



Adding small species to the big picture: Species distribution modelling in an age of landscape scale conservation



Sally Eaton^{a,*}, Christopher Ellis^b, David Genney^c, Richard Thompson^d, Rebecca Yahr^b, Daniel T. Haydon^e

^a Royal Botanic Garden Edinburgh, 20A Inverleith Row, Edinburgh EH3 5LR, UK

^b Royal Botanic Garden Edinburgh, UK

^c Scottish Natural Heritage, Leachkin Road, Inverness IV3 8NW, UK

^d Forest Enterprise Scotland, Glentress Peel, Peebles EH45 8NB, UK

^e Institute of Biodiversity Animal Health and Comparative Medicine, University of Glasgow, G12 8QQ, UK

ARTICLE INFO

Keywords:

Compound variables
Generalised Additive Model (GAM)
Lichen epiphytes
Native Woodland Survey Scotland (NWSS)
Temperate rainforests

ABSTRACT

A recent shift in conservation policy from the site scale to the ecosystem or landscape scale requires underpinning by large-scale species distribution data. This poses a significant challenge in conserving small/less charismatic species (SLCS's) whose often cryptic nature can result in spatially restricted sampling, thus preventing landscape scale conservation projects from being realised for these ecologically important groups.

Species distribution models (SDMs) can provide a powerful tool to bridge this gap. However, in the case of SLCS's (here lichen epiphytes in temperate rainforests of western Scotland are used as a model system), direct predictor variables exist at micro-scales (millimetres to centimetres), which are not extensively available in landscape-scale datasets. Here we identify a group of well-mapped larger-scale 'compound variables' which capture the effect of multiple direct predictors (such as bark pH and topography), and test whether they can be successfully used to predict species distributions at the landscape scale, circumventing the need for direct (micro-scale) predictor data.

By testing the SDMs more widely within western Scotland, accurate predictions of species presence/absence could be made throughout the region for 5 of the 9 lichen epiphytes, making these SDMs extremely valuable as a conservation planning tool.

Species distribution models utilising compound variables as predictors offer a solution to the paucity of species distributional data for SLCS's, and present a valuable resource in conservation planning for such species. The importance of testing the SDMs outside of a training region to prevent prediction error is highlighted however.

1. Introduction

Charismatic megafauna/flora make up just 2.8% of all life on earth (loosely all vascular plants and vertebrates, Jefferies, 2006) yet command the highest share of media exposure, research attention and conservation action (Clark and May, 2002). Because the remaining 97.2% of species are small and/or considered less charismatic (comprising for example nematodes, mollusks, arthropods, fungi and other microorganisms), they receive far less attention from the public, researchers, and policy makers (Jefferies, 2006). Nevertheless, many of these overlooked species are understood to be key providers of

regulating and supporting ecosystem services (MEA, 2005). Obtaining a good understanding of the distribution of ecologically important small/less charismatic species is therefore essential for continued ecosystem service provision (Liu et al., 2013; Polce et al., 2013) and robust conservation planning (Whittaker et al., 2005).

Existing datasets for small/less charismatic species fall into two contrasting categories, being either (i) small scale and highly detailed (Rondinini et al., 2006), often concentrated on protected areas or monitoring plots, or (ii) extensive but coarse grained (Hartley et al., 2004) including standard 10 km mapping schemes (e.g. Seaward, 1995). This leads to a trade-off between spatially restricted (high

Abbreviations: AUC, area under the receiver operating curve; CT, Classification Tree; GAM, Generalised Additive Model; NPMR, nonparametric multiplicative regression; NWSS, Native Woodland Survey Scotland; SLCS's, small/less charismatic species; SDM, species distribution model; TSS, true skill statistic

* Corresponding author.

E-mail addresses: seaton@rbge.ac.uk (S. Eaton), c.ellis@rbge.ac.uk (C. Ellis), David.Genney@snh.gov.uk (D. Genney), Richard.Thompson@forestry.gsi.gov.uk (R. Thompson), R.Yahr@rbge.ac.uk (R. Yahr), Daniel.Haydon@glasgow.ac.uk (D.T. Haydon).

<https://doi.org/10.1016/j.biocon.2017.11.012>

Received 22 June 2017; Received in revised form 30 October 2017; Accepted 10 November 2017

0006-3207/© 2017 Elsevier Ltd. All rights reserved.

resolution) accuracy and spatially extensive (low resolution) generality (McPherson et al., 2006), and thus fails to reflect a new conservation paradigm focused at the landscape-scale. Over the last decade, the species extinction crisis has led to a renewed focus on conserving entire ecosystems at landscape-scales, rather than individual site or single species strategies (Watson et al., 2011). This new approach has demanded species distribution data mapped over correspondingly large geographic areas, though at a resolution that reflects local variability in a species' habitat. This is most challenging for small/less charismatic species which require specialist taxonomic skills, or costly molecular diagnostics for identification. The surveying methods required to gather high resolution data for these species are therefore too limiting to be applied over entire landscapes (Britton et al., 2013; Vanderpoorten et al., 2005). As a result, ecologists are faced with a major challenge in generating species distribution data at a high resolution (at the granularity of the species' habitat) that is extensive (mapped at the landscape scale).

This study draws on species distribution modelling (SDM) as a widely-used technique to derive spatially explicit projections of habitat suitability (Guisan et al., 2013). This approach has made valuable contributions to conservation management and planning (Guisan and Zimmermann, 2000; Nicholls, 1989; Vaughan and Ormerod, 2003) across a wide range of different contexts (see review in Peterson, 2006 and Rodríguez et al., 2007). In the case of small/less charismatic species in particular, SDM provides a potentially cost-effective approach to mapping habitat suitability across large spatial areas (Nicholls, 1989).

The study focused on epiphytic lichens, a group of small cryptic species that lack comprehensive high-resolution distributional data at a landscape scale due to the specialist skills and intensive survey methods required to map them. Previous studies have successfully applied SDMs to project the large-scale epiphyte response to effects such as climate change, at a regional 10 km grid-scale (Ellis et al., 2007), but these approaches have not been applied continuously across a landscape at the resolution of the species' habitat (for epiphytic lichens these habitats are individual trees). Furthermore, the use of SDM in predicting the high-resolution occurrence of these species may be severely limited by the micro-scale over which environmental predictors operate, and which would demand environmental layers at centimetre-scales. For example, micro-niche factors associated with bark e.g. pH (Gauslaa, 1995; Jüriado et al., 2009; Kuusinen, 1996; Lewis and Ellis, 2010) and bark topography/texture (Bates, 1992; Ranius et al., 2008; Fritz et al., 2009) are known to affect the diversity and distribution of lichen epiphytes on individual trees (Ellis et al., 2015). These micro-niche factors will interact with local micro-climatic effects such as humidity/vapour pressure deficit (Hosokawa and Odani, 1957; Rambo, 2010) and light levels (Gustafsson and Eriksson, 1995; Kuusinen, 1994; Uliczka and Angelstam, 1999), to further influence epiphytic lichen distribution giving rise to environmental complexity within individual tree boles. Such fine-scale micro-niche and micro-climatic factors are impossible to map in a resource efficient way that would enable projections of species distribution across entire landscapes. To address this problem, and ultimately to enable the use of SDM to predict the distribution of small/less charismatic species, we test a series of compound variables for their ability to accurately predict species distributions. For example, tree species and size are widely available as digitised layers, and have the potential to capture differences among micro-niche effects such as pH and bark topography. Additionally, the distance to a water source and canopy cover can be mapped remotely, and may be used to represent micro-climatic factors such as humidity and light levels.

The method is tested for a range of lichen epiphyte species (exhibiting a variety of niche specialisms and reproductive modes), in order to understand whether there is variability in the applicability of the method to species exhibiting different ecological traits.

The success of SDMs can be measured by their ability to accurately predict species presence/absence within a region in which the model was fitted, and beyond this 'training region' into spatially removed but

analogous habitat space. Considering the difficulties involved in gathering distributional data for small/less charismatic species, the extent to which a model may be used predictively outside the training region is critical. This extension to the use of a model is referred to as 'generality' (Fielding and Haworth, 1995), or 'transferability' (Kleyeer, 2002), and can be highly variable (Randin et al., 2006) depending on, for example, landscape scale population processes such as mass effects (Pulliam, 2000). Testing of model generality/transferability is therefore required to ensure that SDMs are not applied to environments in which predictive error is unacceptable, but conversely, practical application could be unnecessarily limited by avoiding the wider application of effective SDMs.

This study:

1. Tested the ability of SDMs to accurately assign small/less charismatic species to their suitable micro-environments using compound variables as predictors of species presence/absence. It is based on a training data set of 600 sub-sampled tree stems (representing an approximate basal trunk cross-sectional area of 54 m²);
2. Applies the sub-sampled SDMs to a spatially extensive area relevant to conservation planning (a wooded landscape of 15 km in length) using a widely available digital dataset containing the compound variables;
3. Uses independent data to test the extent to which species distributions may be predicted beyond the wooded landscape, to characterize an entire region (representing a distance of over 200 km), through a comparison with ten widely sampled but comparable sites.

2. Methods

2.1. Study system

The species studied here are oceanic epiphytic lichens occurring in temperate rainforests in Scotland.

A total of nine epiphytic lichen species were selected for field survey according to three criteria: (i) their ability to be identified under field conditions using gross-morphology, (ii) their contrasting niche specialism (niche specialism was derived from local expert opinion), and (iii) their contrasting dispersal modes (derived from Purvis, 1992). Based on prior expectation, two niche generalists were chosen, frequently found in temperate rainforests throughout Scotland (the sexually reproducing *Graphis scripta* sensu lato, (L.) Ach., and the asexually reproducing *Parmelia saxatilis* sensu lato, (L.) Ach.), five niche specialists, frequently-occasionally found in temperate rainforests throughout Scotland (the sexually/asexually reproducing *Lobaria pulmonaria* (L.) Hoffm., the asexually reproducing *Pannaria conoplea* (Ach.) Bory, the sexually reproducing *Pannaria rubiginosa* (Ach.) Bory, the sexually reproducing *Pectenia cyanoloma* (Schaerer) P. M. Jørg., L. Lindblom, Wedin & S. Ekman and the asexually reproducing *Pectenia atlantica* (Degel.) P.M.Jørg., L. Lindblom, Wedin & S. Ekman), and two 'super-specialists', occasionally-rarely found in temperate rainforests throughout Scotland (the asexually reproducing *Nephroma parile* (Ach.) Ach., and the sexually reproducing *Nephroma laevigatum*, Ach.).

The study area comprised a total of 11 sites lying within the temperate rainforest bioclimatic zone of Western Scotland, including the training site and ten test sites (see Fig. 1). The training site was located within Glen Creran, which is a north easterly aligned valley, comprising a sea loch with steep wooded slopes of native broadleaved semi natural woodland. The ten test sites included analogous areas of semi-natural ancient woodland within the Scottish National Forest Estate.

2.2. Collecting the training dataset

Data from the Native Woodland Survey of Scotland (NWSS, Forestry Commission Scotland, 2014) were used to stratify 181 individual

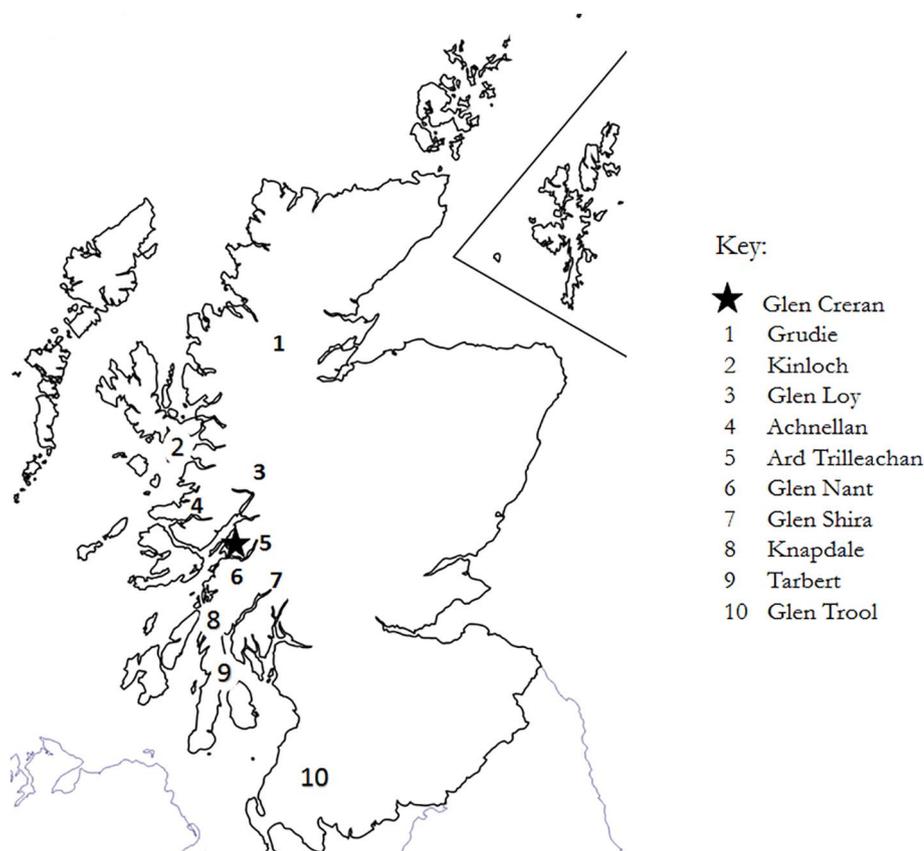


Fig. 1. Map showing the location of the training site (Glen Creran) and the ten test sites. Basemap courtesy of D-maps, 2017.

woodland polygons occurring within Glen Creran (training site), according to habitat type, stand maturity, and canopy cover. Each of the polygons was thus assigned to one of 30 unique woodland ‘types’ e.g. ‘Oakwood, Mature, 80%–100% canopy cover’ (see Supplementary Material E for a list). The largest polygon of each of the different woodland types was selected for survey. Twenty points were randomly generated within each polygon (total 600 sampling points) using the “mapprools” package in R. For each of the sampling points, the nearest tree was subject to a presence/absence survey for the nine lichen epiphyte species. In addition, a variety of compound variables (representing factors known to influence lichen epiphyte presence/absence such as bark pH, light levels, and humidity (Ellis, 2012; Ellis et al., 2015)) were recorded, namely; tree species, tree circumference, canopy cover, mean distance to the nearest tree, distance to water, and presence/absence of an adjacent canopy gap.

2.3. Trialling and evaluating SDM techniques

Full details of how the optimum SDMs were selected are provided in the supplementary material (A). In brief, the three contrasting SDM techniques were applied to the training data: Classification Trees (CT’s) (Breiman et al., 1984), Nonparametric Multiplicative Regression (NPMR) (McCune, 2006), and Generalised Additive Modelling (GAM) (Guisan et al., 2002). The three techniques were evaluated using a threshold independent method, the Area Under the Curve (AUC) of a Receiver Operating Curve (ROC). Once the optimum SDM technique had been selected, its performance in predicting species’ presence/absence was further evaluated by first, cross-validating the threshold independent AUC score and second calculating a threshold dependent score of model accuracy (the True Skill Statistic).

2.4. Applying SDM at the landscape scale

Spatial data from the Native Woodland Survey of Scotland (NWSS,

Forestry Commission Scotland, 2014) comprised tree species, tree maturity, canopy cover and tree density which were used to simulate a spatially realistic wooded landscape of 214,530 trees in ArcGIS v. 10.1 (ESRI, UK). Probability of species presence was generated for each tree within the simulated landscape, using the predict function of ‘mgcv’ (type = response) in R. The optimum thresholds identified in model evaluation (see Supplementary material A) were used to convert probability of presence to species presence/absence at the tree level. Landscapes of species presence/absence could then be plotted in ArcMap. Density maps showing the number of potentially colonised trees per hectare were visualised using DivaGIS.

2.5. Scaling SDM to biogeographic region

At each of the ten test sites, a minimum of 50 trees were identified for survey, comprising the maximum number of species/maturity combinations (according to the NWSS classification) the site would yield (identified from a walkover survey to ensure the full extent of the site and its trees had been explored). Individual trees were surveyed as in the training region (see [Collecting the training dataset](#), above).

The optimum SDMs developed for each of the nine lichen species at the training site (Glen Creran) were used to generate a probability of presence for each lichen species, for each tree surveyed in the test sites. These predictions were then evaluated, using both threshold dependent (True Skill Statistic (TSS), specificity and sensitivity) and threshold independent (AUC) methods. Threshold dependent scores were calculated according to the optimum thresholds identified for the training site.

3. Results

3.1. Study system

A lichen species’ observed frequency conformed to the a priori

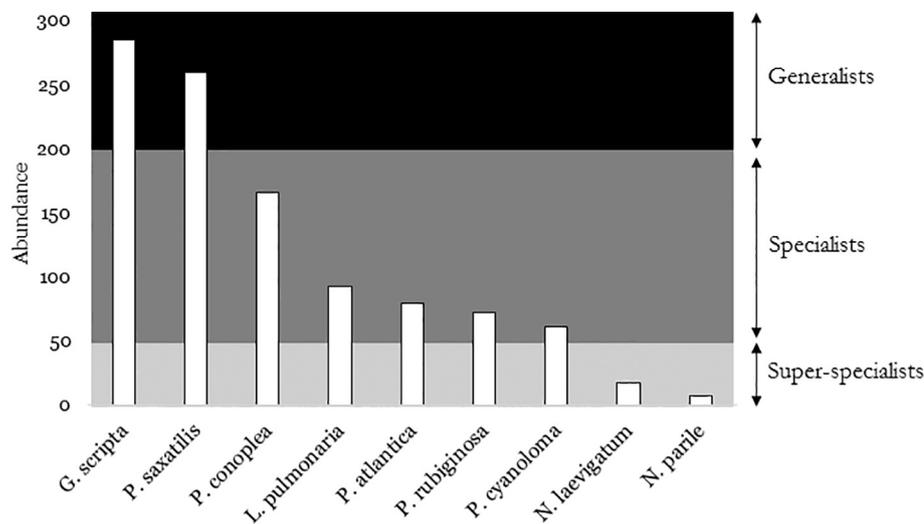


Fig. 2. Frequencies of the epiphytes under study (no. trees colonised), arranged by niche specialism.

expectations of niche breadth, with proposed niche generalists being the most frequent at the training site, and the super-specialists the least frequent (Fig. 2).

3.2. Trialling and evaluating SDM techniques

Full details of SDM evaluation are presented in Supplementary material A. In brief, each of the techniques showed a consistently high performance (except in the case of CT's when applied to the rarest species *N. parile* and *N. laevigatum*). Overall, the GAMs provided the highest AUC scores, ranging from a “good” 0.80 to an “excellent” 0.95 according to Swets, 1988 (see Supplementary material A, Figs. A.1 and A.4), and were therefore selected for final characterisation of species response (see Supplementary material B for visualisations of the species response). The explanatory variables most often used in the final optimised models were ‘tree species’ and ‘circumference’ (see Supplementary material A, Fig. A.2), and were the only variables needed in seven of the nine GAMs (with *P. saxatilis* also requiring the variable distance to water and *N. laevigatum* also requiring the variable canopy gap). Further testing (including cross validation and threshold dependent tests) confirmed that the GAMs provided acceptable levels of accuracy (see Supplementary material A, Figs. A.1 and A.6). Tests of spatial autocorrelation in the underlying data (Moran's I) found statistically significant clustering in four of the nine species (see Supplementary material A Fig. A.1), though the strength of this clustering was weak (Moran's I 0.02 to 0.05), and the remaining five species showed no significant clustering at all.

3.3. Applying SDM at the landscape scale

An example SDM projection onto the wooded landscape of Glen Creran was visualised for the common species *Lobaria pulmonaria*, at a 1 ha resolution (Fig. 3). Supplementary material C shows the projections of all remaining species.

3.4. Applying SDM at the regional scale

When applied to the 10 test sites in the wider biogeographic region, the results of the threshold independent tests (AUC score) and threshold dependent tests (overall sensitivity, specificity and TSS) were mixed (Figs. 4A, B, 5A and B). In 26 of the 90 species-test site combinations, an AUC score could not be calculated because the species did not occur in the survey sample of the relevant test sites. Of the 64 AUC scores that could be calculated, the AUC scores ranged from “fail” 0.32 to “excellent” 0.99 (Swets, 1988). Sixty-six percent of species-test site

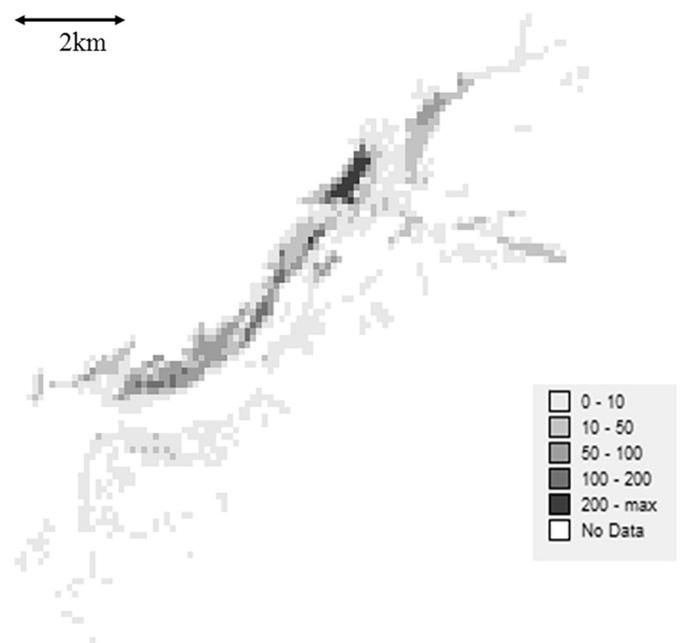


Fig. 3. Predicted species distribution of *Lobaria pulmonaria* within Glen Creran, as number of colonised trees per hectare.

combinations were calculated at a value of 0.7 or above, indicating that the models predicted probability of a species' presence to an acceptable standard (Swets, 1988). Species-test site combinations (see Fig. 4A and B) with a mean AUC scores of below 0.7 comprised the sites of Knapdale and Glen Trool (which had mean site AUC scores of 0.58 and 0.66 respectively), and in the cases of both the rarest and most common species; *N. parile*, *N. laevigatum*, and *G. scripta* (with mean AUC scores of 0.57, 0.44 and 0.69 respectively).

The TSS scores (see Fig. 5A and B) were interpreted according to Landis and Koch (1977) as almost perfect > 0.81, substantial > 0.61, moderate > 0.41, fair > 0.21 and fail < 0.21. Scores ranged from 0 to 0.86 with 58% of cases showing an acceptable or ‘moderate’ level of accuracy (at 0.41 or over), 23% showing a ‘substantial’ level of accuracy (0.61 or over) and just 1% showing an ‘almost perfect’ level of accuracy. Sites/species demonstrating an unacceptable level of accuracy (with median scores falling below 0.41, see Fig. 5A) include Knapdale and Glen Trool (with median scores of 0.30 and 0.27 respectively), and the species *N. parile*, *G. scripta* and *P. saxatilis* (with median scores – 0.19, 0.15 and 0.315). Only three species/sites (Glen

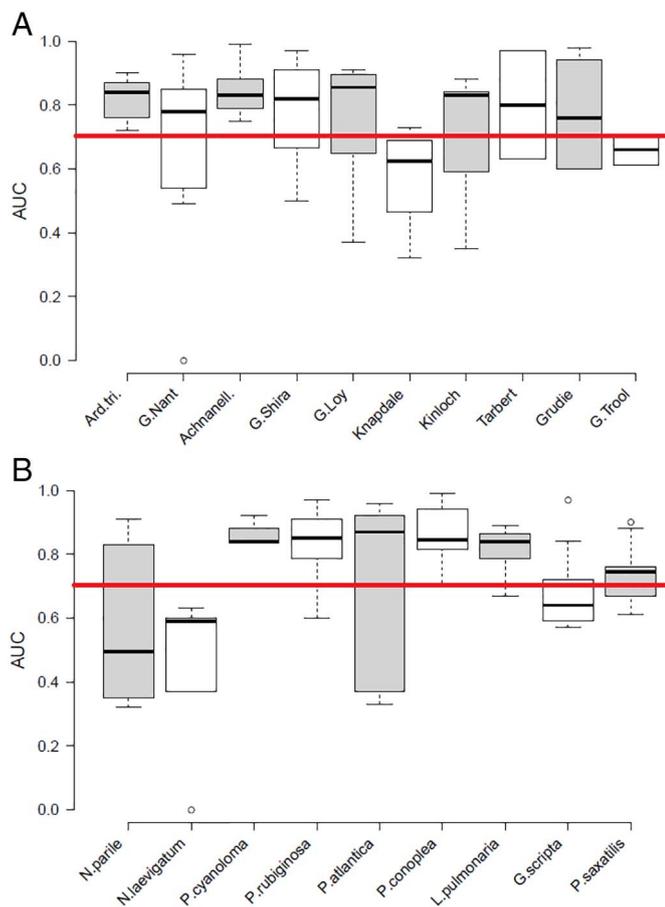


Fig. 4. A. AUC scores arranged per site (arranged by increasing distance from training site) and Panel B. AUC scores per species (arranged by increasing species abundance). Centre lines show the medians; box limits indicate the 25th and 75th percentiles; whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles, outliers are represented by dots. The horizontal line represents the threshold above which models are considered to provide an acceptable level of accuracy.

Shira, *P. cyanoloma* and *P. rubiginosa*) demonstrate a ‘substantial’ level of accuracy, none of the sites demonstrate an almost perfect level of accuracy.

4. Discussion

In this study a range of lichen epiphytes were used to test whether a) compound variables may be used to as environmental predictors of their distribution, b) SDMs can successfully produce robust projections of habitat suitability and predictions of species presence/absence at the landscape scale, and c) whether these SDM’s may be transferred outside of the test site to be used more widely. We found that:

1. A consistent set of ecologically realistic compound predictor variables (tree species and size) can successfully predict species presence/absence for lichen epiphytes within a training region, circumventing the need for micro-scale direct environmental predictor data;
2. By combining SDMs with data from a spatially extensive dataset containing the compound variables of tree species and size (from the Native Woodland Survey of Scotland), predictions of species distribution could be made across the entire training landscape. Easily recorded, widely studied larger-scale variables (such as trees) are therefore valuable resources in predicting the distribution of small/less charismatic species;
3. By testing the SDMs more widely within the ancient temperate

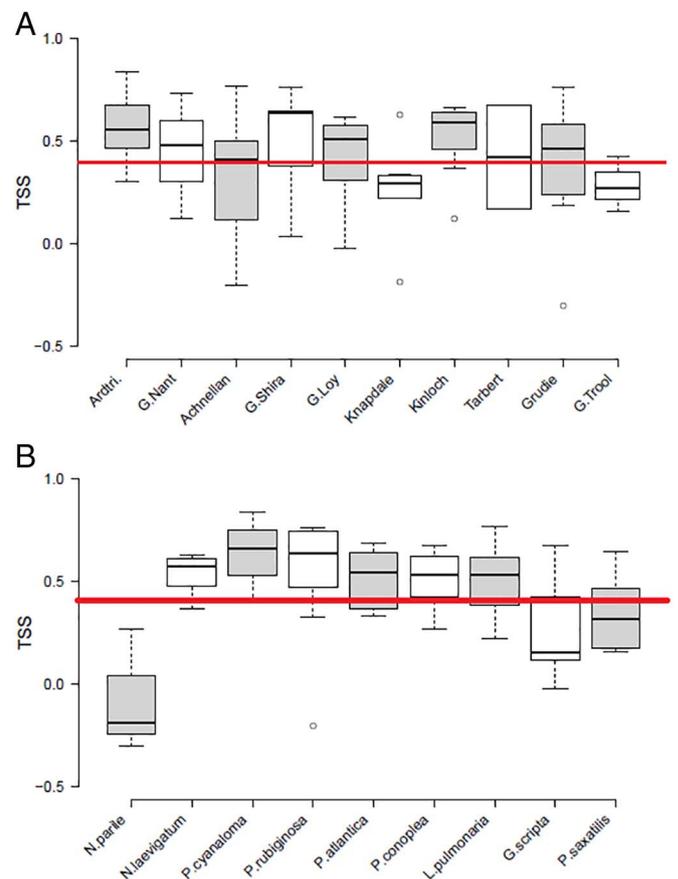


Fig. 5. A. TSS scores arranged per site (arranged by increasing distance from training site) and Panel B. TSS scores per species (arranged by increasing species abundance). Centre lines show the medians; box limits indicate the 25th and 75th percentiles; whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles, outliers are represented by dots. The horizontal line represents the threshold above which models are considered to provide an acceptable level of accuracy.

rainforests of western Scotland, predictions of species presence/absence could be made throughout the region for many lichen epiphyte species, making the SDMs still more valuable as a conservation planning tool. In particular there was variability in model transferability among species exhibiting different niche specificities, with models for niche specialists transferring to new sites well and those for niche generalists and super-specialists transferring poorly.

4.1. The epiphyte study system

The a priori selection of lichen epiphytes aimed to represent organisms with differing niche breadth; the surveyed abundance at our training site showed that these assumptions were met. *Parmelia saxatilis* and *G. scripta* each occupied over 30% of all trees surveyed (niche generalists), while *N. parile* and *N. laevigatum* occupied fewer than 10% of trees (niche super-specialists).

Identifying appropriate environmental predictors is one of the most crucial of SDM operations (Hirzel and Le Lay, 2008), and so becomes particularly challenging for data deficient species that respond to microhabitat variability. Of the previous factors explored in SDMs for lichens, climatic factors are known to be important drivers of species larger-scale distribution (Dymytrova et al., 2016), with oceanic species in Scotland showing a preference for the comparatively warm and consistently wet climate along the Atlantic coast (Ellis et al., 2007). The need to include these climatic factors within the models built here has been negated however, by restricting model application to a narrow band of western Scotland, comprising the temperate rainforest

bioclimatic zone (Ellis, 2016). Operating within this oceanic zone we can have high confidence in the tree level environmental predictors selected in our models. First, there was consistency of predictor variables identified among each of the three SDM techniques, notably tree species and circumference. Second, in agreement with the variables identified here, tree level factors such as species and size have repeatedly been inferred across epiphytic studies as predictors of lichen species occurrence (e.g., Ellis et al., 2015; Johansson et al., 2009; Király et al., 2013; Mežaka et al., 2012; Uliczka and Angelstam, 1999).

Nevertheless, direct niche factors controlling epiphyte distributions comprise gradients that operate at microhabitat scales, such as the physical and chemical properties of tree bark, and light and humidity effects (Ellis, 2012). As it is impossible to map these gradients at the required resolution (millimetre or centimetre scale) over entire landscapes, tree species and size can act as useful compound variables that capture these microhabitat effects in a robust and predictable way, e.g. different tree species have contrasting bark pH (Harmer et al., 2010) or bark textures (Ellis et al., 2015). These compound variables are more easily recorded and mapped, increasing their value in SDM. A spatially extensive data system such as the Native Woodland Survey of Scotland (NWSS) is shown to be an incredibly valuable resource for predicting the distributions of lichen epiphytes and potentially other under-recorded, tree-dependent small/less-charismatic species (such as fungi, bryophytes and invertebrates).

4.2. Testing SDM accuracy – method choice and validation

Previous studies have shown that different SDM techniques can yield inconsistent predictions (Pearson et al., 2006; Thuiller et al., 2003), making the method choice for SDM a critical consideration. However here, the predictions based on three mathematically contrasting techniques (GAMS, NPMR and CT) showed consistently high correlations (particularly in the case of NPMRs and GAMS). The only exception was for CT predictions for the rarest species (*N. parile* and *N. laevigatum*). The lower accuracy of CT's in comparison to GAM's confirms previous findings (Thuiller et al., 2003) and poor performance of the CT's seems to be a result of the very low number of presence records in the dataset.

The most common approach in generating predictions of species presence/absence from probabilistic SDM, is the application of a threshold (Liu et al., 2005). To test the accuracy of the GAM predictions within the training site, both threshold dependent (Sensitivity, Specificity and TSS) and threshold independent evaluation methods (AUC) were used. All evaluation methods found GAMs to produce what could be considered accurate predictions overall, with AUC scores showing “good” to “excellent” discrimination ability across species, and TSS scores showing “moderate” to “perfect” model accuracy across species. This finding is in line with similar comparative studies where GAMS have been identified as reliable model frameworks in SDM (e.g. Randin et al., 2006; Thuiller et al., 2003).

Model robustness was investigated by the additional testing of GAM models through cross-validation. Most cross-validated species models showed at least “good” discrimination ability according to AUC scores, however two of the nine species, *P. atlantica* and *G. scripta*, showed only “satisfactory” discriminatory ability, despite having “very good” non-cross validated AUC scores. The low cross-validated AUC scores indicate a degree of model sensitivity to small changes in the dataset used to build the models, and these cases were expected to display lower transferability in novel environments.

Finally, the spatial structure of model residuals was investigated in order to check for spatial autocorrelation. This can occur because of the role of dispersal limitation in creating a source-sink dynamic, including a ‘mass effect’ (Shmida and Ellner, 1984). The limited signature of spatial autocorrelation in the results underscores the value of the SDMs in capturing a species response to patterns of suitable/unsuitable habitat (unconfounded by spatial constraints such as dispersal-limitation

(Snäll et al., 2003)), allowing them to be projected out-with the training dataset (Guisan and Thuiller, 2005).

4.3. Applying SDM at the landscape scale

The SDMs generated from GAMs were applied systematically within the entire landscape of the training site (Glen Creran), scaling from 600 trees of the training dataset, to a total of over 200,000 trees.

Existing SDMs for epiphytes have tended to focus at the scale of the most easily accessible variables such as interpolated climate data, or larger-scale forest structure such as generic forest types (Bolliger et al., 2007; Ellis et al., 2007; Rubio-Salcedo et al., 2013). Consequently, SDM has been used to predict the distribution of lichen epiphytes at coarse-grained resolutions, such as 1 km² (Bolliger et al., 2007), 500 m² (Wiersma and Skinner, 2011) or 100 m² (Dymytrova, 2016). Though such outputs prove important for example in predicting species range (Rubio-Salcedo et al., 2013), or making predictions about climate change impacts (Ellis et al., 2007), they are of less value to land managers who need to make decisions about the specific detail of forest structure and tree composition within a wooded site. Matching SDM's that use compound variables (trees species and size), with data from the NWSS, it is possible to produce extremely high-resolution inference of species distribution over an entire landscape. It is hoped that spatially extensive high-resolution species distribution maps such as these will provide useful information to land managers and decision makers in the future.

4.4. Scaling SDM to a biogeographic region

The ability of models to be used outside of the region in which they were trained - known as model transferability - is highly species dependent (Kleyeer, 2002; Segurado and Araujo, 2004), e.g. in a transferability study of 54 vascular plant SDMs, 60% of the models were found to fail according to threshold independent methods, with AUC scores < 0.7 (Randin et al., 2006). In this study, large differences in transferability between species were found. According to our use of AUC and TSS, the SDMs for habitat specialists retained high accuracy throughout the test sites where as those for habitat generalists and super-specialists demonstrated lower transferability.

Previous evidence has also demonstrated poor transferability, and low accuracy, in SDMs developed for habitat generalists (Allouche et al., 2006; Guisan and Hofer, 2003; Segurado and Araujo, 2004). For example, in the case of lichen epiphytes, Dymytrova et al. (2016) found model accuracy to be significantly lower in habitat generalists than habitat specialists, which they attributed to the wide substrate tolerance of the more common species. In agreement with this, it is likely that the environment encountered within the training site represents a small portion of the full environmental tolerance of the habitat generalists. Projection to new spatial locations (the test sites) is therefore likely to include aspects of suitable habitat not encountered within the training site, thus leading to low accuracy scores outside within the test sites.

At the opposite end of the scale, the super-specialists also showed poor transferability according to the accuracy measures; AUC and TSS. The low AUC scores can here be explained as a statistical artefact of the AUC metric, which represents a discrimination index testing if the likelihood that a presence will have a higher probability of presence, than absence (Hosmer and Lemeshow, 2000). Thus, if probabilities of presence are only moderately higher than those for absence, even a well fitted model can show poor discrimination ability (Lobo et al., 2008). This is consistent with the low probability of presence predictions for these rare species, e.g. a highly suitable habitat for *N. parile* (such as a hazel stem of 50 cm circumference), only has a probability of presence of 0.07. When combined with the few presences in the test datasets, these low values greatly reduce the discriminatory ability of the models.

Reproductive mode was not found to be useful in explaining the

variability in model transferability, with neither sexually reproducing nor asexually reproducing species appearing to show higher TSS or AUC scores. This suggests that reproductive mode has less of an effect than niche specialism on the ability of a species to occupy available niche space.

In contrast to the differences in model accuracy among species, differences in model accuracy among the test sites can be attributed to factors such as historic or current land use practice, or biotic effects such as the degree of autogenic competition occurring in epiphyte communities, which may differ between sites (Fielding and Haworth, 1995; Randin et al., 2006). According to TSS, the models appear to perform particularly poorly at Glen Trool and Knapdale. However, close inspection of the data shows that, at Glen Trool the only species contributing to the accuracy measures are the habitat generalists *G. scripta* and *P. saxatilis*. Reasons for the poor transferability of habitat generalists have been discussed above, and consequently this result is unsurprising. The TSS scores for Knapdale however encompassed 8 of the 9 species under study, weakening the argument that habitat generalists account for a poor model performance at this site. Knapdale however differed from other sites in its high overall abundance of lichen epiphytes, suggesting a higher occupancy rate of suitable niche space. For example, *Lobaria pulmonaria* appears to have an unusual pattern of occurrence, occupying a high proportion (38%) of the trees surveyed (making it more common than the niche generalist *P. saxatilis*) and occupying tree species it is not often found on such as Birch (*Betula sp.*) and young Rowan (*Sorbus aucuparia*). Such differences could be attributable to mass effects (Pulliam, 2000), a result of high species abundance causing proximal yet unsuitable habitats to become colonised, or strong local effects of topography and soil chemistry, which can interact with tree species identity to shift bark chemical properties (Gustafsson and Eriksson, 1995).

4.5. Conclusions

Where SDMs can be accurately applied to extensive areas, as is the case for five of the nine species studied here, they become extremely valuable tools in conservation planning. However, the variability in model transferability shown in this study highlights the importance of testing SDM's outside of their training region prior to application in the wider landscape to prevent prediction errors.

SDMs will be most useful when predictions of species distribution can be reliably made over entire regions at a resolution of use to conservation planning; compound variables offer a solution to the paucity of species distributional data for small/less charismatic species, and present a valuable resource for predicting the distribution of such species. The importance of testing the SDMs outside of a training region to prevent prediction error is highlighted however.

Acknowledgements

This project was supported by grants from Scottish Natural Heritage (grant number 14503) and Forestry Commission Scotland. We sincerely thank the owners of the studied woodland for giving us access to their private lands.

Author contributions

CE, DH, DG, RY, RT and SE conceived the ideas and designed the methodology, SE collected the data, SE analysed the data with support and guidance from CE and DH, SE led writing of the manuscript with support and guidance from CE and DH. All authors contributed critically to the drafts and gave final approval for publication.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://>

doi.org/10.1016/j.biocon.2017.11.012.

References

- Allouche, O., Tsoar, A., Kadmon, R., 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *J. Appl. Ecol.* 43, 1223–1232.
- Bates, J.W., 1992. Influence of chemical and site factors on *Quercus* and *Fraxinus* epiphytes at Loch Sunart, western Scotland: a multivariate analysis. *J. Ecol.* 80, 163–179.
- Bolliger, J., Bergamini, A., Stofer, S., Kienast, F., Scheidegger, C., 2007. Predicting the potential spatial distributions of epiphytic lichen species at the landscape scale. *Lichenologist* 39, 279–291.
- Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., 1984. *Classification and Regression Trees*. CRC Press.
- Britton, A.J., Mitchell, R.J., Potts, J.M., Genney, D.R., 2013. Development of a surveillance scheme for priority lichens in Scotland. Scottish Natural Heritage Commissioned Report No. 567.
- Clark, J.A., May, R.M., 2002. Taxonomic bias in conservation research. *Science* 297 (5579), 191–192.
- D-maps, 2017. http://www.d-maps.com/carte.php?num_car=15861&lang=en (Accessed 22.06.2017).
- Dymytrova, L., Stofer, S., Ginzler, C., Breiner, F.T., Scheidegger, C., 2016. Forest-structure data improve distribution models of threatened habitat specialists: implications for conservation of epiphytic lichens in forest landscapes. *Biol. Conserv.* 196, 31–38.
- Ellis, C.J., 2012. Lichen epiphyte diversity: a species, community and trait-based review. In: *Perspectives in Plant Ecology, Evolution and Systematics*. 14. pp. 131–152.
- Ellis, C.J., 2016. Oceanic and temperate rainforest climates and their epiphyte indicators in Britain. *Ecol. Indic.* 70, 125–133.
- Ellis, C.J., Coppins, B.J., Dawson, T.P., Seaward, M.R., 2007. Response of British lichens to climate change scenarios: trends and uncertainties in the projected impact for contrasting biogeographic groups. *Biol. Conserv.* 140, 217–235.
- Ellis, C., Eaton, S., Theodoropoulos, M., Elliott, K., 2015. *Epiphyte Communities and Indicator Species. An Ecological Guide for Scotland's Woodlands*. Royal Botanic Garden, Edinburgh.
- Fielding, A.H., Haworth, P.F., 1995. Testing the generality of bird-habitat models. *Conserv. Biol.* 9, 474–499.
- Forestry Commission Scotland, 2014. *The Native Woodland Survey Scotland*. <http://scotland.forestry.gov.uk/supporting/strategy-policy-guidance/native-woodland-survey-of-scotland-nwss> (Accessed December 2014).
- Fritz, O., Brunet, J., Caldiz, M., 2009. Interacting effects of tree characteristics on the occurrence of rare epiphytes in a Swedish boreal forest. *Bryologist* 112, 488–505.
- Gauslaa, Y., 1995. The Lobarion, an epiphytic lichen community of ancient forests threatened by acid rain. *Lichenologist* 27, 59–76.
- Guisan, A., Hofer, U., 2003. Predicting reptile distributions at the mesoscale: relation to climate and topography. *J. Biogeogr.* 30, 1233–1243.
- Guisan, A., Thuiller, W., 2005. Predicting species distributions: offering more than simple habitat models. *Ecol. Lett.* 8, 993–1009.
- Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. *Ecol. Model.* 135, 147–186.
- Guisan, A., Edwards, T.C., Hastie, T., 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecol. Model.* 157, 89–100.
- Guisan, A., Tingley, R., Baumgartner, J.B., Naujokaitis-Lewis, I., Sutcliffe, P.R., Tulloch, A.L., Regan, T.J., Brotons, L., McDonald-Madden, E., Mantyka-Pringle, C., Martin, T.G., 2013. Predicting species distributions for conservation decisions. *Ecol. Lett.* 16, 1424–1435.
- Gustafsson, L., Eriksson, I., 1995. Factors of importance for the epiphytic vegetation of aspen *Populus tremula* with special emphasis on bark chemistry and soil chemistry. *J. Appl. Ecol.* 32, 412–424.
- Harmer, R., Kerr, G., Thompson, R., 2010. *Managing Native Broadleaved Woodland*. The Stationary Office Ltd (TSO).
- Hartley, S., Kunin, W.E., Lennon, J.J., Pocock, M.J.O., 2004. Coherence and discontinuity in the scaling of species distribution patterns. *Proc. R. Soc. B Biol. Sci.* 271, 81–88.
- Hirzel, A.H., Le Lay, G., 2008. Habitat suitability modelling and niche theory. *J. Appl. Ecol.* 45, 1372–1381.
- Hosmer, D.W., Lemeshow, S., 2000. *Applied Logistic Regression*, 2nd edn. John Wiley & Sons, New York.
- Hosokawa, T., Odani, N., 1957. The daily compensation period and vertical ranges of epiphytes in a beech forest. *J. Ecol.* 45, 901–906.
- Jefferies, M.J., 2006. *Biodiversity and Conservation*, 2nd Edition. Routledge, New York.
- Johansson, V., Bergman, K.O., Lättman, H., Milberg, P., 2009. Tree and site quality preferences of six epiphytic lichens growing on oaks in south-eastern Sweden. *Ann. Bot. Fenn.* 46, 496–506.
- Jürriado, I., Liira, J., Paal, J., Suija, A., 2009. Tree and stand level variables influencing diversity of lichens on temperate broad-leaved trees in boreo-nemoral floodplain forests. *Biodivers. Conserv.* 18, 105–125.
- Király, I., Nascimbene, J., Tinya, F., Ódor, P., 2013. Factors influencing epiphytic bryophyte and lichen species richness at different spatial scales in managed temperate forests. *Biodivers. Conserv.* 22, 209–223.
- Kleyer, M., 2002. Validation of plant functional types across two contrasting landscapes. *J. Veg. Sci.* 13, 167–178.
- Kuusinen, M., 1994. Epiphytic lichen diversity on *Salix caprea* in old-growth southern and middle boreal forests of Finland. *Ann. Bot. Fenn.* 31, 77–92.
- Kuusinen, M., 1996. Epiphyte flora and diversity on basal trunks of six old-growth forest

- tree species in southern and middle boreal Finland. *Lichenologist* 28, 443–463.
- Landis, J., Koch, G., 1977. The measurement of observer agreement for categorical data. *Biometrics* 33, 159–174.
- Lewis, J.E.J., Ellis, C.J., 2010. Taxon-compared with trait-based analysis of epiphytes, and the role of tree species and tree age in community composition. *Plant Ecol. Divers.* 3, 203–210.
- Liu, C., Berry, P.M., Dawson, T.P., Pearson, R.G., 2005. Selecting thresholds of occurrence in the prediction of species distributions. *Ecography* 28, 385–393.
- Liu, C., White, M., Newell, G., Griffioen, P., 2013. Species distribution modelling for conservation planning in Victoria, Australia. *Ecol. Model.* 249, 68–74.
- Lobo, J.M., Jiménez-Valverde, A., Real, R., 2008. AUC: a misleading measure of the performance of predictive distribution models. *Glob. Ecol. Biogeogr.* 17, 145–151.
- McCune, B., 2006. Non-parametric habitat models with automatic interactions. *J. Veg. Sci.* 17, 819–830.
- McPherson, J.M., Jetz, W., Rogers, D.J., 2006. Using course-grained occurrence data to predict species distributions at finer spatial resolutions – possibilities and limitations. *Ecol. Model.* 192, 499–522.
- MEA, 2005. Millennium Ecosystem Assessment Findings. Millennium Ecosystem Assessment. Ecosystems and Human Well-being: Biodiversity Synthesis. World Resources Institute, Washington, DC.
- Mežaka, A., Brūmelis, G., Piterāns, A., 2012. Tree and stand-scale factors affecting richness and composition of epiphytic bryophytes and lichens in deciduous woodland key habitats. *Biodivers. Conserv.* 21 (12), 3221–3241.
- Nicholls, A.O., 1989. How to make biological surveys go further with generalized linear models. *Biol. Conserv.* 50, 51–75.
- Pearson, R.G., Thuiller, W., Araújo, M.B., Martínez-Meyer, E., Brotons, L., McClean, C., Miles, L., Segurado, P., Dawson, T.P., Lees, D.C., 2006. Model-based uncertainty in species range prediction. *J. Biogeogr.* 33, 1704–1711.
- Peterson, A.T., 2006. Uses and requirements of ecological niche models and related distribution models. *Biodivers. Inform.* 3, 59–72.
- Polce, C., Termansen, M., Aguirre-Gutiérrez, J., Boatman, N.D., Budge, G.E., Crowe, A., Garratt, M.P., Pietravalle, S., Potts, S.G., Ramirez, J.A., Somerwill, K.E., 2013. Species distribution models for crop pollination: a modelling framework applied to Great Britain. *PLoS One* 8, 76308.
- Pulliam, H.R., 2000. On the relationship between niche and distribution. *Ecol. Lett.* 3, 349–361.
- Purvis, O.W., 1992. Lichen Flora of Great Britain and Ireland. Natural History Museum Publications in Association With the British Lichen Society.
- Rambo, T.R., 2010. Habitat preferences of an arboreal forage lichen in a Sierra Nevada old-growth mixed-conifer forest. *Can. J. For. Res.* 40, 1034–1041.
- Randin, C.F., Dirnbock, T., Dullinger, S., Zimmermann, N.E., Zappa, M., Guisan, A., 2006. Are niche-based distribution models transferable in space. *J. Biogeogr.* 33, 1689–1703.
- Ranius, T., Eliasson, P., Johansson, P., 2008. Large-scale occurrence patterns of red-listed lichens and fungi on old oaks are influenced both by current and historical habitat density. *Biodivers. Conserv.* 17, 2371–2381.
- Rodríguez, J.P., Brotons, L., Bustamante, J., Seoane, J., 2007. The application of predictive modelling of species distribution to biodiversity conservation. *Divers. Distrib.* 13, 243–251.
- Rondinini, C., Wilson, K.A., Boitani, L., Grantham, H., Possingham, H.P., 2006. Tradeoffs of different types of species occurrence data for use in systematic conservation planning. *Ecol. Lett.* 9, 1136–1145.
- Rubio-Salcedo, M., Martínez, I., Carreno, F., Escudero, A., 2013. Poor effectiveness of the Natura 2000 network protecting Mediterranean lichen species. *J. Nat. Conserv.* 21, 1–9.
- Seaward, M.R.D. (Ed.), 1995. Lichen Atlas of the British Isles. British Lichen Society, London.
- Segurado, P., Araújo, M.B., 2004. An evaluation of methods for modelling species distributions. *J. Biogeogr.* 31, 1555–1568.
- Shmida, A., Ellner, S., 1984. Coexistence of plant species with similar niches. *Vegetatio* 58, 29–55.
- Snäll, T., Ribeiro Jr., P.J., Rydin, H., 2003. Spatial occurrence and colonisations in patch-tracking metapopulations: local conditions versus dispersal. *Oikos* 103, 566–578.
- Swets, J.A., 1988. Measuring the accuracy of diagnostic systems. *Science* 240, 1285–1293.
- Thuiller, W., Araújo, M.B., Lavorel, S., 2003. Generalized models v's classification tree analysis: predicting spatial distributions of plant species at different scales. *J. Veg. Sci.* 14, 669–680.
- Uliczka, H., Angelstam, P., 1999. Occurrence of epiphytic macrolichens in relation to tree species and age in managed boreal forest. *Ecography* 22, 396–405.
- Vanderpoorten, A., Sotiaux, A., Engels, P., 2005. A GIS-based survey for the conservation of bryophytes at the landscape scale. *Biol. Conserv.* 121, 189–194.
- Vaughan, I.P., Ormerod, S.J., 2003. Improving the quality of distribution models for conservation by addressing shortcomings in the field collection of training data. *Conserv. Biol.* 17, 1601–1611.
- Watson, R., Albon, S., Aspinall, R., Austen, M., Bardgett, B., Bateman, I., Berry, P., Bird, W., Bradbury, R., Brown, C., Bullock, J., 2011. UK National Ecosystem Assessment: Understanding Nature's Value to Society. Synthesis of Key Findings.
- Whittaker, R.J., Araújo, M.B., Paul, J., Ladle, R.J., Watson, J.E.M., Willis, K.J., 2005. Conservation biogeography: assessment and prospect. *Divers. Distrib.* 11, 3–23.
- Wiersma, Y.F., Skinner, R., 2011. Predictive distribution model for the boreal felt lichen *Erioderma pedicellatum* in Newfoundland, Canada. *Endanger. Species Res.* 15, 115–127.