

1 **EMPIRICAL REALIZED NICHE MODELS FOR BRITISH HIGHER AND**  
2 **LOWER PLANTS – DEVELOPMENT AND PRELIMINARY TESTING**

3

4 Simon M. Smart <sup>1\*</sup>, W. Andrew Scott <sup>1,7</sup>, Jeanette Whittaker <sup>1,8</sup>, Mark O. Hill <sup>2,9</sup>,  
5 David B. Roy <sup>2,10</sup>, C. Nigel Critchley <sup>3</sup>, Lorenzo Marini <sup>4</sup>, Christopher Evans <sup>5,11</sup>,  
6 Bridget A. Emmett <sup>5,12</sup>, Edwin C. Rowe <sup>5,13</sup>, Andrew Crowe <sup>1,14</sup>, Mike Le Duc <sup>6,15</sup>,  
7 Robert H. MARRS <sup>6,16</sup>

8

9 1 NERC Centre for Ecology and Hydrology, Library Avenue, Bailrigg, Lancaster,  
10 LA1 4AP, UK

11 7 Email [was@ceh.ac.uk](mailto:was@ceh.ac.uk)

12 8 Email [jhart@ceh.ac.uk](mailto:jhart@ceh.ac.uk)

13 14 Email [acrowe@ceh.ac.uk](mailto:acrowe@ceh.ac.uk)

14

15 2 NERC Centre for Ecology and Hydrology, Wallingford, Crowmarsh Gifford,  
16 Benson Lane, Cambridgeshire PE28 2LS, UK

17 9 Email [moh@ceh.ac.uk](mailto:moh@ceh.ac.uk)

18 10 Email [dbr@ceh.ac.uk](mailto:dbr@ceh.ac.uk)

19

20 3 ADAS Woodthorne, Wergs Road, Wolverhampton WV6 8TQ, UK

21 Email [Nigel.critchley@adas.co.uk](mailto:Nigel.critchley@adas.co.uk)

22

23 4 Department of Environmental Agronomy and Crop Production, University of  
24 Padova, Viale dell'Università 16 - 35020 Legnaro, Padova, Italy

25 Email [Lorenzo.marini@unipd.it](mailto:Lorenzo.marini@unipd.it)

1

2 5 Centre for Ecology and Hydrology, Environment Centre Wales, Deiniol Road,

3 Bangor, Gwynedd, LL57 2UW, UK

4 11 Email [cev@ceh.ac.uk](mailto:cev@ceh.ac.uk)

5 12 Email [bae@ceh.ac.uk](mailto:bae@ceh.ac.uk)

6 13 Email [ecro@ceh.ac.uk](mailto:ecro@ceh.ac.uk)

7

8 6 Applied Vegetation Dynamics Laboratory, School of Biological Sciences,

9 Biosciences Building, Crown Street, University of Liverpool, Liverpool L69 7ZB, UK

10 15 Email [mled@liverpool.ac.uk](mailto:mled@liverpool.ac.uk)

11 16 Email [calluna@liverpool.ac.uk](mailto:calluna@liverpool.ac.uk)

12 \* Corresponding author. Tel: +44 1524 595823. Email address: [ssma@ceh.ac.uk](mailto:ssma@ceh.ac.uk)

13

14

15 **Abstract**

16 Question: Can useful realised niche models be constructed for British plant species  
17 using climate, canopy height and mean Ellenberg indices as explanatory variables?

18 Location: Great Britain.

19 Methods: General Linear Models were constructed using occurrence data covering all  
20 major natural and semi-natural vegetation types (n=40683 quadrat samples). Paired  
21 species and soil records were only available for 4% of the training data (n=1033) so  
22 modelling was carried out in two stages. First, multiple regression was used to express  
23 mean Ellenberg values for moisture, pH and fertility, in terms of direct soil  
24 measurements. Next, species presence/absence was modelled using mean indicator  
25 scores, cover-weighted canopy height, three climate variables and interactions

1 between these factors but correcting for the presence of each target species in training  
2 plots to avoid circularity.

3 Results: 803 higher plants and 327 bryophyte were modelled. 13% of the niche  
4 models for higher plants were tested against an independent survey dataset not used to  
5 build the models. Models performed better when predictions were based only on  
6 indices derived from the species composition of each plot rather than measured soil  
7 variables. This reflects the high variation in vegetation indices that was not explained  
8 by the measured soil variables.

9 Conclusions: The models should be used to estimate expected habitat suitability rather  
10 to predict species presence. Scenario testing at large scales is also possible using input  
11 from process models, yet least uncertainty attaches to their use as risk assessment and  
12 monitoring tools on nature reserves because they can be solved using mean  
13 environmental indicators calculated from the existing species composition, with or  
14 without climate data.

15

16 **Keywords**: Biodiversity, Conservation, Climate change, Ellenberg values, GLMM,  
17 Great Britain, Large-scale, Logistic regression, Risk assessment, Spatial  
18 autocorrelation

19

## 20 **Introduction**

21 In human dominated regions, changes in ecosystem function are increasingly  
22 recognised as being caused by multiple drivers such as land management, air pollution  
23 and climate change (Sala et al. 2000). Unravelling the relative contributions of these  
24 drivers, environmental covariates and their interactions relies on several approaches  
25 including experiments (Clark & Tilman 2008), retrospective signal attribution studies

1 (e.g. Stevens et al. 2004; Smart et al. 2006) and prospective modelling (Wamelink et  
2 al. 2003).

3         Although the development of process-based models is important, species and  
4 habitat conservation policies require the assessment of impacts on a large number of  
5 individual species (van Dobben et al. 2004; UK Biodiversity Steering Group 1995).  
6 However, developing population dynamic models for each species of concern is not  
7 possible or at least would require unrealistic resourcing.

8         In northern Europe, the existence of large national species occurrence databases  
9 provide opportunities for empirical, data-intensive modelling of species' realised  
10 niches along climatic and other abiotic gradients (eg. Guisan et al. 2002; Bakkenes et  
11 al. 2002; Randin et al. 2006; Araújo & New 2006). We used multiple logistic  
12 regression to develop General Linear Models (GLM) of the realised niche for British  
13 higher and lower plants. These models quantified the response of each species to four  
14 abiotic gradients; soil fertility, soil pH, soil moisture and successional status as  
15 inferred from mean Ellenberg fertility (N), soil reaction (R) and moisture (F) values  
16 and cover-weighted canopy height respectively. These four gradients reflect the major  
17 constraints on biomass production for plants, namely light, water and nutrients, and  
18 therefore the principal axes along which niches differentiate (Geider et al. 2001).  
19 Since the distributions of many British species are known to be climatically  
20 constrained (Preston & Hill 1997; Preston 2007) three climate variables were also  
21 included.

22         Niche models were produced for as many higher and lower plant species as  
23 possible by combining botanical databases representing the abiotic preferences of  
24 species in Britain across all major natural, semi-natural and artificial plant  
25 assemblages. Since soil measurements were only available for a subset of the total

1 training dataset mean unweighted Ellenberg indices (Ellenberg et al. 1991) were used  
2 as explanatory variables in addition to climate and canopy height, but translation  
3 between soil data and mean Ellenberg indices was achieved by constructing multiple  
4 regression models predicting the latter in terms of the former. Ideally, Ellenberg  
5 values would be completely replaced by direct soil measurements (e.g. Wamelink et  
6 al. 2005a) but the cost of pairing soil data with plant species occurrence for most of a  
7 national flora is usually prohibitive. Thus, in Britain, sufficient paired measurements  
8 of soil properties and vegetation are only available for a limited number of plots and  
9 species but all major abiotic gradients are covered. Using this subset, multiple  
10 regression equations were constructed to enable the mean unweighted Ellenberg  
11 indices, used as explanatory variables in the niche models, to be expressed in terms of  
12 a range of measured soil variables. This calibration step allows mean Ellenberg values  
13 rather than measured soil properties to be used as explanatory variables in turn  
14 allowing the complete training dataset of quadrat samples to be used in the  
15 construction of niche models. The benefits of this approach are two-fold: First,  
16 models can be generated for many more species than just those present in plots with  
17 soil data, and second, predictions of habitat suitability for each modelled species can  
18 be readily solved using just the species composition of the vegetation, with or without  
19 climate data. The disadvantage is that each species' response to measured gradients of  
20 soil conditions can only be expressed indirectly via the intermediate multiple  
21 regression models linking mean Ellenberg values to soil variables. We sought to test  
22 whether this disadvantage was outweighed by the ability to produce models for a large  
23 number of species and where each model could be readily solved at small scales using  
24 minimal environmental measurements.

25 In summary our objectives were as follows (see Fig 1):

- 1 1. Build simple realised niche models for as many higher and lower plants as  
2 possible in the British flora, incorporating small-scale abiotic (B1) and larger-  
3 scale climate variables (B2).
- 4 2. In the absence of soil measurements from all model training plots, use mean  
5 Ellenberg indicator values as explanatory variables for soil pH, soil moisture  
6 and fertility (B1).
- 7 3. Based on a smaller random subset of plots, derive multiple regression  
8 equations that explain mean Ellenberg values in terms of measured soil data  
9 (B3).
- 10 4. Decompose the variation in mean Ellenberg values explained by measured soil  
11 data into unique and overlapping components (B3).
- 12 5. Test niche models against independent observations (T1,T2).
- 13 6. Compare the performance of models solved using mean Ellenberg values  
14 based on observed species composition (T1) versus mean Ellenberg values  
15 predicted from measured soil data (T2).

16

## 17 **Methods**

18

### 19 *Selection of botanical data*

20 Quadrat data were used from four large-scale botanical surveys of the plant  
21 communities of Great Britain (England, Wales and Scotland) (Table 1). The largest  
22 dataset was that used to construct the national synoptic catalogue of plant  
23 communities known as the National Vegetation Classification (NVC) (Rodwell 1991).  
24 These data were sampled using a phytosociological approach and contributed by far  
25 the largest number of quadrat samples. The targeting of homogenous stands of

1 vegetation could however, constitute a bias toward sampling the regions close to a  
2 species' abiotic optimum whilst under-sampling sub-optimal conditions associated  
3 with ecotones or heterogenous, disturbed vegetation (Lájer 2007; Braunisch et al.  
4 2008). Therefore three other vegetation survey datasets were also included. These  
5 were all based on stratified, random sampling within landscapes or habitat type (Table  
6 1).

7

#### 8 *Niche model construction*

9 Each GLM model used presence/absence data for each plant species in each  
10 plot as the response variable. Unweighted mean Ellenberg values were calculated for  
11 each plot based on the list of values recently updated for the British flora (Hill et al.  
12 1999; Hill et al 2000). Mean cover-weighted canopy height was used as an index of  
13 the disturbance/successional status of the vegetation in each quadrat. Average canopy  
14 height for each species was based on values taken from Grime et al. (1995) and Stace  
15 (1997). Bryophytes were excluded from these calculations. To avoid circularity,  
16 during model construction, indicator values of the species being modelled were  
17 excluded from calculation of the mean indices used as explanatory variables.

18 Variable selection was carried out by first testing the explanatory power of  
19 each variable separately (both linear and quadratic models) and then entering all those  
20 that were significant, and their interactions, into a stepwise elimination procedure  
21 (proc logistic; SAS Institute, 1999b). Significance was tested using likelihood ratio  
22 tests (Bio 2000). All analysis and data manipulation steps were carried out using SAS  
23 procedures and the SAS macro language (SAS Institute, 1999a).

24 After final models were selected a repeat analysis was carried out to quantify  
25 the influence of correlation between abiotic variables and spatial structure in the

1 training data (Dormann 2007). Given the debate that still surrounds the relative  
2 importance of spatial autocorrelation in species distribution modelling (Hawkins et al.  
3 2007; Diniz-Filho et al. 2003), we adopted a simple, precautionary approach to the  
4 problem: To estimate the importance of spatial correlation with abiotic variables,  
5 spatial location was quantified by seven variables for each quadrat in the training data  
6 (x and y coordinates and their squared terms and interactions) (Borcard et al. 1998;  
7 Corney et al. 2006). These terms were forced into each species model prior to  
8 backward selection of those significant variables that already formed the basis of each  
9 species model. This approach ensured that variance in the species presence data that  
10 could be explained by spatial trends was removed before re-testing combinations of  
11 abiotic variables. Abiotic explanatory variables that were no longer significant after  
12 having first fitted spatial terms were then highlighted in the final table of model  
13 equations (Appendix 3).

14 A second subset of models were also produced where, in addition to the three  
15 mean Ellenberg indices and cover-weighted canopy height, three climate variables  
16 were also available for stepwise selection. These were long-term annual averages  
17 (1961-1999) for maximum July temperature, minimum January temperature and  
18 precipitation for the 5km<sup>2</sup> containing each quadrat location. Data were extracted from  
19 the UK Climate Change Impacts Program database at  
20 [www.metoffice.gov.uk/climatechange/science/monitoring/ukcip.html](http://www.metoffice.gov.uk/climatechange/science/monitoring/ukcip.html). Spatial  
21 autocorrelation analyses were not carried out on this subset of models because climate  
22 gradients across Britain are already known to be strongly spatially structured (Corney  
23 et al. 2006).

24

25 *Explaining mean Ellenberg values with soil variables*

1 Multiple regression was used to model mean Ellenberg indices in terms of five  
2 soil variables; % organic carbon, % organic nitrogen, % soil moisture, soil pH and  
3 Olsen's extractable phosphorus. All were measured in a subset of the quadrats  
4 (n=1033) in the Countryside Survey dataset recorded in 1998 and 1999 (see Black et  
5 al. 2001 for full details). The best minimum adequate model (Crawley 2007) was  
6 determined by fitting sequences of soil variables plus their quadratic and interaction  
7 terms and comparing the performance of pairs of simple and more complex models by  
8 deviance reduction (F) tests (e.g. Manning et al. 2005). This process was carried out  
9 for log<sub>e</sub>, logit and untransformed response variables. Because the calibration dataset  
10 of paired soil measurements and mean Ellenberg indices consisted of quadrats nested  
11 in the 1km sample squares of the Countryside Survey, the 1km square was treated as a  
12 random effect in a general linear mixed model (GLMM) implemented in SAS proc  
13 mixed (Little et al. 2000). The degrees of freedom attaching to each model were  
14 downweighted according to the approximation of Satterthwaite (1946).

15 The importance of each explanatory variable in each final model was expressed by  
16 determining the unique (partial) contribution of each variable in the presence of all  
17 other selected variables and then comparing this to its explanatory power when  
18 entered as the only variable in each model (Singer 1998).

19

#### 20 *Calculation of model pmax values*

21 At the scale of small vegetation sampling plots, species that are inherently less  
22 frequent even when abiotic conditions are optimal, will have lower predicted  
23 maximum probabilities than more abundant species (Manel et al. 2001; Liu et al.  
24 2005). Hence, differences in probability of occurrence between species will reflect  
25 overall patterns of scarcity in the training dataset as well as reflecting the favourability

1 of abiotic conditions. If predicted probabilities are divided by the maximum  
2 probability ( $p_{max}$ ) possible then these effects are removed enabling probabilities to be  
3 used as standardised indices of habitat suitability. Two methods were used to  
4 determine  $p_{max}$  values (see Appendices 1 and 2).

5

### 6 *Model testing*

7 A subset of niche models were tested against independent species presence  
8 data recorded as part of a botanical monitoring program for conservation management  
9 of agriculturally managed land in Britain. Monitoring plots were located in six  
10 geographically dispersed regions designated as Environmentally Sensitive Areas  
11 (ESA). Test data consisted of 244 4m<sup>2</sup> quadrats recorded in 1995 or 1996 in a range  
12 of species-rich grassland and heathland communities. In each quadrat a full list of all  
13 vascular plants and bryophytes was made. In addition, soil was sampled immediately  
14 adjacent to each quadrat and analysed to provide representative values of soil  
15 variables (see Critchley et al. 2002 for full details).

16 106 plant species occurred in greater than 5 quadrats. These were selected for  
17 model testing. Two series of predictions were made. First, niche models were used to  
18 predict the occurrence of each of the 106 plant species in each ESA quadrat using the  
19 mean unweighted Ellenberg N, R and F values and cover-weighted canopy height  
20 derived from the plant species composition of each quadrat. Second, measured soil  
21 pH, %C, %N and estimated % soil moisture were used to predict corresponding mean  
22 Ellenberg values instead of calculating them directly from each species list. Average  
23 soil moisture content was not measured directly but was inferred from soil texture  
24 class using the equivalence table in Anon (1998). The AUC statistic was then  
25 computed across the range of predicted values for each species using a macro written

1 in base SAS. Recent concerns have been raised over the sensitivity of the AUC  
2 statistic to the inclusion of unoccupied plots far from the favourable niche space of  
3 each species (Lobo et al. 2008). Since our test dataset was constrained to sample  
4 habitats for which the test species were broadly characteristic, this issue is not likely  
5 to undermine interpretation of AUC statistics as model performance measures.

6

7 **Results**

8 *Explaining mean Ellenberg values with soil variables*

9 Multiple regression models were constructed between all three mean  
10 unweighted Ellenberg scores and soil variables (Table 2). The percentage of observed  
11 variation explained was moderately high. Ellenberg F was best explained solely by %  
12 soil moisture content while selection of the best minimum adequate model vindicated  
13 treatment of mean Ellenberg R and N values as composite indices best explained by  
14 multiple soil measurements (Table 2). Mean Ellenberg R was best explained by a  
15 combination of % soil moisture, soil pH and % carbon. The sign of the regression  
16 coefficients (Table 2) indicated that species associated with lower substrate pH (low  
17 Ellenberg R) tend to be favoured by wetter substrates with a higher carbon content.  
18 Examination of the independent versus partial contribution of selected variables  
19 showed very considerable overlap so that the total explained variation mainly  
20 comprised joint variation in %C, %soil moisture and soil pH (Fig 2a). This reflects  
21 their correlation across the training dataset, in part related to obvious mechanistic  
22 relationships. For example, waterlogging and lack of aeration inhibits decomposition  
23 but enhances base cation mobilisation and leaching and hence, low base status and  
24 low pH. The highest unique explanatory power was attributable to substrate pH as

1 expected since Ellenberg R values were designed to track a substrate pH gradient (Fig  
2 2a).

3 Mean Ellenberg N was best explained by % soil moisture, soil pH, % carbon  
4 and % nitrogen (Fig 2b). Quadratic terms for moisture content and pH captured  
5 expected curvi-linear relationships with productivity. Hence, mean Ellenberg N values  
6 tend to be lowest at the extremes of soil moisture and pH, and highest under mesic  
7 conditions (Emmett et al., 2004). However, again, the majority of the explained  
8 variation was shared between predictors, whilst soil moisture content  
9 (linear+quadratic terms) had the highest unique explanatory power overall (Fig 2b).  
10 Overall therefore, the explanatory power of any one variable could not be completely  
11 separated from the others once more reflecting mechanistic relationships between soil  
12 moisture, carbon and nitrogen content and pH and their inter-correlation across the  
13 training data. %N had the lowest unique explanatory power of all terms selected for  
14 the final model (Fig 2b).

15

### 16 *Model building*

17 Of 2059 higher and lower plant species recorded in the training dataset, 1252  
18 occurred in more than 10 quadrats. Of these a GLM model with at least one  
19 significant explanatory variable could be constructed for all except 11 species. A  
20 further 111 species were modelled but then rejected from further validation and  
21 testing. These species have a coastal distribution (defined following Hill et al. 2004).  
22 Because we lacked quadrat-scale data on salt load either from the atmosphere or  
23 through inundation, we could not incorporate quantitative response along this  
24 important gradient. These species models were therefore considered inadequately  
25 specified and results are not presented. In total, models were produced for 327

1 bryophytes and 803 higher plants (Appendix 3 – supplementary material). This  
2 amounts to 26% of the UK bryophyte flora as listed in BRYOATT (Hill et al. 2007)  
3 and 44% of the UK higher plant flora as listed in PLANTATT (Hill et al. 2004).  
4 Among higher plants, 15% of alien casuals and recent introductions (neophytes), and  
5 55% of native species have models. For 639 of these species (286 bryophytes and 383  
6 higher plants), at least one climatic gradient explained significant variation in the  
7 presence of other abiotic variables (Appendix 4). As expected, rarer plants are much  
8 less likely to have models than more widely distributed species (see Appendix 5).

9

#### 10 *Model testing*

11       Of the 106 plant species models that were tested (13% of the higher plant  
12 models produced), results differed depending on whether niche models were solved  
13 using mean Ellenberg indices derived from the observed species composition in the  
14 test quadrats or where mean Ellenberg indices were predicted from observed soil data  
15 using the multiple regression equations. AUC statistics showed that predictions of  
16 habitat suitability based on mean Ellenberg values derived from the observed species  
17 composition resulted in 61% of model performance classified as excellent or good  
18 (Swets, 1988) but this dropped to only 22% when soil variables were used to predict  
19 mean Ellenberg values (Fig 3).

20

#### 21 **Discussion**

##### 22 *Application and interpretation of static realised niche models*

23       We have produced niche models for a large proportion of the British native  
24 higher plant flora and a substantial number of infrequent to common bryophyte  
25 species. While they constitute a useful, statistical description of the realised niche of

1 each species they should be applied cautiously since, as with any empirical model  
2 based on observed spatial patterns, their formulation will partly reflect biases in the  
3 training data. For example, our models are largely based on phytosociological data.  
4 Such data are collected to census the total variation in plant communities across a  
5 territory such that rare and common communities are all represented, however there is  
6 a bias towards homogenous stands and this may under-represent less favourable loci  
7 for some species thus overestimating the rate at which a species' habitat suitability  
8 declines moving away from its optimum (Hirzel & Guisan 2002).

9         Just as spatial bias is likely, the choice of years of survey is limited by data  
10 availability yet could influence niche description. The majority of the training datasets  
11 depict species and environment relationships across Britain from the 1950s through to  
12 the late 80s. Exploiting data from this time interval coupled with the exclusion of  
13 most alien casuals and rare species should ensure that the modelled response surfaces  
14 do not miss occupied niche space because of lack of range expansion or regional  
15 extinction (Walker 2007). More importantly, many of the models may depict  
16 equilibrium conditions that could increasingly break down as populations exhibit slow  
17 delayed responses to habitat fragmentation (Lindborg & Eriksson 2004) and recent  
18 climate change (Guisan & Theurillat 2000; Thuiller et al. 2005). Such problems  
19 should not prevent the application and development of models based on extensive  
20 databases of species occurrence but predictions ought to be treated as first  
21 approximations from which deviations in newly observed data are to be expected.  
22 Hence, interpretation of model inaccuracy ultimately increases understanding of the  
23 constraints on each species occurrence in particular situations (Barry & Elith 2006;  
24 Araújo et al 2005).

1           The development of any ecological model trades off local realism against  
2 generality. Our niche models reflect the responses of plant species to principal abiotic  
3 gradients across British habitats, expressed as species presence in relatively small  
4 plots. At this scale, even though abiotic conditions are favourable, a species may be  
5 absent because of dispersal history (Svenning & Skov 2004), current dispersal  
6 limitation (Ozinga et al. 2005), the absence of appropriate gaps (Britton et al. 2000;  
7 Rasran et al. 2007), cyclic response to pathogens (Strengbom et al. 2002; Terry et al.  
8 2004) or other multi-scale constraints on the density of individuals across patchy  
9 landscapes (Tyre et al. 2001; Liebold et al. 2004). An inability to predict such regional  
10 and local influences is less of a problem if rescaled predicted probabilities are  
11 interpreted as indices of abiotic suitability rather than predictions of probable presence  
12 (Hirzel et al. 2001; Barry & Elith 2006).

13

#### 14 *Combining climate and plot-scale abiotic conditions in niche models*

15           Attempts to predict species responses to climate change often show that  
16 models trained only on climate variables perform poorly when tested on independent  
17 data. Results vary between species but highlight the importance of the conditioning  
18 effects of other abiotic conditions related to land-use impacts, landscape structure and  
19 natural variation in topography and soils below the resolution of gridded climate  
20 datasets (Broennimann et al. 2007; Beale et al. 2008; Araújo et al. 2005). Hence, the  
21 best models of biodiversity change are likely to include interactions and main effects  
22 between these additional factors and climate variables. This is especially so across  
23 densely populated and intensively farmed landscapes where restoration and  
24 maintenance of habitats frequently relies on management intervention to control  
25 successional processes, soil moisture and nutrient availability. The inclusion of

1 interaction terms between climate and local patch conditions also allows models to  
2 reflect shifts in optima and tolerances near to range edges.

3 In 57% of the British plant species niche models, at least one climate variable  
4 explained significant additional variation in species presence. In some cases the  
5 absence of climate variables in a species model will reflect weaker correlations  
6 between climate and species simply because of the less precise spatial match between  
7 coarser scale climatic variables that cannot vary within a 5km<sup>2</sup> versus species  
8 occupancy that could vary greatly between the plots within a 5km<sup>2</sup> (Trivedi et al  
9 2008). Clearly, inclusion of climate variables in species models also depends upon the  
10 occurrence of clines in species occupancy that are correlated with climate gradients  
11 across the domain sampled (Randin et al. 2006). For many British species, favourable  
12 climate is encountered throughout the sampling domain so that modelling the climatic  
13 range edge would require occupancy data from the near continent of Europe. Many of  
14 the species without climate variables in their models are likely to fall into this  
15 category.

16

17 *Using mean Ellenberg values as explanatory variables; a necessary evil?*

18 Our approach differed from previous attempts to explain mean Ellenberg  
19 values in terms of measured data (e.g. Ertsen et al. 1998) in that multiple soil  
20 variables were allowed to predict each mean Ellenberg index rather than a calibration  
21 relationship sought from a hypothesised link between a single dependent soil variable  
22 and a single mean Ellenberg value; for example attempting to model soil C/N in terms  
23 of just mean Ellenberg N (fertility) or soil moisture in terms of just mean Ellenberg F  
24 (wetness). Here the mean Ellenberg value was treated as the dependent variable which  
25 we sought to model in terms of multiple explanatory soil variables.

1           While the calibration equations between soil data and mean Ellenberg indices  
2 solve an important problem, they contribute uncertainty related to the fact that mean  
3 Ellenberg scores are generally poorly explained by single soil variables. This leads to  
4 high residual variation (Ertsen et al. 1998; van Dobben et al. 2004; Wamelink et al.  
5 2002). Hill & Carey (1997) however, found that hay yield explained 83% of the  
6 variation in mean Ellenberg N between the treatments of the Rothamsted Park Grass  
7 experiment, thus Ellenberg N is better treated as an index of above-ground  
8 productivity rather than available nitrogen. Since annual primary production is a  
9 function of a complex of intercorrelated factors including soil macro-nutrient status,  
10 soil moisture and seasonal climatic effects (Silvertown et al. 1994; Dunnett et al.  
11 1998) it is reasonable to treat mean Ellenberg values as composite indices that ought  
12 to be more fully explained by multiple factors. Using this approach we have produced  
13 much higher  $r^2$  values than reported for previous attempts to model single abiotic  
14 measurements in terms of single mean Ellenberg values (Ertsen et al. 1998; van  
15 Dobben et al. 2004). However, a feature of the resulting multiple regression models  
16 for mean Ellenberg N and R is that the majority of the total explanatory power is  
17 shared between contributing terms. That is, the amount of unique explanatory power  
18 contributed by a single term reduces greatly when another term is present. The models  
19 thus have a high degree of redundancy despite all terms contributing sufficient partial  
20 explanatory power to justify inclusion in each final model.

21           Regarding the desirability of modelling species presence directly from soil  
22 variables, it is notable that model testing on independent data indicated the much  
23 poorer performance of models solved by predicting mean Ellenberg indices from a  
24 minimum set of soil variables. This emphasises the extent to which mean Ellenberg  
25 indices are not well explained by the available soil variables yet the models based

1 solely on mean Ellenberg values calculated from the species composition itself  
2 performed much better. This cannot have been due to circularity between modelled  
3 species presence and mean Ellenberg value in the training data because explanatory  
4 variables for the focal species' were always excluded from occupied plots during  
5 model building despite the possibility that this could lead to substantial reductions in  
6 information content in species poor plots. The high shared variance between measured  
7 soil variables used to predict mean Ellenberg values reflects predictable relationships  
8 between soil pH, soil moisture, %C and %N but their poor combined ability to predict  
9 mean Ellenberg values must highlight the importance of other factors such as primary  
10 productivity and other controls on local species composition including species pool  
11 effects and the legacy of land-use history (Smart & Scott 2004). The expected link  
12 between soil conditions and indicator values can also be weakened by managed  
13 disturbance (Diekmann 2003). This may also explain the very poor performance of  
14 the soil and mean Ellenberg regression models when applied to the test data since a  
15 large proportion of these data sampled managed grasslands. It is also possible that the  
16 estimation of volumetric soil moisture content from soil texture values reduced the  
17 accuracy of the soil-derived predictions. Extended testing on a wider range of datasets  
18 is clearly essential.

19

20 *GLM as a tool for capturing niche properties - skewness, variable tolerances and*  
21 *bimodality*

22 Despite the recent demonstration of an array of more sophisticated modelling  
23 techniques (Guisan et al. 2006; Elith et al. 2006), GLM remains an attractive member  
24 of the family of species distribution modelling methods (eg. Manel et al. 1999; Bailey  
25 et al. 2002; Berg et al. 2004). Given its simplicity, GLM performs well in comparison

1 with other techniques (Ellith et al. 2006) and its lower propensity to overfit training  
2 datasets appears to favour application in regions other than the one used to train the  
3 models (Randin et al. 2006; Stockwell & Peterson, 2002 but also see Araújo et al.  
4 2005). The most persuasive reason for using GLM was that interaction terms could be  
5 explicitly modelled and the empirical realised niche for each species described using a  
6 single, simple equation readily applicable to new situations, and amenable to the  
7 derivation of *pmax* values and the optimum abiotic conditions associated with them  
8 (Roy et al., 2000; Coudun & Gégout, 2006). Incorporation of interaction terms was  
9 especially important because many plant species are known to exhibit conditional  
10 responses to abiotic conditions on one gradient that depend upon their position along  
11 another (Diekmann 2003; Smart et al. 2006; Pakeman et al. 2008). By modelling  
12 quadratic and interaction terms along multiple gradients, we found that important  
13 aspects of species' responses can be recovered, such as skewed optima, multiple  
14 optima and varying tolerances (Minchin 1987; Huston 2002; Oksanen & Minchin  
15 2002; Austin et al. 2006). The GLM models satisfactorily captured differences in  
16 tolerance around species optima and therefore reflected varying degrees of specialist  
17 versus generalist behaviour along different abiotic gradients (Fig 4 a-b). Skewed  
18 optima with asymmetric response curves are typical of species associated with abiotic  
19 extremes. These patterns are also recoverable by the GLM approach (Fig 4 c-e).

20         When applied to multiple gradients and their interactions, GLM can also  
21 capture bimodality even if 3<sup>rd</sup> or 4<sup>th</sup> order terms are absent. *Cirsium dissectum* for  
22 example, has two apparent peaks of occurrence along the wetness (Ellenberg F) axis  
23 and these are expressed in a model with no 3<sup>rd</sup> or 4<sup>th</sup> order terms but with significant  
24 interaction terms for Ellenberg R\*Ellenberg N, Ellenberg F\*canopy height and  
25 Ellenberg F\*Ellenberg N (Fig 4 f and Appendix 3). This species is associated with a

1 wider range of productivity in its somewhat drier situations but only with highly  
2 infertile conditions in its wettest habitats where the vegetation also tends to be shorter.  
3 The interaction between Ellenberg R and N appears to reflect the fact that its near  
4 neutral loci tend to be less productive than its higher pH situations.

5 Including multiple explanatory variables in combination with their interaction  
6 terms can clearly improve the fit of each model along each individual gradient without  
7 invoking additional parameters to capture multiple or skewed modes. Hence models  
8 appear capable of capturing situations where the separation of different optima  
9 depends upon the conditional influence of other gradients (Pakeman et al. 2008).

10

#### 11 *Prospects for modelling the impact of multiple drivers at small scales*

12 Current global ecological change involves multiple human drivers and may  
13 lead to novel environmental states (Steffen et al. 2004). These aspects emphasise the  
14 importance of models that can be used to reliably explore future scenarios of change  
15 as well as being used to attribute variation in historical observations to alternative  
16 causes. However, scientific and policy interest still centres on the fate of many  
17 individual species even though it is not feasible to build a process model for each.  
18 Pragmatic approaches to the problem are required. For example, the impact of  
19 multiple drivers on ecosystem functions such as primary production and nutrient  
20 cycling, can be modelled dynamically by the best available process models and the  
21 outputs then used to solve empirical niche models over a range of time steps thus  
22 generating trajectories of change in habitat suitability for selected indicator species  
23 including those known to be present in local species pools (van Dobben et al. 2004;  
24 Wamelink et al. 2005b).

1           In terms of model application and future development, two conclusions can be  
2 drawn from our model tests. The better performance of niche models solved using  
3 mean indicator values derived from the observed species composition rather than  
4 solved using soil variables indicates that factors in addition to soil measurements need  
5 to be taken into account. However, soil conditions ought to be powerful predictors of  
6 species presence. It is possible that ineffective sampling rather than low explanatory  
7 power *per se* is responsible. For example many soil measurements rely on small  
8 bulked samples taken at one point in time. Thus any one soil sample may poorly  
9 represent the average spatial and temporal conditions experienced by the range of  
10 species censused in the plot and more effectively integrated by the mean Ellenberg  
11 values (Schaffers & Šýkora 2000; Diekmann 2003; Hill & Carey 1997).

12           At present the niche models are most reliably applied at the vegetation patch  
13 scale. Thus habitat suitability indices can be generated using observed species  
14 composition and canopy height with or without local climate data. Such applications  
15 could quantify changes in the appropriateness of current conditions for maintenance  
16 or reintroduction of target species, or help evaluate the impact of observed or  
17 expected changes in species composition on habitat suitability for possible invaders.  
18 Linking the niche models to outputs from biogeochemical models is also required to  
19 test scenarios of the impact of changing atmospheric nitrogen deposition on plant  
20 assemblages across Europe (DeVries et al. in press). Since, soil variables such as pH,  
21 %C and %N are outputs from these models, improvements in the accuracy of such  
22 model chains will depend on better modelling of species presence in terms of  
23 measured soil variables. This can be achieved in two ways; by either developing  
24 better predictive models of mean indicator values, which we have shown can  
25 effectively model presence of many British species, or by directly predicting species

1 presence in terms of soil and other data as a result of an extended campaign of new  
2 environmental measurements at the places species grow. For many national territories  
3 and their floras the second option is likely to be prohibitively costly. We suggest that  
4 modelling via mean indicator values is a more efficient tactic because it is likely to be  
5 less costly to improve the prediction of mean indicator values along each abiotic  
6 gradient via new targeted soil sampling, than to implement species-specific campaigns  
7 of new soil sampling to improve the predictive power of each separate species model.  
8 Having established that mean indicator gradients provide effective explanatory power  
9 for a large number of species, we suggest that new sampling should focus on a larger  
10 number of soil factors sampled with greater temporal and spatial resolution along each  
11 abiotic gradient.

12

### 13 **Acknowledgements**

14 We thank the following for advice Ian Strachan, Mike Ashmore, Mike Hornung,  
15 Samantha Baker, Andrew Stott, Peter Costigan, Clare Whitfield, Keith Kirby, Richard  
16 Smithers, Bob Bunce, Richard Brand-Hardy, Geoff Kirby, Philip Corney, Simon  
17 Bareham, Clive Walmsley, John Rodwell, Wieger Wamelink, Han van Dobben, Beth  
18 Purse, Dan Chapman and Sarah Honour. Three referees made valuable suggestions on  
19 a previous draft. Financial support for part of this work was provided by the  
20 Department of Environment, Food and Rural Affairs and the UK Natural  
21 Environment Research Council.

22

### 23 **Supplementary material**

24 Appendix 1 - Mathematical estimation of species optima ( $p_{max}$ ) from fitted logistic  
25 regressions with multiple gradients, quadratic and interaction terms

1

2 Appendix 2 - Empirical estimation of species optima (*pmax*) from multiple logistic  
3 regressions equations

4

5 Appendix 3 – GLM model equations for 1130 British plant species

6

7 Appendix 4 – GLM model equations for 639 species incorporating abiotic patch-scale  
8 explanatory variables and three climate variables at 5km<sup>2</sup> scale.

9

10 Appendix 5 – Graphs and notes on patterns of commonness and rarity among  
11 modelled species compared to the whole of the UK native higher and lower plant  
12 flora.

13

1 **References**

2 Anon. 1988. *Agricultural Land Classification of England and Wales. Revised*  
3 *Guidelines and Criteria for Grading the Quality of Agricultural Land*. MAFF  
4 Publications, Alnwick, UK.

5

6 Araújo, M.B. & New, M. 2006. Ensemble forecasting of species distributions. *Trends*  
7 *in Ecology and Evolution* 22: 42-47.

8

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

11

12 Austin, M.P., Belbin, L., Meyers, J.A., Doherty, M.D. & Luoto, M. 2006. Evaluation  
13 of statistical models used for predicting plant species distributions: role of artificial  
14 data and theory. *Ecological Modelling* 199: 197-216.

15

16 Bailey, S-A., Haines-Young, R.H. & Watkins, C. 2002. Species presence in  
17 fragmented landscapes: modelling of species requirements at the national level.  
18 *Biological Conservation* 108: 307-316.

19

20 Bakkenes, M., Alkemade, J.R.M., Ihle, F., Leemans, R. & Latour, J.B. 2002.  
21 Assessing effects of forecasted climate change on the diversity and distribution of  
22 European higher plants for 2050. *Global Change Biology* 8: 390-407.

23

24 Barry, S. & Elith, J. 2006. Error and uncertainty in habitat models. *Journal of Applied*  
25 *Ecology* 43: 413-423.

1

2 Beale, C.M., Lennon, J.J. & Gimona, A. 2008. Opening the climate envelope reveals  
3 no macroscale associations with climate in European birds. *Proceedings of the*  
4 *National Academy of Sciences of the USA* 105: 14908-14912.

5

6 Berg, Å., Gärdenfors, U. & von Poschwitz, T. 2004. Logistic regression models for  
7 predicting occurrence of terrestrial molluscs in southern Sweden – importance of  
8 environmental data quality and model complexity. *Ecography* 27: 83-93.

9

10 Bio, A.M.F. 2000. *Does vegetation suit our models? Data and model assumptions*  
11 *and the assessment of species distribution in space*. PhD thesis. Utrecht, Netherlands.  
12 195 pp.

13

14 Black, H.I.J. , Garnett, J.S. , Ainsworth, G., Coward, P.A., Creamer, R., Ellwood, S.,  
15 Horne, J., Hornung, M., Kennedy, V.H., Monson, F., Raine, L., Osborn, D., Parekh,  
16 N.R., Parrington, J., Poskitt, J.M., Potter, E., Reeves, N., Rowland, A.P., Self, P.  
17 Turner, S., Watkins, J., Woods, C. & Wright, J. 2001. *MASQ: Monitoring And*  
18 *Assessing Soil Quality In Great Britain. Countryside Survey Module 6: Soils and*  
19 *Pollution*. R&D Technical Report E1-063/TR. Environment Agency, London, UK.  
20 185 pp.

21

22 Borcard, D., Legendre, P. & Drapeau, P. 1992. Partialling out the spatial component  
23 of ecological variation. *Ecology* 73: 1045–1055.

24

- 1 Braunisch, V., Bollman, K., Graf, R.F. & Hirzel, A.H. 2008. Living on the edge –  
2 Modelling habitat suitability for species at the edge of their fundamental niche.  
3 *Ecological Modelling* 214:, 153-167.
- 4
- 5 Britton, A.J., Carey, P.D., Pakeman, R.J. & Marrs, R.H. 2000. A comparison of  
6 regeneration dynamics following gap creation at two geographically contrasting  
7 heathland sites. *Journal of Applied Ecology* 37: 832-844.
- 8
- 9 Broennimann, O., Treier, U.A., Müller-Schärer, H., Thuiller, W., Peterson, A. T. &  
10 Guisan, A. 2007. Evidence of climatic niche shift during biological invasion. *Ecology*  
11 *Letters* 10: 701-709.
- 12
- 13 Clark, C.M. & Tilman, D. 2008. Loss of plant species after chronic low-level nitrogen  
14 deposition to prairie grasslands. *Nature* 451: 712-715.
- 15
- 16 Corney, P.M., Le Duc, M.G., Smart, S.M., Kirby, K.J., Bunce, R.G.H. & Marrs, R.H.  
17 2006. Relationships between the species composition of forest field-layer vegetation  
18 and environmental drivers assessed using a national survey. *Journal of Ecology* 94:  
19 383-401
- 20
- 21 Coudun, C. & Gégout, J.C. 2006. The derivation of species response curves with  
22 Gaussian logistic regression is sensitive to sampling intensity and curve  
23 characteristics. *Ecological Modelling* 199: 164-175.
- 24
- 25 Crawley, M.J. 2007. *The R Book*. John Wiley & sons, Chichester, UK, p 327.

1  
2 Critchley, C.N.R., Chambers, B.J., Fowbert, J.A., Bhogal, A., Rose, S.C. &  
3 Sanderson, R.A. 2002. Plant species richness, functional type and soil properties of  
4 grasslands and allied vegetation in English Environmentally Sensitive Areas. *Grass &*  
5 *Forage Science* 57, 82-92.  
6  
7 De Vries, W., Wamelink, W., Van Dobben, H., Kros, H., Reinds, G.J., Mol-Dijkstra,  
8 J., Smart, S., Evans, C., Rowe, E., Belyazid, S., Sverdrup, H., Van Hinsberg, A.,  
9 Posch, M., Hettelingh, J.-P., Spranger, T., & Bobbink, R. In press. Use of dynamic  
10 soil-vegetation models to assess impacts of nitrogen deposition on plant species  
11 composition and to estimate critical loads: an overview. *Ecological Applications*.  
12  
13 Diekmann, M. 2003. Species indicator values as an important tool in applied plant  
14 ecology - a review. *Basic and Applied Ecology* 4: 493-506.  
15  
16 Diniz-Filho, J.A.F., Bini, L.M. & Hawkins, B.A. 2003. Spatial autocorrelation and red  
17 herrings in geographical ecology. *Global Ecology and Biogeography* 12: 53-64.  
18  
19 Dodd, M.E., Silvertown, J., McConway, K., Potts, J. & Crawley, M.J. 1994. Stability  
20 in the plant communities of the Park Grass experiment: the relationships between  
21 species richness, soil pH and biomass variability. *Philosophical Transactions of the*  
22 *Royal Society London. Series B* 346: 185-193.  
23  
24 Dormann, C.F. 2007. Effects of incorporating spatial autocorrelation into the analysis  
25 of species distribution data. *Global Ecology and Biogeography* 16: 129-138.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24

Dunnett, N.P., Willis, A.J., Hunt, R. & Grime, J.P. 1998. A 38-year study of relations between weather and vegetation dynamics in road verges near Bibury, Gloucestershire. *Journal of Ecology* 86:, 610-623.

Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberon, J., Williams, S., Wisz, M.S. & Zimmermann, N.E. 2006. Novel methods improve prediction of species distributions from occurrence data. *Ecography* 29: 129-151.

Ellenberg, H., Weber, H.E., Dull, R., Wirth, V., Werner, W. & Paulissen, D. 1991. Zeigerwerte von Pflanzen in Mitteleuropa. *Scripta Geobotanica* 18: 1-248.

Emmett, B.A., Beier, B., Estiarte, M., Tietema, A., Kristensen, H.L., Williams, D., Penuelas, J., Schmidt, I. & Sowerby, A. 2004. The response of soil processes to climate change: results from manipulation studies of shrublands across an environmental gradient. *Ecosystems* 7: 625-637.

Ertsen, A.C.D., Alkemade, J.R.M. & Wassen, M.J. 1998. Calibrating Ellenberg indicator values for moisture, acidity, nutrient availability and salinity in the Netherlands. *Plant Ecology* 135: 113-124.

1 Geider, R.J., Delucia, E.H., Falkowski, P.G., Finzi, A.C., Grime, J.P., Grace, J., Kana,  
2 T.M., La Roche, J., Long, S.P., Osborne, B.A., Platt, T., Prentice, I.C., Raven, J.A.,  
3 Schlesinger, W.H., Smetacek, V., Stuart, V., Sathyendranath, S., Thomas, R.B.,  
4 Vogelmann, T.C., Williams, P. & Woodward, F.I. 2001. Primary productivity of  
5 planet earth: biological determinants and physical constraints in terrestrial and aquatic  
6 systems. *Global Change Biology* 7: 849-882.  
7  
8 Grime, J.P., Hodgson, J.G., Hunt, R., Thompson, K. 1995. *The Electronic*  
9 *Comparative Plant Ecology*, Chapman & Hall, London, UK.  
10  
11 Guisan, A. & Theurillat, J-P. 2000. Equilibrium modelling of alpine plant distribution:  
12 how far can we go. *Phytocoenologia* 30, 353-384.  
13  
14 Guisan, A., Thomas, C., Edwards, C. & Hastie, T. 2002. Generalized linear and  
15 generalized additive models in studies of species distributions: setting the scene.  
16 *Ecological Modelling* 157: 89-100.  
17  
18 Guisan, A., Lehmann, A., Ferrier, S., Austin, M.P., Overton, J.M.C., Aspinall, R. &  
19 Hastie, T. 2006. Making better biogeographical predictions of species' distributions.  
20 *Journal of Applied Ecology* 43: 386-392.  
21  
22 Hawkins, B.A., Diniz-Filho, J.A.F., Bini, L.M., De Marco, P. & Blackburn, T.M.  
23 2007. Red herring revisited: spatial autocorrelation and parameter estimation in  
24 geographical ecology. *Ecography* 30: 375-384.  
25

1 Hill, M.O. & Carey, P.D. 1997. Prediction of yield in the Rothamsted Park Grass  
2 Experiment by Ellenberg indicator values. *Journal of Vegetation Science* 8: 579-586.  
3  
4 Hill, M.O., Mountford, J.O., Roy, D.B. & Bunce, R.G.H. 1999. *Ellenbergs' indicator*  
5 *values for British plants. ECOFACT Volume 2, Technical Annex.* Department of the  
6 Environment, Transport and the Regions, London, UK. 47pp.  
7  
8 Hill, M.O., Roy, D.B., Mountford, J.O. & Bunce, R.G.H. 2000. Extending  
9 Ellenbergs' indicator values to a new area: an algorithmic approach. *Journal of*  
10 *Applied Ecology* 37: 3-15.  
11  
12 Hill, M.O., Preston, C.D. & Roy, D.B. 2004. *PLANTATT Attributes of British and*  
13 *Irish Plants: Status, Size, Life History, Geography and Habitats.* Centre for Ecology  
14 and Hydrology, Monkswood, Abbots Ripton, Huntingdon, UK. 78pp.  
15  
16 Hill, M.O., Preston, C.D., Bosanquet, S.D.S. & Roy, D.B. 2007. *BRYOATT Attributes*  
17 *of British and Irish Mosses, Liverworts and Hornworts.* Centre for Ecology and  
18 Hydrology, Monkswood, Abbots Ripton, Huntingdon, UK. 36pp.  
19  
20 Hirzel, A., Helfer, V. & Me'tral, F. 2001. Assessing habitat suitability models with a  
21 virtual species. *Ecological Modelling* 145: 111-121.  
22  
23 Hirzel, A. & Guisan, A. 2002. Which is the optimal sampling strategy for habitat  
24 suitability modelling. *Ecological Modelling* 157: 331-341  
25

1 Hornung, M.H. 1996. *The Key Habitats Survey of Britain*. Final report to DEFRA.  
2 Centre for Ecology and Hydrology, Lancaster, UK.  
3  
4 Kirby, K.J., Smart, S.M., Black, H.I.J., Bunce, R.G.H., Corney, P.M. & Smithers, R.J.  
5 2005. *Long-term ecological change in British woodland (1971-2001)*. English Nature  
6 Research report 653. Peterborough, UK. 137pp.  
7  
8 Lájér, K. 2007. Statistical tests as inappropriate tools for data analysis performed on  
9 non-random samples of plant communities. *Folia Geobotanica* 42: 115-122.  
10  
11 Leibold, M.A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J.M., Hoopes,  
12 M.F., Holt, R.D., Shurin, J.B., Law, R., Tilman, D., Loreau, M. & Gonzalez, A. 2004.  
13 The metacommunity concept: a framework for multi-scale community ecology  
14 *Ecology Letters* 7, 601–613.  
15  
16 Lindborg R. & Eriksson, O. 2004. Historical landscape connectivity affects present  
17 plant species diversity. *Ecology* 85: 1840-1845.  
18  
19 Little, R.C., Milliken, G.A., Stroup, W.W. & Wolfinger, R.D. 2000. *SAS System for*  
20 *Mixed Models*. 4<sup>th</sup> edn. SAS Institute Inc. Cary, North Carolina, USA, pp.437-432.  
21  
22 Liu, C., Berry, P.M., Dawson, T.P. & Pearson, R.G. 2005. Selecting thresholds of  
23 occurrence in the prediction of species distributions. *Ecography* 28: 385-393.  
24

1 Lobo, J.M., Jiménez-Valverde, A. & Real, R. 2008. AUC: a misleading measure of  
2 the performance of predictive distribution models. *Global Ecology and Biogeography*  
3 17: 145-151.

4

5 Luoto, M., Virkkala, R. & Heikkinen, R.K. 2007. The role of land cover in  
6 bioclimatic models depends on spatial resolution. *Global Ecology and Biogeography*  
7 16: 34-42.

8

9 Manel S., Dias, J.M., Buckton, S.T. & Ormerod, S.J. 1999. Alternative methods for  
10 predicting species distribution: an illustration with Himalayan birds. *Journal of*  
11 *Applied Ecology* 36: 734-747.

12

13 Manel, S., Williams, H. C. & Ormerod, S. J. 2001. Evaluating presences-absence  
14 models in ecology: the need to account for prevalence. *Journal of Applied Ecology*  
15 38: 921-931.

16

17 Manning, P., Putwain, P.D. & Webb, N.R. 2005. Formulating a general statistical  
18 model for *Betula* spp. invasion of lowland heath ecosystems. *Journal of Applied*  
19 *Ecology* 42: 1031-1041.

20

21 Minchin, P.R. 1987. Simulation of multidimensional community patterns: towards a  
22 comprehensive model. *Vegetatio* 71: 145–156.

23

24 Oksanen, J. & Minchin, P.R. 2002. Community theory revisited: what shape are  
25 species responses along ecological gradients? *Ecological Modelling* 157: 119-129.

1

2 Ozinga, W.A., Schaminee, J.H.J, Bekker, R.M., Poschlod, P., Tackenberg, O.,  
3 Bakker, J. & van Groenendael, J.M. 2005. Predictability of plant species composition  
4 from environmental conditions is constrained by dispersal limitation. *Oikos* 108: 555-  
5 561.

6

7 Pakeman, R.J., Reid, C.L., Lennon, J.J. & Kent, M. 2008. Possible interactions  
8 between environmental factors in determining species optima. *Journal of Vegetation*  
9 *Science* 19: 201-208.

10

11 Preston, C.D. 2007. Which vascular plants are found at the northern or southern edges  
12 of their European range in the British Isles? *Watsonia* 26: 253-269.

13

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

16

17 Randin, C.F., Dirnbock, T., Dullinger, S., Zimmerman, N.E., Zappa, M. & Guisan, A.  
18 2006. Are niche-based species distribution models transferable in space? *Journal of*  
19 *Biogeography* 33: 1689-1703.

20

21 Rasran, L, Vogt, K. & Jensen, K. 2007. Effects of litter removal and mowing on  
22 germination and establishment of two fen-grassland species along a productivity  
23 gradient. *Folia Geobotanica* 42: 271-288.

24

1 Rodwell, J.S. 1991. *Woodland & Scrub. British Plant Communities. Vol 1.* Cambridge  
2 University Press, Cambridge, UK.  
3  
4 Roy, D.B., Hill, M.O., Rothery, P. & Bunce, R.G.H. 2000. Ecological indicator values  
5 of British species: an application of Gaussian logistic regression. *Annales Botanisk*  
6 *Fennici* 37: 219-226.  
7  
8 Sala, E.O., Chapin, F.S., Armesto, J.J., Berlow, E, Bloomfield, J, Dirzo, R, Huber-  
9 Sanwald, E., Huenneke, L.F., Jackson, R.B., Kinzig, A, Leemans, R, Lodge, D.M.,  
10 Mooney, H.A., Oesterheld, M., Poff, N.L., Sykes, M.T., Walker, B.H., Walker, M. &  
11 Wall, D.H. 2000. Biodiversity – Global scenarios for the year 2100. *Science* 287:  
12 1770-1774.  
13  
14 SAS Institute Inc. 1999a. *SAS Macro Language: Reference, Version 8.* Cary, NC:  
15 SAS Institute Inc., USA, 301 pp.  
16  
17 SAS Institute Inc. 1999b. *SAS/STAT User's Guide Version 8.* Cary, NC: SAS Institute  
18 Inc., USA. 3864 pp.  
19  
20 Satterthwaite, F.E. 1946. An approximate distribution of estimates of variance  
21 components. *Biometrics* 2: 110-114.  
22  
23 Schaffers, A.P. & Sýkora, K.V. 2000. Reliability of Ellenberg indicator values for  
24 moisture, nitrogen and soil reaction: comparison with field measurements. *Journal of*  
25 *Vegetation Science* 11: 225-244.

1

2 Silvertown, J., Dodd, M.E., McConway, K., Potts, J. & Crawley, M. 1994. Rainfall,  
3 biomass variation and community composition in the Park Grass experiment. *Ecology*  
4 75: 2430-2437.

5

6 Singer, J.D. 1998. Using SAS PROC MIXED to fit multi-level models, hierarchical  
7 models and individual growth models. *Journal of Educational and Behavioural*  
8 *Statistics* 24: 323-355.

9

10 Smart, S.M., Clarke, R.T., van de Poll, H.M., Robertson, E.J., Shield, E.R., Bunce,  
11 R.G.H. & Maskell, L.C. 2003. National-scale vegetation change across Britain; an  
12 analysis of sample-based surveillance data from the Countryside Surveys of 1990 and  
13 1998. *Journal of Environmental Management* 67: 239-254.

14

15 Smart, S.M., Scott, W.A. 2004. Bias in Ellenberg indicator values – problems with  
16 detection of the effect of vegetation type. *Journal of Vegetation Science* 15: 843-846.

17

18 Smart, S.M., Le Duc, M., Marrs, R.H., Rossall, M.J., Bunce, R.G.H., Thompson, K.  
19 & Firbank, L.G. 2006. Spatial relationships between intensive land cover and residual  
20 plant species diversity in temperate, farmed landscapes. *Journal of Applied Ecology*  
21 43: 1128-1137.

22

23 Steffen, W., Sanderson, A., Tyson, P.D., Jager, J., Matson, P.A., Moore III, B.,  
24 Oldfield, F., Richardson, K., Schellnhuber, H.J., Turner, B.L., Wasson, R.J. 2004.

1 *Global change and the earth system: a planet under pressure*. Springer-Verlag,  
2 Berlin, Germany. 332 pp.  
3  
4 Stevens, C.J., Dise, N.B., Mountford, J.O. & Gowing, D.J. 2004. Impacts of nitrogen  
5 deposition on the species richness of grasslands. *Science* 303: 1876-1879.  
6  
7 Stockwell, D.R.B. & Peterson, A.T. 2002. Effects of sample size on accuracy of  
8 species distribution models. *Ecological Modelling* 148: 1-13.  
9  
10 Strengbom, J., Nordin, A., Nasholm, T., Ericson, L. 2002. Parasitic fungus mediates  
11 change in nitrogen-exposed boreal forest vegetation. *Journal of Ecology* 90: 61-67.  
12  
13 Svenning, J.-C. & Skov, F. 2004. Limited filling of the potential range in European  
14 tree species. *Ecology Letters* 7: 565-573.  
15  
16 Swets, J. 1988. Measuring the accuracy of diagnostic systems. *Science* 240, 1285-  
17 1293.  
18  
19 Terry, A.C., Ashmore, M.R., Power, S.A., Allchin, E.A. & Heil, G.W. 2004.  
20 Modelling the impacts of atmospheric nitrogen deposition on *Calluna*-dominated  
21 ecosystems in the UK. *Journal of Applied Ecology* 41: 897-909.  
22  
23 Thuiller, W., Lavorel, S., Araújo, M.B., Sykes, M.T. & Prentice, I.C. 2005. Climate  
24 change threats to plant diversity in Europe. *Proceedings of the National Academy of*  
25 *Sciences of the USA* 102: 8245-8250.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25

Trivedi, M.R., Berry, P.M., Morecroft, M.D. & Dawson, T.P. 2008. Spatial scale affects bioclimate model projections of climate change impacts on mountain plants. *Global Change Biology* 14: 1089-1103.

Tyre, A.J., Possingham, H.P. & Lindenmayer, D.B. 2001. Matching observed pattern with ecological process: can territory occupancy provide information about life history parameters? *Ecological Applications* 11: 1722-1738.

UK Biodiversity Steering Group. 1995. *Biodiversity: The UK Steering Group Report – Volume I: Meeting the Rio Challenge*. HMSO, London, UK. 93 pp.

Van Dobben, H.F., Schouwenberg, E.P.A.G., Mol, J.P., Wieggers, H.J.J., Jansen, M.J.W., Kros, J. & de Vries, W. 2004. *Simulation of critical loads for nitrogen for terrestrial plant communities in the Netherlands*. Alterra-rapport 953. Wageningen, Netherlands. 84 pp.

Wamelink, G.W.W., Joosten, V., van Dobben, H.F. & Berendse, F. 2002. Validity of Ellenberg indicator values judged from physico-chemical measurements. *Journal of Vegetation Science* 13: 269-278.

Wamelink, G.W.W., Ter Braak, C.J.F. & Van Dobben, H.F. 2003. Changes in large-scale patterns of plant biodiversity predicted from environmental economic scenarios. *Landscape Ecology* 18: 513-527.

- 1 Wamelink, G.W.W., Goedhart, P.W., van Dobben, H.F. & Berendse, F. 2005a. Plant  
2 species as predictors of soil pH: replacing expert judgement with measurements.  
3 *Journal of Vegetation Science* 16: 461-470.  
4
- 5 Wamelink, G.W.W., de Jong, J.J., van Dobben, H.F. & Van Wijk, M.N. 2005b.  
6 Additional costs of nature management caused by deposition. *Ecological Economics*  
7 52: 437-451.  
8
- 9 Walker, K.J. 2007. The last thirty five years: recent changes in the flora of the British  
10 Isles. *Watsonia* 26: 291-302.

1 Table 1. Plant species occurrence data used to train British niche models for higher  
 2 and lower plants.  
 3

<b>Survey source</b>	<b>Sample size</b>	<b>Sampling method</b>	<b>Number of samples</b>	<b>Date of collection</b>
National Vegetation Classification (Rodwell, 1991 et seq)	4m <sup>2</sup> to 250m <sup>2</sup>	Phytosociological – targeted homogenous stands to census full range of plant community variation across Britain	31266	1950 - 1980
The Key Habitats surveys of Britain (Hornung, 1996)	4m <sup>2</sup>	Random within a random selection of 1km squares stratified by three landscape GIS masks for upland, calcareous grassland and lowland heathland ecosystems	548	1992
Broadleaved woodland survey of Britain (Kirby et	200m <sup>2</sup>	16 plots at random within each of 103 representative sites	1648	1971

al., 2005; Corney et		Britain		
al., 2006)				
Countryside Survey	4 to 200m <sup>2</sup>	Random plots	7221	1998/'99
of Great Britain		within a random		
(Smart et al., 2003)		selection of 569		
		1km squares across		
		Britain stratified by		
		ITE Land Class		

---

1

1 Table 2. Calibration equations that predict values of mean unweighted Ellenberg  
 2 values from four soil variables; MC (%soil moisture content), soil pH, C (%organic  
 3 carbon) and N (%organic nitrogen). See text for further details.

4

<b>Unweighted mean</b>	<b>%</b>	<b>Equation</b>
<b>Ellenberg score</b>	<b>variance</b>	<b>explained</b>
F (wetness)	70.1	$=\ln ((MC/(100-MC))+3.27)/0.55$
R (substrate pH)	77.9	$=-0.5293-0.02503(MC)+1.665(pH)-0.1061(pH^2)-$ $0.00566(C)$
N (substrate fertility)	78.2	$=\exp(0.7751-0.00006(MC)-0.00009(MC^2)-$ $0.01475(C)+0.000099(C^2))+0.2639(pH)-$ $0.01684(pH^2)+0.1908(N)$

5

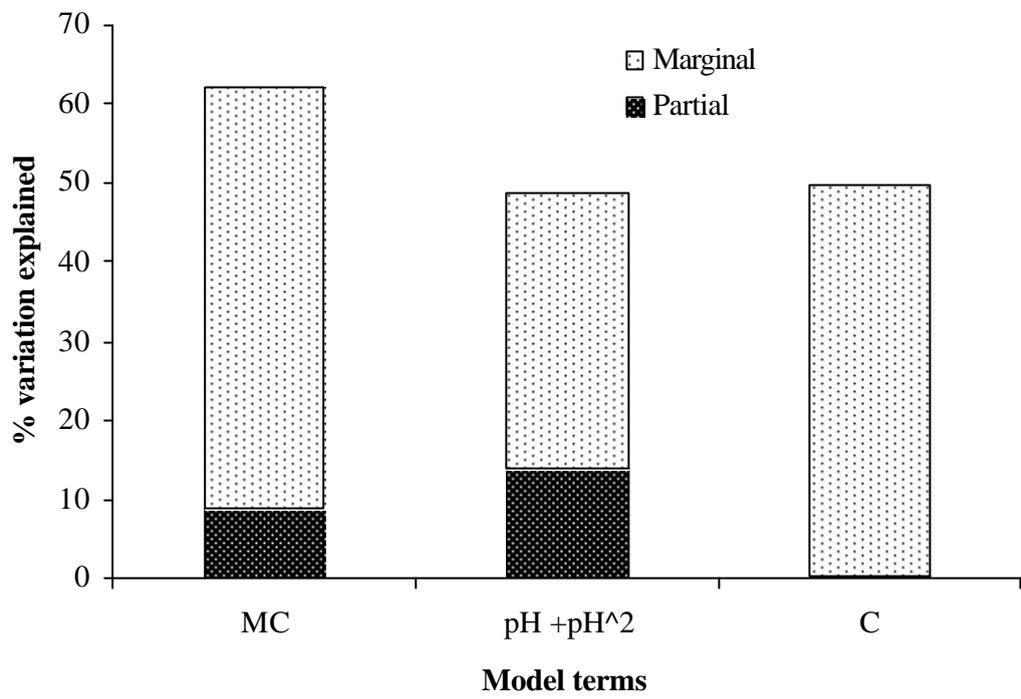
6

- 1 Figure1. Flow diagram showing model building (B) and testing (T) steps. Species
- 2 presence was measured as occupancy in plots that ranged from 2 to 200m<sup>2</sup>. Climate
- 3 variables were based on interpolated data for the 5km<sup>2</sup> containing each plot. Other
- 4 explanatory variables were measured at the plot scale.

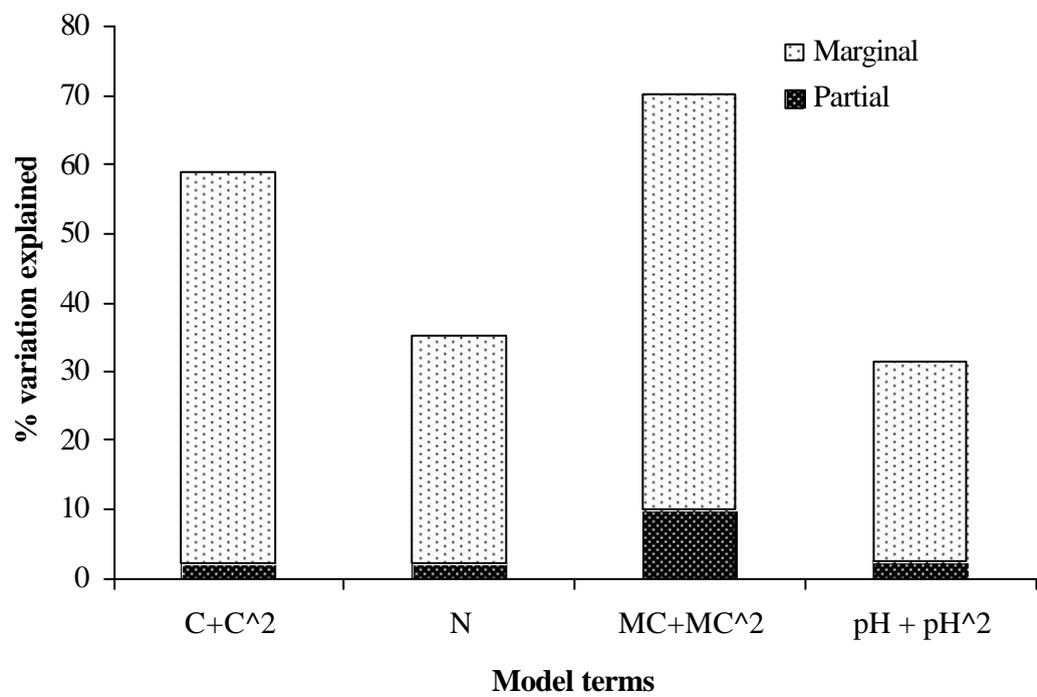


1 Figure 2. Partitioning observed variation in mean Ellenberg indices among  
2 explanatory variables selected for inclusion in the best minimum adequate model for  
3 each index. a. Mean Ellenberg R, b. Mean Ellenberg N. MC = % soil moisture, C = %  
4 organic carbon, N = % organic nitrogen. Marginal=variation uniquely explained by  
5 each variable with no other variables present. Partial=variation uniquely explained  
6 when other variables were included.

1 FIG 2.  
2  
3 a.



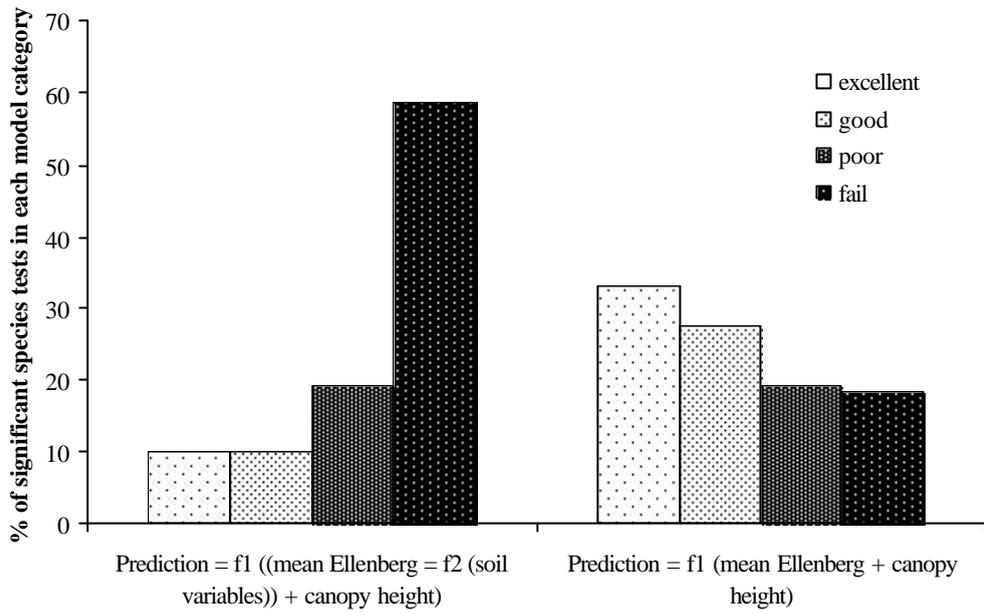
4  
5  
6 b.  
7



8

1 Figure 3. AUC results for GLM model predictions of the presence of 106 higher  
2 plant species in independent vegetation plot data from English Environmentally  
3 Sensitive Areas. Models were solved using cover-weighted canopy height plus either  
4 mean Ellenberg values calculated from the observed species composition of each plot  
5 or using mean Ellenberg values predicted from observed soil data for each plot  
6 (n=244). AUC values classified according to the guidance in Swets (1988); excellent  
7  $\geq 0.9$ , good 0.8 to 0.9, , poor 0.7 to 0.8, fail  $< 0.7$  .

1 FIG 3.  
2  
3



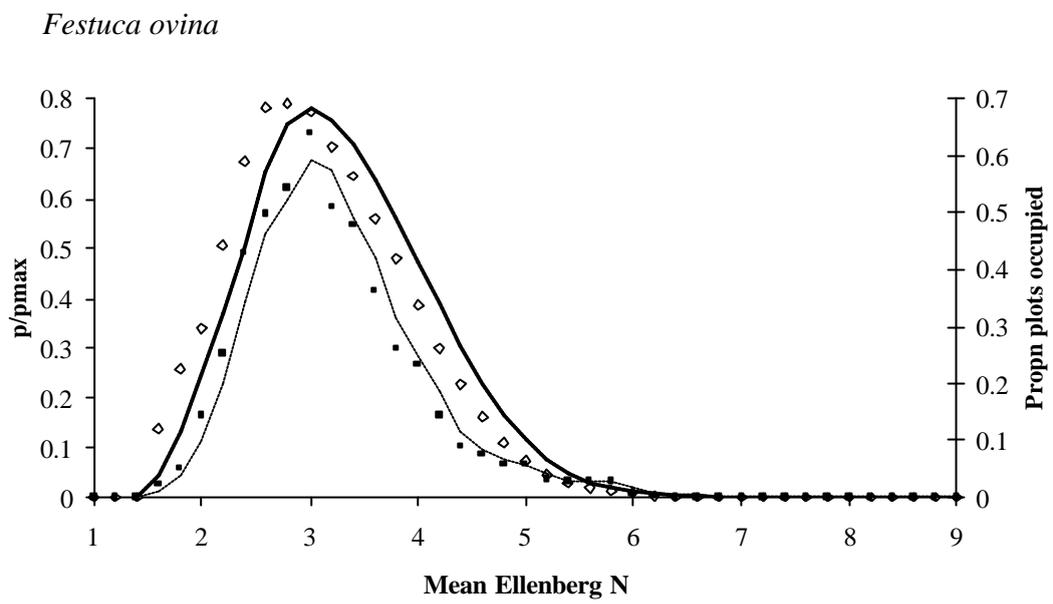
4

1 Figure 4. Examples showing the fit of species models to the GB training dataset. Open  
2 diamonds and the solid line (moving average) show the model prediction of habitat  
3 suitability along each gradient. GLM models were solved at intervals of 0.2 Ellenberg  
4 index units along the x-axis. Terms for other abiotic gradients were then solved in  
5 each interval using the average values for the other explanatory variables in plots  
6 occupied by the species concerned within each interval. In each graph, probabilities  
7 were rescaled by the *pmax* value for each species. Black squares and the dashed line  
8 (moving average) indicate the proportion of plots in each interval actually occupied  
9 by the species.

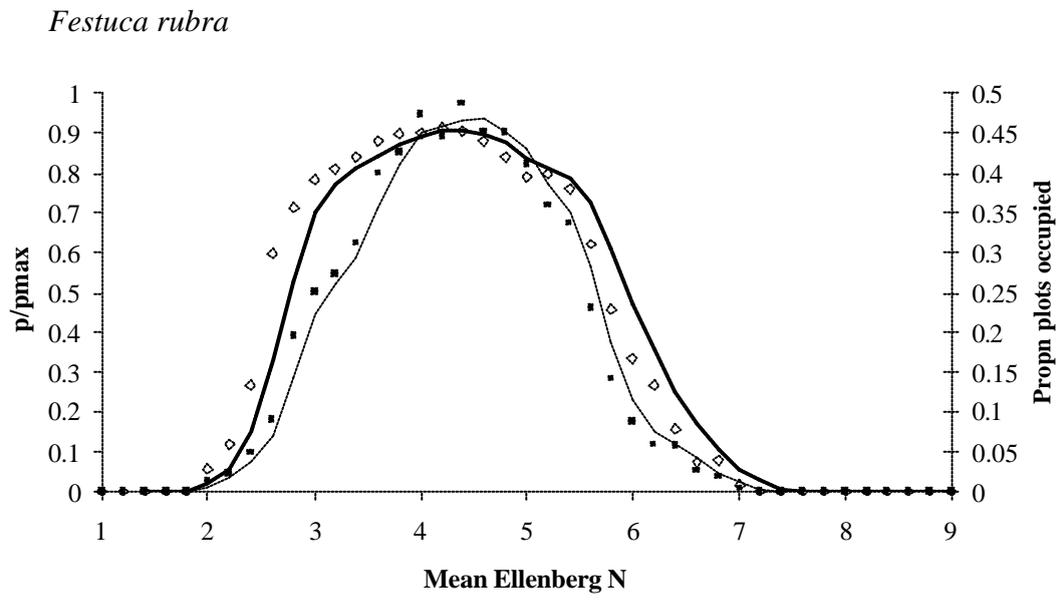
10

11

1 FIG 4.  
2  
3 a.



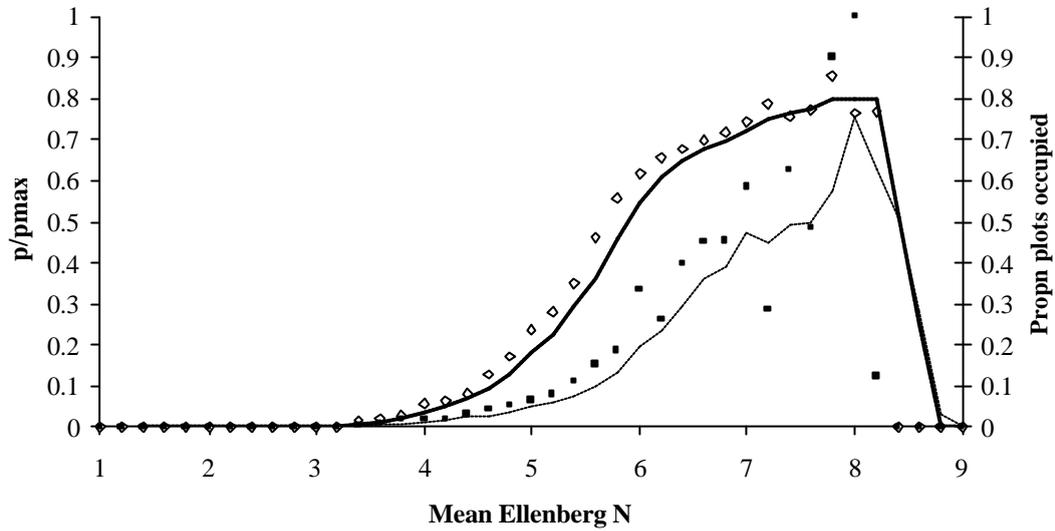
4  
5 b.



6  
7  
8

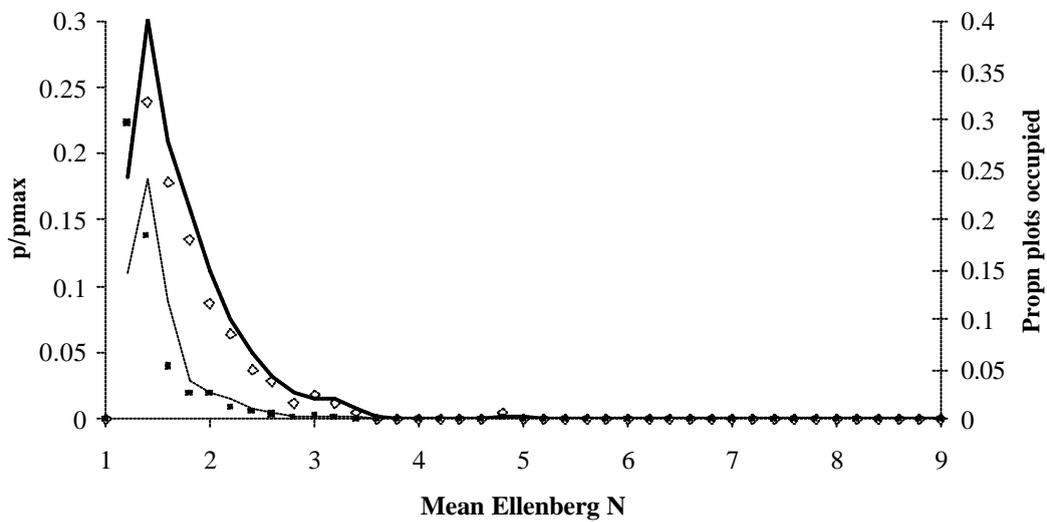
- 1 FIG 4 contd.
- 2
- 3 c.

*Urtica dioica*



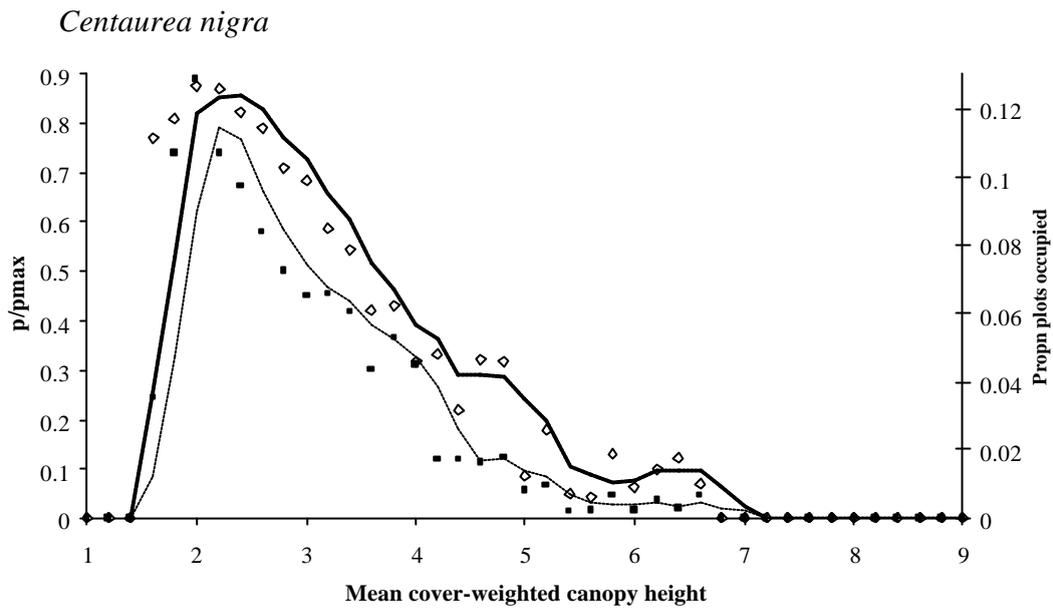
- 4
- 5
- 6 d.

*Vaccinium oxycoccus*

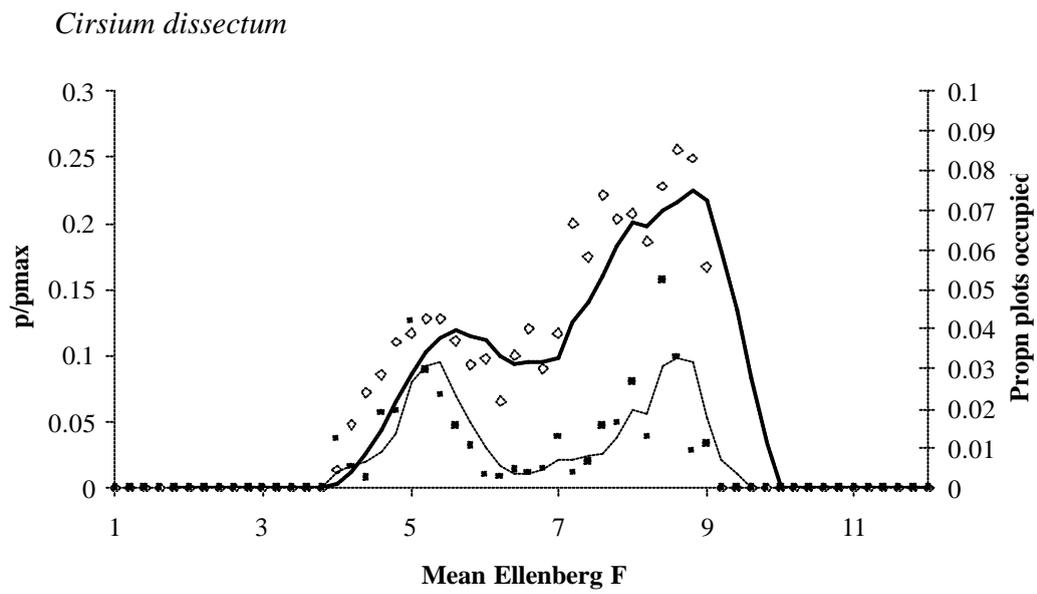


- 7
- 8
- 9
- 10
- 11
- 12
- 13
- 14
- 15
- 16

1 e.



2  
3  
4 f.



5  
6