Global Ecology & Biogeography (2003) **12**, 313–325

RESEARCH PAPER

Large-scale environmental correlates of forest tree distributions in Catalonia (NE Spain)

WILFRIED THUILLER*†, JORDI VAYREDA‡, JOAN PINO‡, SANTI SABATE‡, SANDRA LAVOREL† and CARLES GRACIA‡

†Centre d'Ecologie Fonctionnelle et Evolutive, Centre National de Recherche Scientifique, 1919 route de Mende, 34293 Montpellier cedex 5, France, E-mail: wilfried.thuiller@cefe.cnrs-mop.fr and ‡Centre de Recerca Ecològica i Aplicacions Forestals, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain

ABSTRACT

Aim To explore the environmental correlates of tree species distributions in Catalonia according to the chorological status of the species.

Location The study area is the region of Catalonia, in north-eastern Spain.

Methods We used presence-absence data for 24 species, sampled in random plots distributed throughout forests of Catalonia. A climate model for the Catalonia region provided environmental variables. We used classification tree analysis to explore the environmental correlates of the realized niches of tree species. The predictive accuracy of the models was assessed using the ROC curve approach. Potential distribution maps of tree species were generated for the whole Catalonia region.

Results Models were ranked from low to high accuracy for the 24 species. Differences in accuracy among species were related to the chorological status of species. Zonal species, or species at the core of their range (Mediterranean and Sub-Mediterranean species), were generally well predicted, while extrazonal species, or species at the edge of their range, were predicted only moderately well. Mediterranean species distributions showed good correlations with extreme temperatures and annual precipitation.

Main conclusions The above trends confirmed the difficulty of identifying the realized niche of species at the edges of their ranges. In contrast, Mediterranean and Sub-Mediterranean species, which were at the core of their range, were well-predicted, confirming the importance of extremes of temperature and annual precipitation as effective surrogates for variables having more direct physiological roles in limiting the ability of plants to survive and grow. Maps of potential tree distributions allowed us to define suitable habitats and to highlight areas where species have been planted outside their natural distribution.

Key words chorology, classification tree analysis, environmental gradients, forest trees, potential distribution maps, realized niche, Spain.

INTRODUCTION

The Mediterranean area is considered a hotspot of biodiversity (Myers *et al*., 2000), mainly for its plant richness. Among plant species, trees are often considered as the basis for defining habitat representing many unsampled organisms. Conservation planning and projections of future distributions of plant and animal diversity therefore require efforts to determine the ecological requirements of forest tree species from their current distributions.

Tree species distributions are affected by biotic and abiotic factors that define the realized niches of species. The realized niche is quantified from what determines the current spatial distribution of a species in a community or a region. Attempts to predict and explain species distributions should be based on the modelling of the realized niche (Austin & Smith, 1989). As the biotic factors limiting species can often not be measured over large scales, the realized niche is usually predicted using environmental variables taken as surrogates for physiologically relevant variables. Environmental variables might operate, however, at contrasting scales. Whereas large-scale climatic factors are usually responsible for the distribution of potential zonal communities, local topographic and geological factors determine the presence of extrazonal and azonal communities (Woodward, 1987; Ellenberg, 1988). In

---

* Corresponding author.

addition, distributions of all species are altered by human activity and natural disturbance. Large-scale climatic factors are commonly used to predict current plant distributions (Heikkinen & Birks, 1996; Burke, 2001; Frescino *et al*., 2001) and to project future distributions, or species turnover in response to global change (Huntley *et al*., 1995; Guisan *et al*., 1998; Duckworth *et al*., 2000; Sykes, 2001). However, at a finer resolution, climatic factors are likely to be overridden by local factors, which may be more relevant for explaining and predicting the current distributions of plants or animals (Huston, 1994). Finally, when working on limited geographical extents it is often difficult to capture the entire realized niche of species, because distributions span beyond the boundaries of study areas. In particular, if a species is at the edge of its range (i.e. extrazonal or azonal species), large-scale climatic factors may not explain its distribution, whilst more local-scale factors (topographical and geological) providing conditions prevailing outside of the study region (at the core of the species range) could be more relevant.

The present paper explores the relevance of biogeographical modelling of tree distributions in Catalonia (NE of Spain) in terms of their chorological status (Mediterranean, Sub-Mediterranean, Eurosiberian and Boreoalpine status) and in respect of the environmental variables potentially determining these distributions. Because of its geographical situation, at the borderline of the Eurosiberian and the Mediterranean phytogeographical regions, and containing Boreoalpine habitats in the Pyrenees range, Catalonia provides an interesting case study of the effects of biogeography on the accuracy of species modelling, with a great number of tree species with a variety of chorologies resulting in zonal, extrazonal and azonal situations. This analysis was made possible by the availability of data for forest tree diversity in Catalonia, collected as part of the Ecological Forest Inventory of Catalonia (IEFC), an extensive database of 10638 sampling points that covers all the forestry area in Catalonia at a density of 1 plot per km^2 (Gracia *et al*., 2000a,b,c, 2001a,b).

This study addresses questions about the importance of environmental variables and forest type for plant survival and growth, and specifically examines which environmental variables are most suitable for modelling the realized niche of plant species in Catalonia. We apply classification tree analysis (CTA) in order to study the environmental correlates of the distributions of a set of important tree species. The comparison of the resulting models with current distribution maps is used to quantify the importance of environmental factors on the distribution of species according to their chorological status. Specifically we address the following questions:

- Are there differences of accuracy between models derived for species from different chorological classes?
- Are environmental requirements of species within a chorological class more similar than requirements of species in different classes?

MATERIALS AND METHODS

Study area

Catalonia is a region of the NE of Spain covering 32 098 km^2 (Fig. 1). The region exhibits a high climatic diversity resulting from its geomorphologic and topographical situation, receiving Mediterranean, Atlantic and even Saharan influences. The majority of the area has a Mediterranean climate, with a decrease in rainfall from north to south and from the coast to the central plains. A continentality gradient is also present from the coastal to inland areas, resulting in a gradual climatic variation from moist temperate coastal climates to dry continental inland climates. The north corresponds to the boundary of the Mediterranean and the Eurosiberian phytogeographical regions. The Pyrenees represent a sharp topographic-climatic gradient where Mediterranean or Eurosiberian-type biomes gradually change to subalpine and alpine types, corresponding to the Boreoalpine region.

Forest currently occupies 38% of Catalonia (Gracia *et al*., in press), although its potential area covers up to 61%. Evergreen forests of *Quercus ilex* and *Q. suber* potentially constitute the late successional stage in Mediterranean lowland areas (Barbéro *et al*., 1998). Deciduous oak (*Q. gr. humilis*) forests would substitute the evergreen forests in Sub-Mediterranean transition areas between the Mediterranean and the Eurosiberian regions. Eurosiberian to Boreoalpine forests are concentrated in the Pyrenees and their surrounding ranges, with deciduous forests of *Fagus sylvatica* and *Quercus petraea*, and Boreoalpine coniferous forests with *Abies alba* and *Pinus uncinata*.

The structure of Catalonian landscapes reflects a long and complex interaction between humans, forest and climate, with a historical background of forest regression since the beginning of the classic civilizations and a recent (over 100 year) reversion of this situation, with progressive crop abandonment, afforestation and forest recovery (Peix, 1999). Late successional species have been partially substituted by secondary forest species, mainly native conifers (*Pinus halepensis*, *P. pinaster*, *P. nigra*, *P. sylvestris*), but also alien ones (*P. pinea*, *P. radiata*), which represent the dominant fraction of Catalan forests. Large lowland forests are now dominated by *Pinus halepensis*, while *P. sylvestris* is the most abundant species in mountain areas. Natural disturbances also play a significant role in landscape structure, especially through frequent fires in the north-east of Catalonia.

Data set

The Ecological Forest Inventory of Catalonia (IEFC) is an extensive database of 10 638 circular sampling plots with a radius of 10 m. The field survey was carried out by the Centre for Ecological Research and Forestry Applications in Barcelona (CREAF). The IEFC includes a total of 95 tree

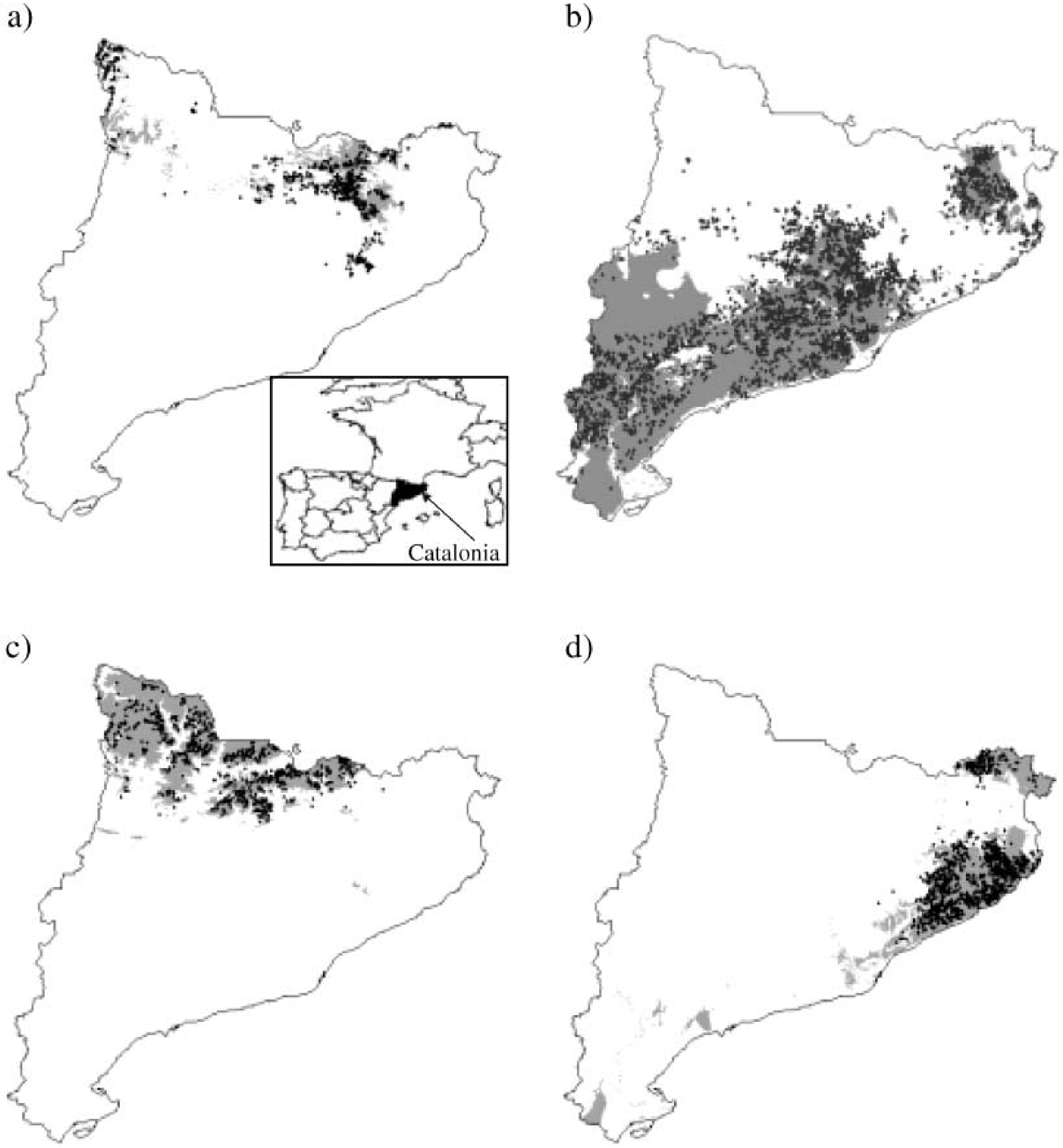

Fig. 1 Potential distribution map for the species very well modelled by CTA. (a) *Fagus sylvatica* (b) *Pinus halepensis* (c) *Pinus uncinata* and (d) *Quercus suber*. Points represent the sampling plots where the species was present and shaded areas are the areas modelled as suitable for the species.

Table 1 List of the selected environmental variables

Environmental variables	Abbreviation
Annual Precipitation	MAP
Summer Precipitation	SumPrec
Winter Precipitation	WintPrec
Autumn Precipitation	AutPrec
Spring Precipitation	SprPrec
Mean Annual Temperature	MAT
Mean Temperature of the Coldest Month (January)	MTC
Mean Temperature of the Hottest Month (July)	MTW
Mean Annual Maximum Temperature	MAMxT
Mean Maximum Temperature of Hottest Month (July)	MMTW
Mean Annual Minimum Temperature	MAMiT
Mean Minimum Temperature of the Coldest Month (January)	MMTC
Mean Annual Radiation	MAR
Slope	Slope
Geology (factor)	Geol
category 1: Siliceous rocks	
category 2: Carbonated rocks	
category 3: Limestone	
category 4: Calcareous sediments	
category 5: Siliceous sediments	

species (of which 5 are present in more than 10% of the plots), including the most abundant *Pinus* (7 spp.) and *Quercus* (12 spp.) species.

For this study, we selected a set of 24 forest tree species representative of the compositional diversity of Catalonia. This set includes common, intermediate and less common species, from 12 genera: *Quercus* (4 spp.), *Pinus* (6 spp.), *Acer* (2 spp.), *Juniperus* (2 spp.), *Populus* (2 spp.), *Abies*, *Castanea*, *Corylus*, *Fagus*, *Salix*, *Sorbus* and *Ulmus* (1 sp.). These species are also a representative sample of the existing variety of phytogeographical patterns, including Boreoalpine (2 spp.), Eurosiberian (10 spp.), Sub-Mediterranean (6 spp.) and Mediterranean (6 spp.) species. These different phytogeographical patterns correspond with species at the edge of their distribution ranges (Boreoalpine and Eurosiberian) compared to species at the core of their ranges (Mediterranean and sub-Mediterranean).

Environmental variables

Environmental variables were selected according to their relevance to tree survival and growth (Table 1). Although the main environmental data required were climate data, given the gradients of altitude and topography in Catalonia, we also included slope and geology. Climatic variables were taken as surrogates for variables having more direct physiological roles in limiting the ability of plants to survive and grow, such as the number of growing degree days or actual evapotranspiration (Bartlein *et al*., 1986; Prentice *et al*., 1991; Huntley *et al*., 1995). Geology is an indirect variable and was considered as a surrogate for soil type, with direct impacts on nutrient and water availability for plant growth (Austin & Smith, 1989).

Climatic variables (temperature and precipitation) were generated by a climate model (Ninyerola *et al*., 2000). This was used to generate layers of monthly precipitation and monthly mean, minimum and maximum temperatures for the entire study area. The climate model used a network of weather stations in Catalonia (257 stations for precipitation and 160 stations for temperatures), and was based on a multiple regression analysis between the meteorological variables and a set of geographical variables (latitude, solar radiation, altitude, continentality and cloudiness factors). Layers of rainfall and temperature were then derived from the original climate model (Ninyerola *et al*., 2000). Topographic layers (elevation, slope and solar radiation) were provided by a DEM (180-m resolution) generated by the Cartographic Institute of Catalonia (ICC).

The data were used in the raster format of ARCVIEW (ESRI Inc., Redlands, California). We combined raster and sampling locations for each species data to determine the values of each environmental variable in each plot.

Statistical analysis

Distribution data sets

Models were calibrated using a random sample of the data (70%: calibration data) and evaluated with the remaining 30% (evaluation data).

Model

We used classification tree analysis (*tree* in Splus software: Anon, 1999) to study and analyse the relationships between environmental factors and species distributions (Breiman *et al.*, 1984). The tree is built by repeatedly splitting the calibration data, according to 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. The heterogeneity of a node is defined with a deviance notion that can be interpreted as the deviance of a Gaussian model (regression tree) or multinomial model (classification tree) (Breiman *et al.*, 1984). To control the length of the tree, we used the '*prune*' function of S-plus (Anon, 1999). The program builds a nested sequence of subtrees of the main tree by recursively snipping off the less important splits in terms of explained deviance.

CTAs have only recently received increased interest for biogeographical studies (Franklin, 1998; Iverson & Prasad, 1998; Vayssières *et al.*, 2000; Rouget *et al.*, 2001) as compared to other modelling techniques like generalized linear and generalized additive models (GLM and GAM, respectively). Several studies have shown that CTA offer a good compromise, being as accurate as GLM or GAM (Franklin, 1998; Vayssières *et al.*, 2000), and offering the following useful properties:

Facility of implementation: CTAs accept mixes of categorical and continuous data (Breiman *et al.*, 1984; Iverson & Prasad, 1998), as also possible with GLM, but not with GAM.

Tree representation provides a hierarchical view of the relationships between species and environmental variables and makes it possible to identify which variable is the major correlate of the presence of a species (Iverson *et al.*, 1999; Vayssières *et al.*, 2000).

CTAs are one of the few methods that can model interactive effects of two or more variables, and can represent them in an easily readable fashion (Iverson & Prasad, 1998; Rouget *et al.*, 2001). Interactions factors can also be added to GLM, but with an a *priori* choice of interactions or using stepwise procedures. Conversely, GAM is purely additive and cannot deal explicitly with interaction terms.

Assessing prediction accuracy

We used the Receiver Operating Characteristic curve (ROC curve) to evaluate the accuracy of our models (Hanley & McNeil, 1982, 1983). This graphical method makes it possible to represent the relationship between the percentage of presences correctly predicted (sensitivity) and 1 minus the percentage of the absences correctly predicted (specificity). The area comprised between a 45° line and the curve measures the discrimination, that is, the ability of the model to classify correctly a species as present or absent in a given plot (area under the curve: AUC). The confidence intervals of AUC at 95% were derived for each species (Pearce & Ferrier, 2000; Vayssières *et al.*, 2000). A rough guide for classifying the accuracy is: 0.50–0.60 = fail; 0.60–0.70 = poor; 0.70–0.80 = fair; 0.80–0.90 = good; 0.90–1 = excellent (Swets, 1988).

Potential distributions

Given the response surface fitted to the observed patterns of occurrence in the sampled plots and the environmental variables for the 24 selected species, it is possible to map the potential distributions of these species for the entire Catalonia region. To plot potential distributions of species, probability values of presence from CTAs were converted into binary form using a threshold maximizing the sensitivity and specificity of the predictions (Pearce & Ferrier, 2000) for each species.

RESULTS

Accuracy of classification tree analysis

We considered four classes of model accuracy (Table 2). The low accuracy (AUC < 0.7) class included *Juniperus communis* and *Ulmus minor* (AUC = 0.646 and AUC = 0.645, respectively). *Ulmus minor* had the lowest accuracy of all species. Six species populated the fair accuracy class (0.7 < AUC < 0.8), including species with many occurrences (*Quercus humilis*, *Q. ilex*) and some scarce species (*Salix caprea*, *Ilex aquifolium*). Good accuracy (0.8 < AUC < 0.9) was assigned to 10 species belonging to 7 genera, including some highly abundant (e.g. *Pinus sylvestris*, 3217 occurrences) and other scarcer species (e.g. *Acer opalus*, 240 occurrences). Within this class two subgroups were considered according to levels of accuracy (according to AUC and confidence intervals). A first subgroup (*Salix caprea*, *Pinus pinaster*, *Juniperus oxycedrus*, *Quercus petraea*, *Sorbus aria* and *Acer opalus*) had AUC ranging from 0.808 to 0.835, and a second subgroup (*Pinus sylvestris*, *Populus tremula*, *Pinus nigra* and *Pinus pinea*) had AUC ranging from 0.875 and 0.891. The high accuracy class (AUC > 0.9) included six species (*Castanea sativa*, *Abies alba*, *Pinus halepensis*, *Fagus sylvatica*, *Quercus suber* and *P. uncinata*). *P. uncinata* was the best-predicted species, with a narrow confidence interval (0.977 < AUC < 0.982) and an extremely high percentage of presences (90.6%) and absences (94.7%) correctly predicted. Accuracy of CTA predictions increased from Eurosiberian species, with only 50% of species being scored in high or good accuracy classes, to Mediterranean ones, which showed 83% of species in these classes. Sub-Mediterranean trees exhibited an intermediate situation, with 67% of species falling into high or good accuracy classes.

Table 2 Prediction accuracy for the 24 species

Species	Chorology	*n*	Calibration data	Evaluation data				
			AUC	Pres	Abs	AUC	Lo. 0.95	Up. 0.95
High accuracy								
Pinus uncinata	B	787	0.976	90.6	94.7	0.979	0.977	0.982
Quercus suber	M	1034	0.981	92.0	93.6	0.956	0.947	0.966
Fagus sylvatica	E	567	0.958	85.8	85.4	0.941	0.933	0.949
Pinus halepensis	M	2712	0.950	87.1	86.8	0.941	0.936	0.945
Abies alba	B	229	0.979	93.0	89.6	0.937	0.915	0.959
Castanea sativa	SM	246	0.957	90.0	82.5	0.932	0.918	0.946
Good accuracy								
Pinus pinea	M	997	0.892	88.8	74.8	0.891	0.883	0.899
Pinus nigra	SM	1971	0.910	89.6	75.5	0.890	0.888	0.897
Populus tremula	E	151	0.899	75.7	84.9	0.886	0.864	0.904
Pinus sylvestris	SM	3217	0.898	75.8	78.7	0.875	0.868	0.882
Acer opalus	SM	240	0.907	78.7	77.8	0.835	0.814	0.857
Sorbus aria	E	140	0.888	77.1	80.0	0.832	0.796	0.869
Quercus petraea	E	262	0.878	81.4	71.2	0.817	0.787	0.848
Juniperus oxycedrus	M	146	0.815	65.8	86.3	0.813	0.775	0.8
Pinus pinaster	M	302	0.848	78.5	81.2	0.81	0.783	0.836
Salix caprea	E	91	0.914	79.3	81.0	0.808	0.753	0.863
Fair accuracy								
Quercus humilis	SM	1847	0.825	79.0	68.7	0.797	0.786	0.808
Populus nigra	E	102	0.882	59.4	86.5	0.796	0.751	0.841
Acer monspessulanum	SM	157	0.885	64.9	72.1	0.785	0.750	0.820
Quercus ilex	M	4070	0.786	74.5	68.7	0.779	0.771	0.788
Ilex aquifolium	E	102	0.856	76.2	71.6	0.765	0.730	0.800
Corylus avenalla	E	182	0.919	73.7	63.5	0.752	0.709	0.796
Low accuracy								
Juniperus communis	E	212	0.819	68.9	59.1	0.646	0.610	0.682
Ulmus minor	E	115	0.942	63.9	65.2	0.645	0.592	0.703

Abbreviations for chorology: B, Boreoalpine; E, Eurosiberian; M, Mediterranean; SM, Sub-Mediterranean. *n* = number of occurrences in the sample sites. Pres: mean percentage of presences correctly predicted. Abs: mean percentage of absences correctly predicted. AUC, Area Under the Curve of the Relative Operating Characteristic Curve (ROC curve). Lo. 95 and Up. 95 represent the lower and upper limits of the confidence interval of AUC.

Potential distribution area of the species

Species with highly accurate predictions exhibited potential distributions very close to the observed ones, with almost all observed points enclosed in the potential area (Figs 1, 2 and 3). Well-modelled (Fig. 1) species included contrasting types of distributions, ranging from widespread species such as *Pinus halepensis* to localized species such as *Pinus uncinata* and *Abies alba*, which grow exclusively in the Pyrenees, and *Quercus suber*, which is limited by siliceous areas in the north-east. Dominant species belonging to good and fair accuracy classes (*Pinus nigra*, *P. sylvestris*, *Quercus ilex*, *Q. humilis*) had similar potential and observed distributions (Fig. 2), although there were many outliers, mainly in peripheral areas. In the case of *Quercus ilex*, the model did not predict accurately the inland distribution of the species. There were still greater differences between potential and observed distributions for several minor species of these two accuracy classes (Fig. 3), probably because of the limited number of sampling points that did not permit an accurate modelling in large parts of their potential area of distribution.

Environmental correlates for the distribution of well-modelled species

Classification trees for several well-modelled species (*Fagus sylvatica*, *Pinus uncinata*, *Pinus halepensis* and *Quercus suber*) are shown in Fig. 4. Tree-based models can identify and express in a relatively simple form nonlinear and non-additive relationships with quantitative thresholds from presence to absence. For instance, altitude above 1335 m was

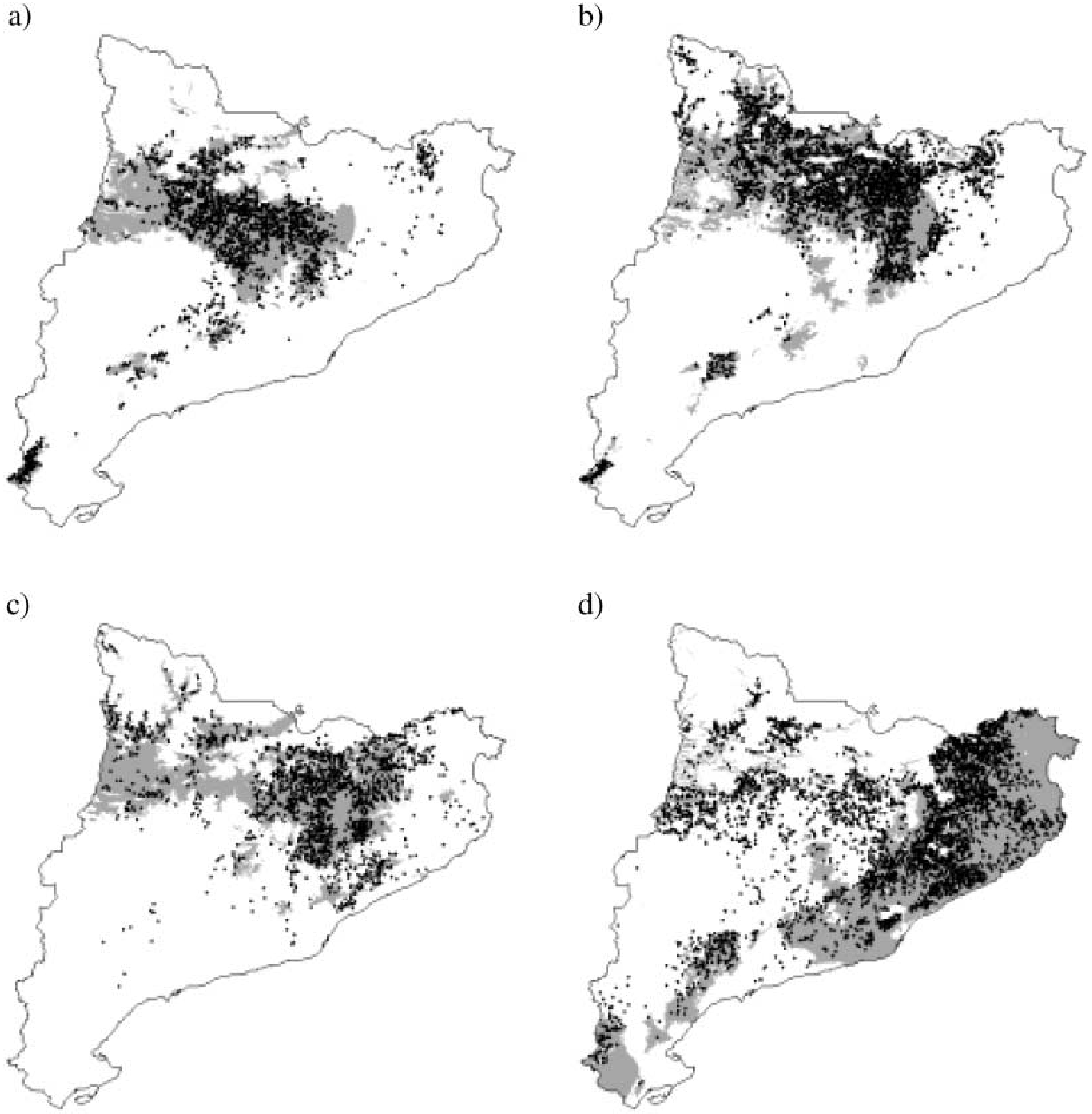

Fig. 2 Potential map of species well modelled by CTA. (a) *Pinus nigra* (b) *Pinus sylvestris* (c) *Quercus humilis* and (d) *Quercus ilex*. Points represent the sampling plots where the species was present, and shaded areas are the areas modelled as suitable for the species.

identified as the main correlate for the presence of *P. uncinata*, but temperature and rainfall appeared as important factors in the lower part of this range below 1720 m. Mean temperature of the coldest month was identified as the main limiting factor for *Quercus suber*, while siliceous soil came second. The model predicted the presence of the species on calcareous soil only in high precipitation areas. The presence of *Pinus halepensis* was mainly correlated with annual precipitation, although this apparent constraint might be compensated by the interaction with many other variables, including other

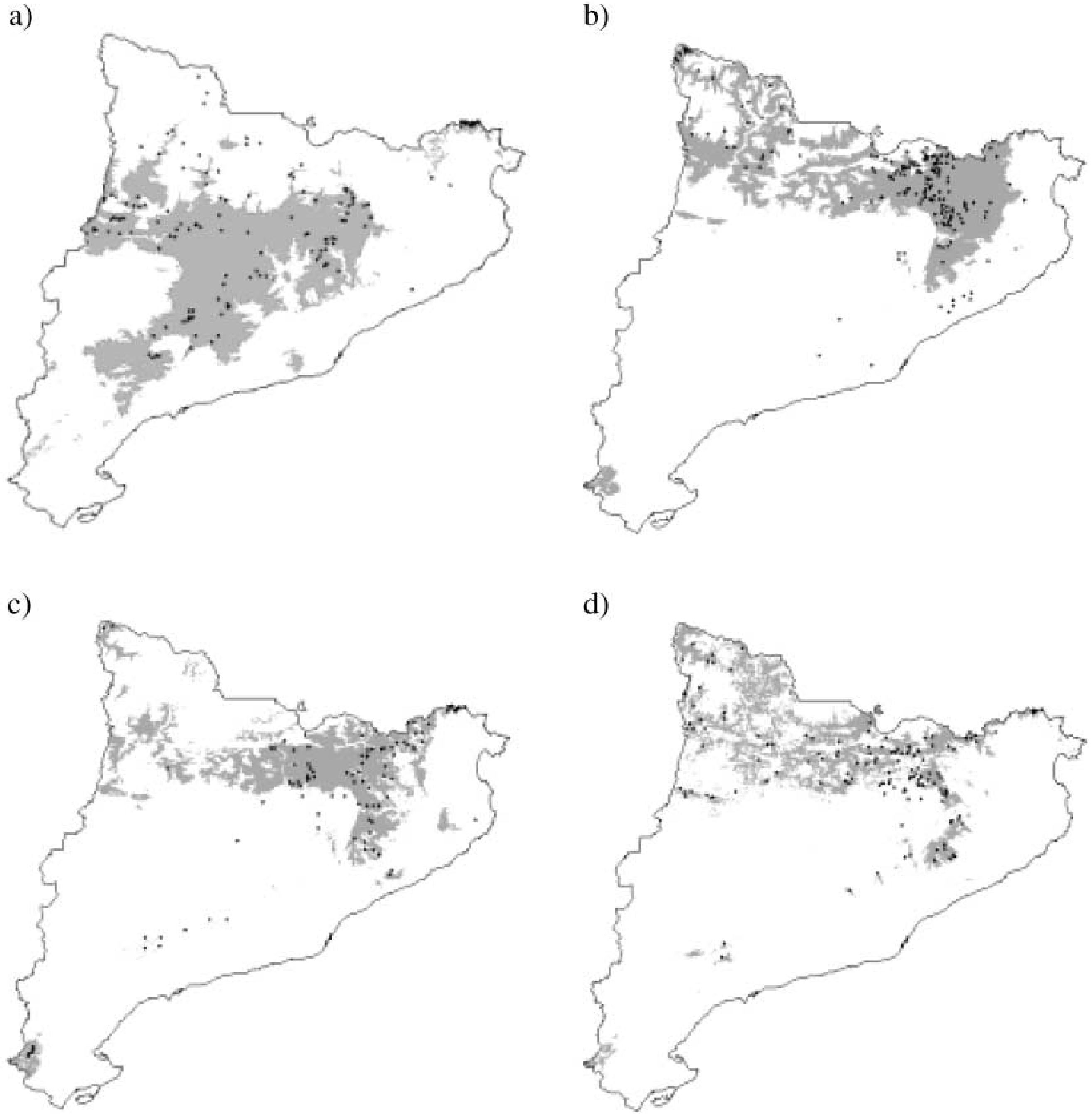

Fig. 3 Potential map of the rare species fairly modelled by CTA. (a) *Acer monspessulanum* (b) *Corylus avellana* (c) *Ilex aquifolium* and (d) *Sorbus aria*. Points represent the sampling plots where the species was present, and shaded areas are the areas modelled as suitable for the species.

components of precipitation, temperature and geology. *Fagus sylvatica* appeared mainly restricted to areas with > 950 mm of annual rainfall, within which, in areas with less than 1050 mm, the distribution was related to winter and summer precipitation.

Main correlates of tree species distributions

The models most often selected six variables as important variables (Table 3): three precipitation variables (annual, summer and winter), two temperature variables (temperature

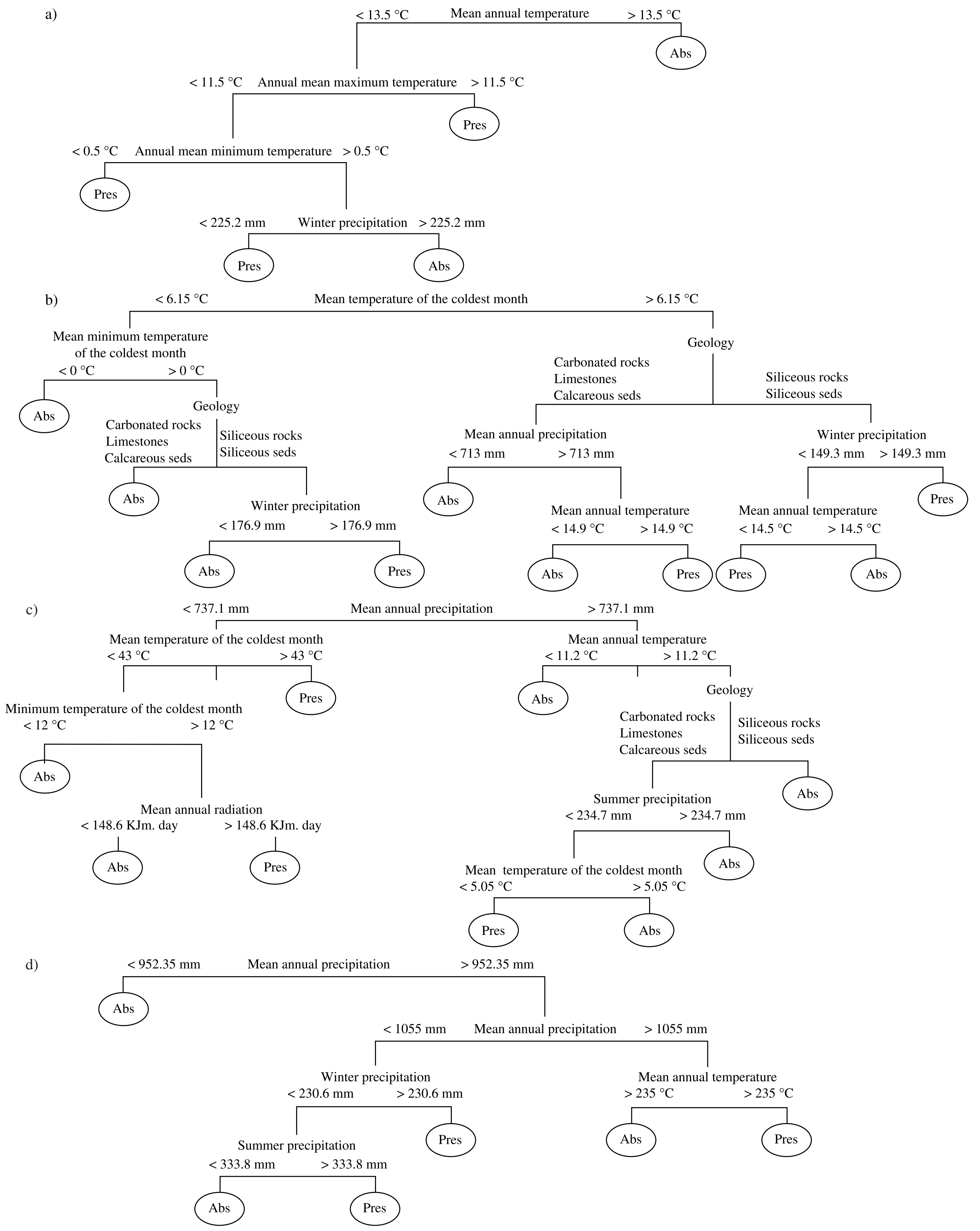

Fig. 4 Main predictors of: (a) *Pinus uncinata* (b) *Quercus suber* (c) *Pinus halepensis* and (d) *Fagus sylvatica* distribution as modelled by CTA. The combination of variables that favour the occurrence of a species are represented by Pres. The dendogram uses the change in deviance to guide the vertical positioning of each pair of nodes.

Table 3 List of the first four environmental variables (in terms of explained deviance) selected by the classification tree analysis for each species. The last row presents the relative importance of each variable for the 24 species. For each species, an index has been attributed to the variables following the order of selection by the model. The value of 4 denotes the first variable, 3 the second, 2 the third and 1 the fourth. Phytogeographical classification of the species is according to Bolòs *et al.* (1990). See Table 1 for variable abbreviations

		Win	Spr	Sum	Aut		MA					MM			
	MAP	Prec	Prec	Prec	Prec	MAT	MiT	MTC	MMTC	MAMxT	MTW	TW	MAR	Slope	Geology
Boreoalpine															
Abies alba		2		3							1		4		
Pinus uncinata		3					2			1					
Eurosiberian															
Corylus avellana	1		2			3									4
Fagus sylvatica	1	3		4						2					
Ilex aquifolium	1							3			4		2		
Juniperus communis		3		1	4		2								
Populus nigra		4										3		1	2
Populus tremula	1		2							3			4		
Quercus petraea			1	3						2		4			
Salix caprea				2		1			4						3
Sorbus aria					3					4	1			2	
Ulmus minor				2				1		3			4		
Submediterranean															
Acer monspessulanum	4				1	2							3		
Acer opalus			3						4		1				2
Castanea sativa	3								2				4		1
Pinus nigra		1	3	4			2								
Pinus sylvestris	3	4		1		2									
Quercus humilis				1	4				2	3					
Mediterranean															
Juniperus oxycedrus					1			2	4			3			
Pinus halepensis	1							2	3				4		
Pinus pinaster		2			4		3	1							
Pinus pinea	3			4			2	1							
Quercus ilex		2		3						1					4
Quercus suber		4						1	2						3
First position	5	1	1	3	2	1	0	4	0	2	3	0	0	1	1
Second position	0	3	2	2	0	2	4	2	3	2	1	0	1	1	2
Third position	3	3	2	3	1	1	1	1	1	3	0	2	1	0	2
Fourth position	1	3	0	2	2	0	0	0	3	4	4	1	4	0	2
Importance of variables	27	22	14	26	12	12	14	24	14	24	19	5	9	7	16

of the coldest month and maximum annual temperature), and geology. Among these six main variables, annual precipitation and mean temperature of the coldest month were most frequently selected as primary factors (in first or second position), while geology, winter and summer precipitation were selected at a variable rank. These six variables had varying weights across species, even within a given genus. For example, the models selected mean winter precipitation as a main determinant for the distribution of all *Pinus* species, except for *P. halepensis*, which appeared to be mainly affected by annual precipitation. Mean temperature of the coldest month was a main environmental constraint for *P. pinea*, *P. pinaster* and *P. halepensis*, but not for the remaining species of the genus. Table 3 reveals patterns within chorological classes rather than taxonomic classes. Mediterranean species are distributed mainly in relation to temperature and secondly are related to precipitation. In contrast, Sub-Mediterranean and Eurosiberian species seem to be influenced mainly by precipitation and secondly by temperature. It is more difficult to find a general pattern for the two Boreoalpine species, both of which are at the edge of their range.

DISCUSSION

Classification trees

As already depicted by Franklin (1998), Iverson & Prasad (1998) and Vayssières *et al.* (2000), CTA provides an interesting tool for analysing the main correlates of species distribution. CTA displays a hierarchical easily understood view of species-environment relationships. Several studies have shown their relatively good predictive performance (Harvey, 1996; Vayssières *et al*., 2000) and most of them underline the power of CTA to find interactions and hierarchical relations among environmental variables (Hastie & Tibshirani, 1990; Austin, 2002). However, generalized linear and additive models may be more powerful tools, providing more meaningful relationships between species and explanatory variables (e.g. bell-shaped curves) (Austin, 2002). As already shown by several studies, there is no universal method and the choice of a modelling method should be made according to the aim of the study. In our case, CTA seemed to be the more suitable tool. Alternatively, dynamic vegetation models (Prentice & Leemans, 1990; Sykes & Prentice, 1996) incorporate explicit representations of key ecological processes (tree growth, establishment, competition, succession) and have been widely used to simulate the potential response of vegetation to past and future climate change (Prentice *et al*., 1991; Sykes & Prentice, 1995; Korzukhin *et al*., 1996; Starfield & Chapin, 1996). However, since such models require extensive knowledge of the species involved, as well as of the historical factors of the studied area, they can only be developed for well-studied species and habitats (Peng, 2000). In our case, perfect knowledge of species ecology and historical factors over the entire Catalonia region were missing, precluding the use of such approaches.

Correlates of tree species distribution

Classification Tree Analyses succeeded as a method of identifying the main correlates of actual species distributions and as a method for predicting the potential distributions of 67% (high or good accuracy) of the selected species. Model accuracy increased from Eurosiberian species (50% of species scored as high or good accuracy classes) through Sub-Mediterranean (67%) to Mediterranean (83%). Trends in accuracy mirrored the decreasing extrazonality or azonality from Eurosiberian (species at the edge of range) to Mediterranean species (species at the core of range) (Ellenberg, 1988). Catalonia is located in the borderline of the Eurosiberian and the Mediterranean regions and the general climate is Mediterranean. We confirmed the importance of annual precipitation and extreme temperatures in explaining regional patterns of distribution for Mediterranean species, in accordance with other studies at similar scales (Barbéro *et al*., 1998; Rouget *et al*., 2001). For example, *Pinus halepensis*, which belongs to the Klaus (1989) group of Mediterranean shore and island pines, is expected to occur in semihumid to semiarid habitat (Barbéro *et al*., 1998). Our model corroborated such expectations, showing that the distribution of *P. halepensis* is related to intermediate annual precipitation values and is restricted to mild winter temperatures (Fig. 4). This species is well adapted to Mediterranean conditions and the study area is situated at the core of its distribution, making for accurate predictions of its potential distribution in Catalonia. Conversely, Eurosiberian and Borealpine species at the edge of their range are restricted to specific local climatic and topographic situations, where landscape factors (aspect, convexity, water table height) can be much more important than the average climate. This is particularly true for riparian species such as *Populus nigra* and *Ulmus minor*, which therefore showed fair to low model accuracies. The two Boreoalpine species (*Abies alba* and *Pinus uncinata*) are restricted to the Pyrenees and have specific local requirements (climatic and topographic), hence our relatively good predictions. The relative failure for species at the edge of their distribution range could be improved by working at a finer scale and with more topographic and geomorphologic variables (Reed *et al*., 1993; Bridge & Johnson, 2000). However, models can also show high accuracy for some species at the edge of their range, like *Fagus sylvatica* (AUC = 0.94). This species is continental and occurs principally in central Europe, Catalonia being its southern limit. Huntley *et al*. (1989) and Prentice & Helmisaari (1991) showed that *F. sylvatica* is found mainly in sites with annual precipitation close to 1200 mm and is absent from places where the mean temperature of the coldest month is less than –3 °C. Our study suggested the same restrictions due to cold limits and annual precipitation, with values close to those reported by Prentice & Helmisaari (1991).

Finally, there is no general pattern in terms of accuracy for azonal species. Further developments on the chorology and modelling of numerous azonal species are needed to find a straightforward explanation of such patterns.

Land use and biogeographical issues

Logically, the best-modelled species were those growing in a large proportion of their potential habitats. This is the case for *Pinus halepensis*, which, thanks to its high colonizing ability, is only absent from its potential distribution area in the arid inland zone, devoted to agriculture, and in the extreme south-east, as a result of frequent wildfires. Other well-modelled species are dominant trees in forests growing in relatively extreme habitats where human impact has been traditionally low. This is the case for subalpine forests of *Pinus uncinata* and *Abies alba*, and to a lesser degree for mountain forests of *Fagus sylvatica*, which are all located in cold and moist north-facing areas unsuitable for agriculture. The same can also apply to *Quercus suber* and *Castanea sativa*, growing in siliceous, poor soils, which are also inadequate for cropping.

Model accuracy was lower for the remaining species. Among them, forestry pine species (*Pinus nigra*, *P. pinea*, *P. sylvestris*, and *P. pinaster*) were modelled with good accuracy, indicating that they have been planted or have spread mainly within their potential areas. Although we would expect that their current distributions would be strongly influenced by land use history and current management, in the absence of data for these variables bioclimatic models seemed sufficient. These commercial species have been planted at a range of sites known to be climatically suitable. Consequently, climate may provide a reasonable surrogate for land management effects. However, models rightly detected as false negatives occurrences of *P. sylvestris* and *P. nigra* in north-eastern Catalonia, which has a maritime climate, and where these species are not native but planted (Fig. 2). False negatives typically reflect the inability of static equilibrium models such as CTA to deal with historical factors and dynamic behaviour involved when species are planted outside their natural environmental range (Guisan & Zimmermann, 2000).

Minor Eurosiberian species were in the good to low accuracy categories, indicating a variable dependence on average environmental conditions. Some of these species (*Populus tremula*, *Sorbus aria* and *Ilex aquifolium*) are found in large forests of southern Catalonia (Pino *et al*., 2001), at the border of their range, where only small and localized populations persist. In these areas they behave as forest interior species (Forman, 1998) where they are protected from interactions with adjacent human habitats. Alternatively the fair or poor accuracy of models for such azonal species could be a result of the sampling strategy of the IEFC. Sampling favours abundant species and scarce species are likely to have been under-sampled. This explanation is also plausible for widespread but scarce species like *Acer monspessulanum* or *Sorbus aria*. *Q. ilex* and *Q. humilis* make up a third group of species with models of fair accuracy. *Q. ilex* forests would be the dominant vegetation in non siliceous lowland areas in Catalonia and southern France, and *Q. humilis* forests would dominate the Sub-Mediterranean areas bordering the Eurosiberian region (Bolòs *et al*., 1990; Barbéro *et al*., 1998). However, in the majority of their potential area they have been substituted by pine forests, scrublands, grasslands, croplands and urban areas and therefore occupy a small proportion of their potential distribution. This pattern is the likely cause for the low accuracy of models for these species. In addition the model did not accurately predict the inland distribution of *Q. ilex*, where the typical subspecies *Q. ilex* ssp. *ilex* is substituted by another (*Q. ilex* ssp. *ballota*) with contrasting ecological requirements.

ACKNOWLEDGMENTS

This research was funded by the European Union's ATEAM (EVK2-CT-2000–00075) project. Species and environmental data were kindly made available to us by the Centre for Ecological Research and Forestry Applications in Barcelona. We are indebted to Lluis Brotons, Miguel Araújo, John Thompson, Martin Sykes and two anonymous reviewers for comments on successive versions of the manuscript. This paper contributes to the Montpellier — Barcelona European Laboratory (LEA, CNRS) 'Mediterranean Ecosystems in a Changing World'.

REFERENCES

Anon (1999) *S-PLUS 2000 User's Guide*. Data Analysis Products Division, MathSoft, Seattle.

Austin, M.P. (2002) Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling*, **157**, 101–118.

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

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

Bartlein, P.J., Prentice, I.C. & Webb, T. (1986) Climatic response surfaces from pollen data for some eastern North American taxa. *Journal of Biogeography*, **13**, 35–57.

Bolòs, O.D., Vigo, J., Masalles, R.M. & Ninot, J.M. (1990) *Flora manual dels països. Catalans* Pòrtic, Barcelona.

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

Bridge, S.R.J. & Johnson, E.A. (2000) Geomorphic principles of terrain organization and vegetation gradients. *Journal of Vegetation Science*, **11**, 57–70.

Burke, A. (2001) Classification and ordination of plant communities of the Naukluft Mountains, Namibia. *Journal of Vegetation Science*, **12**, 53–60.

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. *Journal of Biogeography*, **27**, 347–358.

Ellenberg, H. (1988) *Vegetation ecology of Central Europe*. Cambridge University Press, Cambridge.

Forman, R.T.T. (1998) *Land mosaics: the ecology of landscapes and regions*. Cambridge University Press, Cambridge.

Franklin, J. (1998) Predicting the distribution of shrub species in southern California from climate and terrain-derived variables. *Journal of Vegetation Science*, **9**, 733–748.

Frescino, T.S., Edwards,T.C. Jr & Moisen, G.G. (2001) Modeling spatially explicit forest structural attributes using Generalized Additive Models. *Journal of Vegetation Science*, **12**, 15–26.

Gracia, C., Burriel, J.A., Ibàñez, J.J., Mata, T. & Vayreda, J. (2000a) *Inventari ecològic i forestal de Catalunya*. Regió Forestal V. CREAF, Bellaterra.

Gracia, C., Burriel, J.A., Ibàñez, J.J., Mata, T. & Vayreda, J. (2000b) *Inventari ecològic i forestal de Catalunya*. Regió Forestal VI. CREAF, Bellaterra.

Gracia, C., Burriel, J.A., Ibàñez, J.J., Mata, T. & Vayreda, J. (2000c) *Inventari ecològic i forestal de Catalunya*. Regió Forestal IV. CREAF, Bellaterra.

Gracia, C., Burriel, J.A., Ibàñez, J.J., Mata, T. & Vayreda, J. (2001a) *Inventari ecològic i forestal de Catalunya*. Regió Forestal II. CREAF, Bellaterra.

Gracia, C., Burriel, J.A., Ibàñez, J.J., Mata, T. & Vayreda, J. (2001b) *Inventari ecològic i forestal de Catalunya*. Regió Forestal III. CREAF, Bellaterra.

Gracia, C., Ibàñez, J.J. & Rodà, F. (in press) Estat dels Boscos de Catalunya. *'Estat Del Medi Ambient a Catalunya', Generalitat de Catalunya* (ed. by Departament de Medi Ambient de la Generalitat de Catalunya), DMA Generalitat de Catalunya, Barcelona.

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. & Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147–186.

Hanley, J.A. & McNeil, B.J. (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, **143**, 29–36.

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

Harvey, L.E. (1996) *Third international conference/workshop on integrating GIS and environmental modelling, Santa Fe, NM Santa Barbara, CA*. National Center for Geographic Information and Analysis, Santa Barbara.

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

Heikkinen, R.K. & Birks, H.J.B. (1996) Spatial and environmental components of variation in the distribution patterns of subartic plant species at Kevo, N Finland — a case study at the meso-scale level. *Ecography*, **19**, 341–351.

Huntley, B., Bartlein, P.J. & Prentice, I.C. (1989) Climatic control of the distribution and abundance of beech (*Fagus* L.) in Europe and North America. *Journal of Biogeography*, **16**, 551–560.

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. *Journal of Biogeography*, **22**, 967–1001.

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

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

Iverson, L.R., Prasad, A. & Schwartz, M.W. (1999) Modelling potential future individual tree-species distributions in the eastern United States under a climate change scenario: a case study with *Pinus virginiana*. *Ecological Modelling*, **115**, 77–93.

Klaus, W. (1989) Mediterranean pines and their history. *Plant Systematics and Evolution*, **162**, 133–163.

Korzukhin, M.D., TerMikaelian, M.T. & Wagner, R.G. (1996) Process versus empirical models: which approach for forest ecosystem management? *Canadian Journal of Forest Research*, **26**, 879–887.

Myers, N., Mittermeier, R.A., Mittermeier, C.G., Da Fonseca, G.A.B. & Kent, J. (2000) Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853–858.

Ninyerola, M., Pons, X. & Roure, J.M. (2000) A methodological approach of climatological modelling of air temperature and precipitation through GIS techniques. *International Journal of Climatology*, **20**, 1823–1841.

Pearce, J. & Ferrier, S. (2000) Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling*, **133**, 225–245.

Peix, J. (1999) *Programa de gestió del risc d'incendi forestal*. Generalitat de Catalunya, Barcelona.

Peng, C. (2000) From static biogeographical model to dynamic global vegetation model: a global perspective on modelling vegetation dynamics. *Ecological Modelling*, **135**, 33–54.

Pino, J., Vayreda, J. & Rodà, F. (2001) Distribución de espécies arbóreas eurosiberianas en relación al tamaño del bosque. *'Montes Para la Sociedad Del Nuevo Lineio III Congreso Forestal Español'*, Vol. 3 (ed. by Consejerća de Medio Ambiente de la Junta de Andalucía), pp. 202–207, Junta de Andalucía, Sevilla.

Prentice, I.C., Bartlein, P.J. & Webb, T. (1991) Vegetation and climate change in eastern North America since the last glacial maximum. *Ecology*, **72**, 2038–2056.

Prentice, I.C. & Helmisaari, H. (1991) Silvics of north European trees: compilation, comparisons and implications for forest succession modelling. *Forest Ecology and Management*, **42**, 79–93.

Prentice, I.C. & Leemans, R. (1990) Pattern and process and the dynamics of forest structure: a simulation approach. *Journal of Ecology*, **78**, 340–355.

Reed, R.A., Peet, R.K., Palmer, L.W. & White, P.S. (1993) Scale dependence on vegetation-environment correlations: a case study of a North Carolina piedmont woodland. *Journal of Vegetation Science*, **4**, 329–340.

Rouget, M., Richardson, D.M., Lavorel, S., Vayreda, J., Gracia, C. & Milton, S.J. (2001) Determinants of distribution of six *Pinus* species in Catalonia, Spain. *Journal of Vegetation Science*, **12**, 491–502.

Starfield, A.M. & Chapin, F.S. (1996) Model of transient changes in arctic and boreal vegetation in response to climate and land use change. *Ecological Applications*, **6**, 842–864.

Swets, K.A. (1988) Measuring the accuracy of diagnostic systems. *Science*, **240**, 1285–1293.

Sykes, M.T. (2001) Modelling the potential distribution and community dynamics of lodgepole pine (*Pinus contorta* Dougl. ex. Loud.) in Scandinavia. *Forest Ecology and Management*, **141**, 69–84.

Sykes, M.T. & Prentice, I.C. (1995) Boreal forest futures: modelling the controls on tree species range limits and transient responses to climate change. *Water, Air and Soil Pollution*, **82**, 415–428.

Sykes, M.T. & Prentice, I.C. (1996) Climate change, tree species distributions and forest dynamics: a case study in the mixed conifer/northern hardwoods zone of northern Europe. *Climatic Change*, **34**, 161–177.

Vayssières, M.P., Plant, R.E. & Allen-Diaz, B.H. (2000) Classification trees: an alternative non-parametric approach for predicting species distributions. *Journal of Vegetation Science*, **11**, 679–694.

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

BIOSKETCHES

The authors of this paper participate in the European Union's ATEAM project.

Wilfried Thuiller and **Sandra Lavorel** are involved within ATEAM in the 'Biodiversity and nature conservation' sector. The main interest of the lead author is to study the determinants of species diversity at the biogeographical level in order to forecast the impacts of future global change on the distribution of diversity. He is currently working on a PhD examining patterns and causes of diversity distribution in Europe and South Africa.

Santi Sabate, Carles Gracia, Joan Pino and **Jordi Vayreda** contribute to the 'Forestry' sector of ATEAM.