the resulting regression models may have greater statistical explanatory power, this is of little value in helping to understand the ecological and landscape processes that determine species richness if the components cannot be reliably interpreted. Regression model goodness of fit was compared using the adjusted coefficient of determination (R^2). Prior to analysis all continuous variables were tested for normality and where appropriate log transformed to normalise their distribution. In all cases unless otherwise stated $n = 35$. All analyses were performed using MINITAB version 15.1.

Results and Discussion

Simple models: Species richness regressed against landscape variables.

Ghyll woodland area did not explain species richness independently of other variables. Ghyll area only became a significant factor influencing total species richness when in combination with other variables. Total species richness was significantly related to ghyll length (GL), the average distance to the nearest fives ghylls (AD5G) and ghyll woodland area (GA). The multiple regression model incorporating these three variables was significant (R^2 0.24, $p < 0.005$, $n = 35$; Table 2). This relationship was only significant when total ghyll woodland species richness was used as the dependent variable. The relationship was not significant if the contribution of the field layer species to total diversity was removed. No significant predictive relations were found between canopy or understory tree diversity with either individual or combinations of ghyll and landscape variables. Canopy tree and understory species composition varied little between ghyll woodlands (Burnside *et al.* 2006). Even though site area varied between 1.272 and 61.648 hectares the total number of both canopy and understory species only ranged from 2 to 10. Canopy and understory species composition reflects regional patterns of woodland species abundance and the common occurrence of NVC type W10 and W8 woodlands (Rodwell, 1991) with a dominance of *Quercus* sp. and *Fraxinus* sp. respectively in the canopy. Regionally frequent and common canopy and understory woodland species were also common and frequent within ghyll woodlands. Ghyll canopy and understory species richness and composition would appear to be largely determined by regional biogeography with local landscape scale factors having relatively little effect.

Field layer species richness: Environmental factors.

In contrast to the canopy and understory, field layer diversity and composition varied substantially between sites. Some of this variation can be related to ghyll and landscape variables. Field layer species richness was not correlated with area but it was significantly correlated with ghyll length ($r = 0.319$, $R^2 = 10.2\%$, $p < 0.05$). This may reflect the linear nature of ghyll woodlands. Ma *et al.* (2002) examined the relationship between plant species

richness and area for long narrow riparian buffer grasslands in Finland. They found that for these essentially linear grassland fragments species richness was more strongly related to habitat length, than to either width or area.

Table 2. Table of fitted models. Total species diversity (ST), Field layer diversity (SFL), Average distance to the nearest 5 ghylls (AD5G), Ghyll area (GA), Ghyll Length (GL), principal component scores (PC1, PC2, PC3).

Variables included in the model	Model	R2%	P
Models linking species richness with habitat diversity			
Modified Choros model. SFL predictor GAxPC1	$\log(\text{SFL}) = 0.9709 + 0.08315 \log(\text{GA} \times \text{PC1})$	45%	P<0.005
Choros model SFL predictor GAxHD	$\log(\text{SFL}) = 0.499 + 0.223 \log(\text{GA} \times \text{HD})$	13.2%	0.05
SFL predictor HD.	$\log \text{SFL} = 0.093 + 0.879 \log \text{HD}$	31.1%	P<0.005
Mixed models: Multiple regression models incorporating PCA scores and landscape variables.			
SFL predictors, AD5G, GL, GA, PC1, PC2, PC3 scores.	$\log \text{SFL} = 1.44 - 0.159 \log \text{AD5G} + 0.187 \log \text{GA} - 0.000453 \text{GL} + 0.0539 \text{PC1} + 0.0187 \text{PC2} + 0.0410 \text{PC3}$	48.5%	P<0.005
SFL predictors, AD5G, GL, GA, PC1, PC3 scores.	$\log \text{SFL} = 1.44 - 0.187 \log \text{AD5G} + 0.196 \log \text{GA} - 0.000370 \text{GL} + 0.0594 \text{PC1} + 0.0334 \text{PC3}$	49.6%	P<0.005
SFL predictors, AD5G, GL, GA, PC1, PC2 scores.	$\log \text{SFL} = 1.39 - 0.193 \log \text{AD5G} + 0.204 \log \text{GA} - 0.000279 \text{GL} + 0.0669 \text{PC1} - 0.0021 \text{PC2}$	46.3%	P<0.005
SFL predictors, AD5G, GL, GA, PC1 scores.	$\log \text{SFL} = 1.39 - 0.190 \log \text{AD5G} + 0.203 \log \text{GA} - 0.000286 \text{GL} + 0.0664 \text{PC1}$	48.1%	P<0.005
SFL predictors PC1, GA.	$\log \text{SFL} = 0.8777 + 0.083 \text{PC1} + 0.203 \log \text{GA}$	47.7%	P<0.005
PCA models: Based on regressing species richness against principle component scores.			
SFL predictor PC1, PC2, PC3 scores	$\log \text{SFL} = 0.960 + 0.0925 \text{PC1} - 0.0099 \text{PC2} + 0.0166 \text{PC3}$	35.8%	0.001
SFL predictor PC1, PC2 scores.	$\log \text{SFL} = 0.960 + 0.0925 \text{PC1} + 0.0166 \text{PC2}$	37.5%	P<0.005
SFL predictor PC1 scores.	$\log \text{SFL} = 0.960 + 0.0925 \text{PC1}$	40.2%	P<0.005
Simple models: Species richness regressed against landscape variables.			
SFL predictors GL, AD5G, GA.	$\log \text{SFL} = 2.191 - 0.0005 \text{GL} - 0.45 \log \text{AD5G} + 0.186 \log(\text{GA})$	38.8%	P<0.005
ST predictors GL, AD5G, GA.	$\log \text{ST} = 1.987 - 0.00026 \text{GL} - 0.26 \text{AD5G} + 0.1 \log \text{GA}$	24.3%	P<0.005

The best fit regression model incorporating non-correlated variables identified three key variables. In this model field layer species richness was negatively related to ghyll length (GL) and isolation (isolation measured as the average distance to the nearest 5 ghyll woodlands; ADN5G) and positively related to ghyll area (GA). This model accounted for approximately 38% of the variation in field layer species richness (Table 2).

PCA models: Based on regressing species richness against principle component scores.

The PCA was performed on a matrix of 16 environmental and landscape variables. These variables related to distinct aspects of the ghyll woodlands and the surrounding adjacent landscape matrix (see Table 1). The first three principle components accounted for approximately 51% of the variation, with PC1 accounting for approximately 20%, PC2 for 16% and PC3 for 13.9% (eigenvalues = 3.259, 2.605 and 2.218 respectively). Inspection of the principal component loading values suggests that high PC1 scores are primarily associated with predominantly shady conditions and range of valley types present. PC2 provided a combined measure of the physical nature of ghylls with high PC2 scores strongly positively correlated with ghylls in broad flat valleys, with open rides and wide water channels. Low PC2 scores were associated with ghylls occurring in steep sided narrow valleys. High PC3 scores were associated with ghylls that had open glades and were geologically uniform but had a high degree of topographical variation (*i.e.* vertical difference between the highest and lowest points).

The species diversity of the field layer vegetation was most strongly correlated with PC1 scores. The linear regression model with PC1 ghyll scores as the single explanatory variable explained approximately 40% of the variation in field layer species. Incorporating PC2 and PC3 scores as additional explanatory variables failed to substantially increase the explanatory power for the resulting linear regression models (Table 2). This suggests that the extent of shade and range of valley profile types within a ghyll are important determinants of field layer diversity. As area was not included in these models, the highly significant relationship between species richness and principal component scores indicated that environmental heterogeneity, independent of area, is a major determinate of ghyll field layer diversity. These results also show that within the context of this study and more generally within landscape ecology, the use of PCA to simplify multivariate data sets and to generate proxy environmental variables can be a useful investigative tool.

Mixed models: Multiple regression models incorporating PCA scores and landscape variables.

When the proxy environmental variables generated from the PCA (PC1, PC2 and PC3 scores) were incorporated into models along with variables excluded from the PCA such as isolation, area and length, the explanatory power of the models are increased substantially (Table 2). For example a model incorporating PC1 and ghyll area (GA) as explanatory variables accounted for over 50% of the variation in field layer species richness. The inclusion of PC2 and PC3, along with the average nearest distance to ancient woodland fragments and ghyll length, only marginally increased fit of the models, adjusted R^2 values increasing from 47.7 to 49.6% (Table 2). These results suggest that while environmental or habitat variation as measured by the PC1 scores is a major determinate of field layer species richness, to a lesser extent there are also area and isolation effects.

Models linking species richness with habitat diversity.

The 'Choros' model (the product of area and habitat number; Triantis *et al.* 2003; 2008) provides a better predictor of species richness than area for a range of different taxons among true island systems and also for calcareous grassland fragments in Southern England (Waite pers. ob.). Here we defined habitat diversity (HD) as the total number of different habitats, *i.e.*

types of shade conditions, woodland, woodland management, presence of glades, rides and different types of wetland habitats occurring in a ghyll. The logarithm of field layer species richness was significantly correlated with the logarithm of habitat diversity ($r = 0.558$, $p < 0.0005$). Replacing habitat diversity by the product of habitat diversity and ghyll area did not increase the strength of the correlation ($r = 0.361$, $n = 35$, $p < 0.03$). Thus in this study the Choros model provided a better fit than the standard species area relationship (Equation 1) which was not significant, but habitat diversity alone was a better predictor of field layer species richness.

One potential issue with the application of the Choros model is the subjective way in which habitats and habitat diversity are quantified. As habitat types are frequently defined with reference to the structure of their vegetation, this issue is likely to be particularly marked when using the Choros model approach to predict plant species diversity. In these circumstances the dependent variable, species diversity, may not be truly independent of the explanatory variable, the product of area and habitat diversity. To overcome this problem, PC1 scores were used in place of habitat diversity. These provide an objective proxy variable for habitat and environmental diversity. Thus, when PC1 scores are used as the habitat diversity scaling factor (F) in the modified Choros model (i.e., Equation 2) the model fit is improved (R^2 increasing from 13.3% ($p < 0.05$) to 46.7% ($p < 0.0005$)).

Conclusions

The canopy and understory species diversity varied little between ghyll woodlands and was not related to ghyll area or other measured ghyll characteristics. Canopy and Understory appear to be determined by, and to reflect, regional woodland biogeography. Field layer diversity varied between ghylls, but did not fit the standard species area relationship (Equation 1). The standard Choros model (Equation 2) while providing a significant fit for the variation between ghylls had only weak explanatory power. The modified Choros model (Equation 2, with $F = PC1$), where PC1 scores were used in place of habitat diversity, provided a good fit for the field layer diversity data. Principal component scores provide an objective multivariate measure of habitat diversity. This result confirms the value and validity of using PCA to avoid colinearity and simplify complex datasets allowing the extraction of useful proxy variables. Overall the results show that among the samples of ghylls used in the study, field layer diversity was primarily determined by habitat and environmental diversity, and to a lesser extent woodland isolation, ghyll length and area. The study demonstrates the potential predictive power of models incorporating measures of fragment area and habitat diversity to describe spatial patterns of species richness in an increasingly fragmented landscape.

References

- Boecklen, W.J. & Gotelli, N.J. (1984).** Island Biogeographic Theory and Conservation Practice: Species-Area or Specious-Area Relationships? *Biological Conservation* 29:63-80.
- Bruun H.H (2002)** Determinants of species richness in patches of grassland and heathland in Himmerland (Denmark). *Nordic Journal of Botany*. **21**: 607- 614.
- Burnside, N.G., Carrett, N.J., Metcalfe, D., & Waite, S. (2002)** *GIS analysis of ancient and ghyll woodlands within East Sussex*. Biogeography & Ecology Research Group, University of Brighton, Brighton,UK.

- Burnside, N.G., Metcalfe, D.J, Smith R.F & S. Waite (2006)** Ghyll woodlands of the Weald: characterisation and conservation. *Biodiversity and Conservation*. **15**: 1319-1338.
- Connor, E.F. & McCoy, E.D. (1979)**. The statistics and biology of the species-area relationship. *Amateur Naturalist*. **113**: 791-833.
- Fattorini S. (2010)** The use of cumulative area curves in biological conservation: A cautionary note. *Acta Oecologica*. **36**: 255-258.
- Forestry Commission (2001)** *National Inventory of Woodland and Trees: England*. Forestry Commission, Edinburgh, UK.
- Krauss J., Klein A.M., Steffan-Dewenter I. & Tschardt T. (2004)** Effects of habitat area, isolation, and landscape diversity on plant species richness of calcareous grasslands. *Biodiversity & Conservation*. **13**: 1427-1539.
- Laurance W. F (2008)** Theory meets reality: How habitat fragmentation research has transcended island biogeographic theory. *Biological Conservation* **141**:1731-1744
- Ma, M., Tarmi, S. & Helenius, J., (2002)** Revisiting the species-area relationship in a semi-natural habitat: floral richness in agricultural buffer zones in Finland. *Agriculture Ecosystems & Environment*. **89**:137-148.
- MacArthur R.H. & Wilson E.O. (1967)** *The theory of island biogeography*, Monographs in population biology, Princeton, New Jersey, USA.
- McGarigal, K., and Marks, B.J. (1994)** FRAGSTATS Spatial pattern analysis program for quantifying landscape structure. Version 2, Forest Science Department, Oregon State University, Corvallis.
- Peterken, G.F. (1981)** Wood anemone in central Lincolnshire: an ancient woodland indicator? *Transactions Lincolnshire Natural Union* **20**: 78-82
- Powledge, F. (2003)**. Island biogeography's lasting impact. *BioScience* **53**:1023-1
- Rackham, O. (1980)** *Ancient Woodland*. Arnold, London.
- Ratcliffe, D.A. (1968)** An ecological account of the Atlantic bryophytes in the British Isles. *New Phytologist* **67**: 365-430.
- Rodwell, J.S. (1991)** *British Plant Communities Volume 1: Woodlands and scrub*, Cambridge University Press, Cambridge, UK.
- Rose, F., and Patmore, J.M. (1997)** *Gill woodlands in the weald*. English Nature, Peterborough,UK.
- Rosenzweig, M.L. (1995)** *Species diversity in space and time*. Cambridge University Press, Cambridge, UK.
- Rosenzweig, M. L. (2004)** Applying species-area relationships to the conservation of diversity, In: M.V. Lomolino and L.R. Heaney (eds) *Frontiers of biogeography: new directions in the geography of Nature*. Sinauer Associates, Sunderland, MA., pp.325 – 343.
- Smith A.B. (2010)**. Caution with curves: Caveats for using the species – area relationship in conservation. *Biological Conservation*. **143**: 555-564.
- Triantis, K.A., Nogués-Bravo, D., Hortal, J., Borges, A.V.P., Adersen, H., Fernandez-Palacios, M., Araujo, M.B. & Whittaker, R.J. (2008)** Measurement of area and the (island) species-area relationship: new directions for an old pattern. *Oikos* **117**: 1555-1559.
- Triantis, K.A. Mylonas, M., Lika, K., & Vardinoyannis, K. (2003)** A model for the species-area-habitat relationship. *Journal of Biogeography*. **30**: 19-27.
- Waite, S. (2000)** *Statistical Ecology: A practical guide*, Prentice Hall, Harlow, UK.
- Waller, M.P., and Marlow, A.D. (1994)** Flandrian vegetational history of southeastern England - stratigraphy of the Brede Valley and pollen data from Brede Bridge. *New Phytologist* **126**(2): 369-392
- Whittaker R.J. & Fernández-Palacios J.M. (2007)** *Island Biogeography: Ecology, Evolution, and Conservation* (2nd Ed), Oxford University Press, Oxford, UK.