Biological Conservation 142 (2009) 2501–2509

Contents lists available at ScienceDirect

Biological Conservation

journal homepage: www.elsevier.com/locate/biocon

Using summed individual species models and state-of-the-art modelling techniques to identify threatened plant species hotspots

Miia Parviainen [a,*], Mathieu Marmion [a], Miska Luoto [a], Wilfried Thuiller [b], Risto K. Heikkinen [c]

[a] Department of Geography, University of Oulu, P.O. Box 3000, FIN-90014 Oulu, Finland
[b] Laboratoire d'Ecologie Alpine, UMR CNRS 5553, Université Joseph Fourier, BP 53, 38041 Grenoble Cedex 9, France
[c] Finnish Environment Institute, Research Department, P.O. Box 140, FIN-00251 Helsinki, Finland

ARTICLE INFO

Article history:
Received 12 November 2008
Received in revised form 18 May 2009
Accepted 24 May 2009
Available online 21 June 2009

Keywords:
Hotspot
Regression techniques
Rule-based methods
Taiga
Threatened species
Uncertainty

ABSTRACT

Reliable identification of hotspot areas with high numbers of threatened plant species has a central role in conservation planning. We investigated the potentiality of identifying the distribution, richness and hotspots of threatened plant species at a 25 ha resolution using eight state-of-the-art modelling techniques (GLM, GAM, MARS, ANN, CTA, GBM, MDA and RF) in a taiga landscape in north-eastern Finland. First, the individual species models developed based on occurrence records of 28 species in the 1677 grid squares and derived from different statistical techniques were extrapolated to the whole study area of 41 750 km^2. Second, the projected presence/absence maps were then combined to create species richness maps, and the top 5% of grid cells ranked by species richness were classified as hotspots. Finally, we created an overall summary map by combining the individual hotspot maps from all eight modelling techniques and identified areas where the individual hotspots maps overlapped most. There were distinguishing differences in projections of the geographic patterns of species richness and hotspots between the modelling techniques. Most of the modelling techniques predicted several hotspot locations sporadically around the study area. However, the overall summary map showed the highest predictive performance based on Kappa statistics, indicating that the locations where the hotspot maps from the eight models coincided most harboured highest observed species richness. Moreover, the summary map filtered out the patchy structures of individual hotspot maps. The results show that the choice of modelling technique may affect the accuracy and prediction of hotspot patterns. Such differences may hamper the development of useful biodiversity model applications for conservation planning, and thus it is beneficial if the conservation decision-making can be based on sets of alternative maps and overlaying of predictions from multiple models.

1. Introduction

Recognition and understanding of geographical patterns of species distributions have long been central themes in ecology and biogeography (Franklin, 1995; Wu and Smeins, 2000; Seoane et al., 2003; Araújo and Guisan, 2006), and such information is frequently used in different ecological applications especially in the field of nature conservation and management planning (Gioia and Pigott, 2000; Wu and Smeins, 2000; Loiselle et al., 2003; Vaughan and Ormerod, 2003). One important attribute of biological diversity that has a significant role in conservation and management strategies is concentrations (hotspots) of threatened species (Eyre and Rushton, 1989; Gaston, 1994). Information of where hotspots of threatened species richness are located, or are most likely to occur, is thus critical for successful nature conservation (Margules and Pressey, 2000).

Predictive models of single species geographical distributions and species richness provide an attractive alternative for incomplete or spatially biased survey data as a basis for conservation planning (Fleishman et al., 2002; Lehmann et al., 2002; Vaughan and Ormerod, 2003; Heikkinen et al., 2007a; Rodriguez et al., 2007; Parviainen et al., 2008). If reliable models for occurrences and richness of (rare or threatened) species can be produced, they can enhance the capacity for land-use planning and assist managers in meeting different objectives (Fleishman et al., 2002), and the development of rapid and cost-effective methods to map areas in order to assess their biological value for nature conservation (Margules and Austin, 1991; Gaston, 1996; Debinski et al., 1999; Nagendra and Gadgil, 1999; Ottaviani et al., 2004).

A traditional way to develop spatial projections of species richness has been direct measurement of numbers of species from surveyed sites and relation of this information to the environmental variables measured from the same sites or derived from different digital layers of geographical information (GIS) data, to produce models that yield predictions of richness to unsampled sites

---

* Corresponding author. Tel.: +358 503048517.
E-mail address: miia.parviainen@oulu.fi (M. Parviainen).

0006-3207/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.biocon.2009.05.030

(Gould, 2000; Luoto et al., 2004; Moser et al., 2005). A less often employed way of simulating richness patterns is to model each species separately, and then to identify trends and hotspots of species richness by summing up the spatial predictions of individual species (Cumming, 2000; Gioia and Pigott, 2000; Lehmann et al., 2002; Pineda and Lobo, 2009). This approach can provide some interesting advantages, such as better control for poorly modelled species and unacceptable prediction maps (Gioia and Pigott, 2000), and easier identification of the set of the most important explanatory variables and the response shapes between of species and their environment in certain subgroups of species. However, there are also potential limitations in this approach. The broad array of methods currently available to modellers and managers for modelling patterns of species distribution can be confusing (Franklin, 1995; Guisan and Thuiller, 2005; Elith et al., 2006), and evaluation of the relative performance of different techniques remains a continuous challenge in ecology and conservation biology (Loiselle et al., 2003; Thuiller, 2003; Lawler et al., 2006; Pearson et al., 2006; Elith and Leathwick, 2007; Heikkinen et al., 2007a; Marmion et al., 2009a). Moreover, very little is known concerning whether the differences between various methods projections for separate species accumulate in summed richness maps, which may make the planning decisions even more sensitive to differences between model outputs. It is possible that the spatial overlaying of simulated richness maps from several different techniques, and examination of the locations where they overlap, could help in generating more robust identification of richness hotspots. However, this question has been insufficiently studied (but see Loiselle et al., 2003).

In northern Europe, intensive forest management has caused drastic changes in forests during recent decades (Kouki et al., 2001). This has resulted in the decline and endangerment of many forest species. For example, 80% of all red-listed (threatened or near-threatened) species recorded in Finland are primarily threatened by habitat changes (Rassi et al., 2001). The distributions of many threatened species groups, especially vascular plants, are in many areas rather well-known in Finland. Thus these species provide a useful study setting for investigating the potentiality of identifying threatened species hotpots in more remote and insufficiently areas in the landscape, by summing up the predictions of separate species models and using several different modelling techniques.

In this study, we aimed to identify the distribution and hotspots of threatened plant species using eight modelling techniques, applied to each of the 28 threatened vascular plant species in order to predict current species distribution, richness and threatened plant species hotspots. The study is an extension based on the work by Marmion et al. (2009a) in which a thorough comparison of the performance of the same eight modelling techniques was provided. Here, we specifically focused on topics that were not explored in Marmion et al. (2009a), namely (1) to investigate and compare the differences between the simulated hotspot maps of threatened plant species among the eight techniques, and (2) to explore the potentiality of identifying locations of the threatened plant species hotspots in taiga landscapes by overlaying the predictions from multiple models.

2. Materials and methods

2.1. Study area

The study area (41,750 km^2) is located between 31–32°45′E and 65–67°50′N in north-eastern Finland (Fig. 1). Phytogeographically the study area lies within the northern boreal zone (Ahti et al., 1968), where pine- and spruce-dominated forests prevail. Wetlands, numerous lakes and rivers characterize the landscape of the study area. The bedrock is calcium-rich in many places, providing favourable conditions for species-rich plant communities (Parviainen et al., 2008). The mean annual temperature is ca. −0.5 °C, the growing season lasts ca. 130 days, and the difference between the mean temperature of the coldest (January, ca. −14 °C) and the warmest (July, ca. 15 °C) month is ca. 29 °C (Atlas of Finland, 1987).

2.2. Plant species data

Due to the long-standing naturalist tradition, the species distribution patterns of threatened vascular plant species are relatively well known in many parts of Finland. In this study we used occurrence records from the national database of threatened species maintained by the Finnish Environment Institute. The comprehensive field records made by voluntary amateur and professional botanists are the most important data sources for this data base, but information on species occurrences has also been gathered from the scientific literature and herbaria.

The species data for our study area consisted of presence records of 28 species with ten or more records in the 1677 grid squares with an area of 25 ha (Table 1). These 1677 grid squares were spread across the whole study area and used as the starting data set in the model building (Parviainen et al., 2008). The database of the threatened plant species does not include records of the absence of the species. However, we assumed that the absence of a record in any of the 1677 grid squares corresponded to true absence of the species (Eyre et al., 2004), given the quasi-exhaustive sampling strategy. According to the IUCN classification (Gärdenfors et al., 2001), 24 (86%) of the plant species were defined as vulnerable and 4 (14%) as endangered (EN) (Table 1).

2.3. Environmental variables

In total, three climate, four topography, three geology and six land cover variables were used to explain threatened plant species distribution (see Parviainen et al., 2008). The correlations between explanatory variables used in the modelling experiments were low (Spearman's correlation coefficient, $|r| < 0.70$) The explanatory variables are described in detail in Parviainen et al. (2008).

2.4. Modelling techniques

We simulated the distribution of the 28 threatened plant species and their hotspots using the BIOMOD tool (Thuiller, 2003), as implemented for R software. The following eight techniques based on binomial data were used in modelling analyses: generalized linear models (GLM), generalized additive models (GAM), multivariate adaptive regression splines (MARS), artificial neural networks (ANN), classification tree analysis (CTA), general boosting method (GBM), mixture discriminant analysis (MDA), and random forests (RF). The modelling techniques are described and discussed in Thuiller (2003), Elith et al. (2006) and Marmion et al. (2009a, in press). Marmion et al. (2009a) provides a detailed comparison of the accuracy of these eight modelling techniques with the present species data, and thus those results are only briefly touched upon here.

First, for all eight modelling methods, we calibrated and validated the models using the species data in the 1677 grid squares with an area of 25 ha. The derived models were then used to produce simulated distribution maps for individual species at a spatial resolution of 25 ha in the whole study area (i.e. 166,968 grid cells each 25 ha in size) showing the area where the environmental requirements for the species were met (Guisan and Thuiller, 2005). Second, to create richness maps we combined the presence/absence maps for individual species and used a simple summation of the predicted occurrences (Lehmann et al., 2002). Third, we identified richness hotspots as the top 5% of grid cells ranked by species richness (see Prendergast et al., 1993; Williams

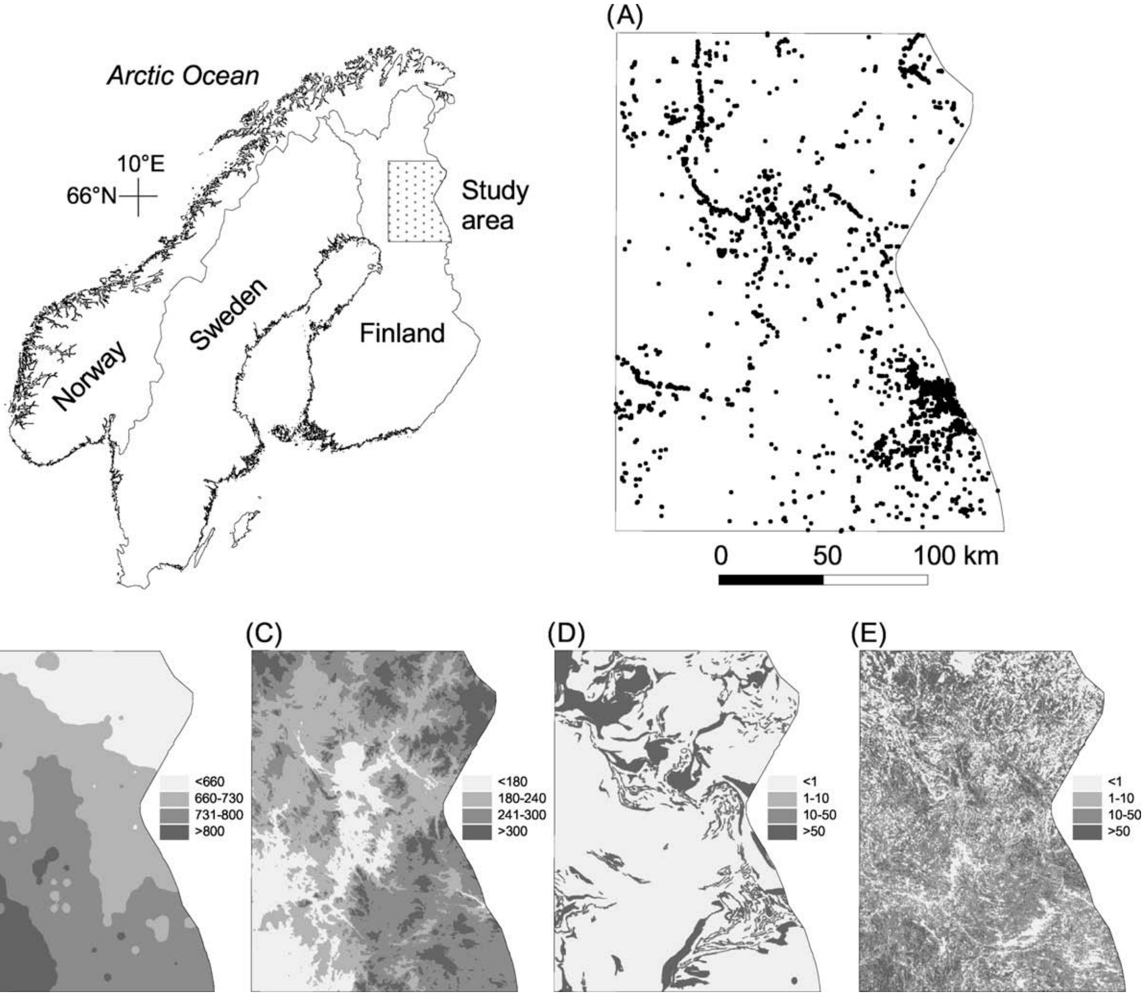

Fig. 1. (A) A distribution map with the observational points of the studied threatened plant species in the 1677 grid squares used in the model building and environmental variables calculated for the whole study area: (B) growing degree days (gdd), (C) elevation (m), and percentage cover (%) of (D) calcareous rock and (E) open mire.

et al., 1996). Finally, we created an overall summary map by combining the individual hotspot maps from all eight modelling techniques. The resulting overall summary hotspot map is a summation of the eight individual hotspots maps and any location on this summary map can have a hotspot value between 0 and 8. In other words, high values in the summary maps indicate the locations where the hotspots identified by the eight models coincide most.

For a comparison, we produced a map of predicted species richness, using the traditional direct species richness modelling approach by relating the observed species richness to the same predictor variables as used in the single-species models, using a GAM and Poisson distribution and logarithmic link function (Heikkinen, 1998). The hotspot map (5% of grid cells ranked by species richness) produced with this approach was then compared with the hotspot maps obtained by summing individual species models.

2.5. Model evaluation

To evaluate the models, the original data set of 1677 grid squares was randomly divided into model training (70%, $n = 1174$) and model evaluation sets (30%, $n = 503$) (split-sample approach; see Guisan and Zimmermann, 2000). The predictions and observed data from the subsets of data were combined and used to measure the performance of models by calculating the area under the curve (AUC) of a receiver operating characteristic (ROC) plot (Fielding and Bell, 1997) to assess the agreement between the presence–absence records and the predictions. The AUC ranges generally from 0.5 for models with no discrimination ability to 1 for models with perfect discrimination.

Finally, we evaluated how well the different modelling approaches (a hotspot map derived from direct species richness model provided by GAM, hotspot maps based on the eight individual models, and the overall summary map based on the eight hotspot maps) identified the observed species richness hotspot in the model evaluation set. The predictive performance of the different approaches was here assessed by Kappa (κ) statistics using the modelled and the observed hotspot information. The κ value expresses the agreement not obtained randomly between two qualitative variables (Congalton, 1991). The κ index is based on the misclassification matrix. The range of κ is from -1.0 to 1.0.

We acknowledge here that our evaluation data set does not represent a totally independent test set for assessing the predictive abilities of different models (cf. Araújo et al., 2005; Randin et al., 2006; Heikkinen et al., 2007b). However, as the 25 ha grid cells in both our model calibration and evaluation data sets were distributed rather sparsely across the whole study area (grid cells used in modelling covered only ca. 1% of the whole study area), we assume that the effect of spatial autocorrelation was rather small. Moreover, in the recent paper by Parviainen et al. (2008) the effect of spatial autocorrelation on the modelling results of the same data set was tested by including autocovariate terms in

Table 1
Prevalence (the ratio of presence squares to the total sample in the original data set of 1677 grid squares), main habitats of species and conservation status for the studied 28 threatened plant species.

Species	Prevalence (%)	Main habitats	Status
Arnica angustifolia	1.85	R	VU
Asplenium ruta-muraria	2.62	R	VU
Botrychium boreale	3.88	C, R, S	VU
Botrychium lanceolatum	2.74	S, C	VU
Calypso bulbosa	17.11	F, M	VU
Carex appropinquata	3.58	M, A, S	VU
Carex heleonastes	9.90	M, A	VU
Carex lepidocarpa ssp. jemtlandica	1.61	M	VU
Carex microlochin	1.25	S, M, R	EN
Carex viridula var. bergrothii	2.98	M, S	VU
Cerastium alpinum (*ssp. alpinum*)	4.00	R	EN
Cypripedium calceolus	16.82	F, M, C	VU
Dactylorhiza traunsteineri	6.74	M	VU
Dactylorhiza lapponica	1.01	M	VU
Dactylorhiza incarnata ssp. cruenta	4.83	M, S, C	VU
Elymus fibrosus	6.14	S, C	VU
Epilobium laestaedii	1.49	M, A, C	EN
Epipogium aphyllum	1.31	F, M	VU
Gentianella amarella	3.64	C, S, R	VU
Gypsophila fastigiata	2.03	F, R	VU
Lonicera caerulea	0.72	S, F	EN
Lychnis alpina var. serpentinicola	2.03	K	VU
Minuartia biflora	1.25	R	VU
Moehringia lateriflora	12.82	F, S	VU
Primula stricta	2.74	S	VU
Saxifraga hirculus	22.06	M, A	VU
Schoenus ferrugineus	1.91	M	VU
Silene tatarica	6.02	S, C	VU
Mean ± std	5.18 ± 5.54		

Main habitats of the species: A = aquatic, C = cultural, F = forest, M = mire, R = rocky, S = shore. Conservation status: EN = endangered, VU = vulnerable (Rassi et al., 2001).

the GAMs. Inclusion of the autocovariate had only a minor effect on the importance of the environmental variables and the shape of predictor–response curves. Thus, we consider that our study setting is not considerably impaired by spatial autocorrelation.

3. Results

The number of threatened species recorded in 1677 grid cells each 25 ha in size varied between 1 and 8, with a mean of 1.47 species. The average AUC values of the evaluation set varied from 0.730 (CTA) to 0.856 (RF) with a mean value of 0.801 (Appendix A). Although the variation of the AUC values of the evaluation data was high (ranging from 0.482 to 0.998), almost half (47%) of the models had good or excellent model performance (AUC > 0.8). The environmental variables selected into models are presented in Appendix B.

As in Marmion et al. (2009a), also in the present results there were partly marked differences in the predicted distributions between different modelling techniques. Predictions of threatened plant species richness and richness hotspots based on summation of single-species predictions and direct species richness approach are shown in Figs. 2 and 3. Based on visual examination and statistical evaluation, also here there was considerable variation between modelling methods. The correlations between the projected richness patterns and hotspots were highest between GLM and GAM (Spearman's correlation coefficient, $r = 0.578$ and Kappa = 0.300; Table 2). GAM and GLM illustrated the greatest hotspot concentration in the eastern part of study area, whereas MARS suggested hotspot squares scattered all over the study area. Very dissimilar richness maps were produced by ANN and GBM, as indicated by the negative correlation ($r = -0.119$) and very low Kappa-value (0.033) (Table 2) between their projections.

The geographic overlap of threatened plant species hotspots between different modelling techniques is represented in Fig. 4. A total of 1695 squares (1.02%) of the whole study area were identified as hotspots by at least four modelling techniques, but only 13 grid (0.01%) squares by all eight modelling techniques. Visual interpretation showed rather clearly the bimodal distribution pattern of the summarized hotspots in the eastern and northern parts of the study area. Based on Kappa statistics, the overall summary map based on eight hotspot maps showed the highest predictive performance to identify observed species richness hotspots in the evaluation data (Fig. 5). Moreover, the overall summary map filtered out the patchy structures of individual hotspot maps improving the visual interpretation of the model outputs.

4. Discussion

4.1. Examination of threat spots

Total species richness and biodiversity hotspots are widely used measures of biological diversity (Prendergast et al., 1993; Purvis and Hector, 2000). However, developing representative assessments of biodiversity for a given area also requires examination of threatened species, i.e. species that are considered to be at risk of extinction in the wild. Threatened species often have restricted geographical distributions and are not well-represented in overall species richness analysis (Jetz and Rahbek, 2002). Examination and reliable modelling of regions with high numbers of threatened species, threat spots, may thus have an essential role in successful conservation strategies (Gaston, 1994, 1996; Kivinen et al., 2008). This is because although regions with a high number of species may contain many rare species, rarity and species richness often provide independent and complementary measures of biological diversity (Prendergast et al., 1993; Ceballos and Ehrlich, 2006).

Classical way to model total species richness or threatened species richness is based on GLM or GAM techniques using Poisson distribution and logarithmic link function (Heikkinen, 1998). However, modelling of the spatial distribution of threatened species and developing useful models for their richness patterns generally poses different challenges than modelling of species with large ranges. This is because the performance of the models depends on the characteristics of the species (see Pearce et al., 2001; Seoane et al., 2005; Carrascal et al., 2006; McPherson and Jetz, 2007; Pöyry et al., 2008). Whereas species with large ranges are often generalist species with wide ecological tolerances showing continuous distribution ranges, threatened species distribution patterns are usually geographically much more limited and tend to be particularly at finer resolutions spatially sporadic due to their strict ecological requirements for certain habitat characteristics (e.g. soil types, microclimate). For modelling studies carried out at landscape scale, including the present study, the sporadic distribution patterns of threatened species may cause a certain problem. Namely, data on the summed number of threatened species in the surveyed locations across the study area are often zero inflated. Zero inflated data are defined as data that has a larger proportion of zeros than expected from pure count (Poisson) data (Barry and Welsh, 2002). If we ignore this feature of the threatened species richness data and apply standard Poisson error models problems with inference can occur. These problems arise because the Poisson assumption is not always an adequate approximation to the conditional distribution of the data. In that case modelling species distributions separately for each species and then summing the predicted presences could be a useful strategy to assess the spatial variation in species richness for conservation purposes, particularly for poorly sampled regions (Cumming, 2000; Pineda and Lobo, 2009).

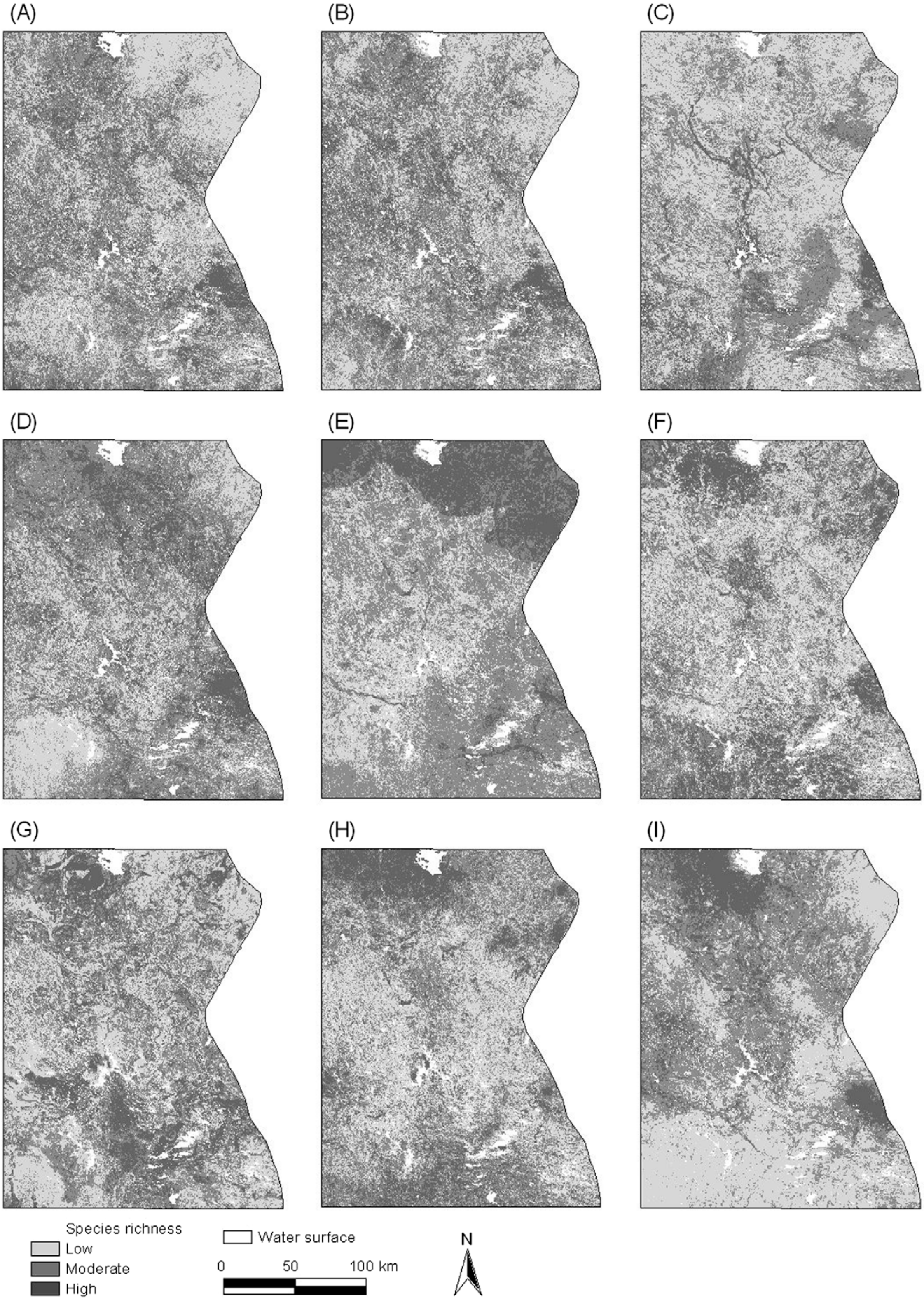

Fig. 2. Spatial predictions of species richness for the whole study area based on eight different modelling techniques: (A) GLM, (B) GAM, (C) MARS, (D) ANN, (E) CTA, (F) GBM, (G) MDA, (H) RF and (I) direct species richness approach based on Poisson GAM.

4.2. Interpretation of the prediction maps and multiple models

Maps representing the locations of the projected hotspots are important to conservation managers, landscape planners and researchers. However, there is currently a broad array of approaches available to modellers and conservation managers, and thus it can be difficult to choose the most suitable technique from several competing methods (Cumming, 2000; Pineda and Lobo, 2009). This study makes a contribution to this arena by summing up of predictions from multiple

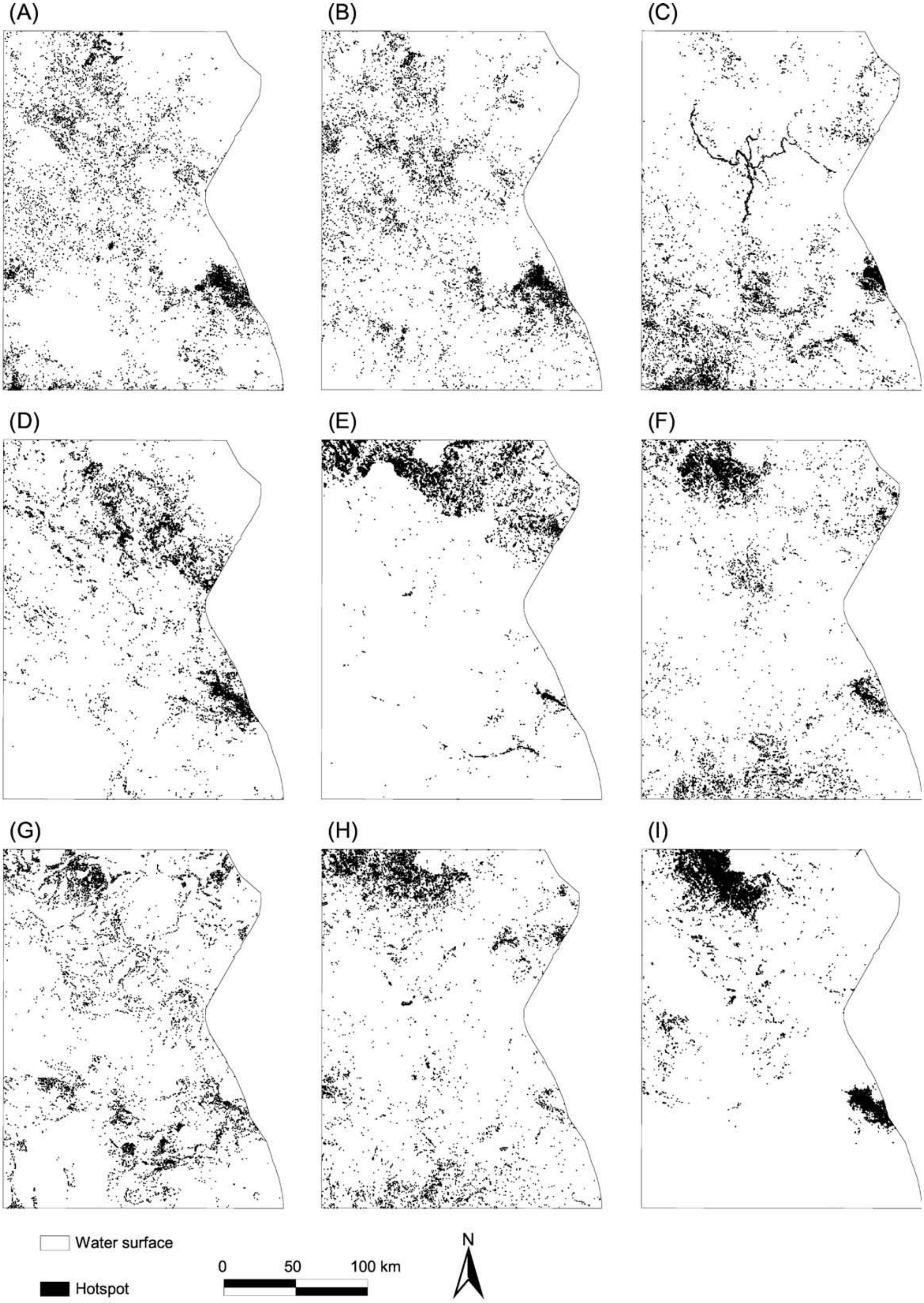

Fig. 3. Spatial predictions of threatened plant species hotspots based on eight different modelling techniques: (A) GLM, (B) GAM, (C) MARS, (D) ANN, (E) CTA, (F) GBM, (G) MDA, (H) RF and (I) direct species richness approach based on Poisson GAM.

occurrence models in identifying hotspots of threatened plant species.

Based on our modelling results, the spatial patterns of threatened plant species hotspots in the overall summary map indicate that there were two distinct hotspot concentrations, in the eastern and the northern part of study area, but minor concentrations of threatened plant species were also found in other parts of the study area, especially near the rivers. It is known that threatened species diversity can be encompassed in relatively small areas of the landscape (Reid, 1998). In general, hotspots of threatened plant

Table 2
Pairwise correlation coefficients (Spearman rank correlation) for the predicted species richness between eight modelling techniques and direct richness modelling approach based on GAM. The value between brackets denotes Kappa values for the spatial agreement between the observed and predicted hotspots. The correlations are based on predicted species richness, whereas the Kappa-values are based on the binary information of the predicted hotspots.

	GAM Poisson	GLM	GAM	MARS	ANN	CTA	GBM	MDA
GLM	0.281 (0.160)							
GAM	0.274 (0.173)	0.578 (0.300)						
MARS	−0.043 (0.033)	0.097 (0.040)	0.065 (0.032)					
ANN	0.520 (0.266)	0.152 (0.131)	0.080 (0.121)	0.058 (0.018)				
CTA	0.084 (0.216)	−0.106 (0.000)	−0.032 (0.019)	−0.021 (−0.006)	0.108 (0.021)			
GBM	−0.006 (0.312)	0.181 (0.067)	0.291 (0.099)	0.053 (0.049)	−0.119 (0.033)	0.426 (0.294)		
MDA	0.064 (0.146)	0.177 (0.072)	0.239 (0.079)	−0.022 (0.014)	0.065 (0.039)	0.240 (0.133)	0.264 (0.152)	
RF	0.065 (0.252)	0.053 (0.049)	0.071 (0.044)	0.045 (0.026)	0.041 (0.018)	0.455 (0.285)	0.534 (0.294)	0.249 (0.135)

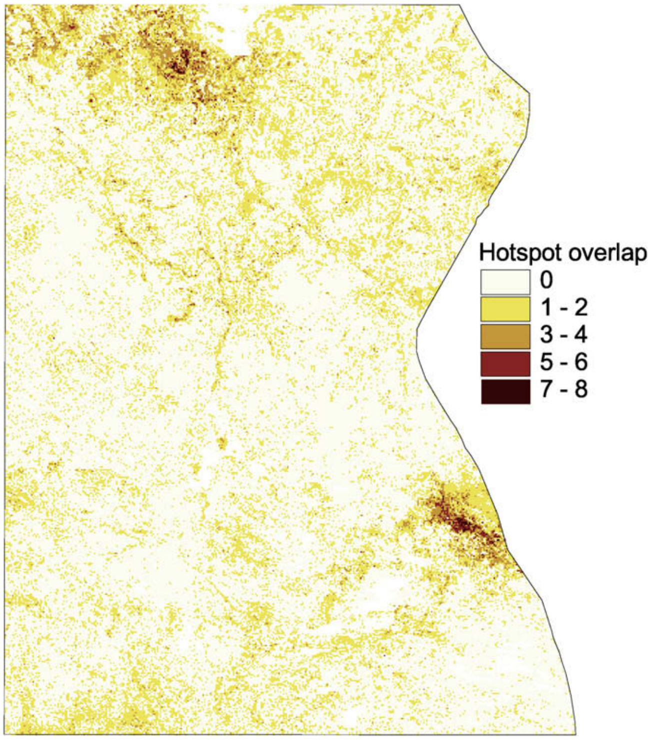

Fig. 4. Overall summary map of threatened plant species hotspots based on the combination of the individual hotspot maps from all eight modelling techniques (see Fig. 3), showing the degree of the overlap between the individual hotspots maps.

species in our study area were located in areas characterized by rather continental climate and at relatively high altitudes with sun-drenched cliffs, where calcareous rich sandy soils prevail (see Parviainen et al., 2008).

Similarly as in the previous study (Marmion et al., 2009a), we found also here marked variation in modelling success between different modelling techniques. This emphasizes the fact that it is beneficial to consider prediction outputs from multiple models while making assessments of species richness trends and hotspot patterns (Thuiller, 2003; Segurado and Araújo, 2004; Elith et al., 2006). However, it is also imperative to realize that good modelling performance with observed distributions does not guarantee that the model produces the ecologically most plausible distribution maps, especially for non-sampled parts of a given area (Pearson and Dawson, 2003; Pearson et al., 2006; Randin et al., 2006). Due to the large differences between predictions from alternative modelling techniques, it is usually not possible to single out the "best" technique for identification of potential hotspot locations. This has important consequences for the applied use of predictive models e.g., for conservation purposes, where the accuracy of the models is a critical factor distinguishing the hotspots of threatened species (Heikkinen, 1998).

An important aspect of model evaluation is the visual interpretation of the probability maps. It is noteworthy that although AUC yielded rather similar results across techniques, there were nevertheless non-negligible differences in the spatial predictions made for the whole study area. This indicates that small differences between two modelling techniques in terms of evaluation statistics can result in large differences in predictions (Thuiller, 2003, 2004). The marked variation in results between species and different modelling techniques supports the argument of Elith and Leathwick (2007) that the fitting of reliable models is not always possible, and that it is important to evaluate the output as rigorously and comprehensively as possible.

The correlations among richness maps and hotspot maps created with different techniques were rather weak. Somewhat surprisingly, although the predictive performance of GAM and GLM was only moderate, the richness and hotspot maps produced by these models were nevertheless best supported by the biological knowledge of threatened plant species richness. Our findings also reinforce the argument that different modelling techniques do not necessarily produce equivalent and equally plausible hotspot predictions; particularly projections by ANN and GBM did not generally coincide. Most of the modelling techniques predicted several hotspot locations sporadically around the study area, and thus the individual predicted hotspot distribution maps for the study area were remarkably patchy. However, the patterns of predicted hotspot locations became much more clearly defined when individual species richness prediction maps were combined. Thus, overlaying of predictions from multiple models may significantly enhance filtering out of the patchy structures of individual hotspot maps and identification of the distinct hotspot concentrations in a given area by summarizing agreements among projections generated by different models. In addition, the overall summary map of hotspots provided a more accurate projection of hotspots, that coincided better with the locations of high observed species richness that the forecasts from the traditional approach of modelling species richness directly using a GAM with a Poisson distribution assumption.

Predictive distribution modelling of species and associated richness maps offer a useful tool for identification of threatened species hotspots and help in directing attention to critical geographical locations for conservation and management planning. However, caution should be exercised also in interpretation of the summed richness, because the separate species distribution models forming the basis for such richness maps may be sensitive to many different problems, ranging from the selection of predictor variables and scaling problems into sampling biases and species geocoding problems (see Gioia and Pigott, 2000; Gutzwiller and Barrow, 2001; Elith et al., 2002; Vaughan and Ormerod, 2003).

In the North European taiga landscape, there is a great deal of pressure to increase resource extraction, mining, tourism industry

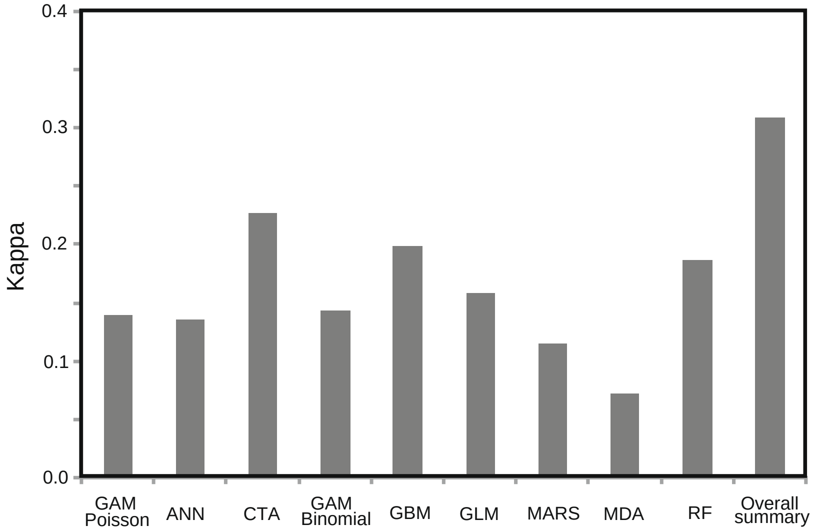

Fig. 5. The spatial agreement between the observed and predicted hotspots in direct species richness approach, the eight individual hotspot maps and overall summary map in the evaluation data. The degree of spatial agreement was assessed by Kappa statistics.

and road building, to generate power from rivers, to protect native heritage and economies, and to preserve wilderness areas (Kouki et al., 2001). All these activities require an analysis of potential ecological resources and effects on an area. Our results show that species distribution models and the summed richness maps from multiple models can provide powerful means to delimit valuable nature conservation areas and focus management efforts in order to ensure the preservation of biological diversity in taiga landscapes.

5. Conclusions

Identification of the locations harbouring several threatened plant species (hotspots) based on summation of the predictions from individual species distribution models provides an attractive alternative for ‘direct’ modelling of species richness (Cumming, 2000; Gioia and Pigott, 2000; Lehmann et al., 2002). In addition, instead of production of a single richness or hotspot map for conservation prioritization (Redford et al., 2003), it is beneficial if the broader conservation decision-making can be based on sets of alternative maps based on varying parameters and choices (Whittaker et al., 2005). The summed richness and hotspots patterns and maps such as presented here can be readily divided into different sub-categories. This enables the modellers and managers to scrutinize the predictions for species with, for example, different endangerment status, different responses to alternative management measures, or species with different ecological traits. Information on species compositional trends is also easier to extract from separate species models than from traditional species richness maps (Luoto et al., 2004). Most importantly, differences between the spatial predictions among different models may hamper the reliable identification of richness hotspots of threatened species (or other species groups of interest) and development of useful applications for conservation and management planning. A more solid basis for planning can be built by overlaying several different model outputs. Hotspots identified in this way represent potentially important areas for the targeting of survey and conservation efforts.

Acknowledgments

A study of this nature would not have been possible without the hundreds of volunteers who contributed their data to the threatened plant species database. The comments of two anonymous referees improved the manuscript greatly. Michael Bailey helped with correction of the English text. Different parts of this study were funded by the Academy of Finland (Project Grant 116544) and EC FP6 Integrated Project ALARM (GOCE-CT-2003-506675) (Settele et al., 2005). WT received support from European Commission’s FP6 MACIS (Minimisation of and Adaptation to Climate change Impacts on biodiversity No. 044399) and ECOCHANGE (Challenges in assessing and forecasting biodiversity and ecosystem changes in Europe, No. 066866 GOCE) projects.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.biocon.2009.05.030.

References

Ahti, T., Hämet-Ahti, L., Jalas, J., 1968. Vegetation zones and their sections in northwestern Europe. Annales Botanici Fennici 5, 169–211.

Araújo, M.B., Guisan, A., 2006. Five (or so) challenges for species distribution modelling. Journal of Biogeography 33, 1677–1688.

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

Atlas of Finland, 1987. Climate, folio 131. National Board of Survey and Geographical Society of Finland, Helsinki, p. 58.

Barry, S.C., Welsh, A., 2002. Generalized additive modeling and zero inflated count data. Ecological Modelling 157, 179–188.

Carrascal, L.M., Seoane, J., Palomino, D., Alonso, C.L., Lobo, J.M., 2006. Species-specific features affect the ability of census-derived models to map winter avian distribution. Ecological Research 21, 681–691.

Ceballos, G., Ehrlich, P.R., 2006. Global mammal distributions, biodiversity hot spots, and conservation. Proceedings of Natural Academy of Sciences, USA 103, 19374–19379.

Congalton, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. Remote Sensing of Environment 37, 35–46.

Cumming, G.S., 2000. Using habitat models to map diversity: pan-African species richness of ticks (Acari: Ixodida). Journal of Biogeography 27, 425–440.

Debinski, D.M., Kindscher, K., Jakubauskas, M.E., 1999. A remote sensing and GIS-based model of habitats and biodiversity in the Greater Yellowstone Ecosystem. International Journal of Remote Sensing 17, 3281–3291.

Elith, J., Burgman, M.A., Regan, H.M., 2002. Mapping epistemic uncertainties and vague concepts in predictions of species distributions. Ecological Modelling 157, 313–329.

Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T., Phillips, S.J., Richardson, K.S., Scachetti-Pereira, R., Schapire, R.E., Sobero′n, J., Williams, Wisz, M.S., Zimmermann, N.E., 1996. Novel methods improve prediction of species' distributions from occurrence data. Ecography 29, 129–151.

Elith, J., Leathwick, J., 2007. Predicting species distributions from museum and herbarium records using multiresponse models fitted with multivariate adaptive regression splines. Diversity and Distributions 13, 265–275.

Eyre, M., Rushton, S., Luff, M., Telfer, M., 2004. Predicting the distribution of ground beetle species (Coleoptera, Carabidea) in britain using land cover variables. Journal of Environmental Management 72, 163–174.

Eyre, M.D., Rushton, S.P., 1989. Quantification of conservation criteria using invertebrates. Journal of Applied Ecology 26, 159–171.

Fielding, A., Bell, J., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation 24, 38–49.

Fleishman, E., Murphy, D.D., Sjögren-Gulve, P., 2002. Modeling species richness and habitat suitability for taxa of conservation interest. In: Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A., Samson, F.B. (Eds.), Predicting Species Occurrences: Issues of accuracy and scale. Island Press, Covelo, California, pp. 507–517.

Franklin, J., 1995. Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. Progress in Physical Geography 19, 474–499.

Gaston, K.J., 1994. Rarity. Chapman & Hall, London.

Gaston, K.J., 1996. Species richness: measure and measurement. In: Gaston, K.J. (Ed.), Biodiversity: A Biology of Numbers and Difference. Blackwell Science, Oxford, pp. 77–113.

Gioia, P., Pigott, J.P., 2000. Biodiversity assessment: a case study in predicting richness from the potential distributions of plant species in the forests of south-western Australia. Journal of Biogeography 27, 1065–1078.

Gould, W.A., 2000. Remote sensing of vegetation, plant species richness, and regional diversity hotspots. Ecological Applications 10, 1861–1870.

Guisan, A., Thuiller, W., 2005. Predicting species distribution: offering more than simple habitat models. Ecology Letters 8, 993–1009.

Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. Ecological Modelling 135, 147–186.

Gutzwiller, K.J., Barrow Jr., W.C., 2001. Bird-landscape relations in the Chihuahuan Desert: coping with uncertainties about predictive models. Ecological Applications 11, 1517–1532.

Gärdenfors, U., Hilton-Taylor, C., Mace, G.M., Rodríguez, J.P., 2001. The application of IUCN Red List criteria at regional levels. Conservation Biology 15, 1206–1212.

Heikkinen, R.K., 1998. Can richness patterns of rarities be predicted from mesoscale atlas data? A case study of vascular plants in the kevo reserve. Biological Conservation 83, 133–143.

Heikkinen, R.K., Luoto, M., Kuussaari, M., Toivonen, T., 2007a. Modelling the spatial distribution of a threatened butterfly: impacts of scale and statistical technique. Landscape and Urban Planning 79, 347–357.

Heikkinen, R.K., Luoto, M., Virkkala, R., Pearson, R.G., Korber, J.-H., 2007b. Biotic interactions improve prediction of boreal bird distributions at macro-scales. Global Ecology and Biogeography 16, 754–763.

Jetz, W., Rahbek, C., 2002. Geographic range size and determinants of avian species richness. Science 297, 1548–1551.

Kivinen, S., Luoto, M., Heikkinen, R.K., Saarinen, K., 2008. Threat spots and environmental determinants of red-listed plant, butterfly and bird species in boreal agricultural environments. Biodiversity and Conservation 17, 3289–3305.

Kouki, J., Löfman, S., Martikainen, P., Rouvinen, S., Uotila, A., 2001. Forest fragmentation in Fennoscandia: linking habitat requirements of wood-associated threatened species to landscape and habitat changes. Scandinavian Journal of Forest Research 16, 27–37.

Lawler, J.J., White, D., Neilson, R.P., Blaustein, A.R., 2006. Predicting climate-induced range shifts: model differences and model reliability. Global Change Biology 12, 1–17.

Lehmann, A., Overton, J.M., Austin, M.P., 2002. Regression models for spatial prediction: their role for biodiversity and conservation. Biodiversity and conservation 11, 2085–2092.

Loiselle, B.A., Howell, C.A., Graham, C.H., Goerck, J.M., Brooks, T., Smith, K.G., Williams, P.H., 2003. Avoiding pitfalls of using species distribution models in conservation planning. Conservation Biology 17, 1591–1600.

Luoto, M., Virkkala, R., Heikkinen, R.K., Rainio, K., 2004. Predicting bird species richness using remote sensing in boreal agricultural–forest mosaic. Ecological Applications 14, 1946–1962.

Margules, C.R., Austin, M.P., 1991. Nature conservation: cost effective biological surveys and data analysis. CSIRO, Australia. 207 p.

Margules, C.R., Pressey, R.L., 2000. Systematic conservation planning. Nature 405, 243–253.

Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R.K., Thuiller, W., 2009a. Evaluation of consensus methods in predictive species distribution modelling. Diversity and Distributions 15, 59–69.

Marmion, M., Luoto, M., Heikkinen, R. K., Thuiller, W., (in press). The performance of state-of-the-art modelling techniques depends on geographical distribution of species. Ecological Modelling. doi:10.1016/j.ecolmodel.2008.10.019.

McPherson, J.M., Jetz, W., 2007. Effects of species' ecology on the accuracy of distribution models. Ecography 30, 135–151.

Moser, D., Dullinger, S., Englisch, T., Niklfeld, H., Plutzar, C., Sauberer, N., Zechmeister, H.G., Grabherr, G., 2005. Environmental determinants of vascular plant species richness in the Austrian Alps. Journal of Biogeography 32, 1117–1127.

Nagendra, H., Gadgil, M., 1999. Satellite imagery as a tool for monitoring species diversity: an assessment. Journal of Applied Ecology 36, 388–397.

Ottaviani, D., Lasinio, G.J., Boitani, L., 2004. Two statistical methods to validate habitat suitability models using presence-only data. Ecological Modelling 179, 417–443.

Parviainen, M., Luoto, M., Ryttäri, T., Heikkinen, R.K., 2008. Modelling the occurrence of threatened plant species in taiga landscapes: methodological and ecological perspectives. Journal of Biogeography 35, 1888–1905.

Pearce, J., Ferrier, S., Scotts, D., 2001. An evaluation of the predictive performance of distributional models for flora and fauna in north-east New South Wales. Journal of Environmental Management 62, 171–184.

Pearson, R.G., Dawson, T.P., 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? Global Ecology and Biogeography 12, 361–371.

Pearson, R.G., Thuiller, W., Araújo, M.B., Martinez-Meyer, E., Brotons, L., McClean, C., Miles, L., Segurado, P., Dawson, T.P., Lees, D.C., 2006. Model-based uncertainty in species range prediction. Journal of Biogeography 33, 1704–1711.

Pineda, E., Lobo, J.M., 2009. Assessing the accuracy of species distribution models to predict amphibian species richness patterns. Journal of Animal Ecology 78, 182–190.

Pöyry, J., Luoto, M., Heikkinen, R.K., Saarinen, K., 2008. Species traits contribute to the uncertainty of bioclimatic models. Global Ecology and Biogeography 17, 403–414.

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

Purvis, A., Hector, A., 2000. Getting the measure of biodiversity. Nature 405, 212–219.

Randin, C.F., Dirnböck, T., Dullinger, S., Zimmerman, N.E., Zappa, M., Guisan, A., 2006. Are species distribution models transferable in space. Journal of Biogeography 33, 1689–1703.

Rassi, P., Alanen, A., Kanerva, T., Mannerkoski, I. (Eds.), 2001. The 2000 red list of Finnish species. The II Committee for the monitoring of threatened species in Finland. Ministry of the Environment & Finnish Environment Institute, Helsinki, Finland, p. 32 (In Finnish with English summary).

Redford, K.H., Coppolillo, P., Sanderson, E.W., Da Fonseca, G.A.B., Dinerstein, E., Groves, C., Mace, G., Maginnis, S., Mittermeier, R.A., Noss, R., Olson, D., Robinson, J.G., Vedder, A., Wright, M., 2003. Mapping the conservation landscape. Conservation Biology 17, 116–131.

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

Rodriguez, J.P., Brotons, L., Bustamante, J., Seoane, J., 2007. The application of predictive modelling of species distribution to biodiversity conservation. Diversity and Distributions 13, 243–251.

Segurado, P., Araújo, M., 2004. An evaluation of methods for modelling species distributions. Journal of Biogeography 31, 1555–1569.

Seoane, J., Carrascal, L.M., Alonso, C.L., Palomino, D., 2005. Species-specific traits associated to prediction errors in bird habitat suitability modelling. Ecological Modelling 185, 299–308.

Seoane, J., Vinuela, J., Díaz-Delgado, R., Bustamante, J., 2003. The effects of land use and climate on red kite distribution in the Iberian peninsula. Biological Conservation 11, 401–414.

Settele, J., Hammen, V., Hulme, P., Karlson, U., Klotz, S., Kotarac, M., Kunin, W., Marion, G., O'Connor, M., Petanidou, T., Peterson, K., Potts, S., Pritchard, H., Pysek, P., Rounsevell, M., Spangenberg, J., Steffan-Dewenter, I., Sykes, M., Vighi, M., Zobel, M., Kuhn, I., 2005. Alarm: assessing large-scale environmental risks for biodiversity with tested methods. GAIA – Ecological Perspectives for Science and Society 14, 69–72.

Thuiller, W., 2003. BIOMOD – optimizing predictions of species distributions and projecting potential future shifts under global change. Global Change Biology 9, 1353–1362.

Thuiller, W., 2004. Patterns and uncertainties of species' range shifts under climate change. Global Change Biology 10, 2020–2027.

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. Conservation Biology 17, 1601–1611.

Whittaker, R.J., Araújo, M.B., Paul, J., Ladle, R.J., Watson, J.E.M., Willis, K.J., 2005. Conservation biogeography: assessment and prospect. Diversity and Distributions 11, 3–23.

Williams, P., Gibbons, D., Margules, C., Rebelo, A., Humphries, C., Pressey, R., 1996. A comparison of richness hotspots, rarity hotspots and complementary areas for conserving diversity using British birds. Conservation Biology 10, 155–174.

Wu, X.B., Smeins, F.E., 2000. Multiple-scale habitat modeling approach for rare plant conservation. Landscape Urban Planning 51, 11–28.