

Remote sensing-based predictors improve distribution models of rare, early successional and broadleaf tree species in Utah

N. E. ZIMMERMANN*, T. C. EDWARDS JR†, G. G. MOISEN‡,
T. S. FRESCINO‡ and J. A. BLACKARD‡

*Swiss Federal Research Institute WSL, 8903 Birmensdorf, Switzerland; †USGS Utah Cooperative Fish and Wildlife Research Unit, Utah State University, Logan, UT 84322–5290, USA; and ‡Rocky Mountain Research Station, USDA Forest Service, Ogden, Utah 84401, USA

OnlineOpen: This article is available free online at www.blackwell-synergy.com

Summary

1. Compared to bioclimatic variables, remote sensing predictors are rarely used for predictive species modelling. When used, the predictors represent typically habitat classifications or filters rather than gradual spectral, surface or biophysical properties. Consequently, the full potential of remotely sensed predictors for modelling the spatial distribution of species remains unexplored. Here we analysed the partial contributions of remotely sensed and climatic predictor sets to explain and predict the distribution of 19 tree species in Utah. We also tested how these partial contributions were related to characteristics such as successional types or species traits.

2. We developed two spatial predictor sets of remotely sensed and topo-climatic variables to explain the distribution of tree species. We used variation partitioning techniques applied to generalized linear models to explore the combined and partial predictive powers of the two predictor sets. Non-parametric tests were used to explore the relationships between the partial model contributions of both predictor sets and species characteristics.

3. More than 60% of the variation explained by the models represented contributions by one of the two partial predictor sets alone, with topo-climatic variables outperforming the remotely sensed predictors. However, the partial models derived from only remotely sensed predictors still provided high model accuracies, indicating a significant correlation between climate and remote sensing variables. The overall accuracy of the models was high, but small sample sizes had a strong effect on cross-validated accuracies for rare species.

4. Models of early successional and broadleaf species benefited significantly more from adding remotely sensed predictors than did late seral and needleleaf species. The core-satellite species types differed significantly with respect to overall model accuracies. Models of satellite and urban species, both with low prevalence, benefited more from use of remotely sensed predictors than did the more frequent core species.

5. *Synthesis and applications.* If carefully prepared, remotely sensed variables are useful additional predictors for the spatial distribution of trees. Major improvements resulted for deciduous, early successional, satellite and rare species. The ability to improve model accuracy for species having markedly different life history strategies is a crucial step for assessing effects of global change.

Key-words: core-satellite species hypothesis, K-fold cross-validation, Landsat TM, partial regression, predictive habitat distribution models, species traits, variation partitioning

Journal of Applied Ecology (2007) **44**, 1057–1067
doi: 10.1111/j.1365-2664.2007.01348.x

Introduction

Predictive habitat distribution modelling is a powerful approach for conservation planning and for applied and theoretical ecological study (Guisan & Zimmermann 2000; Austin 2002a; Guisan & Thuiller 2005). It is used for a wide variety of ecological applications such as the detection (Edwards *et al.* 2005; Guisan *et al.* 2006) or prediction of rare species (Edwards *et al.* 2004; Engler, Guisan & Reichsteiner 2004; Welch & MacMahon 2005), the estimation of species richness (e.g. Heikkinen 1996; Ferrier *et al.* 2004), the testing of ecological niche concepts (Austin & Smith 1989; Austin 2002b), for biogeographical hypotheses (Leathwick 1998), the assessment of potential invasion risks (Thuiller *et al.* 2005) or for reserve selection and conservation planning (Araujo *et al.* 2005).

Increasingly, these models are being used to assess the potential consequences of global change on species distributions (Guisan, Theurillat & Spichiger 1995; Iverson & Prasad 1998; Peterson *et al.* 2002; Dirnböck, Grabherr & Dullinger 2003; Bomhard *et al.* 2005; Thuiller *et al.* 2006), thus including the effects of climate change as well as land use change and alterations of the global nitrogen cycle (Vitousek 1994). By linking climate projections with the known physiological tolerances of many species, it is possible to model direct and indirect consequences of global change scenarios on species and ecological systems. However, such future projections require two important additional components. The first is the proper understanding of the species–environment relationship, often couched in terms of niche theory. The second is the ability to relate these species–environment relationships to structural habitat properties in the form of digital geographical information system (GIS)-based layers. Projections will thus rely ideally on the melding of basic life history information with current and evolving remote-sensing techniques.

The conceptual background for habitat distribution modelling is rooted in the niche concept (Hutchinson 1957). The distinction of both indirect (distal) as well as direct or resource variables (proximal) was important for theoretical advances in this field, and for testing assumptions based on the species' niche (Austin 1980). These types of models evaluate not only the shape of the species response to the environment (Austin 1987; Yee & Mitchell 1991), but also describe the distributions of species along environmental gradients. In the past, such gradients were evaluated primarily using climate and topographic variables, but land cover remained important, especially for animal species (Manel, Buckton & Ormerod 2000; Venier *et al.* 2004). Consequently, the development of new environmental predictors of direct or resource variables (e.g. Prentice *et al.* 1992; Zimmermann & Kienast 1999; Dirnböck *et al.* 2003; Parra, Graham & Freile 2004; Edwards *et al.* 2005; Randin *et al.* 2006), and the evaluation of new statistical techniques for modelling (Austin 1987; Yee & Mitchell 1991; Elith *et al.* 2006), have been important research issues.

Competition is one reason why a species is not able to occupy all physiologically possible sites (i.e. the fundamental niche), but rather only a fraction thereof, i.e. the realized niche. The realized niche of a species is often complex (Austin & Nicholls 1997; Oksanen & Minchin 2002), and only some of the relevant predictors can be translated accurately into spatial depictions within a GIS. Therefore, information such as derived from remotely sensed data is used frequently as a predictor. These data, considered surrogates that integrate many ecological relationships, are often used at coarse spatial resolution (e.g. McPherson, Jetz & Rogers 2004; Parra *et al.* 2004; Venier *et al.* 2004), and often exist in the form of, or are derived from, land cover classifications (Pearson, Dawson & Liu 2004; Thuiller, Araujo & Lavorel 2004). Spatially explicit information on more direct predictors of species distribution, such as soil characteristics, is frequently unavailable and rarely used in habitat distribution modelling.

Also, in predictive modelling studies carried out at higher spatial resolution, the use of vegetation classifications is usually not desirable, especially for plant species where circularity would be introduced when using plant communities as predictors. Rather, subtle differences in the vegetation/soil properties, or in phenological characteristics, are of higher interest for discriminating between suitable and unsuitable sites within the potential distribution (Guisan, Theurillat & Kienast 1998; Vetaas 2002; Chefaoui, Hortal & Lobo 2005). Thus, remote sensing data can be used in better ways than simply for filtering pixels to avoid predictions in unsuitable areas. Rather, we can retrieve continuous gradient predictors to improve the calibration of the species' niche compared to topographic and bioclimatic predictors alone. These gradients do not only resolve habitat structural properties, but allow us to address additionally biophysical and process-level information such as canopy chlorophyll (Bicheron & Leroy 1999; Ustin *et al.* 2004), nitrogen and lignin content (Ollinger *et al.* 2002) or growth capacity (Waring *et al.* 2006). This challenge of better linking remote-sensing to underlying ecological relationships must be addressed if species habitat distribution models are to be improved for assessing effects of global change.

The goal of our study was to explore the potential of remote sensing for enhancing the predictive power of habitat distribution models of 19 tree species in Utah, USA. We tested a variety of multitemporal remotely sensed spectral gradients (reflectance of individual bands and surface temperatures) and indices [normalized difference vegetation index (NDVI), as well as wetness, greenness and brightness from tasselled cap transformations] with no loss of information due to classification. Our specific objectives were to explore: (a) to what degree remotely sensed predictors may improve a model calibrated from topographic and bioclimatic predictors; (b) to what degree the variation of a full model can be partitioned into fractions explained by remote sensing predictors alone, topo-climatic predictors alone, and a

joint fraction explained by both predictor sets; and (c) to what degree the addition of remotely sensed predictors affected models of species with different characteristics. We examined these objectives using variation partitioning approaches similar to those described by Borcard, Legendre & Drapeau (1992). Realizing these objectives will shed light on whether habitat distribution models can bridge the gap successfully between more direct, bioclimatic variables, and more indirect variables such as those obtained through remotely sensed information. We expected to find the largest proportion of the variance/deviance explained to be shared between both predictor sets, but with highly variable single contributions of remote sensing predictors among species (see reasoning below).

We also evaluated several hypotheses regarding the partitioned percentage of explained variation of the habitat distribution models in relation to species characteristics. First, we expected early successional species to benefit more from using remotely sensed predictors, because climate variables alone cannot be used to distinguish possible reflectance differences between early and late seral species and thus often predict the temporally more stable late seral species. Secondly, we expected predictive models of the rare species to benefit more from remotely sensed predictors. There are many reasons why a species may be rare, and not all these reasons may be modelled from topo-climatic predictors (Edwards *et al.* 2005). However, remotely sensed predictors may capture subtle differences in either the soil/vegetation characteristics or in the phenology linked with the rare species. Finally, we hypothesized that broadleaf trees might benefit more from adding multitemporal remote sensing predictors due to their more distinct phenology compared to needleleaf trees.

Materials and methods

STUDY AREA

The study area encompasses the forested and mountainous area of the eastern part of the Great Basin of the Interior West, USA. The specific area analysed is the USGS zone 16 (Fig. 1), as defined by the national mapping efforts within the Multi-Resolution Land Characteristics project (Homer & Gallant 2001). Zone 16 is a biogeographical region situated primarily in Utah. It encompasses roughly 6 million hectares of heterogeneous mountainous terrain, including a wide variety of vegetation types ranging from shrubsteppe through forests to alpine communities. We restricted our analysis to the forested parts of zone 16.

DEPENDENT VARIABLES AND SAMPLING

For our analyses we used data from the Forest Inventory and Analysis Program (FIA) of the United States Forest Service, which conducts inventories in forested ecosystems nationwide (<http://www.fia.fs.fed.us/>). A

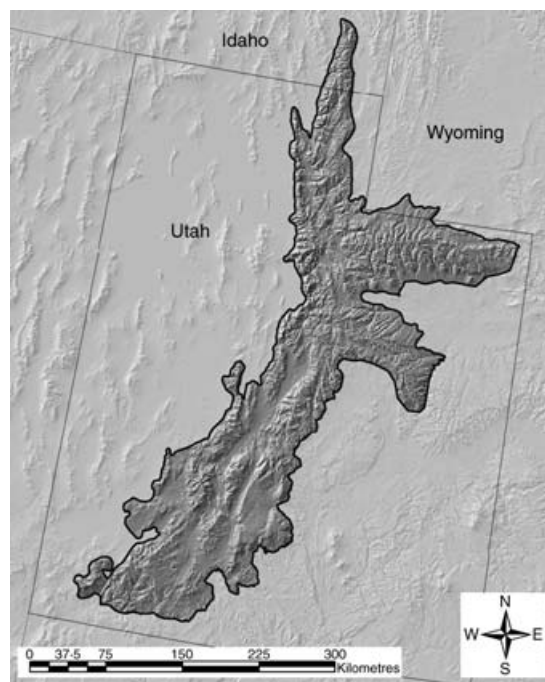


Fig. 1. The study area of zone 16 spans across Utah and stretches into Wyoming and Idaho.

network of sample plots has been established across the country at an intensity of approximately one plot per 2400 hectares, and data collection is conducted under an annual rotating panel system. Of the 3456 plots available in zone 16, only 1941 forested and single-condition plots were used in the analyses. FIA collects extensive stand- and tree-level measurements at each sample plot, which are compiled and combined with stand-level variables to produce plot-level summaries. These summaries include total tree basal area by each of 19 tree species available for our modelling (Table 1). Each sample plot was then characterized as having a tree species present if any basal area occurred, and absent otherwise.

PREDICTOR VARIABLES

We derived two sets of predictors; a first set of topographic and bioclimatic variables and a second set of remote sensing-based variables. For both sets we examined all available variables, finally reducing both sets to eight variables each in order to keep the number of variables equal, reasonably small and easy to interpret (see Appendix S1 in Supplementary material for derivation of these data sets).

The topo-climatic predictor set was generated at a 90 m spatial resolution, originating from a downscaling procedure of the DAYMET 1 km monthly climate maps (<http://www.daymet.org>). These climate maps were developed using the procedures of Thornton, Running & White (1997). The derivation of the final predictor set included a variable reduction approach. For each variable set of 12 monthly maps, we calculated the annual mean and the difference between

Table 1. Tree species used in the modelling analyses. The species characteristics are abbreviated as follows: N = needleleaf, B = broadleaf, E = evergreen, D = deciduous

| Species | Frequency $n_{\text{tot}} = 1941$ | Leaf type | Leaf longevity | Succession type | Core-satellite species type |
|-------------------------------|--------------------------------------|--------------|-------------------|--------------------|--------------------------------|
| <i>Abies concolor</i> | 233 | N | E | Late | Urban |
| <i>Abies lasiocarpa</i> | 429 | N | E | Late | Core |
| <i>Acer glabrum</i> | 16 | B | D | Early | Urban |
| <i>Acer grandidentatum</i> | 119 | B | D | Early | Urban |
| <i>Cercocarpus ledifolius</i> | 147 | B | E | Early | Urban |
| <i>Juniperus osteosperma</i> | 473 | N | E | Late | Core |
| <i>Juniperus scopulorum</i> | 230 | N | E | Early | Satellite |
| <i>Picea engelmannii</i> | 357 | N | E | Late | Core |
| <i>Picea pungens</i> | 25 | N | E | Late | Urban |
| <i>Pinus aristata</i> | 12 | N | E | Late | Urban |
| <i>Pinus contorta</i> | 230 | N | E | Early | Urban |
| <i>Pinus edulis</i> | 405 | N | E | Late | Core |
| <i>Pinus flexilis</i> | 96 | N | E | Late | Satellite |
| <i>Pinus monophylla</i> | 29 | N | E | Late | Satellite |
| <i>Pinus ponderosa</i> | 173 | N | E | Late | Satellite |
| <i>Populus angustifolia</i> | 8 | B | D | Early | Satellite |
| <i>Populus tremuloides</i> | 623 | B | D | Early | Core |
| <i>Pseudotsuga menziesii</i> | 417 | N | E | Late | Core |
| <i>Quercus gambelii</i> | 273 | B | D | Early | Core |

summer and winter climates. The final eight selected topoclimatic predictors included: annual degree-days of growing season using a 0 °C threshold (DDEG), summer-to-winter difference in (i) average daily minimum temperature (TMIN.d) and (ii) relative humidity (RELH.d), yearly means of (i) daily potential global solar radiation (SFMM.y), (ii) relative humidity (RELH.y) and (iii) precipitation sum (PRCP.y), as well as SLP and TOPO. The selection of the final eight predictors was made so that correlations among variables were < 0.7 in order to minimize collinearity problems.

The remote sensing-based predictors were developed using Landsat ETM+ imagery obtained from the USGS Multi-Resolution Land Characteristics consortium of 2001 (<http://www.mrlc.gov/>). Imagery was collected for three different time periods representing the temporal dynamics of vegetation; early (spring), peak (summer) and late growing seasons (autumn). In order to distinguish between predictors of the respective seasons, we use the abbreviation ‘.sp’, ‘.su’ and ‘.au’ hereafter. We prepared bands 1–5 and 7, as well as derived indices of NDVI, of surface temperature, and of tasselled cap transformations originating from a principal component analysis of all seven bands (see Appendix S1 for details). All indices were re-sampled to 90 m, in order to (i) match the spatial resolution of the topo-climatic predictors and (ii) to cover an area that is at least the full spatial extent of the dependent forest inventory plot data. Similar to the topo-climatic predictors, we carried out a variable reduction procedure, stopping after having selected eight final remote sensing-based model parameters: green vegetation index (GVI.su), the wetness index (WI.sp), the normalized difference vegetation index (NDVI.sp., NDVI.au), the soil brightness index (SBI.su), the surface temperature index (T9.su, T9.au) and spring

season band 3 (B3.sp.). With the exception of band 3, only transformed indices remained after the variable selection, and correlations among variables remained < 0.7.

SPECIES CHARACTERISTICS

We selected four attributes for each tree species to test if the partial model contributions of the two predictor sets differed as a function of species characteristics (Table 1). The first two characteristics examined were leaf longevity (evergreen vs. deciduous) and leaf type (broadleaf vs. needleleaf). We hypothesized that a deciduous species is easier to discriminate from all others using multitemporal remote sensing data because its specific phenology can be remotely sensed. In contrast, evergreen species do not exhibit easily detectable phenological characteristics from remotely sensed imagery. We did not evaluate the effects of leaf type independently from leaf longevity because the evergreen broadleaf *Cercocarpus ledifolius* Nutt. in T. & G. was the only species where longevity was not a function of leaf type.

The third characteristic considered was the successional type of the species. We reclassified all tree species into early and late successional status, depending on their behaviour in stand development subsequent to moderate-to-severe disturbance. While most species are clearly of early or late successional type, there are species that may switch their status depending on the environmental conditions. *Pinus contorta* Dougl. ex Loudon is a late seral species on poor volcanic soils, but in Utah it is primarily an early successional species responding to fire. *Quercus gambelii* Nutt. can be early or late, depending on the water availability. We classified it as early in zone 16 forests. Finally, *Populus angustifolia* James is a riparian species occurring on sand

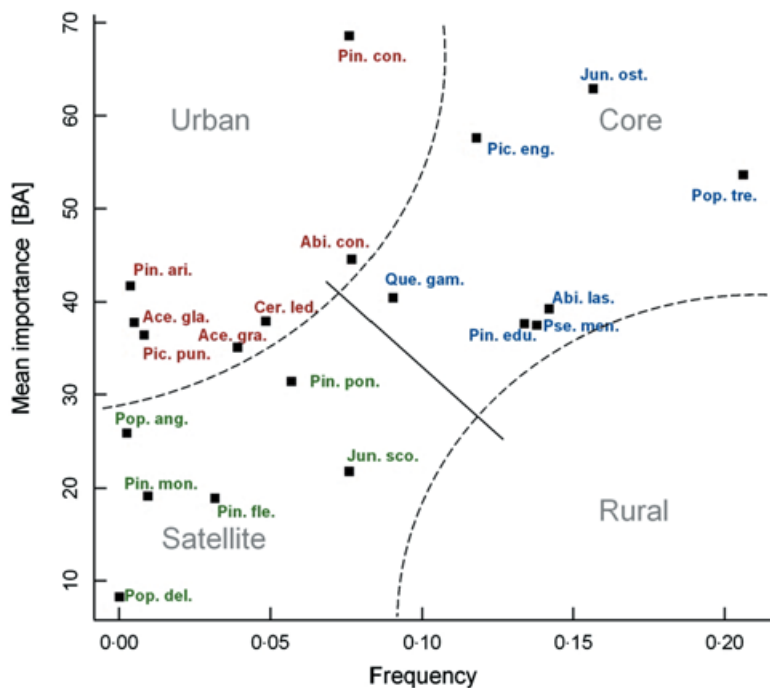


Fig. 2. Allocation of species types based on the extended core-satellite species hypothesis (Hanski 1982; Collins *et al.* 1993). We used mean basal area per plot as importance measure and the frequency among all FIA plots used.

bars and terraces of creeks and streams, and rarely becomes shaded out. Still, we classified it as early.

Last, we categorized the species according to the extended core-satellite hypothesis first proposed by Hanski (1982), and later extended by Collins, Glenn & Roberts (1993), who added urban and rural types. Core species are those with high frequency across the landscape and high abundance per plot, whereas satellite species are rare with low average abundance. Urban species are comparably infrequent, but show high abundance where they occur. Finally, rural species are low in abundance, but occur frequently. To assign species types, we analysed the species frequency of occurrence in the landscape and average abundance (cover) per plot using FIA data (Fig. 2). Because we analysed large gradients across many vegetation types with elevations spanning more than 3000 m of relief, we did not expect any of the tree species to occur everywhere. Thus, we classified core species at much lower frequencies, starting at around 10% of sites occupied.

DATA ANALYSIS

We used generalized linear models (GLMs; McCullagh & Nelder 1989) with logit links to relate the species presences to the topo-climatic and remotely sensed predictor sets. We fitted three regression models per species: two partial models, using (1) only the topo-climatic predictors; (2) only the remote sensing-based predictors; and a full model (3) using the combined topo-climatic and remote sensing-based predictors. Both predictor sets had eight variables each. In each case, we started from a complete model with all variables

included with linear and quadratic powers. Interactions were modelled only for DDEG and SFMM, *y* in order to allow the detection of mixed radiation and thermal energy effects. We then applied stepwise regression procedures, optimizing the models based on the Aikake's information criterion (AIC). The order of the predictors entering the model was according to the deviance explained (D^2 , a measure analogous to the R^2 in ordinary regression) of the predictors when fitted individually, starting with the best predicting variable first. For the full model, we alternated topo-climatic and remotely sensed predictor variables in the same descending order starting with DDEG, the best climatic predictor. The models were fit using the R statistical package (R Development Core Team 2004) (<http://www.r-project.org/>).

To evaluate the model fit, we calculated the adjusted D^2 ($adj.D^2$) following Weisberg (1980). This approach corrects the D^2 (deviance explained) for the number of fitted regression parameters and the number of observations, thus considering the degrees of freedom (Guisan & Zimmermann 2000). It yields similar corrections to the method by Liao & McGee (2003). A 10-fold cross-validation was applied to test the model accuracy. Cross-validation was set up so that (1) the model was fixed and only the parameters were re-adjusted and (2) the original prevalence in the data set was maintained in each fold. Where the number of observed presences was less than 30 for a species, we used jack-knife procedures as a measure of cross-validation (five species, see Table 1). Model accuracy was assessed by two measures, Cohen's kappa (Cohen 1960) and the area under the receiver characteristic curve (ROC) (AUC; Fielding & Bell 1997). We evaluated kappa in 5% steps, determining the threshold where the highest kappa value was obtained for each species.

We employed variation partitioning techniques (Borcard *et al.* 1992) to partition out the individual and joint contribution of both predictor sets relative to the full model. We first calculated the three partial models as discussed above, estimating the $adj.D^2$ for each model type per species. Next, for each species we subtracted the $adj.D^2$ values of (1) the topo-climatic model and (2) the remote sensing-based model from the $adj.D^2$ of the full model, yielding the partial fractions of the full model not included in the remote sensing-based and topo-climatic model, respectively. Finally, we subtracted the sum of the two partial contributions from the full model, which yields the fraction explained jointly by both predictor sets. We used the same method as Lobo, Castro & Moreno (2001). It differs from the method by Borcard *et al.* (1992) in that we did not use the residuals of the partial models for calibrating an additional model for the respective other predictor set, as the logit-transformed residuals in GLMs are not suitable for processing as was performed by Borcard *et al.* (1992) using canonical correspondence analysis. This was pointed out recently by Araujo & Guisan (2006), and further statistical development is necessary to adopt the same procedure.

Table 2. Summary of model fit (deviance explained) and crossvalidated accuracy. Kappa and AUC are derived from the 10-fold cross-validated models, while the model fit was evaluated from the stepwise optimized models. Adjusted D^2 values are listed for the models containing both predictor sets (FULL), the topo-climatic predictors only (CLIM), and the remote sensing-based predictors (RS), respectively. n : the number of observed presences in the data set used in each species-specific model

| Species | n | Kappa FULL | AUC FULL | $adj.D^2$ FULL | $adj.D^2$ CLIM | $adj.D^2$ RS |
|-------------------------------|-----|---------------|-------------|-------------------|-------------------|-----------------|
| <i>Abies concolor</i> | 233 | 0.46 | 0.88 | 0.35 | 0.26 | 0.19 |
| <i>Abies lasiocarpa</i> | 429 | 0.57 | 0.90 | 0.44 | 0.39 | 0.30 |
| <i>Acer glabrum</i> | 16 | 0.14 | 0.80 | 0.47 | 0.31 | 0.25 |
| <i>Acer grandidentatum</i> | 119 | 0.50 | 0.93 | 0.48 | 0.38 | 0.26 |
| <i>Cercocarpus ledifolius</i> | 147 | 0.42 | 0.87 | 0.36 | 0.20 | 0.21 |
| <i>Juniperus osteosperma</i> | 473 | 0.76 | 0.95 | 0.61 | 0.54 | 0.45 |
| <i>Juniperus scopulorum</i> | 230 | 0.35 | 0.83 | 0.26 | 0.19 | 0.14 |
| <i>Picea engelmannii</i> | 357 | 0.71 | 0.95 | 0.60 | 0.56 | 0.41 |
| <i>Picea pungens</i> | 25 | 0.22 | 0.84 | 0.35 | 0.29 | 0.12 |
| <i>Pinus aristata</i> | 12 | 0.23 | 0.88 | 0.56 | 0.32 | 0.32 |
| <i>Pinus contorta</i> | 230 | 0.74 | 0.97 | 0.67 | 0.53 | 0.48 |
| <i>Pinus edulis</i> | 405 | 0.72 | 0.96 | 0.62 | 0.56 | 0.48 |
| <i>Pinus flexilis</i> | 96 | 0.32 | 0.87 | 0.32 | 0.23 | 0.19 |
| <i>Pinus monophylla</i> | 29 | 0.58 | 0.88 | 0.62 | 0.47 | 0.41 |
| <i>Pinus ponderosa</i> | 173 | 0.61 | 0.93 | 0.51 | 0.37 | 0.32 |
| <i>Populus angustifolia</i> | 8 | 0.16 | 0.72 | 0.57 | 0.34 | 0.21 |
| <i>Populus tremuloides</i> | 623 | 0.66 | 0.92 | 0.50 | 0.32 | 0.33 |
| <i>Pseudotsuga menziesii</i> | 417 | 0.49 | 0.86 | 0.32 | 0.24 | 0.16 |
| <i>Quercus gambelii</i> | 273 | 0.63 | 0.94 | 0.52 | 0.39 | 0.25 |
| Average | | 0.49 | 0.89 | 0.48 | 0.36 | 0.29 |

We used Kruskal–Wallis tests to evaluate the effects of species characteristics upon partial and total deviance explained and upon model accuracy. We used the Mann–Whitney test for the two-level leaf and successional types. Linear regression was used to assess the effect of sample size upon partial and total deviance explained as well as upon model accuracy.

Results

MODEL PERFORMANCE

The full models developed from the combined topo-climatic (direct) and remote sensing-based predictors (indirect) had both good model fits ($adj.D^2$) and high accuracies (AUC). When testing the accuracy of the fitted models by a 10-fold cross-validation, we obtained a mean AUC value of 0.89, with values ranging from 0.72 to 0.97 for the species modelled (Table 2). Cross-validated kappa values averaged 0.49, and ranged from 0.14 to 0.76.

Model fit was highest when using both predictor sets, with an average $adj.D^2$ of 0.48 and values ranging from 0.26 and 0.67. Both individual predictor sets alone resulted in lower model fits. The average $adj.D^2$ value of the topo-climatic models was 0.36, which is higher than the average value of 0.29 for the RS-based models. The rank order in model qualities between model types was generally the same for all individual species models, with the exception of *Cercocarpus ledifolius*, *Pinus aristata* Engelm. and *Populus tremuloides* Michx., where both predictor sets reveal comparable model qualities. Adjusted D^2 values differed significantly between all three model types ($P < 0.001$).

The cross-validated accuracy assessments confirmed this trend as well (Fig. 3). When comparing the cross-validated and stepwise optimized model accuracies using AUC, it becomes obvious that the number of observations had an influence on model performance. A minimum of 200 observations was needed to generate comparably stable models (Fig. 3a). Fewer observations resulted in a considerable loss in accuracy when tested by cross-validation. Further, the number of observations had an effect on model accuracy irrespective of predictor set used for the model calibration (Fig. 3b). Models calibrated from remote sensing data showed low cross-validated model accuracies when the number of observed presences was low.

PARTIAL CONTRIBUTIONS OF THE TWO PREDICTOR SETS

The partitioning of the adjusted deviance explained revealed clear differences between the two parameter sets (Table 3). On average, roughly 20% of the overall deviance was explained by the topo-climatic predictors alone (mean = 0.19), and 10% was explained by remote sensing predictors alone (mean = 0.11). A further 20% of the deviance were explained jointly by both predictor sets (mean = 0.18). There were also considerable differences among species, resulting in high standard deviations.

On average, the broad-leaved species showed high percentages explained by the remote sensing predictors, with the exception of *Acer grandidentatum* Nutt. in T. & G. There was also a high correlation between the deviance explained by the full models and the deviance

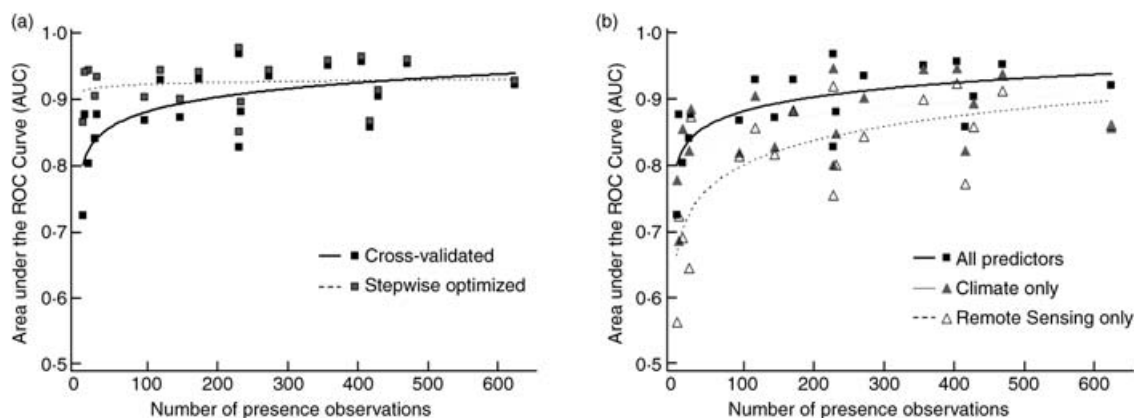


Fig. 3. Model accuracies of all tree species as a function of observed frequencies. (a) AUC of stepwise optimized (open boxes) and additionally cross-validated (closed boxes) models. (b) AUC of cross-validated models calibrated from both predictor sets (closed boxes), from topo-climatic (grey triangles), and from remote sensing-based (open triangles) predictors.

Table 3. Partitioning of the deviance explained by the two predictor sets. The first and the third column list the proportion of deviance explained exclusively by the topo-climatic, and by the remote sensing predictors, respectively. The second column lists the deviance explained jointly by both predictor sets. The total deviance explained represents the *adjusted D*² of the full model

| Species | CLIM alone | CLIM and RS | RS alone | Total expl. | Total unexpl. |
|-------------------------------|------------|-------------|----------|-------------|---------------|
| <i>Abies concolor</i> | 0.16 | 0.11 | 0.09 | 0.35 | 0.65 |
| <i>Abies lasiocarpa</i> | 0.13 | 0.26 | 0.04 | 0.44 | 0.56 |
| <i>Acer glabrum</i> | 0.22 | 0.09 | 0.16 | 0.47 | 0.53 |
| <i>Acer grandidentatum</i> | 0.21 | 0.18 | 0.09 | 0.48 | 0.52 |
| <i>Cercocarpus ledifolius</i> | 0.14 | 0.07 | 0.15 | 0.36 | 0.64 |
| <i>Juniperus osteosperma</i> | 0.16 | 0.38 | 0.07 | 0.61 | 0.39 |
| <i>Juniperus scopulorum</i> | 0.12 | 0.08 | 0.07 | 0.26 | 0.74 |
| <i>Picea engelmannii</i> | 0.19 | 0.37 | 0.04 | 0.60 | 0.40 |
| <i>Picea pungens</i> | 0.23 | 0.07 | 0.05 | 0.35 | 0.65 |
| <i>Pinus aristata</i> | 0.24 | 0.09 | 0.23 | 0.56 | 0.44 |
| <i>Pinus contorta</i> | 0.19 | 0.34 | 0.14 | 0.67 | 0.33 |
| <i>Pinus edulis</i> | 0.14 | 0.43 | 0.05 | 0.62 | 0.38 |
| <i>Pinus flexilis</i> | 0.13 | 0.10 | 0.09 | 0.32 | 0.68 |
| <i>Pinus monophylla</i> | 0.20 | 0.28 | 0.14 | 0.62 | 0.38 |
| <i>Pinus ponderosa</i> | 0.18 | 0.19 | 0.14 | 0.51 | 0.49 |
| <i>Populus angustifolia</i> | 0.36 | -0.02 | 0.23 | 0.57 | 0.43 |
| <i>Populus tremuloides</i> | 0.16 | 0.17 | 0.17 | 0.50 | 0.50 |
| <i>Pseudotsuga menziesii</i> | 0.16 | 0.09 | 0.07 | 0.32 | 0.68 |
| <i>Quercus gambelii</i> | 0.26 | 0.13 | 0.13 | 0.52 | 0.48 |
| Average | 0.19 | 0.18 | 0.11 | 0.48 | 0.52 |
| Standard deviation | 0.06 | 0.13 | 0.06 | 0.12 | 0.12 |

explained jointly by the two predictor sets (Spearman's $r = 0.64$, $P < 0.004$; Fig. 4). *Populus angustifolia* had no joint deviance explained by the two predictor sets, despite its comparably high model accuracy. On the other hand, the *Pinus edulis* Engelm. model had almost the full deviance explained jointly with little unique contribution by the two predictor sets.

SPECIES CHARACTERISTICS

Accuracy as measured by kappa differed significantly among core-satellite species types ($P = 0.032$; Table 4). Core species – the most abundant in the data set – had highest kappa values (Fig. 5a). Sample size also explained significantly variations in kappa tested in a regression ($P < 0.001$). AUC, on the other hand, could be explained only from the number of observations

($P = 0.014$), while core-satellite types did not show significant differences among AUC values of all modelled species. The overall *adj. D*² of all models did not differ between any of the species characteristics analysed. However, the species characteristics did differ significantly in the percentages explained by the two predictor sets used. Remote sensing-based predictors improved the deviance explained significantly more for broadleaf trees than for conifers ($P = 0.03$, Fig. 5c). In contrast, topo-climatic predictors showed significant differences in improving the explained deviance among leaf longevity types ($P = 0.01$), with deciduous trees showing higher gains from this predictor set (Fig. 5d). Finally, we observed a considerable difference among successional types in the deviance explained by the remote sensing-based predictors, although not significant ($P = 0.07$; Fig. 5b).

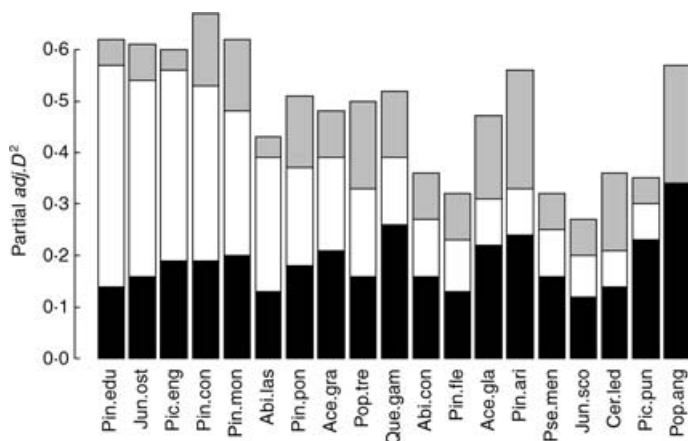


Fig. 4. Partial deviance explained by the two predictor sets for all tree species modelled. Species are ordered by descending fraction of joint adjusted deviance explained ($adj.D^2$) from both predictor sets.

Table 4. Significance levels for the effects of species characteristics upon model accuracy and model fit. The effect of the number of observations (n) on model output was measured by linear models. Leaf type effects were measured by Mann–Whitney tests, whereas the effects of successional and core-satellite types were measured by Kruskal–Wallis test. See Table 3 for a description of the $adj.D^2$ origin and Table 1 for the description of species characteristics

| | n | Leaf type | Leaf longevity | Successional type | Core-satellite type |
|-------------------------------------|-----|-----------|----------------|-------------------|---------------------|
| Model accuracy | | | | | |
| Kappa | *** | ‡ | ‡ | ‡ | * |
| AUC | * | ‡ | ‡ | ‡ | † |
| Model fit and partial contributions | | | | | |
| $adj.D^2$ – total explained | ‡ | ‡ | ‡ | ‡ | ‡ |
| $adj.D^2$ – RS alone | † | * | * | † | ‡ |
| $adj.D^2$ – RS and Clim | * | ‡ | ‡ | ‡ | † |
| $adj.D^2$ – Clim alone | † | † | * | ‡ | ‡ |

*** < 0.001; ** < 0.01; * < 0.05; † < 0.1; ‡ > 0.1.

Discussion

MODEL PERFORMANCE

The high accuracies of the full models based on remotely sensed and topo-climatic predictors indicate that the models are reliable. An evaluation of other studies reveals that our models compare favourably with respect to overall model accuracy and deviance explained. Randin *et al.* (2006) obtained medians for cross-validated kappa values of 0.28 and 0.33, and AUC values of 0.76 and 0.78, respectively, for models of alpine plants in two different areas of the European Alps. Thuiller and coworkers (Thuiller 2003; Thuiller *et al.* 2004) built a range of models for plant and animal species in different areas across Europe, based partly on remotely sensed predictors. For trees, they obtained average kappa and AUC values of 0.66 and 0.94, respectively. McKenzie *et al.* (2003) modelled 14 tree species in the Pacific North-west using three different tree data sets, using predictors and tree species similar to ours. Model D^2 values ranged from 0.11 to 0.51 (mean = 0.30), and AUC values from 0.71 to 0.96 (mean = 0.85).

It is obvious that the rare species models, although yielding high $adj.D^2$ values, were less accurate (Fig. 5). However, even though prevalence generally influenced model accuracy and $adj.D^2$, some rare species such as *Pinus monophylla* Torr. & Frem. in Frem. still yielded high model accuracies. The drop in model accuracy when $n < 200$ observed presences is in agreement with work by Pearce & Ferrier (2000), who observed a severe drop in model accuracy for sample sizes below 250 observations. McPherson *et al.* (2004) also documented a clear drop in the accuracy of rare bird species models, as assessed by kappa and AUC, for sample sizes below 300. In addition, model accuracy dropped in a second

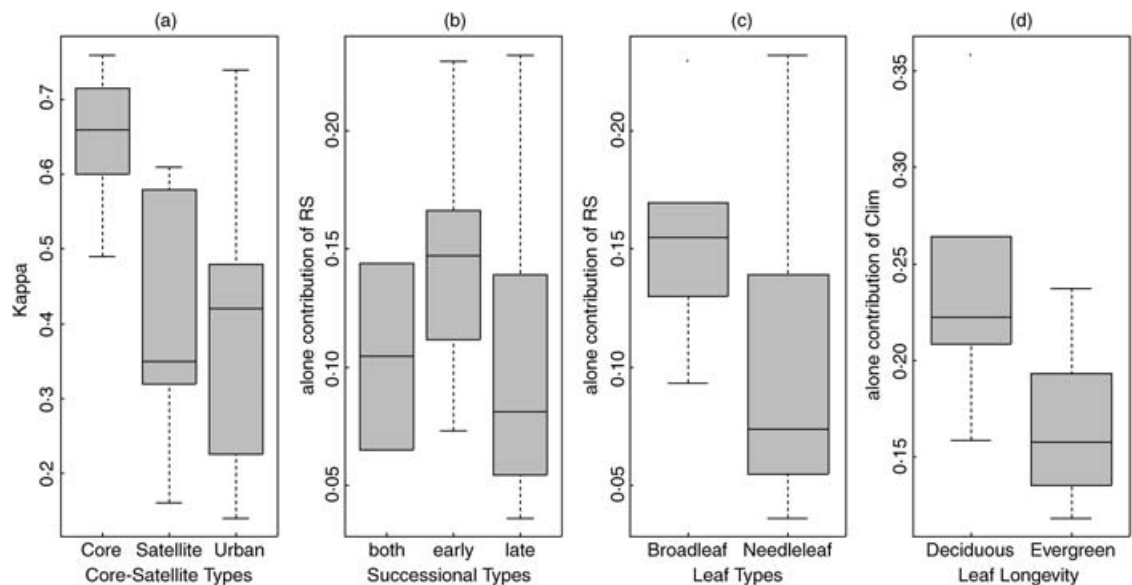


Fig. 5. Linkages between species characteristics and model accuracy and fit. (a) Core-satellite types significantly differ in model accuracy. (b) Remote sensing-based predictors and successional types. (c) Remote sensing-based predictors increase model fit for broadleaf trees more than for conifers. (d) Topo-climatic predictors add more to model fit of deciduous than to evergreen trees. See Table 4 for significance tests. Box and whisker boundaries represent quartiles.

simulation study originating from the same bird data set when the prevalence was either below 10% or above 90% (if the presence-absence classification error is below 1%). However, Stockwell & Peterson (2002) tested a range of statistical models against sample size in 103 abundant Mexican birds and observed a clear drop in average accuracy only if models were calibrated from less than 50 random samples.

REMOTE SENSING AND PARTIAL CONTRIBUTIONS OF PREDICTOR SETS

Several studies included remotely sensed information for predictive distribution modelling. Thuiller *et al.* (2004) investigated the extent to which the remotely sensed PELCOM land cover classification improved the predictive power when added to bioclimatic predictors in models for a range of taxonomic groups. They found that remotely sensed predictors, although clearly improving the fit of individual species' models, did not further improve the cross-validated accuracy of the models.

These findings are in agreement with most of our results. First, *adj. D*² increased significantly when remotely sensed predictors were added to the topo-climatic models (~+25% over the topo-climatic models, $P < 0.01$), with some species clearly benefiting more than others. However, when evaluating the models, we obtained only a weak and non-significant increase in predictive cross-validated accuracy. Secondly, the models built from remotely sensed predictors alone revealed a reasonable fit, but were less accurate than those based solely on topo-climatic predictors. We interpret this as an indication that land cover patterns are highly correlated with bioclimatic gradients; thus, both predictor sets are expected to provide similar prediction accuracies. Thirdly, not all species share the same environmental requirements. Remote sensing addresses primarily vegetation structural and biomass or productivity related properties. We agree with the conclusion by Pearson *et al.* (2004) that the remotely sensed habitat information helps to discriminate between suitable and unsuitable sites that cannot be distinguished from bioclimatic layers alone. However, the addition of remote sensing predictors adds additional noise to the species-environment relationships as measured by bioclimatic variables, perhaps because similar structural features (as seen from satellites) may arise under different topo-climatic conditions.

When comparing the partial contributions of topo-climatic and remotely sensed predictors to the models, we observed somewhat large differences among species. As a general observation, the contribution of remotely sensed predictors towards overall model fit decreased as the overall model fit increased (Fig. 4). The exceptions include primarily rare species (*Acer glabrum* Torr., *Picea pungens* Engelm., *Pinus aristata*, *Populus angustifolia*). When excluding these four species, the Spearman's rho increases to 0.85 ($P < 0.001$). Thus, the models with lower fit gained considerably more from the remotely sensed predictors.

SPECIES CHARACTERISTICS

The extended core and satellite species types partly met our expectations, with core species having the best model accuracies, followed by the urban species. Our core species are not as frequent in the landscape as Collins *et al.* (1993) require for species to be classified in this type, due to the large environmental gradients found in our study region. However, if species are very abundant in a landscape, it may be difficult also to reach high model accuracies (see McPherson *et al.* 2004). The urban species have high variability in cross-validated accuracies. We expected urban species to be similar to core species. Among the seven urban species present in our data set, three have very low prevalence (*Acer glabrum*, *Picea pungens*, *Pinus aristata*). Without these three species, the average kappa value is clearly higher. The satellite species were difficult to model. They are generally rare (low prevalence) and they are often not present where the topo-climatic conditions seem appropriate, making it difficult to model accurately the realized niche of these species.

There are many reasons why a species may be rare (or common), some of which can be captured by predictive models, while others cannot. Edwards *et al.* (2005) hypothesized that among the three species types that are either rare and/or present only with low average abundance (i.e. urban, satellite and rural), only the urban species would be comparably easy to model. We did not have data on rural species, and thus could not test this idea unequivocally; nor did we see a clear difference between the urban and satellite tree species. However, as noted above, we did see a difference between the latter two types if we omitted the very rare species in our analysis. Urban species with a sample size larger than 30 (prevalence > 1.5%) provided reasonable model fits. Thus, we conclude that the hypothesis is partly supported, but it clearly needs to be extended with requirements regarding sample size to generate accurate models (McPherson *et al.* 2004; Edwards *et al.* 2005), as well as better understanding of sample design effects on model performance (Edwards *et al.* 2006).

As hypothesized, early successional and broadleaf, deciduous trees benefited more from adding multitemporal remote sensing predictors than did late seral and needleleaf species. This means that temporal sequences detect more effectively the strong phenological signal in leaf longevity and an improved distinction in spatial pattern of deciduous from evergreen trees. This effect may not be obtained if only data from a single season were used. The effect was less significant for early successional species. Nevertheless, the multitemporal images allowed us to recognize more clearly early successional species that usually have similar climatic requirements as other seral species. The effects of prevalence and sample size upon model accuracy have often been tested in the past. However, we need to explore further the additional effects of species characteristics upon model behaviour if we are to make significant progress in the development of niche-based models in ecology and their application to global change scenarios.

Remotely sensed variables proved to be good predictors of the distribution of tree species in our study area. We expect remote sensing predictors to be even more important if mixed vegetation communities other than only forests are combined in such analyses. The benefit of adding remotely sensed predictors was especially high when applied to rare species, and to species that have low model accuracies (urban and rural species). On one hand, it means that species and biodiversity conservation actions to detect and manage rare and occasional (satellite) species clearly benefit from adding remote sensing predictors to predictive habitat distribution models. On the other hand, considerable efforts are required to develop a better theoretical and conceptual understanding of the ecological meaning of the spectral response from vegetation in order to maximize the benefit from using remote sensing information, which may require the development of new ecological paradigms (Ustin, Smith & Adams 1993).

Acknowledgements

Portions of this research were funded by the 5th and 6th Framework Programme of the European Union (Contract Numbers EVK2-CT-2002-00136 and GOCE-CT-2003-505376), and by the Swiss Federal Research Institute WSL. We thank three anonymous reviewers for helpful comments on an earlier version of this manuscript and Dave Roberts for discussions.

References

- Araujo, M.B. & Guisan, A. (2006) Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, **33**, 1677–1688.
- Araujo, M.B., Thuiller, W., Williams, P.H. & Reginster, I. (2005) Downscaling European species atlas distributions to a finer resolution: implications for conservation planning. *Global Ecology and Biogeography*, **14**, 17–30.
- Austin, M.P. (1980) Searching for a models for use in vegetation analysis. *Vegetatio*, **42**, 11–21.
- Austin, M.P. (1987) Models for the analysis of species' response to environmental gradients. *Vegetatio*, **69**, 35–45.
- Austin, M.P. (2002a) Case studies of the use of environmental gradients in vegetation and fauna modelling: theory and practice in Australia and New Zealand. *Predicting Species Occurrences: Issues of Accuracy and Scale* (eds J.M. Scott, P.J. Heglund, F. Samson, J. Haufler, M. Morrison, M. Raphael & B. Wall), pp. 73–82. Island Press, Covelo, CA.
- Austin, M.P. (2002b) Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling*, **157**, 101–118.
- Austin, M.P. & Nicholls, A.O. (1997) To fix or not to fix the species limits, that is the ecological question: response to Jari Oksanen. *Journal of Vegetation Science*, **8**, 743–748.
- Austin, M.P. & Smith, T.M. (1989) A new model for the continuum concept. *Vegetatio*, **83**, 35–47.
- Bicheron, P. & Leroy, M. (1999) A method of biophysical parameter retrieval at global scale by inversion of a vegetation reflectance model. *Remote Sensing of Environment*, **67**, 251–266.
- Bomhard, B., Richardson, D.M., Donaldson, J.S., Hughes, G.O., Midgley, G.F., Raimondo, D.C., Rebelo, A.G., Rouget, M. & Thuiller, W. (2005) Potential impacts of future land use and climate change on the Red List status of the *Proteaceae* in the Cape Floristic Region, South Africa. *Global Change Biology*, **11**, 1452–1468.
- Borcard, D., Legendre, P. & Drapeau, P. (1992) Partialling out the spatial component of ecological variation. *Ecology*, **73**, 1045–1055.
- Chefaoui, R.M., Hortal, J. & Lobo, J.M. (2005) Potential distribution modelling, niche characterization and conservation status assessment using GIS tools: a case study of Iberian copris species. *Biological Conservation*, **122**, 327–338.
- Cohen, J. (1960) A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, **20**, 37–46.
- Collins, S.L., Glenn, S.M. & Roberts, D.W. (1993) The hierarchical continuum concept. *Journal of Vegetation Science*, **4**, 149–156.
- Dirnböck, T., Dullinger, S. & Grabherr, G. (2003) A regional impact assessment of climate and land-use change on alpine vegetation. *Journal of Biogeography*, **30**, 401–417.
- Edwards, T.C. Jr, Cutler, D.R., Zimmermann, N.E., Geiser, L. & Moisen, G.G. (2006) Effects of underlying sample survey designs on the utility of classification tree models in ecology. *Ecological Modelling*, **199**, 132–141.
- Edwards, T.C. Jr, Cutler, D.R., Geiser, L., Alegria, J. & McKenzie, D. (2004) Assessing rarity of species with low detectability: lichens in Pacific Northwest forests. *Ecological Applications*, **14**, 414–424.
- Edwards, T.C. Jr, Cutler, D.R., Zimmermann, N.E., Geiser, L. & Alegria, J. (2005) Model-based stratifications for enhancing the detection of rare ecological events. *Ecology*, **86**, 1081–1090.
- Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R., Soberón, J., Williams, S., Wisz, M.S. & Zimmermann, N.E. (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.
- Engler, R., Guisan, A. & Rechsteiner, L. (2004) An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. *Journal of Applied Ecology*, **41**, 263–274.
- Ferrier, S., Powell, G.V.N., Richardson, K.S., Manion, G., Overton, J.M., Allnutt, T.F., Cameron, S.E., Mantle, K., Burgess, N.D., Faith, D.P., Lamoreux, J.F., Kier, G., Hijmans, R.J., Funk, V.A., Cassis, G.A., Fisher, B.L., Flemons, P., Lees, D., Lovett, J.C. & Van Rompaey, R. (2004) Mapping more of terrestrial biodiversity for global conservation assessment. *Bioscience*, **54**, 1101–1109.
- Fielding, A.H. & Bell, J.F. (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, **24**, 38–49.
- Guisan, A., Broennimann, O., Engler, R., Vust, M., Yoccoz, N.G., Lehmann, A. & Zimmermann, N.E. (2006) Using niche-based models to improve the sampling of rare species. *Conservation Biology*, **20**, 501–511.
- Guisan, A., Theurillat, J.P. & Kienast, F. (1998) Predicting the potential distribution of plant species in an Alpine environment. *Journal of Vegetation Science*, **9**, 65–74.
- Guisan, A., Theurillat, J.-P. & Spichiger, R. (1995) Effects of climate change on alpine plant diversity and distribution: the modeling and monitoring perspectives. *Potential Ecological Impacts of Climate Change in the Alps and Fennoscandian Mountains* (eds A. Guisan, J.I. Holten, R. Spichiger & L. Tessier), pp. 129–135. Edition Conservatoire et Jardin Botanique, Geneva, Switzerland.
- 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.
- Hanski, I. (1982) Dynamics of regional distribution – the core and satellite species hypothesis. *Oikos*, **38**, 210–221.
- Heikkinen, R.K. (1996) Predicting patterns of vascular plant species richness with composite variables: a meso-scale study in Finnish Lapland. *Vegetatio*, **126**, 151–165.

- Homer, C. & Gallant, A. (2001) *Partitioning the Conterminous United States Into Mapping Zones for Landsat TM Land Cover Mapping*. p. 7. USGS Draft White Paper. Available at: <http://www.mrlc.gov/>, accessed 27 June 2007.
- Hutchinson, G.E. (1957) Concluding remarks. *Cold Spring Harbour Symposium on Quantitative Biology*, **22**, 415–427.
- Iverson, L.R. & Prasad, A.M. (1998) Predicting abundance of 80 tree species following climate change in the eastern United States. *Ecological Monographs*, **68**, 465–485.
- Leathwick, J.R. (1998) Are New Zealand's *Nothofagus* species in equilibrium with their environment? *Journal of Vegetation Science*, **9**, 719–732.
- Liao, J.G. & McGee, D. (2003) Adjusted Coefficients of Determination for logistic regression. *American Statistician*, **57**, 161–165.
- Lobo, J.M., Castro, I. & Moreno, J.C. (2001) Spatial and environmental determinants of vascular plant species richness distribution in the Iberian Peninsula and Balearic Islands. *Biological Journal of the Linnean Society*, **73**, 233–253.
- Manel, S., Buckton, S.T. & Ormerod, S.J. (2000) Testing large-scale hypotheses using surveys: the effects of land use on the habitats, invertebrates and birds of Himalayan rivers. *Journal of Applied Ecology*, **37**, 756–770.
- McCullagh, P. & Nelder, J.A. (1989) *Generalized Linear Models*, 2nd edn. Chapman & Hall, London, UK.
- McKenzie, D., Peterson, D.W., Peterson, D.L. & Thornton, P.E. (2003) Climatic and biophysical controls on conifer species distributions in mountain forests of Washington State, USA. *Journal of Biogeography*, **30**, 1093–1108.
- McPherson, J.M., Jetz, W. & Rogers, D.J. (2004) The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact? *Journal of Applied Ecology*, **41**, 811–823.
- Oksanen, J. & Minchin, P.R. (2002) Continuum theory revisited: what shape are species responses along ecological gradients? *Ecological Modelling*, **157**, 119–129.
- Ollinger, S.V., Smith, M.L., Martin, M.E., Hallett, R.A., Goodale, C.L. & Aber, J.D. (2002) Regional variation in foliar chemistry and N cycling among forests of diverse history and composition. *Ecology*, **83**, 339–355.
- Parra, J.L., Graham, C.C. & Freile, J.F. (2004) Evaluating alternative data sets for ecological niche models of birds in the Andes. *Ecography*, **27**, 350–360.
- Pearce, J. & Ferrier, S. (2000) An evaluation of alternative algorithms for fitting species distributions models using logistic regression. *Ecological Modelling*, **128**, 127–147.
- Pearson, R.G., Dawson, T.P. & Liu, C. (2004) Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data. *Ecography*, **27**, 285–298.
- Peterson, A.T., Ortega-Huerta, M.A., Bartley, J., Sanchez-Cordero, V., Soberon, J., Buddemeier, R.H. & Stockwell, D.R.B. (2002) Future projections for Mexican faunas under global climate change scenarios. *Nature*, **416**, 626–629.
- Prentice, I.C., Cramer, W., Harrison, S.P., Leemans, R., Monserud, R.A. & Solomon, A.M. (1992) A global biome model based on plant physiology and dominance, soil properties and climate. *Journal of Biogeography*, **19**, 117–134.
- R Development Core Team (2004) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Randin, C.F., Dirnböck, T., Dullinger, S., Zimmermann, N.E., Zappa, M. & Guisan, A. (2006) Are species distribution models transferable in space? *Journal of Biogeography*, **33**, 1689–1703.
- Stockwell, D.R.B. & Peterson, A.T. (2002) Effects of sample size on accuracy of species distribution models. *Ecological Modelling*, **148**, 1–13.
- Thornton, P.E., Running, S.W. & White, M.A. (1997) Generating surfaces of daily meteorological variables over large regions of complex terrain. *Journal of Hydrology*, **190**, 214–251.
- 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., Araujo, M.B. & Lavorel, S. (2004) Do we need land-cover data to model species distributions in Europe? *Journal of Biogeography*, **31**, 353–361.
- Thuiller, W., Broennimann, O., Hughes, G.O., Alkemade, J.R.M., Midgley, G.F. & Corsi, F. (2006) Vulnerability of African mammals to anthropogenic climate change under conservative land transformation assumptions. *Global Change Biology*, **12**, 424–440.
- Thuiller, W., Richardson, D.M., Pysek, P., Midgley, G.F., Hughes, G.O. & Rouget, M. (2005) Niche-based modelling as a tool for predicting the risk of alien plant invasions at a global scale. *Global Change Biology*, **11**, 2234–2250.
- Ustin, S.L., Roberts, D.A., Gamon, J.A., Asner, G.P. & Green, R.O. (2004) Using imaging spectroscopy to study ecosystem processes and properties. *Bioscience*, **54**, 523–534.
- Ustin, S.L., Smith, M.O. & Adams, J.B. (1993) Remote sensing of ecological processes: a strategy for developing and testing ecological models using spectral mixture analysis. In: *Scaling Physiological Processes: Leaf to Globe* (eds J.R. Ehleringer & C.B. Field), pp. 339–357. Academic Press, San Diego, CA, USA.
- Venier, L.A., Pearce, J., McKee, J.E., McKenney, D.W. & Niemi, G.J. (2004) Climate and satellite-derived land cover for predicting breeding bird distribution in the Great Lakes Basin. *Journal of Biogeography*, **31**, 315–331.
- Vetaas, O.R. (2002) Realized and potential climate niches: a comparison of four rhododendron tree species. *Journal of Biogeography*, **29**, 545–554.
- Vitousek, P.M. (1994) Beyond global warming: ecology and global change. *Ecology*, **75**, 1861–1876.
- Waring, R.H., Milner, K.S., Jolly, W.M., Phillips, L. & McWethy, D. (2006) Assessment of site index and forest growth capacity across the Pacific and Inland Northwest USA with a MODIS satellite-derived vegetation index. *Forest Ecology and Management*, **228**, 285–291.
- Weisberg, S. (1980) *Applied Linear Regression*. Wiley, New York, NY.
- Welch, N.E. & MacMahon, J.A. (2005) Identifying habitat variables important to the rare Columbia spotted frog in Utah (USA): an information-theoretic approach. *Conservation Biology*, **19**, 473–481.
- Yee, T.W. & Mitchell, N.D. (1991) Generalized additive models in plant ecology. *Journal of Vegetation Science*, **2**, 587–602.
- Zimmermann, N.E. & Kienast, F. (1999) Predictive mapping of alpine grasslands in Switzerland: species versus community approach. *Journal of Vegetation Science*, **10**, 469–482.

Received 5 August 2006; revised version accepted 25 April 2007
 Editor: Jack Lennon

Supplementary material

The following supplementary material is available for this article.

Appendix S1. Derivation of predictor variables.

This material is available as part of the online article from: <http://www.blackwell-synergy.com/doi/full/10.1111/j.1365-2664.2007.01348.x>
 (This link will take you to the article abstract).

Please note: Blackwell Publishing is not responsible for the content or functionality of any supplementary material supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.