

Article (refereed) - postprint

Henrys, P.A.; Smart, S.M.; Rowe, E.C.; Jarvis, S.G.; Fang, Z.; Evans, C.D.; Emmett, B.A.; Butler, A. 2015. **Niche models for British plants and lichens obtained using an ensemble approach.**

© Botanical Society of Britain & Ireland 2015

This version available <http://nora.nerc.ac.uk/513147/>

NERC has developed NORA to enable users to access research outputs wholly or partially funded by NERC. Copyright and other rights for material on this site are retained by the rights owners. Users should read the terms and conditions of use of this material at <http://nora.nerc.ac.uk/policies.html#access>

This is an Accepted Manuscript of an article published by Taylor & Francis Group in New Journal of Botany on October 2015 available online:

<http://www.tandfonline.com/doi/full/10.1179/2042349715Y.0000000010>

Contact CEH NORA team at
noraceh@ceh.ac.uk

Niche models for British plants and lichens obtained using an ensemble approach

Summary

Site-occupancy models that predict habitat suitability for plant species in relation to measurable environmental factors can be useful for conservation planning. Such models can be derived from large-scale presence-absence datasets on the basis of environmental observations or, where only floristic data are available, using plant trait values averaged across a plot. However, the estimated modelled relationship between species presence and environmental variables depends on the type of statistical model adopted and hence can introduce additional uncertainty. We used an ensemble-modelling approach to constrain and quantify the uncertainty due to the choice of statistical model, applying generalised linear models (GLM), generalised additive models (GAM), and multivariate adaptive regression splines (MARS). Niche models were derived for over 1000 species of vascular plants, bryophytes and lichens, representing a large proportion of the British flora and many species occurring in continental Europe. Each model predicts habitat suitability for a species in response to climate variables and trait-based scores (evaluated excluding the species being modelled) for soil pH, fertility, wetness and canopy height. An R package containing the fitted models for each species is presented which allows the user to predict the habitat suitability of a given set of conditions for a particular species. Further functions within the package are included so that these habitat suitability scores can be plotted in relation to individual explanatory variables. A simple case study shows how the R package (MultiMOVE) can be used to quickly and efficiently answer questions of scientific interests, specifically whether climate change will counteract any benefits of sheep-grazing for a particular plant community. The package itself is freely available via <http://doi.org/10.5285/94ae1a5a-2a28-4315-8d4b-35ae964fc3b9>.

Keywords: biodiversity; climate change; envelope; niche occupancy; pollution; R package

32 **1. Introduction**

33

34 Biodiversity loss is a pressing global concern, and can be seen as largely driven by declines in
35 habitat suitability and availability for individual species (Hirzel and Le Lay, 2008). In seeking
36 to understand habitat suitability for any given species, it is important that two key
37 components are considered: the spatial distribution of the species (e.g. as in Groom, 2013;
38 Preston et al., 2013); and the relationship between the species and other influencing factors
39 (e.g. Geddes and Miller, 2012). To aid decision making in conservation practice, the current
40 distribution of a species and likely changes to this under different management, pollutant or
41 climate pressures should be understood. It is therefore useful to be able to define the
42 ecological range of a given species in relation to different environmental factors, i.e. its
43 realised niche, to enable efficient and timely decision making.

44

45 Niche models can be developed from species records when these are sufficiently numerous,
46 by relating presence or absence to environmental data where these are recorded alongside
47 occurrence, and/or to proxy variables based on the traits of co-occurring species (Latour,
48 1993; Smart et al., 2010b). Niche modelling approaches based on correlative analyses
49 contrast with population dynamics models, which require detailed parameterisation to
50 represent processes of reproduction and mortality, emigration and immigration (Crone et al
51 2011). While correlative-based niche models cannot simulate impacts of dynamic population
52 processes, they provide useful indications of how the availability of favourable niche space
53 will change (Dormann et al., 2012; de Vries et al., 2013, Thuiller et al., 2008). Ideally, any
54 model attempting to describe the spatial distribution of a species should be based on a large
55 number of observation records representing equilibrium conditions, to ensure that the
56 predictions are robust (Elith et al 2010).

57

58 Many taxa have been the focus of species niche modelling (Elith & Leathwick 2009). It is
59 particularly useful to predict habitat suitability for plant species, since they deliver
60 supporting ecosystem services such as primary production, nectar provision for pollinating
61 insects, genetic variation for crop breeding and cultural significance for wildlife conservation
62 (Alexander et al., 1997; Costanza et al., 2007; Kremen et al., 2007; UK National Ecosystem
63 Assessment, 2011). Plants also underpin the diversity of other taxa by providing habitat
64 structure and a diverse range of food substrates. Here we present an ensemble of empirical
65 niche models for a large number of higher and lower plants in the British flora comprising all
66 major community dominants and a range of subordinates.

67

68 Previous work developed niche models for a similar group of plant species based on multiple
69 logistic regression (Smart et al 2010b); hereafter referred to in the more generic framework
70 of Generalised Linear Models, GLMs. In recent years the diversity of techniques applied to
71 niche modelling has expanded due, in large part, to the need to overcome issues related to
72 model constraints and interpretability (Elith & Leathwick 2009). However, different

73 modelling approaches can result in different representations of the variation in the observed
74 data, leading to differences in model transferability (Munoz and Felicísimo, 2004; Leathwick
75 et al., 2006; Smart et al. 2010a; Wenger & Olden 2012). This has led to the increase in
76 popularity of an ensemble approach (eg Araújo and New, 2007; Thuiller et al., 2009) where
77 the explainable variation and uncertainty relating specifically to model selection are more
78 robustly conveyed based on output across different model types. The work presented here is
79 based on the addition of a further two techniques to the GLM approach applied in Smart et
80 al. 2010b: generalised additive modelling (GAM); and multivariate adaptive regression
81 splines (MARS). The models produced allow spatial and temporal prediction of change in the
82 favourability of niche space for each species based on outputs from the three modelling
83 techniques, conditional upon measured or predicted environmental conditions.

84

85 The models themselves have been bundled into a publicly available R package to allow the
86 wider community of scientists, land managers and conservation policy makers to query,
87 scrutinise and exploit the fitted models for scientific and decision making purposes. The
88 package facilitates understanding and explanation of species' distributions by allowing clear
89 inspection of species responses along environmental gradients. By applying projected
90 changes in input variables, the user can also explore future scenarios of environmental
91 change (eg. Smart et al 2010a). The aim of this paper is to present a brief overview of the
92 fitted models before introducing the R package containing all the model fits and a clear
93 example of how this can be used to provide speedy and efficient answers to policy-relevant
94 questions.

95

96 **2 Methods and Materials**

97 **2.1 Data**

98

99 Fine-grained data on the presence/absence of plant species were available at a large number
100 of locations throughout the UK from four studies: the Countryside Survey (CS) (Smart et al.
101 2003), GB Woodland Survey (Kirby et al., 2005; Corney et al., 2006), the surveys that
102 provided data for Key Habitat Types (Hornung, 1996) and the National Vegetation
103 Classification (NVC) (Rodwell, 1991 et seq.). We pooled the data from the four different
104 surveys, giving a total of 32272 vegetation plots. The NVC surveys represent the largest
105 source of species data and were designed to sample from the full range of UK plant
106 assemblages, so they include more records for scarce species than would a random survey.
107 The NVC design may therefore have resulted in over-sampling under optimal conditions, so
108 data from surveys with a stratified randomized design (or which at least included an element
109 of random plot location) were also included in order to try and provide an unbiased
110 representation of the entire ecological range of a species, importantly including the tails of
111 the distribution moving away from the optimum. Information on plot size was unavailable
112 for NVC quadrats and therefore no standardisation across plot sizes was possible. To
113 overcome this models were, for species with sufficient data (typically n=30 records), re-fitted

114 using CS data only (for which the plots are all of a standard size and the location follows a
115 stratified, random design).

116

117 The plant species modelled exclude the rarest species in our flora and mainly comprise
118 habitat dominants and a large range of subordinates (*sensu* Grime 1998). Dominant species
119 are responsible for the majority of the primary production at a site and strongly underpin
120 other ecosystem functions (Smith & Knapp 2003; Laughlin 2011). Thus the species modelled
121 comprise a disproportionately large fraction of the biomass and cover in British habitats. The
122 list of species modelled includes 97% of the Common Standards Monitoring indicator
123 species (JNCC, 2004) used to judge conservation value of semi-natural habitats. The
124 suitability of conditions for rare species' populations often depends upon the varying cover
125 and persistence of the more abundant species, so even where not directly modelled,
126 information about the prospects for rare species can be inferred from modelling the
127 responses of their more common associates (Smart 2000, Gogol-Prokurat 2011). Non-native
128 species were also excluded from analysis as they are more likely to be undergoing increases
129 in range and colonisation of suitable niche space. As such species are not in equilibrium, the
130 estimated environmental effects from spatially derived models may be confounded with the
131 effects of incomplete dispersal (Svenning & Skov, 2004).

132

133 In choosing environmental characteristics to define the niche, we selected a set of variables
134 representing abiotic and climatic influences. Climate variables included in the models,
135 chosen due to their relationship with plant physiology and growth (Thuiller et al., 2005),
136 were long-term (1961-1990) annual average: rainfall; minimum January temperature; and
137 maximum July temperature (all of which are available from
138 <http://www.metoffice.gov.uk/climatechange/science/monitoring/ukcp09/>). Abiotic variables
139 selected were based on mean values for trait-based indicators. These were: mean un-
140 weighted Ellenberg scores for soil wetness (F), substrate fertility (N) and soil pH (R) for each
141 vegetation plot. Ellenberg scores are ordinal values that are assigned to each plant species.
142 They were originally designed to reflect habitat preferences in central Europe (Ellenberg *et al.*
143 *al.* 1991), but have subsequently been adapted for use with British higher and lower plant
144 species (Hill *et al.*, 1999, 2004, 2007). Mean Ellenberg scores provide a quantitative proxy for
145 the abiotic characteristics of a particular plot (Diekmann, 2003). Similarly, we used species-
146 specific scores for typical canopy height following the ordinal categorisation of Grime et al.
147 (1988) to derive a mean cover-weighted canopy height trait score. This provides a measure
148 of the successional status of the vegetation, which also reflects management intensity or the
149 frequency and severity of biomass removal.

150

151 Using trait scores allowed for models to be constructed with floristic and climate data only,
152 without the additional need for coincident soil measurements which are rarely available.
153 Consistent with Smart et al (2010b), if the species being modelled was recorded in a plot its
154 trait values were removed prior to calculating the trait-based explanatory variables for the

155 plot, to avoid circularity in the model. Where measurements or model estimates of soil
156 variables are available, translation functions can be used to predict the mean Ellenberg
157 scores required to solve the niche models alongside climate and cover-weighted canopy
158 height (Rowe et al., 2011a; Smart et al., 2010b; Rowe et al., 2014b). Mean Ellenberg scores
159 were not weighted by observed species cover, since little information is gained when doing
160 so relative to the noise in the cover estimates (Valentin et al 2012; Kafer & Witte 2004).
161 Canopy height values were cover-weighted, however, to convey the influence of canopy
162 extent as well as height on shade at ground level.

163

164 2.2 Models

165

166 All surveys recorded the presence or absence of individual species within individual plots,
167 and models with a binomial distribution for the response variable were fitted to this data
168 using three modelling techniques – GLMs (McCullagh et al., 1989), GAMs (Hastie and
169 Tibshirani, 1990) and MARS (Friedman, 1991). Previously, GLMs were fitted to the same
170 vegetation data by Smart et al (2010b) and are useful for modelling simple linear or
171 polynomial responses. However, GLMs cannot fit more complex surfaces which may
172 characterise species' niches. GAMs are a much more flexible class of models, allowing the
173 relationship between the response and any individual predictor to have a smoothly varying
174 form. However, estimating such complex relationships is more difficult, so if the response is
175 in fact simple the models may have predictive power for new data (Smart et al., 2010a). The
176 MARS models are similar to GAMs but instead of fitting smoothed terms they fit “hinges”
177 (Friedman, 1991). They are similarly more flexible than GLMs, but their use of piecewise
178 linearity accommodates different types of responses to GAMs. An ensemble of all three
179 models was fitted for each species.

180

181 For each species the full model contained all seven covariates (long term averages for
182 maximum July temperature, minimum January temperature and annual rainfall, canopy
183 height and Ellenberg F, N and R scores) and all of the 21 possible two-way interactions. This
184 is in contrast to Thuiller et al. (2003) who considered only additive effects without
185 interaction. Correlation across all combinations of variables was assessed and evidence of
186 some relationships was found: out of the 21 pairwise correlations, 5 were either greater
187 than 0.5 or less than -0.5 and the largest correlation (between Ellenberg N and Ellenberg R)
188 was 0.75. However, due to the modelling techniques chosen and the implementation of a
189 suitable model selection routine for each species modelled individually, this was not
190 considered a problem. For the GLM models stepwise backwards selection based on
191 minimum AIC was used to define the final set of covariates and two-way interactions to be
192 considered for each species. Main effects were only considered for removal if all interactions
193 containing these variables had already been removed. The variables and interactions that
194 were used within the GAM models were those chosen in the final selected GLM models,
195 because it was found to be computationally infeasible to perform a separate model selection

196 procedure for the GAMS. Within the MARS framework model selection is performed
197 automatically as part of the model fitting process using forward selection, i.e. starting from a
198 null model and adding in those terms that lead to the greatest improvement in fit.

199

200 All models were fitted in the R statistical environment (R Development Core Team, 2009)
201 using the mgcv (Wood, 2006), earth (Milborrow, 2014) and leaps (Lumley, 2009) packages.
202 Nomenclature follows Stace (2010) for vascular plants, the Atlas of British and Irish
203 Bryophytes (Blockeel *et al.* 2014) and the British Lichen Society Taxon Dictionary (2015;
204 <http://www.britishlichensociety.org.uk/resources/lichen-taxon-database>).

205

206 2.3 Model Checking

207

208 Model performance was assessed by comparing the observed presence absence data (0 or
209 1) to predicted values using AUC (Fielding and Bell, 1997), where predicted values in this
210 case were defined as the estimated probability of presence for each of the three model
211 approaches (GLM, GAM, MARS). Values of AUC close to one indicate good levels of
212 predictive performance whereas a model with predictive power that is no better than
213 chance will return an AUC of 0.5. The AUC values should be interpreted with caution
214 because they effectively treat the cost associated with a false positive as being identical to
215 the cost associated with a false negative, and this may not always be appropriate. Other
216 measures are possible, though the low frequency of presences for the majority of species at
217 observed sites makes measures like false omission rate and negative predictive value
218 unhelpful.

219

220 Since an adequate independent test data set was not available, a cross-validation approach
221 is required to investigate the out-of-sample performance of the fitted models. However,
222 computational costs for checking all species models would be prohibitive, so the full set of
223 models was compared and examined using within-sample AUC diagnostics as described
224 above. Out of sample performance was assessed on a much smaller set of 30 species, listed
225 in Table 1. This set was designed to represent a range of distributions and taxa, and explicitly
226 included a number of species which show evidence of a non-linear relationship between
227 abiotic conditions and prevalence. For each of the 30 selected species, the observed data
228 were subset at random into two components: 75% of the data for training, 25% for testing.
229 Each of the three models were then built on the 75% dataset and AUC values were
230 calculated based on the remaining, independent 25% of the data. We repeated this process
231 10 times for each species.

232

233 The predicted values across the range of the training data were also mapped across GB in
234 order to provide an indication of the spatial extent of each species according to the models.
235 This enabled us to check against expert knowledge and previously produced maps, such as

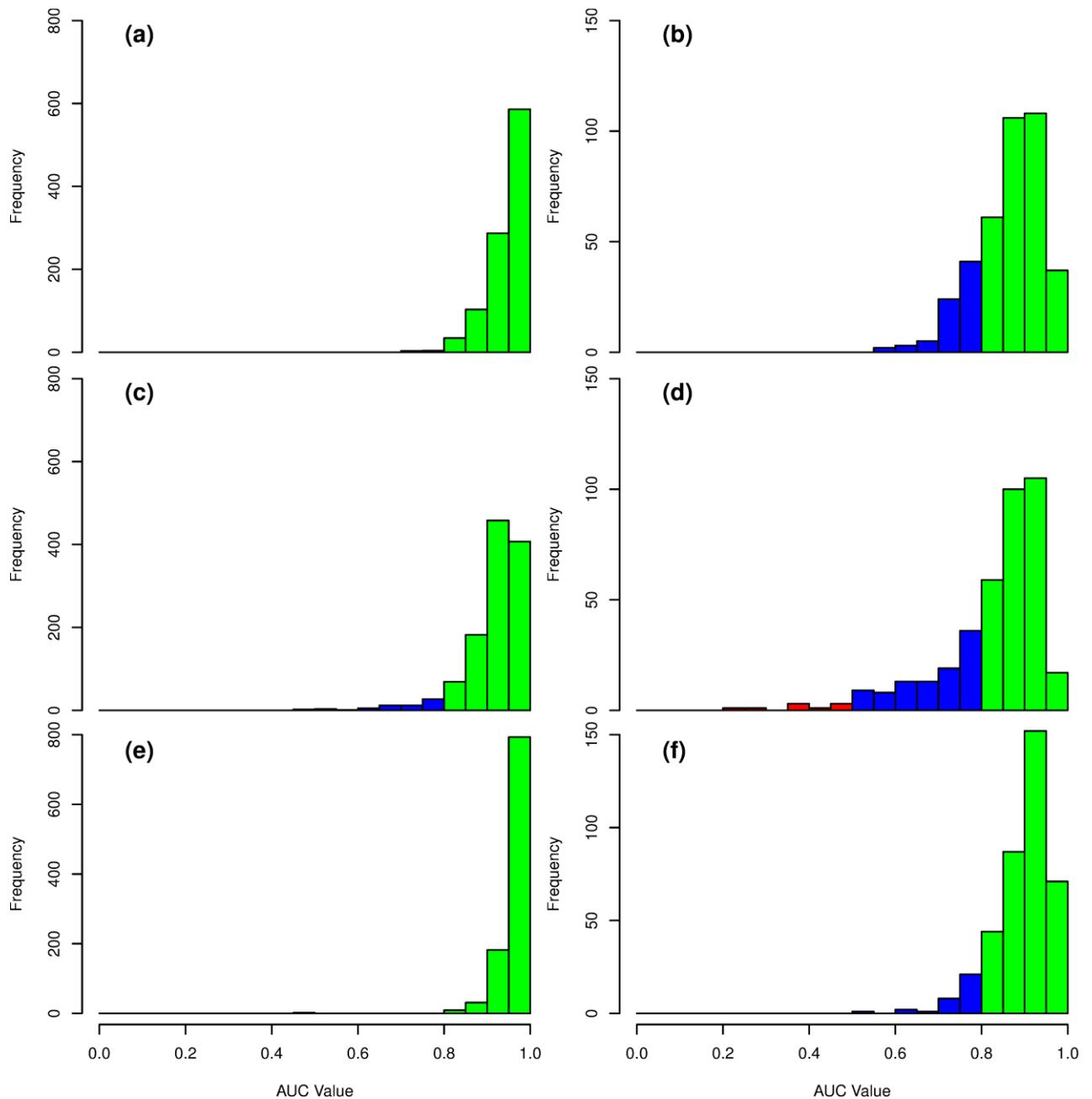
236 those in Preston et al., 2013 and those readily available on the BSBI and BRC websites, how
237 well the fitted models did in characterising the range and extent of species occurrence.

238 **3. Results**

239

240 Within-sample AUC values for each of the fitted models across the three methods built using
241 all data and only CS data respectively are shown as histograms in Figure 1. For the majority
242 of species the overall performance of all methods in fitting the observed data appears to be
243 very good with AUC values in excess of 0.8. AUC values for models built using CS data only
244 were generally slightly lower (Figure 1, Table 1), but still showed good performance with a
245 high percentage of AUC value in excess of 0.8 across all models (Table 1). The lower AUC
246 values resulting from the CS models is likely to be due to the smaller proportion of absences
247 in the CS dataset than in the NVC dataset.

248



249
 250 **Figure 1:** Histograms of AUC values for GLM (a, b), MARS (c, d) and GAM (e, f) models built using all survey data
 251 (a, c, e) and CS data only (b, d, f). AUC values > 0.8 coloured green; AUC values > 0.5 and < 0.8 coloured blue;
 252 AUC values < 0.5 coloured red.
 253

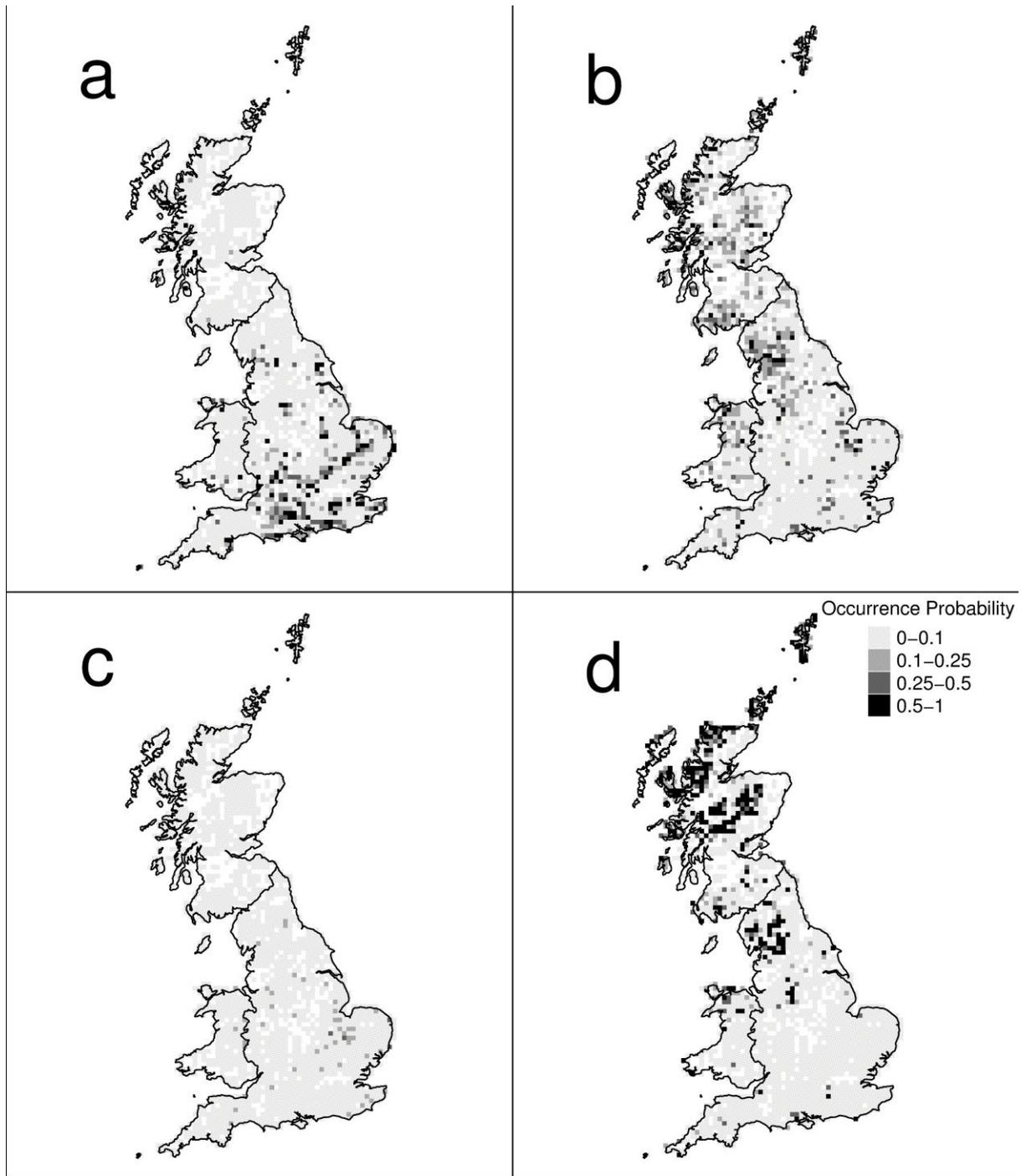
254 Differences between methods were relatively small for many species, but there were
 255 examples where the MARS and GAMS approaches clearly provided a substantially better fit
 256 to the data (at least in terms of AUC) than the GLM approach. Across the full set of species
 257 modelled using all data, only 1% had the highest AUC value for the GLM models compared
 258 with 17% for the MARS and 82% for the GAMs (Table 1). For the models built using CS data
 259 these percentages equate to 12%, 14 % and 73% respectively, though far fewer species
 260 models were fitted using the CS data only due to the availability of sufficient data records
 261 (Table 1).

262 **Table 1:** Number of species models fitted using each method and data source together with a summary of AUC
 263 statistics across each model type showing the proportion of cases where each model type was the “best”
 264 (highest AUC)
 265

	All Veg data			CS only		
	GLM	MARS	GAM	GLM	MARS	GAM
Number of Species Models	1017	1178	1017	387	388	387
% Maximum AUC	0.85	17.23	81.92	12.37	14.18	73.45
% AUC values > 0.8	99.31	94.74	99.80	80.62	72.42	91.47
% AUC values > 0.9	85.84	73.43	95.87	37.47	31.44	57.62

266
 267

268 Spatial predictions assessed against the corresponding maps published in Preston et al. 2013
 269 and the individual species distribution maps on the BSBI website, showed that the models
 270 broadly captured the spatial drivers dictating the distribution of individual species.
 271 Altitudinal, coastal and latitudinal preferences were clearly captured adequately by the
 272 niche models. Some examples of predicted distributions for four species are shown in Figure
 273 2. *Clematis vitalba* is shown to be a species of calcareous soils, with maximum habitat-
 274 suitability on the chalk hills of southern England. *Epilobium montanum* is a species with a
 275 wide range in terms of altitude, precipitation and soil. *Lemna trisulca* is shown to be a
 276 species of wet lowland sites, and *Selaginella selaginoides* is restricted to wet mountains.



277
 278
 279
 280
 281
 282
 283
 284

Figure 2: Probability distribution maps produced using the average fitted probability across the three modelling approaches based on predicting the full GB extent of input data (i.e. the environmental values from the training datasets) for 4 species: a) *Clematis vitalba*; b) *Epilobium montanum*; c) *Lemna trisulca*; d) *Selaginella selaginoides*. White squares represent areas missing from the training data set and hence no prediction could be made there.

285 AUC values obtained from the cross validation exercise showed that within-sample
 286 predictive performance was matched by out-of-sample performance (Table 2). In only two
 287 cases out of a total of 90 (3 models for each of 30 species) did the AUC value drop from
 288 above 0.8 using within-sample prediction to less than 0.8 using the cross-validation. These
 289 were both MARS models for *Senecio erucifolius* and *Teesdalia nudicaulis*. Given the range of
 290 species tested, both rare and abundant species, we conclude that our models provide a
 291 useful way of assessing the favourability of species to environmental conditions.

292

293 **Table 2:** Model fit diagnostics (in the form of AUC statistics) for all three modelling approaches and an estimate
 294 based on the average of the fitted values across these models from a subset of models for 30 test species.

295

Species Name	Within Sample AUC						Out of sample AUC		
	Models built using all survey data			Models built using CS data only			Cross Validation		
	MARS	GAM	GLM	MARS	GAM	GLM	MARS	GAM	GLM
<i>Agrostis capillaris</i>	0.881	NA	NA	0.858	0.818	0.707	0.872	0.852	0.749
<i>Campylopus flexuosus</i>	0.894	0.911	0.905	NA	NA	NA	0.891	0.906	0.904
<i>Campylopus introflexus</i>	0.867	0.965	0.948	NA	NA	NA	0.831	0.952	0.937
<i>Campylopus pyriformis</i>	0.881	0.918	0.903	NA	NA	NA	0.896	0.922	0.91
<i>Carex caryophylla</i>	0.949	NA	NA	0.915	0.948	0.951	0.95	0.963	0.943
<i>Carex limosa</i>	0.995	0.997	0.991	NA	NA	NA	0.924	0.98	0.99
<i>Carex viridula</i> subsp. <i>oedocarpa</i>	0.947	NA	NA	0.916	0.944	0.893	0.935	0.949	0.923
<i>Cochlearia pyrenaica</i>	0.811	0.989	0.978	NA	NA	NA	0.9	0.978	0.972
<i>Conopodium majus</i>	0.914	NA	NA	0.808	0.862	0.787	0.91	0.911	0.85
<i>Cynosurus cristatus</i>	0.899	NA	NA	0.87	0.882	0.865	0.897	0.932	0.894
<i>Dryas octopetala</i>	0.995	0.999	0.992	NA	NA	NA	0.969	0.99	0.991
<i>Helianthemum nummularium</i>	0.964	0.977	0.97	0.893	0.901	0.947	0.964	0.976	0.963
<i>Juncus articulatus</i>	0.928	0.955	0.911	0.893	0.773	0.734	0.912	0.936	0.898
<i>Lolium perenne</i>	0.951	NA	NA	0.942	0.947	0.91	0.939	0.944	0.932
<i>Neottia ovata</i>	0.852	0.937	0.871	NA	NA	NA	0.875	0.937	0.872
<i>Plantago major</i>	0.892	NA	NA	0.861	0.869	0.822	0.872	0.879	0.859
<i>Plantago media</i>	0.954	0.973	0.963	0.827	0.944	0.948	0.944	0.971	0.959
<i>Polytrichastrum alpinum</i>	0.96	0.979	0.974	NA	NA	NA	0.957	0.976	0.974
<i>Ranunculus repens</i>	0.889	NA	NA	0.847	0.879	0.801	0.873	0.899	0.851
<i>Sanguisorba officinalis</i>	0.933	0.974	0.901	NA	NA	NA	0.925	0.942	0.9
<i>Senecio erucifolius</i>	0.816	0.946	0.897	0.838	0.941	0.909	0.765	0.924	0.892
<i>Sphagnum cuspidatum</i>	0.941	0.976	0.972	NA	NA	NA	0.938	0.971	0.969
<i>Sphagnum denticulatum</i> s.l.	0.944	0.95	0.93	NA	NA	NA	0.833	0.917	0.926
<i>Sphagnum fimbriatum</i>	0.95	0.969	0.919	NA	NA	NA	0.919	0.956	0.914
<i>Sphagnum squarrosum</i>	0.967	0.934	0.901	NA	NA	NA	0.957	0.916	0.894
<i>Teesdalia nudicaulis</i>	0.968	0.993	0.982	NA	NA	NA	0.458	0.932	0.971
<i>Trifolium repens</i>	0.909	NA	NA	0.894	0.906	0.861	0.901	0.917	0.876
<i>Vaccinium myrtillus</i>	0.956	NA	NA	0.944	0.95	0.941	0.952	0.962	0.951
<i>Viola hirta</i>	0.922	NA	NA	0.377	0.936	0.951	0.925	0.955	0.937
<i>Viola riviniana</i>	0.893	0.925	0.875	NA	NA	NA	0.901	0.934	0.878

296

297 **4. Model exposition and example application**

298 Any user, applied scientist or policy maker may wish to explore each of the fitted models in
299 detail. For example, the models may be used to investigate relationships and make
300 predictions under different scenarios. However, there are too many individual models to
301 describe in sufficient detail. In order to provide full functionality and interrogation of all
302 models, they were packaged into a user friendly R library created specifically for this
303 purpose - MultiMOVE. As well as containing the actual model fits themselves, the
304 MultiMOVE package created by the authors consists of four key functions to access and
305 query the fitted niche models. These functions enable predictions to be made, covariate
306 relationship to be plotted, raw probabilities to be converted to a rescaled habitat suitability
307 score and the final model formula to be exposed. Here we describe this R package and
308 provide an example of its use in a scenario exploration for a single species.

309 As an example of using the fitted models and the MultiMOVE package to explore
310 relationships and answer scientific questions of interest relating to a species' niche, we
311 consider the question posed by Geddes and Miller (2012): will climate change counteract
312 the benefits of sheep-grazing in conserving a rare alpine dwarf-herb community? We focus
313 on *Festuca ovina agg.*, one of the key species in the study.

314

315 **4.1 Making predictions**

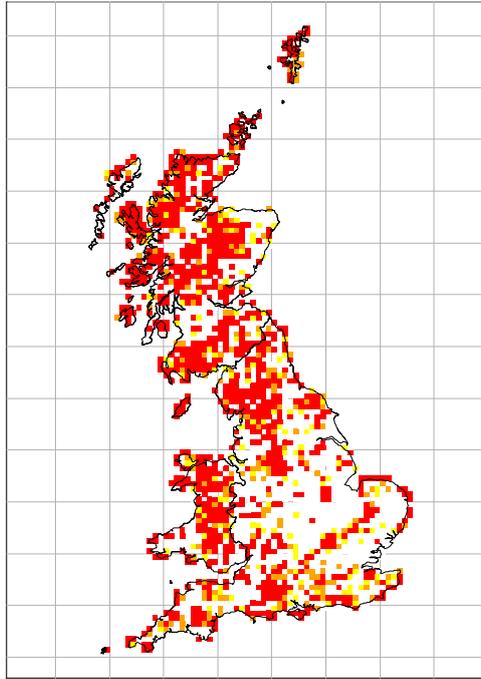
316

317 The prediction function allows the user to access the fitted ensemble of niche models and to
318 make predictions for a given species with specified environmental data located within the
319 modelled covariate space. The *MM_pred* function takes a data frame of the environmental
320 covariates and returns an aligned data frame with model predictions of species occurrence
321 probabilities. Upper and lower confidence limits for the predictions are also returned,
322 though as the models did not account for any residual spatial autocorrelation, extreme
323 caution is advised when using these.

324

325 Running the *MM_pred* function on *Festuca ovina agg.* (BRC number 920821) across the
326 whole of the training data set and mapping the estimated probabilities allows visualisation
327 of the predicted species range. This is useful not only for prediction, but also as a check to
328 see if the models have adequately captured the distributional range and preferences for a
329 given species before further investigation or use of the model. The predicted distribution
330 map for *Festuca ovina agg.* (Figure 3) shows good agreement with the online atlas map.

331



332
333

334 **Figure 3:** Predicted distribution map of *Festuca ovina agg.* using fitted probabilities from the *MM_pred*
335 command on the full training data set. Colour scale: Yellow = $0.1 < \text{probability occurrence} < 0.25$; Orange =
336 $0.25 < \text{probability occurrence} < 0.5$; Red = $0.5 < \text{probability occurrence} < 1$. White areas represent either returned
337 probabilities of 0 or areas entirely absent from the training data.

338

339 4.2 Visualising covariate effects

340

341 The *MM_plot* command feature enables the user to observe the marginal or joint effects of
342 covariates on the species response. When viewing the marginal or joint effects of specified
343 covariates all other covariates are held at their median values. Either one or two covariates
344 can be specified at a time to visualise their effect on species' occurrence probabilities.

345

346 Geddes and Miller, 2012 were specifically interested in the effects of grazing and climate. In
347 the MultiMOVE models these variables are represented by cover-weighted canopy height
348 and maximum July temperature respectively. Therefore, to answer the question as to
349 whether climate change would counteract any advantages grazing would have, we can use
350 the MultiMOVE package to visualise the joint effects of canopy height and July temperature
351 and draw our inference from that. Running the MultiMOVE commands in R as follows,

352

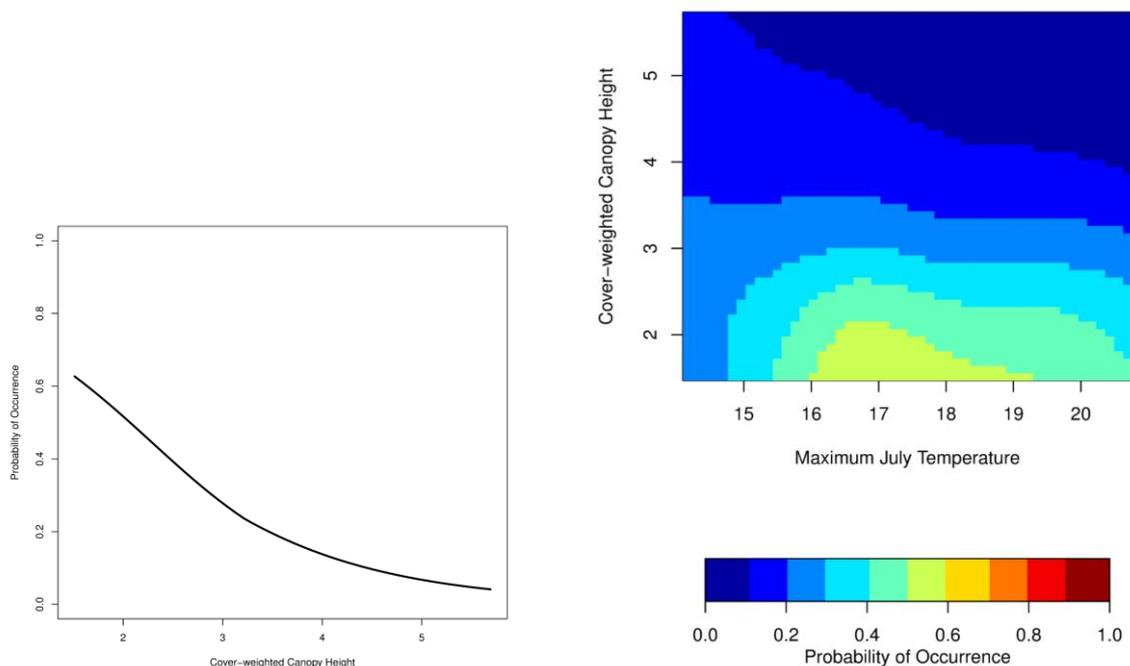
```
353 > MM_plot(input_data, BRC=920821, view_term=c("cov4"), display = "raw"); and  
354 > MM_plot(input_data, BRC=920821, view_term=c("mju", "cov4"), display = "raw"),
```

355

356 returns plots of the fitted marginal effects of canopy height (our proxy for grazing) and a plot
357 of the fitted joint effects between canopy height and July temperature for *Festuca ovina*
358 *agg.*

359

360
361



362
363
364
365
366
367

Figure 4: Output from the `MM_plot` command run on *Festuca ovina agg.* The figure on the left shows the marginal effect of cover weighted canopy height on the occurrence of *Festuca ovina agg.* with each of the model predictions plotted. The plot on the right shows the joint effect of canopy height and maximum July temperature plotted using the average fitted values across the models.

368 Fitted models within MultiMOVE show that *Festuca ovina agg.* favours lower canopy heights
369 and hence that grazing has positive benefits for this species (Figure 4). The interaction
370 between canopy height and temperature indicates that when canopy height is low, if
371 maximum July temperature rises to above 19°C, the likelihood of occurrence decreases. Our
372 conclusion therefore would be that there is evidence to suggest that the potential effects of
373 climate change (increasing July temperature) could potentially counteract the benefits of
374 grazing for *Festuca ovina agg.* This is in keeping with the results from the Geddes and Miller,
375 2012 study.

376

377 **4.3 Converting raw probabilities into habitat suitability scores**

378

379 MultiMOVE also contains a function (`HS_convert`) to transform the fitted probabilities, the
380 values of which are dependent on the tolerance or niche breadth of the species and its
381 prevalence, into a habitat suitability score on a standardised scale which is then directly
382 comparable between species. There are two options for doing this within the function:
383 based on the cumulative distribution of fitted probabilities to the training data; and based
384 on the prevalence of the species within the training data.

385

386 In the first case, the cumulative distribution of probabilities fitted to the training data set is
387 obtained for all model types across all species. Rescaled probabilities are then estimated to

388 be the percentile of the cumulative distribution corresponding to the un-scaled raw
389 occurrence probability. The rescaled “habitat suitability” scores therefore provide an
390 indication of how extreme the raw occurrence probability is across the full sample in the
391 training data i.e. does the un-scaled occurrence probability correspond the edges of a
392 species’ range.

393

394 Using the species prevalence to re-scale the fitted occurrence probabilities follows the
395 approach and formula suggested by Real et al., 2006. The rescaling in this case ensures that
396 a habitat suitability score of 0.5 corresponds to what would be obtained by taking an
397 average of the prevalence across the whole training data set and hence corresponds to
398 equal favourability everywhere.

399

400 **4.4 Obtaining the fitted model**

401

402 The final function contained within the MultiMOVE package, *extract_MM_model*, enables
403 the user to extract the fitted model for a given species, in particular to see which covariates
404 were included in the final model after model selection.

405

406 **5. Discussion**

407

408 Species niche models encapsulate the relationships that drive the spatial distribution of
409 individual species and have many potential uses in assessing the effects of environmental
410 change on habitat suitability for individual species, the consequences this has for
411 biodiversity value and the provision of ecosystem services. However, despite considerable
412 efforts in constructing such distribution models, the application of such models in
413 developing policy and management recommendations has been patchy (Guisan et al., 2013).
414 This can often be due to the complex nature and time involved in re-running models under
415 different scenarios. The large number of models developed and user friendly R package
416 described here make it highly relevant for assessing the likely impact of management and
417 policy decisions and for facilitating efficient and responsive outcomes for conservation
418 management.

419

420 The assessment of the model fits showed good performance based on within sample AUC
421 across all species and out of sample AUC for a subset of test species. AUC statistics
422 suggested that use of an ensemble approach was beneficial as no one model type was
423 optimal across all species. Greater model flexibility requires more data to estimate
424 accurately, so there can be cases with limited data where the more free-form models
425 underperform while a linear (or log-linear) model can adequately capture a pattern of
426 responses (as shown in Smart et al., 2010a). This suggests that there are benefits in using
427 multiple modelling techniques (Araújo & New, 2007).

428 It is important to note that potential residual spatial autocorrelation was not accounted for
429 within any of our models. This should not introduce systematic bias into our parameter
430 estimates (and, therefore, predictions), but it is likely to mean that we substantially
431 underestimate the uncertainty associated with the outputs from our models. This can lead
432 to overfitting, especially with the more complex models, compounding the need for caution.
433 On our small subset of species, we investigated any potential evidence of this by using the
434 pairwise distance sampling technique of Hijmans (2012) to reduce the impact of spatial
435 effects on cross-validated AUC measures. The results suggest that our methods still perform
436 quite well despite the above limitation. It is possible that the methods considered here could
437 be extended to explicitly account for residual spatial autocorrelation and recent software
438 developments (such as the R-INLA package; <http://www.r-inla.org/>) mean that it is now
439 computationally feasible to apply such extensions to relatively large datasets. However, the
440 large number of species across which each model is run means that the inclusion of these
441 additional models will inevitably require a substantial amount of computational effort. Due
442 to the complex nature of spatial statistical models and the large number of sites and species
443 in our data set, fitting and optimising simpler models, as done here, is a sensible first step.
444

445 The MultiMOVE package exploits the extensive datasets describing the occurrence of a large
446 number of vascular plant and bryophyte species across GB to build a picture of how
447 individual species favour different environments. Some care should be taken in interpreting
448 the outputs, since the fitted models will reflect any bias in the training data. A key advantage
449 of the current work has been the use of standardized quadrat data with which to train the
450 models, primarily from the Countryside Survey and the NVC, rather than opportunistic
451 presence-only data. This should greatly reduce the influence of gross spatial heterogeneities
452 in recording effort that often arise in opportunistic datasets. As always, the development of
453 niche models depends heavily on reliable data on which to train the models. For niche
454 models and their outputs to be representative of conditions and niche preferences across
455 Britain, the training data itself should be unbiased and representative of this same
456 population. As such, the fitted models do not apply to environments not included in the
457 training data - those outside of GB for example - and MultiMOVE accounts for this by
458 warning when predictions are attempted using input data outside the covariate space
459 observed in the training data.
460

461 The MultiMOVE package allows exploration of the impact of environmental change on plant
462 species and assemblages across terrestrial ecosystems in Britain, by making use of the large
463 datasets that are available and an ensemble approach to modelling niches. Site managers
464 can quickly assess current suitability for individual species, using freely available climate data
465 (<http://www.metoffice.gov.uk/climatechange/science/monitoring/ukcp09/>) and a list of
466 currently-present species. This makes use of the fact that the species that are present give a
467 considerable amount of information about site conditions. The package also allows the
468 effects of environmental change to be expressed in terms of likely changes in species

469 composition and derived biodiversity indicators (Rowe et al., 2014a). Observed
470 environmental changes can be interpreted in terms of how they are likely to have affected
471 habitat suitability for species. Cautious ecological judgement is needed. The models are built
472 on spatial patterns. When used to project change through time the assumption of space for
473 time substitution is strong and likely to be highly questionable especially when confronting
474 past patterns with potentially novel environments (Williams & Jackson 2007). In addition the
475 uncertainty around the relationship between abiotic conditions and the Ellenberg values,
476 whose means are used to convey these conditions, is propagated through the application of
477 MultiMOVE. The impact of this uncertainty is likely to be especially critical when scenario
478 testing since any lagged changes are not simulated. Changes in habitat suitability can
479 however be driven by the outputs of dynamic models of climate or biogeochemical
480 processes (Rowe et al. 2011b). The opportunity to explore the ecological responses of the
481 UK flora that MultiMOVE provides gives it a large number of potential applications.

482

483 **Access**

484

485 MultiMOVE was built using R 3.1.2 and depends on the following R packages: mgcv, fields,
486 leaps, earth, stringr, gsubfn, randomForest and nnet. In some circumstances, when not
487 automatically installed by MultiMOVE itself, these may need to be installed prior to installing
488 MultiMOVE. The package binary - MultiMOVE_2.0.1.zip – is available via eidc.ceh.ac.uk (doi:
489 <http://doi.org/10.5285/94ae1a5a-2a28-4315-8d4b-35ae964fc3b9>).

490

491 **Supplementary Information**

492

493 A manual for the latest version of the package (v2.0.1) is available in the supplementary
494 material. Please note that this latest version of the MultiMOVE package contains an
495 additional two methods to describe a species' niche: Random Forests; and Neural Networks.

496

497 **Acknowledgements**

498

499 The study was funded by the UK Department for Environment, Food and Rural Affairs (Defra)
500 through the UKREATE programme (<http://ukreate.defra.gov.uk/>). We also thank four
501 anonymous reviewers for their comments and suggestions that resulted in a significantly
502 improved manuscript.

503

504 **References**

505

506 Alexander, S., Ehrlich, P. R., Goulder, L., Lubchenco, J., Matson, P. A., Mooney, H. A., Postel, S.,
507 Schneider, S. H., Tilman, D. & Woodwell, G. M. 1997. *Ecosystem services: benefits supplied to*
508 *human societies by natural ecosystems (Vol. 2)*. Washington (DC): Ecological Society of
509 America.

- 510 Araújo, M. B., & New, M. 2007. Ensemble forecasting of species distributions. *Trends in*
511 *ecology & evolution*, 22(1): 42-47.
- 512 Blockeel, T. L., Bosanquet, S. D. S., Hill, M. O. and Preston, C. D. 2014. *Atlas of British & Irish*
513 *Bryophytes*. Pisces Publications, Newbury.
- 514 Corney, P. M., Le Duc, M. G., Smart, S. M., Kirby, K. J., Bunce, R. G. H. and Marrs, R. H. 2006.
515 Relationships between the species composition of forest field-layer vegetation and
516 environmental drivers assessed using a national survey. *Journal of Ecology*, 94: 383–401.
- 517 Costanza, R., Fisher, B., Mulder, K., Liu, S., & Christopher, T. 2007. Biodiversity and ecosystem
518 services: A multi-scale empirical study of the relationship between species richness and net
519 primary production. *Ecological economics*, 61(2): 478-491.
- 520 Crone, E.E., Menges, E.S., Ellis, M.M., Bell, T., Bierzychudek, P., Ehrlén, J., Kaye, T.N., Knight,
521 T.M., Lesica, P., Morris, W.F., Oostermeijer, G., Quintana-Ascencio, P.F., Stanley, A., Ticktin, T.,
522 Vlaverde, T. & Williams, J.L. 2011. How do plant ecologists use matrix population models?
523 *Ecology Letters*, 14: 1-8.
- 524 de Vries, W., Wamelink, W., van Dobben, H., Kros, H., Reinds, G.-J., Mol-Dijkstra, J., Smart, S.,
525 Evans, C., Rowe, E., Belyazid, S., Sverdrup, H., van Hinsberg, A., Posch, M., Hettelingh, J.-P.,
526 Spranger, T., Bobbink, R., 2010. Use of dynamic soil-vegetation models to assess impacts of
527 nitrogen deposition on plant species composition and to estimate critical loads: an overview.
528 *Ecological Applications*, 20: 60-79.
- 529 Diekmann, M. (2003). Species indicator values as an important tool in applied plant ecology -
530 a review. *Basic and Applied Ecology*, 4: 493-506.
- 531 Dormann, C. F., Schymanski, S. J., Cabral, J., Chuine, I., Graham, C., Hartig, F., Kearney, M.,
532 Morin, X., Römermann, C., Schröder, B. & Singer, A. (2012). Correlation and process in
533 species distribution models: bridging a dichotomy. *Journal of Biogeography*, 39(12): 2119-
534 2131.
- 535 Elith, J., Kearney, M., Phillips, S. 2010. The art of modelling range-shifting species. *Methods in*
536 *Ecology and Evolution* 1, 330-342.
- 537 Elith, J., Leathwick, J. 2009. Species distribution models: Ecological explanation and
538 prediction across space and time. *Annual Review of Ecology, Evolution and Systematics*, 40:
539 677-697
- 540 Ellenberg, H., Weber, H.E., Düll, R., Wirth, V., Werner, W. & Paulissen, D. 1991. Zeigerwerte
541 von Pflanzen in Mitteleuropa. *Scripta Geobotanica*, 18: 1-248.
- 542 Fielding, A. H., & Bell, J. F. 1997. A review of methods for the assessment of prediction errors
543 in conservation presence/absence models. *Environmental conservation*, 24(01): 38-49.

- 544 Friedman, J. H. 1991. Multivariate Adaptive Regression Splines (with discussion), *Annals of*
545 *Statistics*, 19: 1-141.
- 546 Geddes, C., & Miller, G. R. 2012. Will climate change counteract the benefits of sheep-
547 grazing in conserving a rare alpine dwarf-herb community? *New Journal of Botany*, 2(2): 92-
548 99.
- 549 Gogol-Prokurat, M. 2011. Predicting habitat suitability for rare plants at local spatial scales
550 using a species distribution model. *Ecological Applications*, 21(1): 33-47.
- 551 Grime, J. P., Hodgson, J. G., Hunt, R. 1988. *Comparative plant ecology : a functional*
552 *approach to common British species*. London: Unwin Hyman. 752 pp.
- 553 Grime, J.P. 1998 Benefits of plant diversity to ecosystems: immediate, filter and founder
554 effects. *Journal of Ecology*, 86: 902-910.
- 555 Groom, Q. J. 2013. Estimation of vascular plant occupancy and its change using kriging. *New*
556 *Journal of Botany*, 3(1): 33-46.
- 557 Guisan, A., Tingley, R., Baumgartner, J. B., Naujokaitis-Lewis, I., Sutcliffe, P. R., Tulloch, A. I.,
558 Regan, T. J., Brotons, L., McDonald-Madden, E., Mantyka-Pringle, C., Martin, T. G., Rhodes, J.
559 R., Maggini, R., Setterfield, S. A., Elith, J., Schwartz, M. W., Wintle, B. A., Broennimann, O.,
560 Austin, M., Ferrier, S., Kearney, M. R., Possingham, H. P. & Buckley, Y. M. 2013. Predicting
561 species distributions for conservation decisions. *Ecology Letters*, 16(12), 1424-1435.
- 562 Hastie, T. J. and Tibshirani, R. J. 1990. *Generalized Additive Models*. Chapman & Hall/CRC.
- 563 Hijmans, R. J. 2012. Cross-validation of species distribution models: removing spatial sorting
564 bias and calibration with a null model. *Ecology*, 93(3): 679-688.
- 565 Hill, M. O., Mountford, J. O., Roy, D. B. and Bunce, R. G. H. 1999. *Ellenberg's indicator values*
566 *for British plants*. Ecofact Volume 2, Technical Annex, 46 pp.
- 567 Hill, M.O., Preston, C.D., Roy, D.B. 2004. *PLANTATT - attributes of British and Irish plants:*
568 *status, size, life history, geography and habitats*. Abbots Ripton, Centre for Ecology &
569 Hydrology, 73pp.
- 570 Hill, M. O., Preston, C. D., Bosanquet, S. D. S., Roy, D. B. 2007. *BRYOATT: attributes of British*
571 *and Irish mosses, liverworts and hornworts*. Cambridge, Centre for Ecology and Hydrology.
- 572 Hirzel, A. H., & Le Lay, G. 2008. Habitat suitability modelling and niche theory. *Journal of*
573 *Applied Ecology*, 45(5): 1372-1381.
- 574 Hornung, M. H. 1996. *The key habitats survey of Britain*. Final report to DEFRA. Centre for
575 Ecology and Hydrology, Lancaster, UK.
- 576 Kirby, K. J., Smart, S. M., Black, H. I. J., Bunce, R. G. H., Corney, P. M. and Smithers, R. J. 2005.

577 *Long-term ecological change in British woodland (1971–2001)*. English Nature,
578 Peterborough, UK. 137pp.

579 Kremen, C., Williams, N. M., Aizen, M. A., Gemmill-Herren, B., LeBuhn, G., Minckley, R.,
580 Packer, L., Potts, S. G., Roulston, T., Steffan-Dewenter, I., Vázquez, D. P., Winfree, R., Adams,
581 L., Crone, E. E., Greenleaf, S. S., Keitt, T. H., Klein, A.-M., Regetz, J. and Ricketts, T. H. 2007.
582 Pollination and other ecosystem services produced by mobile organisms: a conceptual
583 framework for the effects of land-use change. *Ecology Letters*, 10: 299–314.

584 Latour, J. B., Reiling, R. 1993. *MOVE: a multiple-stress model for vegetation*. The Science of
585 the Total Environment Supplement: 1513-1526.

586 Laughlin, D.C. 2011. Nitrification is linked to dominant leaf traits rather than functional
587 diversity. *Journal of Ecology*, 99: 1091-1099.

588 Leathwick, J. R., Elith, J., Hastie, T. 2006. Comparative performance of generalized additive
589 models and multivariate adaptive regression splines for statistical modelling of species
590 distributions. *Ecological Modelling*, 199: 188-196.

591 Lumley, T. Using Fortran code by Alan Miller. 2009. *leaps: regression subset selection*. R
592 package version 2.9. <http://CRAN.R-project.org/package=leaps>.

593 McCullagh, P., Nelder, J. A., & McCullagh, P. 1989. *Generalized linear models (Vol. 2)*. London:
594 Chapman and Hall.

595 Milborrow, S. Derived from mda:mars by Trevor Hastie and Rob Tibshirani. Uses Alan Miller's
596 Fortran utilities with Thomas Lumley's leaps wrapper. 2014. *earth: Multivariate Adaptive*
597 *Regression Spline Models*. R package version 3.2-7. <http://CRAN.R-project.org/package=earth>.

598 Munoz, J. and Felicísimo, A. M. 2004. Comparison of statistical methods commonly used in
599 predictive modelling. *Journal of Vegetation Science*, 15: 285-292.

600 Preston, C. D., Hill, M. O., Harrower, C. A., & Dines, T. D. 2013. Biogeographical patterns in
601 the British and Irish flora. *New Journal of Botany*, 3(2): 96-117.

602 R Development Core Team 2009. *R: A Language and Environment for Statistical Computing*.
603 R Foundation for Statistical Computing, Vienna. URL <http://www.R-project.org>

604 Real, R., Barbosa, A.M., Vargas, J.M. 2006. Obtaining environmental favourability functions
605 from logistic regression. *Environmental and Ecological Statistics*, 13: 237-245.

606 Rodwell, J. S. 1991. *Woodland & scrub. British plant communities*. Vol 1, Cambridge
607 University Press, Cambridge, UK.

608 Rodwell, J. S., Dring, J. C., Averis, A. B. G., Proctor, M. C. F., Malloch, A. J. C., Schaminée, J.N.J.
609 & Dargie, T.C.D. 2000. *Review of coverage of the National Vegetation Classification*. Joint
610 Nature Conservation Committee Report No. 302. <http://www.incc.gov.uk/page-2312>

611 Rowe, E. C., Emmett, B. A., Smart, S. M., Frogbrook, Z. L. 2011a. A new net mineralizable
612 nitrogen assay improves predictions of floristic composition. *Journal of Vegetation Science*,
613 22: 251-261

614 Rowe, E.C., Jones, M.L.M., Henrys, P.A., Smart, S.M., Tipping, E., Mills, R.T.E., Evans, C.D.
615 2011b. Predicting effects of N pollutant load on plant species based on a dynamic soil
616 eutrophication indicator. *Countryside Council for Wales*. Science Report 977, 39 pp.

617 Rowe, E.C., Ford-Thompson, A., Monteith, D., van Hinsberg, A., Smart, S.M., Henrys, P.A.,
618 Ashmore, M. 2014 A biodiversity metric for interpreting outputs of models of atmospheric
619 nitrogen pollution impacts on habitats. *Final report on Defra project AQ0828 / CEH project*
620 *NEC04988*. Centre for Ecology and Hydrology, Bangor, LL57 2UW, UK. 94 pp.

621 Rowe, E.C., Smart, S.M., Emmett, B.A., 2014. Phosphorus availability explains patterns in a
622 productivity indicator in temperate semi-natural vegetation. *Environmental Science-*
623 *Processes & Impacts* 16: 2156-2164.

624 Smart, S.M. 2000. Ecological assessment of vegetation from a nature reserve using regional
625 reference data and indicator scores. *Biodiversity & Conservation*, 9: 811-832.

626 Smart, S. M., Henrys, P. A., Scott, W. A., Hall, J. R., Evans, C. D., Crowe, A., Rowe, E. C.,
627 Dragosits, U., Page, T., Whyatt, J. D., Sowerby, A., Clark, J. M. 2010a. Impacts of pollution and
628 climate change on ombrotrophic Sphagnum species in the UK: analysis of uncertainties in
629 two empirical niche models. *Climate Research*, 45: 163-177.

630 Smart, S. M., Scott, W. A., Whitaker, J., Hill, M. O., Roy, D. B., Critchley, C. N., Marini, L., Evans,
631 C. D., Emmett, B. A., Rowe, E. C., Crowe, A., Le Duc M., Marrs, R. H. 2010b. Empirical realised
632 niche models for British higher and lower plants - development and preliminary testing.
633 *Journal of Vegetation Science*, 21: 643-656.

634 Smith, M.D., Knapp, A.K. 2003. Dominant species maintain ecosystem function with non-
635 random species loss. *Ecology Letters*, 6: 509-517.

636 Stace, C. 2010. *New flora of the British Isles*. Cambridge University Press.

637 Svenning, J.-C., Skov, F. 2004. Limited filling of the potential range in European tree species.
638 *Ecology Letters*, 7: 565-573.

639 Thuiller, W. 2003. BIOMOD—optimizing predictions of species distributions and projecting
640 potential future shifts under global change. *Global change biology*, 9(10): 1353-1362.

641 Thuiller, W., Lavorel, S., Araújo, M. B., Sykes, M. T., & Prentice, I. C. 2005. Climate change

- 642 threats to plant diversity in Europe. *Proceedings of the National Academy of Sciences of the*
643 *United States of America*, 102(23): 8245-8250.
- 644 Thuiller W., Albert, C., Araujo, M.B., Berry, P.M., Cabeza, M., Guisan, A., Hickler, T., Midgley,
645 G.F., Paterson, J., Schurr, F.M., Sykes, M.T. & Zimmermann, N.E. 2008. Predicting global
646 change impacts on plant species' distributions: Future challenges. *Perspectives in Plant*
647 *Ecology, Evolution and Systematics*, 9: 137-152.
- 648 Thuiller, W., Lafourcade, B., Engler, R., & Araújo, M. B. 2009. BIOMOD—a platform for
649 ensemble forecasting of species distributions. *Ecography*, 32(3): 369-373.
- 650 UK National Ecosystem Assessment. 2011. *The UK National Ecosystem Assessment: Synthesis*
651 *of the Key Findings*. UNEP-WCMC, Cambridge.
- 652 Wenger, S. J., Olden, J. D. 2012. Assessing transferability of ecological models: an
653 underappreciated aspect of statistical validation. *Methods in Ecology and Evolution*, 3: 260-
654 267.
- 655 Williams, J.W. & Jackson, S.T. 2007. Novel climates, no-analog communities and ecological
656 surprises. *Frontiers in Ecology and the Environment*, 5: 475-482.
- 657 Wood, S. N. 2006. *Generalized Additive Models: An Introduction with R*. CRC/Chapman &
658 Hall.