Journal of Vegetation Science 14: 669-680, 2003

Generalized models vs. classification tree analysis: Predicting spatial distributions of plant species at different scales

Thuiller, Wilfried*; Araújo, Miguel B. & Lavorel, Sandra

*Centre d'Ecologie Fonctionnelle et Evolutive, Centre National de Recherche Scientifique, 1919 route de Mende, F-34293 Montpellier Cedex 5, France; *Corresponding author; E-mail wilfried.thuiller@cefe.cnrs-mop.fr*

Abstract. Statistical models of the realized niche of species are increasingly used, but systematic comparisons of alternative methods are still limited. In particular, only few studies have explored the effect of scale in model outputs. In this paper, we investigate the predictive ability of three statistical methods (generalized linear models, generalized additive models and classification tree analysis) using species distribution data at three scales: fine (Catalonia), intermediate (Portugal) and coarse (Europe). Four Mediterranean tree species were modelled for comparison. Variables selected by models were relatively consistent across scales and the predictive accuracy of models varied only slightly. However, there were slight differences in the performance of methods. Classification tree analysis had a lower accuracy than the generalized methods, especially at finer scales. The performance of generalized linear models also increased with scale. At the fine scale GLM with linear terms showed better accuracy than GLM with quadratic and polynomial terms. This is probably because distributions at finer scales represent a linear sub-sample of entire realized niches of species. In contrast to GLM, the performance of GAM was constant across scales being more data-oriented. The predictive accuracy of GAM was always at least equal to other techniques, suggesting that this modelling approach is more robust to variations of scale because it can deal with any response shape.

Keywords: Accuracy assessment; Climate variable; Generalized additive model; Generalized linear model; Mediterranean tree species; Receiver operating characteristics curve.

Abbreviations: GLM = Generalized Linear Model; GAM = Generalized Additive Model; CTA = Classification Tree Analysis; ROC curve = Receiver Operating Characteristic curve; AUC = Area Under the Curve.

Introduction

Modelling plant distributions in relation to environmental factors has gained momentum with recent developments in the fields of GIS and statistical techniques applied to ecological analysis (e.g. Guisan & Theurillat 2000). When they include climate data, these models can be used to make projections of species' distributional changes in response to climate change (Bakkenes et al. 2002; Huntley et al. 1995; Midgley et al. 2003).

One of the common approaches in plant bioclimatic modelling is the statistical fit of correlative models, where plant responses are fitted by regression to some environmental variables. Some of these methods are based on the idea of bell-shaped species responses along environmental gradients as used in regression and ordination-based vegetation response models (e.g. Austin 1985; ter Braak 1987). However, several authors (e.g. Austin & Gaywood 1994; Huisman et al. 1993) have shown that asymmetric and other complex response curves are very frequent. Methods to overcome this problem include generalized linear models (GLM), with skewed response curves fitted by a β-function or third-order cubic polynomial terms (Austin et al. 1994), generalized additive models (GAMs) (Yee & Mitchell 1991), regression and classification tree analyses (RTA and CTA) (Breiman et al. 1984) and other rule-based methods such as neural networks (Olden & Jackson 2002; Pearson et al. 2002) or cellular automata (Carey 1996), for which no assumptions are made regarding species response curves to environmental gradients.

There are few studies comparing the predictive accuracy of statistical and rule-based methods (but see Bio et al. 1998; Franklin 1998; Pearce & Ferrier 2000b; Vayssières et al. 2000; Vetaas 2000; Walker & Cocks 1991). Guisan et al. (1999) compared the predictive accuracy of GLM and canonical correspondence analysis (CCA), and showed that GLM provided better models for single species while CCA provided a broader overview of multiple species assemblages. These results provide an interesting but still limited exploration of which method would perform best for different goals

and types of data. One of the problems with these analyses is that results are very much dependent upon the mathematical procedure used. There are often several procedures to fit the same model.

Model results are also dependent on the adequacy and quality of environmental data. To provide good predictions of current species distribution, caution must be given to the selection of environmental variables for analysis. By order of preference, it is advisable to select fewer relevant climatic variables that are suitable surrogates of direct physiological parameters of plants (Woodward 1987). In addition to these, soil type, slope and elevation are often considered as good indirect variables to predict plant distributions, even if their link with physiological mechanisms is more complex (Burke 2001; Nicholls 1989). The relative importance of environmental factors to determine plant species distributions also varies according to spatial scale (Huston 1994). At the community and landscape scales direct physiological parameters, species dispersal abilities, biotic interactions and local disturbances control plant distributions. At continental and global scales, macroclimate is the major determinant of plant species distributions (Woodward 1987, 1990). At intermediate scales, human impact or management can be particularly important, but few studies have included such variables because of the paucity of adequate digital data sets. Indeed, many regression models have been fitted to site data with a variety of predictors, including climate, land use or intensity of disturbance, but without concern for regional or global prediction. When projections to other sites are intended, complete coverage of the predictors for the region is required but rarely available. Finally, at any scale, plant distribution models are mainly dependent on the quality, adequacy and resolution of the environmental data.

The aim of this study was to compare the predictive accuracy of one familiar parametric approach, generalized linear models (GLM), and two non-parametric approaches that have recently gained popularity in ecology, generalized additive models (GAM) and classification tree analysis (CTA). The comparison was performed using three independent data sets representing three scales and resolutions: fine (Catalonia), intermediate (Portugal) and coarse (Europe). Bioclimatic models were calculated for four common tree species and a range of environmental variables to address the following questions: 1. How do the different methods respond to increases in scale? 2. Which method is more adequate to fine, intermediate or coarse scales? 3. Which are the best environmental variables for a given scale? 4.What methodological framework might be more robust and accurate to generate projections of current plant distributions in Europe?

Data and Methods

Data sets

Study species

We selected four Mediterranean species for analysis: *Juniperus oxycedrus*, *Pinus pinaster*, *Pinus pinea* and *Quercus suber*. The choice of common species took into consideration the number of records per species in our databases, as both very rare and very common species are bound to be more difficult to model.

J. oxycedrus is a dioecious tree reaching 14 m, but is usually encountered as a shrub of oval habit. It occurs between 0 and 1400 m a.s.l. and is also found in more continental locations of Europe and W Asia.

P. pinaster is a larger single-stemmed tree (30-40 m), mainly found in the W Mediterranean. In continental Europe it occurs from 0 to 400 m a.s.l., but reaches up to 900 m in Corsica and to 2000 m in Morocco.

P. pinea is a shorter pine tree (12-25 m) often identified with Mediterranean landscapes. It is probably original to the Iberian Peninsula, but an archaeophyte throughout the Mediterranean region (Barbéro et al. 1998).

Q. suber occurs all over the Mediterranean in open woodlands on hills and lower slopes, generally on acidic soils. It is grown commercially for its thick cork bark for which Portugal is the world's largest producer.

Fine-scale data from Catalonia

Species data: These are from a database of the Forestry Inventory of Catalonia (IEFC) managed by the Centre for Ecological Research and Applied Forestry in Barcelona (CREAF). Sampling areas have a circular shape usually with a radius of 10 m.

Environmental data: A climate model (Ninyerola et al. 2000) was used to generate layers of monthly precipitation and monthly mean, minimum and maximum temperatures for the entire study area (Table 1). This model used a network of weather stations in Catalonia (257 stations for precipitation and 160 stations for temperatures). From these data, a multiple regression analysis was performed between the meteorological variables and a set of geographic variables (latitude, solar radiation, altitude, distance to the coast and cloudiness factors) derived from a 180-m resolution Digital Elevation Model (DEM). Accuracy of the climate surfaces was assessed and corrected with reference to an independent set of weather stations (40% of the initial weather stations). Layers of climatic variables (Table 1) were then derived from the original climate model. Topographic layers (elevation, slope and solar radiation) were provided by a DEM (180 m resolution). A geology layer was provided by the Institute of Cartography of Catalonia (1 : 250000).

The data were used in raster format in ArcView (Vers. 3.2a for Windows, ESRI Corp.) with a resolution of 180 m. We combined raster environmental data and vegetation point data to determine the value of each environmental variable in each vegetation plot (Table 1).

Intermediate-scale data from Portugal

Species data: Data were compiled from an ongoing database project at the University of Évora (www.cea.uevora.pt/umc). Presence-absence data were referenced onto 993 UTM 10 km × 10 km grid cells.

Environmental data: Climatic and geomorphologic data (Table 1) were compiled from available digital layers of the 'Atlas do Ambiente' (http://www.dga.pt). Data were converted from available vector maps and re-sampled to 10 km × 10 km grid maps, using zonal functions in ArcView to calculate the mode of each variable in each cell (Table 1).

Coarse scale from Europe

Species data: Species presence data were a subset of the Atlas Florae Europaeae (AFE) (Jalas & Suominen 1972-1996) digitized by Lahti & Lampinen (1999). Data are referenced onto 4419 UTM 50 km × 50 km grid cells. This grid offers mapping units across Europe.

Environmental data: Seven environmental variables were selected and converted from 0.5° latitude-longitude maps to UTM grid cells (Table 1): mean annual precipitation (Legates & Wilmott 1990a); mean temperature in January and July (Legates & Wilmott 1990b); mean annual potential evapotranspiration and mean potential evapotranspiration in January and July (Ahn & Tateishi 1994) and altitude, obtained from the United Nations Environmental Program grid. Temperature and precipitation were extracted from the National Climatic Data Center (NOAA). Data were re-sampled with the GIS package IDRISI (Eastman 1996).

Comparability of environmental data

Environmental variables used for analysis are not entirely comparable across the three scales (Table 1). However, they can be grouped according to shared impacts on plant growth, development and potential distribution. Following Austin & Smith (1989), we distinguished two types of variables. First, direct variables that have direct physiological impacts on plant growth. Examples include temperature, precipitation or radiation. The three variables were available in the Catalonian environmental data set, while only precipitation and temperature were available within the Portuguese and European environmental data sets. In addition, the Catalonian data contained not only mean annual precipitation (MAP), like the others, but also seasonal precipitation. Mean of the coldest and warmest months (MTC and MTW, respectively) were available in the Catalonian and European data, but not in the Portuguese data, which provided mean annual temperature (MAT) instead. This little difference, in which direct gradients were quantified for each data set made it possible to compare the significance of variables that were either common to all data sets (seasonal precipitation, MTC, MTW) or were exclusive to some of them (MAP, MAT). Second, indirect variables are those such as topography and geology, both of which were represented in our three environmental databases. Environmental data from Catalonia contained Mean Elevation, Slope and Geology. Portuguese data contained four variables related to elevation (Table 1), three variables related to slope (Table 1) and geology (Table 1), while Europe contained only Mean Elevation. Hence the three environmental data contained similar variables that could be related to each other for comparisons.

Table 1. Environmental variables used to construct the different models. prec. = precipitation; temp. = temperature; evap. = evapotranspiration; SD = standard deviation. Cat. = Catalonia; Port. = Portugal; Eur. = Europe.

		Cat.	Port.	Eur.
Mean annual prec.	MAP	X	X	X
Mean summer prec.	SumP	X		
Mean winter prec.	WinP	X		
Mean Autumn prec.	AutP	X		
Mean Spring prec.	SprP	X		
Mean Annual temp.	MAT	X	X	
Mean temp. coldest month	MTC	X		X
Mean temp. warmest month	MTW	X		X
Mean annual max. temp.	MAMxT	X		
Mean max temp. hottest month	MMTW	X		
Mean annual min. temp.	MAMnT	X		
Mean min. temp. coldest month	MMTC	X		
Mean annual radiation	MAR	X		
Annual potential evap.	PET			X
Potential evap. coldest month	PETC			X
Potential evap.warmest month	PETW			X
Elevation	Elev	X	X	X
Max. elevation	MxElev		X	
Min. elevation	MnElev		X	
SD elevation	SdElev		X	
Slope	Slope	X	X	
SD slope	SdSlope		X	
Variation slope coefficient	VCoSlope		X	
Geology	Geol	X		

Statistical models

Statistical analyses were performed using Splus (Vers. 5 for Windows, MathSoft Inc.), with standard functions (*gam* for generalized additive models, *glm* for generalized linear models and *tree* for classification tree analysis) and some custom functions. Species distribution maps were created with ArcView and its Spatial

Analyst extension (Vers. 3.2a for Windows, ESRI Corp.). In order to evaluate the quality of predictions, we divided databases into two subsets: calibration and evaluation. The first, a random sample from 70% of total database, was used to calibrate the models, whereas the second, comprising of the remaining data, was used to evaluate model predictions (Fielding & Bell 1997).

Generalized Linear Models

GLMs are often used to describe the relationship between species and their environment (see Guisan & Zimmermann 2000). GLMs provide a less restrictive form than classic multiple regressions by providing error distributions for the dependent variable other than normal and non-constant variance functions (McCullagh & Nelder 1989). If the response with a predictor variable is not linear, then a transformation can be included; polynomial terms are allowed for the simulation of skewed and bimodal responses (Guisan et al. 1999), β-functions (Austin & Gaywood 1994) or a hierarchical set of models (Huisman et al. 1993). The nature of the relationship between species and environmental gradients has to be known *a priori*.

We developed two kinds of GLM. First, we examined the relationship between the species presence-absence response and environmental predictors. We then performed a simple GLM with linear terms allowing for the possibility of interactions. A complex GLM was performed with the same parameters, to which were added second and third order terms and polynomial terms (second and third orders). In this case, we did not consider interactions between quadratic and polynomial terms because their biological significance would be difficult to interpret.

To select for the most parsimonious model, we used an automatic stepwise model selection using the Akaike Information Criterion (AIC) (Chambers & Hastie 1997). This procedure allowed the removal of redundancy in variables and eliminated multicollinearity problems, even if principal correspondence analysis on environmental data sets did not show strong patterns of correlations.

Generalized Additive Models

GAMs are designed to capitalize on the strengths of GLMs without requiring the problematic steps of postulating a response curve shape or specific parametric response function. They use a class of equations called 'smoothers' that attempt to generalize data into smooth curves by local fitting to subsections of the data. GAMs are therefore useful when the relationship between the variables is expected to be of a more complex form, not easily fitted by standard linear or non-linear models, or where there is no *a priori* reason for using a particular model (Hastie & Tibshirani 1990).

We used a cubic spline smoother, which is a collection of polynomials of degree less than or equal to 3, defined on subintervals. A separate polynomial is fitted for each neighbourhood, thus enabling the fitted curve to join all of the points. The degree of smoothness was automatically selected by cross-validation and restricted to 4. Similarly to GLM, we used an automated stepwise process (the *step.gam* function in Splus) to select the most significant variables for each species.

Regression and Classification Trees

These provide an alternative to regression techniques (e.g. Vayssières et al. 2000; Thuiller et al. 2003). They do not rely on *a priori* hypotheses about the relation between independent and dependent variables. This method consists of recursive partitions of the dimensional space defined by the predictors into groups that are as homogeneous as possible in terms of response. The tree is built by repeatedly splitting the data, defined by a simple rule based on a single explanatory variable. At each split, the data are partitioned into two exclusive groups, each of which is as homogeneous as possible (Thuiller 2003). To control the length of the tree, we use the *prune* function of Splus (Vers. 5 for Windows, MathSoft Inc.). The program builds a nested sequence of subtrees by recursively snipping off the less important splits in terms of explained deviance. The length of the tree was controlled by choosing the best trade-off between explained deviance and tree size. Each predictor could be used several times in the tree if it improved the predictive performance.

Predicting assessment accuracy

To compare different models from different species using records from different databases, we were reluctant to use an accuracy index dependent on any fixed threshold. Instead we used the Receiver Operating Characteristic curve (ROC curve) that is not dependent on the threshold. The use of the ROC curve method is still in its infancy in ecology (Pearce & Ferrier 2000b; Thuiller et al. 2003) as compared to the more widely used κ or Classification Accuracy (e.g. Franklin 1998; Guisan et al. 1998; Huntley 1995). The ROC curve is a graphical method that represents the relation between the False Positive fraction (1 – specificity) and the sensitivity for a range of thresholds. If all predictions were possibly expected by chance, the relation would be a 45° line. Good model performance is characterized by a curve that maximizes sensitivity for low values of (1-specificity), i.e. when the curve passes close to the upper left corner of the plot. The area between the 45° line and the curve measures discrimination, that is, the ability of the model to correctly classify a species as present or absent in a given plot (area under the curve: AUC). A rough

guide for classifying the accuracy of a diagnostic test is the traditional academic point system (Swets 1988): 0.90-1.00 = excellent; 0.80-0.90 = good; 0.70-0.80 = fair; 0.60-0.70 = poor; 0.50-0.60 = fail.

The difference between the areas under ROC curves generated by two or more models provides a measure of comparative discrimination ability of these models when applied to independent evaluation data. In our study, the significance of the difference between two ROC curves (AUC1 and AUC2) can be calculated as a critical ratio test (Hanley & McNeil 1983):

$$Z = \frac{AUC1 \pm AUC2}{\sqrt{Se_{AUC1} + Se_{AUCC2} * Se_{AUC1} Se_{AUC2}}} \qquad (1)$$

Se is the standard error and *r* is the correlation between both areas under the curve. *r* is the mean of the correlation between predictions from both models for the positive events and the correlation between predictions for the negative events, calculated using the Spearman rank correlation coefficient. To project species distributions into binary presence-absence form, we used a probability threshold maximising the percentage of presence and absence correctly predicted.

To address our questions we have articulated the results in two parts, within-scale and across-scale.

Results

Within-scale comparisons

Fine-scale Catalonia

At this scale all models provided generally good results for all species (Table 2). The highest model accuracies were observed for *P. pinea* and *Q. suber* (AUC > 0.9), while slightly lower but still high accuracies were obtained for *J. oxycedrus* and *P. pinaster* (AUC > 0.82 and AUC > 0.85 respectively). *Q. suber* was the species best predicted, with on average 94% of occurrences correctly predicted, while for *J. oxycedrus* only 76% of the occurrences were correctly predicted. For *J. oxycedrus* and *P. pinaster* GLM 'simple' and GAM had significantly higher discrimination ability ($p < 0.05$) than GLM 'complex' and CTA whereas for *P. pinea* and *Q. suber* all types of general models (GLM 'simple', 'complex' and GAM) were significantly better than CTA ($p < 0.05$).

The most significant environmental variables were related to winter climate (coldest month mean temperature and winter precipitation) for *Pinus* and *Q. suber*, but for *J. oxycedrus* the most important variables were related to summer and annual precipitation.

Some differences were found between the discrimination abilities of species. The distribution of *Q. suber* was well simulated, but *J. oxycedrus* and *P. pinaster* were poorly predicted. As an illustration, we plotted the predicted species distribution of *Q. suber* according to GAM (one of the highest AUC values) and CTA (the lowest AUC value) (Fig. 1). Both predicted distributions were similar, slightly over-predicting the distribution of *Q. suber* in the south. The northern distribution margin was better predicted by GAM than by CTA. For example, CTA predicts two presences in northern Catalonia, where *Q. suber* has not been recorded.

A trend of accuracy of models is discernible from the analyses. CTA models had significantly lower discrimination ability than generalized models at this scale. GLM 'simple' and GLM 'complex' had similar results showing that more complex models do not always perform better than simple ones. Ease of interpretation and parsimony would therefore make GLM 'simple' the best model choice at the fine scale.

Intermediate-scale Portugal

For this data set and scale there were strong disparities between models for individual species and these were not consistently related to the type of method used (Table 3). *J. oxycedrus* was the best-predicted species (mean AUC = 0.90), whereas models for *Q. suber* were not accurate (mean AUC = 0.635). Both species of *Pinus* had a fairly accurate prediction (0.741 < AUC < 0.804) across all models. There were no significant differences among models for any species except for *P. pinea,* where GLM 'simple' had a better accuracy than GLM 'complex' and GAM, and for *P. pinaster* where GAM had a better accuracy than GLM 'complex'.

Relevant environmental variables for *J. oxycedrus* and *P. pinaster* were related to annual precipitation, while elevation explained most of the variation for *P. pinea*. For *Q. suber*, different models tended to select inconsistent sets of variables.

We selected the distributions predicted by GLM 'complex' and GAM. *J. oxycedrus* has a aggregated distribution and is restricted to NE Portugal (Fig. 2). Both GLM 'complex' and GAM over-predicted the observed distribution of this species. The major difference between these two models was that GAM had a lower rate of false positive, over-predicting less than GLM 'complex'. The fact that both models over-predicted the distribution of *J. oxycedrus* in the same areas could reflect the potential distribution of the species, which may not yet have colonized this part of Portugal or may be restricted to the northeast by human activities.

Hence there was no general trend in accuracy among models, and no model seemed to be better than others at this scale.

Species	Model	Selected environmental variables	Evaluation		
			se	sp	AUC
Juniperus	GLMs	MAP, SdSlope, Geol, MAP:SdSlope	93.3	92.8	**0.939**
oxycedrus	GLMs	AugP, Geol, MAP, MAP:SumP	76.3	76.3	**0.871** *
	GLMc	pol(AugP,3), pol(MAP,2), Geol, SumP^3	76.3	76.3	**0.859**
	GAM	s(AugP), s(MAP),Geol, s(SumP)	79.0	78.3	**0.865** *
	CTA	AugP, MTC, MMTW	65.8	86.3	**0.813**
Pinus	GLMs	MTC, SprP, WinP, Elev	82.3	82.2	**0.869** *
pinaster	GLMc	pol(MTC,2), WinP^3, pol(SumP,2), pol(MMTC,3)	82.3	81.9	**0.862**
	GAM	s(MTC), s(WinP), s(MAMxT), s(SprP)	82.3	82.6	**0.866** *
	CTA	MTC, WinP, Elev, SumP	83.5	77.1	**0.816**
Pinus	GLMs	MTC, SumP:AugP, MAP:SumP, AugP	83.3	83.2	**0.904** *
pinea	GLMc	MTC^3, pol(SumP,3), MAMxT^3, pol(MMTC,3)	84.1	84.0	**0.911** *
	GAM	s(MTC), s(SprP), s(MMTC), s(SprP)	84.0	84.0	**0.911** *
	CTA	MTC, WinP, MAMxT, MAP	87.1	77.1	**0.880**
Quercus	GLMs	Geol, MTC, WinP, WinP:MTC	95.4	95.5	**0.988** **
suber	GLMc	pol(MTC,3), Geol, pol(WinP,3), MAMnT	94.5	94.6	**0.986** **
	GAM	s(MTC), Geol, s(MAMnT), s(WinP),	94.1	94	**0.985** *
	CTA	MTC, Geol, WinP, MMTW	91.6	93.6	**0.952**
Mean over	GLMs		84.3	84.3	**0.908**
species	GLMc		84.3	84.2	**0.905**
	GAM		84.9	84.7	**0.907**
	CTA		82.0	83.5	**0.865**

Table 2. Prediction accuracy on evaluation data of the different models for each species in the Catalonia data set: The four most important selected environmental variables influencing are presented in order of decreasing deviance explained. Se = sensitivity; sp = specificity (%). * = model significantly better than other models at $p = 0.05$; ** $p = 0.01$. GLMc = GLM 'complex'; GLMs = GLM 'simple'. E.g. pol(MAP,2) generate a matrix of orthonormal polynomials; (MAP+ MAP^2). SumP^3 = variable summer precipi- tation at third degree; MAP: SumP = interaction between mean annual and summer precipitation.

Species	Model	Selected environmental variables	Evaluation		
			se	sp	AUC
Juniperus	GLMs	MAP, SdSlope, Geol, MAP:SdSlope	93.3	92.8	**0.939**
oxycedrus	GLMc	pol(MAP,2), SdSlope^2, pol(MnElev,2), pol(MxElev,3)	80	80	**0.929**
	GAM	s(SdSlope), s(MAP), s(Elev), s(MAT)	86.7	86.7	**0.941**
	CTA	MxElev, MAP, Slope, SdElev	80	90	**0.936**
Pinus	GLMs	MAP, MAP:SdSlope, Geol, SdSlope	70.3	70.5	**0.787**
pinaster	GLMc	pol(MAP,3), pol(VCoElev,3), Geol, pol(SdSlope,2)	71	71.2	**0.782**
	GAM	s(MAP), s(VcoElev), s(Elev), Geol	73.9	73.7	**0.804** *
	CTA	MAP, VCoElev, Elev, MAT	73.9	71.2	**0.785**
Pinus	GLMs	Geol, Elev, SdElev, MAP	70.8	70.3	**0.766****
pinea	GLMc	pol(Elev,3), Geol, SdElev, MAP	70.8	70.7	**0.745**
	GAM	s(Elev), Geol, s(SdElev), s(MAP)	70.8	70.3	**0.748**
	CTA	Elev, Sd.Slope, MAT, SdElev	73.6	70.7	**0.741**
Quercus	GLMs	SdSlope, Elev, MAP:SdSlope, MxElev	60.3	60.2	**0.630**
suber	GLMc	pol(MAP,3), pol(Elev,2), pol(MnElev,2), Elev^2	60.3	60.2	**0.651**
	GAM	s(Elev), s(MAP), s(MnElev), s(VcoElev)	57.4	58	**0.635**
	CTA	VCoElev, Elev, Geol, Slope	47.1	67.3	**0.628**
Mean over	GLMs		73.7	73.5	**0.781**
species	GLMc		70.5	70.5	**0.777**
	GAM		72.2	72.2	**0.782**
	CTA		68.7	74.8	**0.773**

Table 3. Prediction accuracy on evaluation data of the different models for each species in the Portugal data set: The four most important selected environmental variables influencing are presented in order of decreasing deviance explained. See further Table 2.

Species	Model	Selected environmental variables	Evaluation		
			se	sp	AUC
Juniperus	GLMs	PET, MTW:PET, PETC, Elev	87	87.2	**0.936**
oxycedrus	GLMc	pol(PET,3), pol(MTW,3), pol(Elev,3), pol(PETC,2)	88	88.1	**0.941**
	GAM	s(MTW), s(Elev), s(MTC), s(PETW)	88	88.1	**0.937**
	CTA	PET, PETW, MTW, Elev	87	86.8	**0.927**
Pinus	GLMs	MTC, Elev, PETC, MTC:MTW	87.2	87.8	**0.935**
pinaster	GLMc	pol(PET,3), pol(MTW,3), pol (MTC,3), pol(PET,3)	85.1	86	**0.947**
	GAM	s(MTC), s(Elev), s(MTW), s(PET)	87.2	86.5	**0.952** *
	CTA	PET, MTC, MTW, Elev	87.2	87.8	**0.926**
Pinus	GLMs	PET, MTC, MTC:PET, MTC:PETC	82.4	82.8	**0.901**
pinea	GLMc	pol(PET,3), MTW^3, pol(MTW,2), pol(MTC,3)	85.3	85.4	**0.931** *
	GAM	s(MTC), s(PETW), s(PETC)	82.4	83.6	**0.923** *
	CTA	PET, MTC, MTW, PETW	85.3	84.7	**0.879**
Quercus	GLMs	MTC, PET, PETC, MTC:PET	88.7	89	**0.941**
suber	GLMc	pol(PET,3), pol(MTC), pol(MTW), pol(Elev,3)	90.6	90.6	**0.960** *
	GAM	s(MTC), s(MTW), s(PET)	90.6	90.6	**0.962** *
	CTA	PET, MTC, MTW, Elev	92.5	87.9	**0.945**
Mean over	GLMs		86.3	86.7	**0.928**
species	GLMc		87.3	87.5	**0.945**
	GAM		87.1	87.2	**0.944**
	CTA		88.0	86.8	**0.919**

Table 4. Prediction accuracy on evaluation data of the different models for each species in the European data set: The four most important selected environmental variables influencing are presented in order of decreasing deviance explained. See further Table 2

Coarse-scale Europe

Models at the European scale had generally high AUC values (Table 4). For *Q. suber* they were very high (0.945 < AUC < 0.962), and GLM 'complex' and GAM were significantly better than CTA and GLM 'simple'. The same pattern was found for *P. pinea* (0.879 <AUC < 0.931). GAM was also better than GLM 'simple' for *P. pinaster* (0.926 < AUC < 0.952), but there were no significant differences between models for *J. oxycedrus* (0.927 < AUC < 0.936). The two variables most frequently selected by the models were mean temperature of the coldest month and annual potential evapotranspiration. Elevation was important for *J. oxycedrus* and *P. pinaster*, whereas mean temperature of the warmest month was selected in second or third position for *P. pinea* and *Q. suber*.

Though differences between models tended to be small at this scale, two groups could be distinguished: GAM and GLM 'complex' had a better discrimination ability and accuracy (AUC = 0.944 and AUC = 0.945 respectively) than GLM 'simple' and CTA. The predictive ability of different types of models was similar across species although these covered a wide range of occurrence values (*J. oxycedrus*: 326; *P. pinaster*: 157; *P. pinea*: 114; *Q. suber*: 183 occurrence records on 2209 available grid cells) and had rather different geographical distributions. We plotted the predicted distribution of *Q. suber* by GLM 'simple' and GAM models (Fig. 3). Both models over-predicted its actual distribution in Spain. However, the errors generated by the two models showed different spatial patterns. GLM 'simple' over-predicted species distribution in Greece while GAM was closer to reality. Inversely, GAM over-predicted the distribution in southwestern France while GLM reproduced almost perfectly the observed distribution.

Across-scale comparisons

Models

A comparison of models across scales revealed some distinctive patterns in model performance among species. At the finest scale GLM 'simple' and GAM provided better discrimination ability than GLM 'complex' and CTA for all species. In contrast, at the intermediate scale there was no best model (except GLM 'simple' for *P. pinea*). At the largest scale GLM 'complex' and GAM discriminated better and GLM 'simple' seemed to be the least accurate. Hence, differences in performance between GLM with linear terms (and interaction) and GLM with linear (and interaction), quadratic and polynomial terms appeared to be scale-dependent. At the fine scale GLM 'simple' performed better. At the intermediate scale there were no major differences between the two, and at the coarse scale GLM 'complex' was

Fig. 1. Observed and predicted distribution of *Quercus suber* in Catalonia. **A.** Observed distribution; **B.** Predicted distribution using classification tree analysis; **C.** Predicted distribution using generalized additive models.

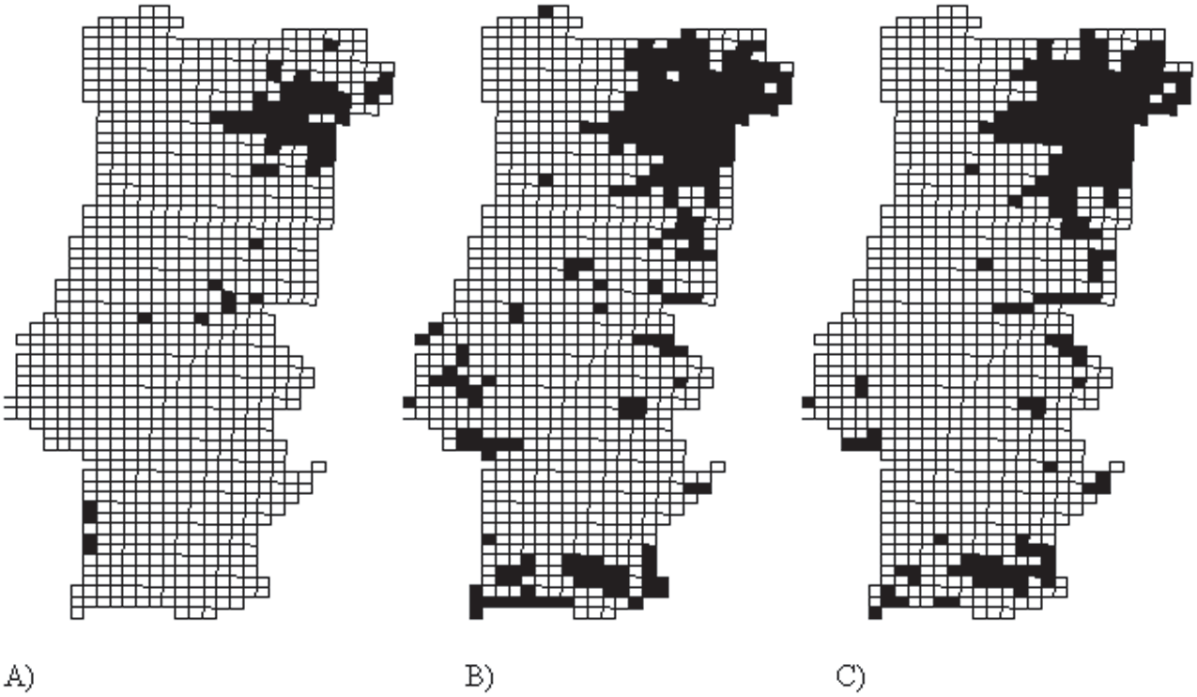

Fig. 2. Observed and predicted distribution of *Juniperus oxycedrus* in Portugal. **A.** Observed distribution; **B.** Predicted distribution using GLM 'complex'; **C.** Predicted distribution using GAM.

better. It appears that GLM requires increasing complexity as the scale of analysis increases, which could be interpreted in terms of species response shape to gradient. Being data-oriented and not influenced by a parametric choice, GAM is useful to explore response shapes (Oksanen & Minchin 2002). The response shape of *Quercus suber* to temperature of the coldest month in Catalonia showed a linear pattern (not presented). It is therefore logical that GLM 'simple', which uses only linear terms, provides as accurate a prediction as GLM 'complex' for this scale. Inversely, if we look at the coarsest scale, the response shape of *Quercus suber* to temperature of the warmest month was approximately a bell-shaped-curve (not presented). In this case, GLM with simple linear term cannot accurately fit the actual species response shape.

Generally, with our data, CTA performed poorly, but their discrimination ability increased with scale. At a finer scale, CTA had the lowest accuracy. At the intermediate scale CTA was nearly equal to the other methods, and at coarser scales CTA was similar to GLM 'simple'. The discrimination ability of GAM did not seem to be affected by changes in scale, suggesting that these kinds of models are relatively robust to scaling.

Species and environmental data

Although specific environmental variables differed across the three databases, they were still useful in allowing simple comparisons to be drawn.

First, selected variables by models differed markedly between Catalonia/Europe and Portugal. Indeed, at fine and coarse scales, climatic variables were always selected as the main determinants of the distribution, whereas in Portugal, geologic and topographic variables were selected as being the most important. Although environmental data were highly related to topography in Portugal, if mean annual temperature and mean annual precipitation were the most important variables, models should have selected them as they did, for example, for *Pinus pinaster*.

In contrast, models for Catalonia and Europe were consistent in the variables they selected. Mean temperature of the coldest month was selected as the main factor for *P. pinaster*, *P. pinea* and *Q. suber* at both scales. In contrast, the variables selected by the models for *J. oxycedrus* were quite different for Catalonia, where precipitation was the most important variable, and Portugal or Europe where mean temperature of the warmest month was the key factor. For the European wide resolution, annual precipitation was never selected as one of the main factors determining species distributions, while mean of the coldest month and annual potential evapotranspiration were the most frequently selected variables.

Discussion

Effects of scale on model selection

The discriminative ability and predictive accuracy of our four modelling techniques showed different patterns as a function of the scale in the data. GAM provided fairly constant results across scales. A strong feature is that GAM is non-parametric and can take any smooth shape (Oksanen & Minchin 2002). This is particularly important as species response shapes are expected to change with scale, particularly for species with wide distributions (e.g. *Quercus suber*). In Catalonia, *Q. suber* has ranges that extend beyond the study area (Fig. 2) and thus the shape of response to a large climate gradient is likely to be monotonic as it only partially reflects the complete realized niche of the species. Hence, at this scale, GLM including only linear terms can be suitable to fit species response shapes. Conversely, at coarse scales, the shape of the response function describes a broader proportion of the realized niche. As shown here for *Q. suber*, and demonstrated by others, the shape of species response is likely to be Gaussian or asymmetric but not linear (Austin & Gaywood 1994; Austin & Meyers 1996; Huisman et al. 1993; Oksanen & Minchin 2002). GLM including more complex functions than linear are more suitable to fit species response shapes as the proportion of the realized niche included in the study increases.

Classification trees provide a different framework to model and predict species responses to environmental gradients. Our analyses showed that at a fine scale CTA had a lower accuracy than at intermediate and coarse scales when compared to generalized models. However, minimizing deviance and maximizing prediction accuracy may be conflicting goals. Indeed CTAs never performed better than generalized models at any of the scales, even if CTA yielded the lowest deviance (Franklin 1998). This lower power of CTA models may be attributable to the lack of interactive effects among environmental variables (Hastie & Tibshirani 1990). The increasing power of CTA with increasing scale may reflect the emergence of more complex relationships among variables. Nevertheless CTA remains an accurate method that makes it possible to describe how species respond to environmental variables, and to summarize species-environment relationships in a readable way.

We analysed real data to compare the models. Hence we cannot know the absolute performance of selected methods because 'truth' remains unknown. To test for absolute performance we could have utilized artificial data. However, results would be too much dependent on the assumptions used to generate the data and if they were unrealistic, then results could also be misleading.

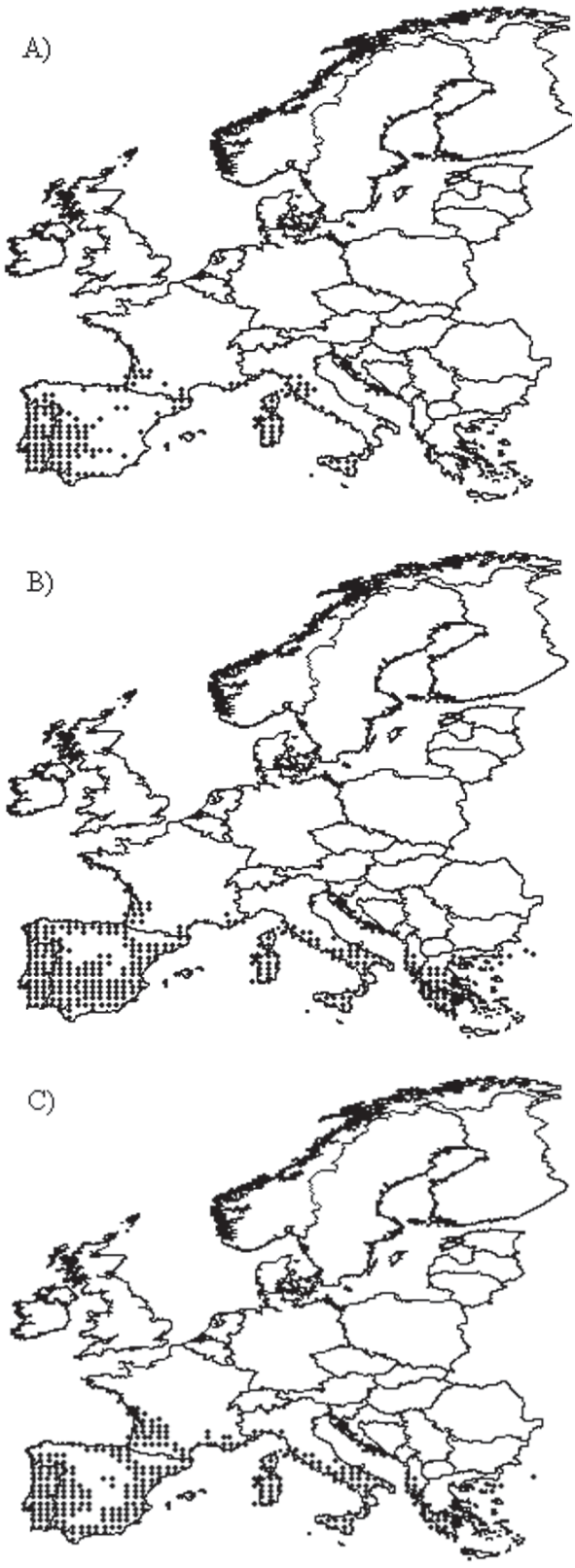

Fig. 3. Observed and predicted distribution of *Quercus suber* in Europe. **A.** Observed distribution; **B.** Predicted distribution using GLM 'simple'; **C.** Predicted distribution using GAM.

Factors influencing model fitting

There are many other parameters that influence differences in model adequacy across scales. First this study has shown that species with compact distributions are more easily modelled than species with scattered distributions. This phenomenon has already been documented (Araújo & Williams 2000) and relates to the previous point concerning the amplitude of gradients captured by models. Indeed, even at a fine scale, a species with a compact distribution is likely to have a greater proportion of its range within the sampled area. It is therefore appropriate to use such models to examine the shape of the response function describing the realized niche or predict range changes (Franklin 1998). Here, at fine and intermediate scales, species with compact distributions were well predicted by all models *(Q. suber* in Catalonia and *J. oxycedrus* in Portugal) and it is easy to explain why GLM 'simple' and GLM 'complex' had similar discrimination abilities.

Second, the assumption of correlative methods that species distributions are at equilibrium with their environment is a questionable one (Guisan & Zimmermann 2000). Non-equilibrium distributions resulting from history, biotic interactions, disturbances, random perturbations and human interferences can make predictions hazardous (Bolliger et al. 2000). Yet, impacts of forest management and other human activities, which do influence current plant distributions, remain hard to evaluate (Duckworth et al. 2000; Schulze & Kunz 1995). For instance, in Catalonia *Pinus pinea* and *P. pinaster* are intensively planted and their distributions are likely not to be at equilibrium with current environmental conditions, possibly explaining why models explained their distributions so poorly (Thuiller et al. in press). However, Bolliger et al. (2000) argues that at the regional scale correlative models are robust in predicting species even if they are not exactly at equilibrium. They also show that equilibrium models performed as well as dynamical models including physiological parameters and life history in predicting regional patterns of forest composition. This is still an unresolved debate.

Relevant environmental variables

Several authors have demonstrated the importance of climate on plant distributions at regional and continental scales (Woodward 1987, 1990), and that these are the result of specific physiological mechanisms (Woodward & Williams 1987). For example, annual or monthly minimum temperatures limit plant distributions by exceeding species lethal threshold for survival, while precipitation or water balance affect plant growth through control on leaf mass. In the Catalonian and European databases, the main determinants of species distribution conformed to this model, and the mean temperature of the coldest month was a key explanatory variable for three of four species at both scales. In addition, seasonal precipitation was a better predictor of species distributions in Catalonia than mean annual precipitation (MAP), while in Europe mean potential evapotranspiration was a better predictor. Large intra-annual heterogeneity in rainfall in Europe and particularly in the Mediterranean is the likely reason why seasonal variables are more relevant to plant growth than MAP.

In contrast, the availability of climatic data for Portugal was limited to MAP and mean annual temperature (MAT), but contrary to the other data sets many other indirect variables such as topographic features were available. Models almost never selected MAT, whereas MAP was selected for all four species. Instead, several indirect variables such as slope and elevation were selected. The generally poor fits obtained suggest that plant distributions cannot be predicted accurately using a collection of indirect variables and only few simple climatic variables. Even when indirect variables are important locally, they may be applied only within a limited geographical extent without significant error, because in a different region the same topographic position may correspond to a different combination of direct and resource factors (Guisan & Zimmermann 2000). As a consequence, models based on indirect environmental variables have limited predictive power and cannot be used to project distributions elsewhere, or to model range shifts according to global change scenarios.

Alternatively, the poor accuracy of models at the intermediate scale (i.e. the Portuguese data set) may have resulted from the influence of human impacts on plant distributions. Using a database comprising 39 additional variables related to land use and land cover (http://www.dga.pt) we calculated new models (not presented) for the four species. Results were significantly better for *Pinus pinaster* (AUC evaluation = 0.85) and *Juniperus oxycedrus* (AUC evaluation = 0.96) only, suggesting that at this particular scale and for this data, species distributions are influenced by a combination of climatic, topographic and land use variables.

At the fine and intermediate scales, competitive interactions could also explain part of the spatial distribution of species. For example, Leathwick & Austin (2001) showed the importance of competition in explaining the spatial structure of forests in New Zealand. They argued that the abundance of *Nothofagus* species decrease the abundance of other conifer and broad-leaved tree species. They propose a framework, using GAM models and incorporating abundance of *Nothofagus* as independent variable to predict conifer and broad-leaved tree species. In our study, competitive

interactions were not directly taken into account. However, it is likely that the inclusion of information for other species as independent variables could improve the fit of the models. Arguably, however, it would be difficult to address competitive interactions of trees at the European scale using a resolution of 50 km × 50 km. Most plants occurring together in such grid cells will not compete with each other.

Extrapolating current and future species distributions

GLM 'complex' and GAM models were the most robust and accurate methods to interpolate current plant distributions at a coarse scale (Europe). Some authors have used non-parametric methods to predict future distributions from three bioclimatic variables (mean temperature of the coldest month, growing degree-days and ratio actual evapotranspiration/ potential evapotranspiration, including local weighted regression (loess) (Huntley et al. 1995; Beerling 1993) and cubic splines (Flannigan & Woodward 1994). Although robust non-parametric methods (GAM, loess) are data dependent and therefore lack underlying theoretical models, their applicability beyond the initial range for which they were developed remains questionable (but see Beerling et al. 1995). Alternatively, a framework using GAM to identify and test current response functions, and then deriving explicit functions to parameterize a GLM (Yee & Mitchell 1991), could be applied to extrapolate current species distributions to other areas or into the future under global change scenarios.

An important problem could arise in attempting to project future distributions, as the realized niche of species is dependent on abiotic factors, but also on biotic factors and particularly on competitive interactions. In the future, the realized niches may change as a result of competition from different species. However, if the modelling objective is to project future distribution of particular species or communities during the early stages of climate change, and if the site is dominated by competitors familiar to the target species, then the realized niche incorporating the average effects of interspecific competition across species' range may be used to a first approximation to future distribution.

Acknowledgements. This research is funded by the European Union (Advanced Terrestrial Ecosystem Analysis and Modelling project EVK2-CT-2000-00075. Species and environmental data for Portugal were kindly supplied by Pedro Segurado, University of Évora; The Centre for Ecological Research and Applied Forestry in Barcelona kindly supplied the species and climatic data for Catalonia. We thank Lluis Brotons, Mathieu Rouget, Guy Midgley, and two anonymous referees for their useful comments on the manuscript.

References

Ahn, C.H. & Tateishi, R. 1994. Development of a global 30-minute grid potential evapotranspiration data set. *J. Jpn. Soc. Photogram. Rem. Sens.* 33: 12-21.

Araújo, M.B. & Williams, P. 2000. Selecting areas for species persistence using occurrence data. *Biol. Conserv.* 96: 331-345.

Austin, M.P. 1985. Continuum concept, ordination methods, and niche theory. *Annu. Rev. Ecol. Syst.* 16: 39-61.

Austin, M.P. & Gaywood, M.J. 1994. Current problems of environmental gradients and species response curves in relation to continuum theory. *J. Veg. Sci.* 5: 473-482.

Austin, M.P. & Meyers, J.A. 1996. Current approaches to modelling the environmental niche of eucalypts: implication for management of forest biodiversity. *For. Ecol. Manage.* 85: 95-106.

Austin, M.P. & Smith, T.M. 1989. A new model for the continuum concept. *Vegetatio* 83: 35-47.

Austin, M.P., Nicholls, A.O., Doherty, M.D. & Meyers, J.A. 1994. Determining species response functions to an environmental gradient by means of a beta-function. *J. Veg. Sci.* 5: 215-228.

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

Barbéro, M., Loisel, R., Quézel, P., Richardson, D.M. & Romane, F. 1998. Pines of the Mediterranean Basin. In: Richardson, D.M. (ed.) *Ecology and biogeography of Pinus*, pp. 450-473. Cambridge University Press, Cambridge, UK.

Beerling, D. J. 1993. The impact of temperature on the northern distribution limits of the introduced species *Fallopia japonica* and *Impatiens glandulifera* in north-west Europe. *J. Biogeogr.* 20: 45-53.

Beerling, D.J., Huntley, B. & Bailey, J.P. 1995. Climate and the distribution of *Fallopia japonica*: use of an introduced species to test the predictive capacity of response surface. *J. Veg. Sci.* 6: 269-282.

Bio, A.M.F., Alkemade, R. & Barendregt, A. 1998. Determining alternative models for vegetation response analysis: a non-parametric approach. *J. Veg. Sci.* 9: 5-16.

Bolliger, J., Kienast, F. & Bugmann, H. 2000. Comparing models for distributions: concept, structures, and behavior. *Ecol. Model.* 134: 89-102.

Breiman, L., Friedman, J.H., Olshen, R.A. & Stone, C.J. 1984. *Classification and regression trees*. Chapman and Hall, New York, NY, US.

Burke, A. 2001. Classification and ordination of plant communities of the Naukluft Mountains, Namibia. *J. Veg. Sci* 12: 53-60.

Carey, P.D. 1996. A cellular automaton for predicting the distribution of species in a changed climate. *Global Ecol. Biogeogr. Lett.* 5: 217-226.

Chambers, J.M. & Hastie, T.J. 1997. *Statistical models in S*. Chapman & Hall, London, UK.

Duckworth, J.C., Bunce, R.G.H. & Malloch, A.J.C. 2000. Modelling the potential effects of climate change on calcareous grasslands in Atlantic Europe. *J. Biogeogr.* 27: 347-358.

Eastman, J.R. 1996. *Idrisi for Windows – User's Guide Ver-*

sion 2.0. (Clark Labs for Cartographic Technology and Geographic Analysis) Clark University, Worcester, MA, US.

Fielding, A.H., & Bell, J.F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ. Conserv.* 24: 38-49.

Flannigan, M.D. & Woodward, F.I. 1994. Red pine abundance: current climatic control and response to future warming. *Can. J. For. Res.* 24: 1166-1175.

Franklin, J. 1998. Predicting the distribution of shrub species in southern California from climate and terrain-derived variables. *J. Veg. Sci.* 9: 733-748.

Guisan, A. & Theurillat, J.-P. 2000. Equilibrium modelling of alpine plant distribution: how far can we go? *Phytocoenologia* 30: 353-384.

Guisan, A. & Zimmermann, N.E. 2000. Predictive habitat distribution models in ecology. *Ecol. Model.* 135: 147-186.

Guisan, A., Theurillat, J.-P. & Kienast, F. 1998. Predicting the potential distribution of plant species in an alpine environment. *J. Veg. Sci.* 9: 65-74.

Guisan, A., Weiss, S.B., & Weiss, A.D. 1999. GLM versus CCA spatial modelling of plant species distribution. *Plant Ecol.* 143: 107-122.

Hanley, J.A. & McNeil, B.J. 1983. A method of comparing the areas under receiver operating characteristic (ROC) curve. *Radiology* 148: 839-843.

Hastie, T.J. & Tibshirani, R. 1990. *Generalized additive models*. Chapman and Hall, London, UK.

Huisman, J., Olff, H. & Fresco, L.F.M. 1993. A hierarchical set of models for species response analysis. *J. Veg. Sci.* 4: 37-46.

Huntley, B. 1995. Plant species' response to climate change: implications for the conservation of European birds. *IBIS* 137: 127-138.

Huntley, B., Berry, P.M., Cramer, W. & McDonald, A.P. 1995. Modelling present and potential future ranges of some European higher plants using climate response. *J. Biogeogr.* 22: 967-1001.

Huston, M.A. 1994. *Biological diversity*. Cambridge University Press, Cambridge, UK.

Iverson, L.R. & Prasad, A. 1998. Predicting abundance for 80 tree species following climate change in the Eastern United States. Ecological Monographs 68(4): 465-485.

Jalas, J. & Suominen, J. 1972-1996. *Atlas Florae Europaeae*. Vol. 1-11. Committee for Mapping the Flora of Europe and Societas Biologica Fennica Vanamo, Helsinki, FI.

Lahti, T. & Lampinen, R. 1999. From dot maps to bitmaps – Atlas Florae Europaeae goes digital. *Acta Bot. Fenn.* 162: 5-9.

Leathwick, J.R. & Austin, M.P. 2001. Competitive interactions between tree species in New Zealand's old-growth indigenous forests. *Ecology* 82: 2560-2573.

Legates, D.R. & Wilmott, C.J. 1990a. Mean seasonal and spatial variability in gauge-corrected, global precipitation. *Int. J. Climatol.* 10: 111-127.

Legates, D.R. & Wilmott, C.J. 1990b. Mean seasonal and spatial variability in global surface ait temperature. *Theor. Appl. Climatol.* 41: 11-21.

McCullagh, P. & Nelder, J.A. 1989. *Generalized linear models*. Chapman & Hall, London, UK.

Midgley, G.F., Hannah, L., Millar, D., Thuiller, W. & Booth, A. 2003. Developing regional and species-level assessments of climate change impacts on biodiversity: A preliminary study in the Cape Floristic Region. *Biol. Conserv.* 112: 87-97.

Nicholls, A.O. 1989. How to make biological surveys go further with generalized linear models. *Biol. Conserv.* 50: 51-75.

Ninyerola, M., Pons, X. & Roure, J.M. 2000. A methodological approach of climatological modelling of air temperature and precipitation through GIS techniques. *Int. J. Climatol.* 20: 1823-1841.

Oksanen, J. & Minchin, P.R. 2002. Continuum theory revisited: what shape are species responses along ecological gradients. *Ecol. Model.* 157: 119-129.

Olden, J.D. & Jackson, D.A. 2002. Illuminating the 'black box': a randomization approach for understanding variable contributions in artificial neural networks. *Ecol. Model.* 154: 16.

Pearce, J. & Ferrier, S. 2000a. Evaluating the predictive performance of habitat models developed using logistic regression. *Ecol. Model.* 133: 225-245.

Pearce, J. & Ferrier, S. 2000b. An evaluation of alternative algorithms for fitting species distribution models using logistic regression. *Ecol. Model.* 128: 127-147.

Pearson, R.G., Dawson, T.P., Berry, P.M. & Harrison, P.A. 2002. SPECIES: A spatial evaluation of climate impact on the envelope of species. *Ecol. Model.* 154: 289-300.

Schulze, R.E. & Kunz, R.P. 1995. Potential shifts in optimum growth areas of selected commercial tree species and subtropical crops in southern Africa due to global warming. *J. Biogeogr.* 22: 679-688.

Swets, K.A. 1988. Measuring the accuracy of diagnostic systems. *Science* 240: 1285-1293.

ter Braak, C.J.F. 1987. The analysis of vegetation-environment relationship by canonical correspondence analysis. *Vegetatio* 69: 69-77.

Thuiller, W., Vaydera, J., Pino, J., Sabaté, S., Lavorel, S. & Gracia, C. 2003. Large-scale environmental correlates of forest tree distributions in Catalonia (NE Spain). *Global Ecol. Biogeogr.* 12: 313-325.

Vayssières, M.P., Plant, R.E. & Allen-Diaz, B.H. 2000. Classification trees: an alternative non-parametric approach for predicting species distributions. *J. Veg. Sci.* 11: 679-694.

Vetaas, O.R. 2000. Comparing species temperature response curves: population density versus second-hand data. *J. Veg. Sci.* 11: 659-666.

Woodward, F.I. 1987. *Climate and plant distribution*. Cambridge University Press, Cambridge, UK.

Woodward, F.I. 1990. The impact of low temperatures in controlling the geographical distribution of plants. *Phil. Trans. R. Soc. Lond.* B 326: 585-593.

Woodward, F.I. & Williams, B.G. 1987. Climate and plant distribution at global and local scales. *Vegetatio* 69: 189-197.

Yee, T.W. & Mitchell, N.D. 1991. Generalized additive models in plant ecology. *J. Veg. Sci.* 2: 587-602.

Received 11 April 2002;
Revision received 15 January 2003;
Accepted 13 February 2003.
Coordinating Editor: N. Kenkel.