DOI: 10.1111/ddi.12736

BIODIVERSITY RESEARCH

WILEY Diversity and Distributions

Modelling the distribution and compositional variation of plant communities at the continental scale

Borja Jiménez-Alfaro[1] | **Susana Suárez-Seoane**[1,2] | **Milan Chytrý**[3] | **Stephan M. Hennekens**[4] | **Wolfgang Willner**[5,6] | **Michal Hájek**[3] | **Emiliano Agrillo**[7] | **Jose M. Álvarez-Martínez**[8] | **Ariel Bergamini**[9] | **Henry Brisse**[10] | **Jörg Brunet**[11] | **Laura Casella**[12] | **Daniel Dítě**[13] | **Xavier Font**[14] | **François Gillet**[15] | **Petra Hájková**[3,16] | **Florian Jansen**[17] | **Ute Jandt**[18,19] | **Zygmunt Kącki**[20] | **Jonathan Lenoir**[21] | **John S. Rodwell**[22] | **Joop H. J. Schaminée**[4] | **Lucia Sekulová**[3] | **Jozef Šibík**[13] | **Željko Škvorc**[23] | **Ioannis Tsiripidis**[24]

[1]Research Unit of Biodiversity (CSIC, UO, PA), Oviedo University, Mieres, Asturias, Spain
[2]Department of Biodiversity and Environmental Management, University of León, León, Spain
[3]Department of Botany and Zoology, Masaryk University, Brno, Czech Republic
[4]Alterra, Wageningen UR, Wageningen, The Netherlands
[5]Vienna Institute for Nature Conservation and Analyses, Vienna, Austria
[6]Department of Botany and Biodiversity Research, University of Vienna, Wien, Austria
[7]Department of Environmental Biology, Sapienza University of Rome, Rome, Italy
[8]Environmental Hydraulics Institute IH, Parque Científico y Tecnológico de Cantabria, Santander, Spain
[9]Biodiversity and Conservation Biology, Swiss Federal Research Institute WSL, Birmensdorf, Switzerland
[10]Aix-Marseille Université, Marignane, France
[11]Southern Swedish Forest Research Centre, Swedish University of Agricultural Sciences, Alnarp, Sweden
[12]ISPRA, Italian National Institute for Environmental Protection and Research, Roma, Italy
[13]Plant Science and Biodiversity Centre, Slovak Academy of Sciences, Institute of Botany, Bratislava, Slovakia
[14]Department of Plant Biology, University of Barcelona, Barcelona, Spain
[15]UMR Chrono-environnement, CNRS – Université Bourgogne Franche-Comté, Besançon Cedex, France
[16]Laboratory of Paleoecology, Department of Vegetation Ecology, The Czech Academy of Sciences, Brno, Czech Republic
[17]Faculty of Agricultural and Environmental Sciences, University of Rostock, Rostock, Germany
[18]Geobotany and Botanical Garden, Institute of Biology, Martin Luther University Halle Wittenberg, Halle (Saale), Germany
[19]German Centre for Integrative Biodiversity Research (iDiv), Halle-Jena-Leipzig, Germany
[20]Botanical Garden, University of Wroclaw, Wrocław, Poland
[21]UR "Ecologie et Dynamique des Systèmes Anthropisés" (EDYSAN UMR 7058, CNRS-UPJV), Université de Picardie Jules Verne, Amiens, France
[22]University of Lancaster, Lancaster, UK
[23]Faculty of Forestry, University of Zagreb, Zagreb, Croatia
[24]Department of Botany, School of Biology, Aristotle University of Thessaloniki, Thessaloniki, Greece

Correspondence
Borja Jiménez-Alfaro, Research Unit of Biodiversity (CSIC/UO/PA), Oviedo University, Mieres, Asturias, Spain.
Email: jimenezalfaro.borja@gmail.com

Abstract

Aim: We investigate whether (1) environmental predictors allow to delineate the distribution of discrete community types at the continental scale and (2) how data completeness influences model generalization in relation to the compositional variation of the modelled entities.

Funding information
European Environment Agency, Grant/Award Number: EEA/NSV/14/006; European Social Fund and the State Budget of the Czech Republic, Grant/Award Number: CZ.1.07/2.3.00/30.0037; Czech Science Foundation, Grant/Award Number: 14-36079G

Editor: Jeremy VanDerWal

Location: Europe.

Methods: We used comprehensive datasets of two community types of conservation concern in Europe: acidophilous beech forests and base-rich fens. We computed community distribution models (CDMs) calibrated with environmental predictors to predict the occurrence of both community types, evaluating geographical transferability, interpolation and extrapolation under different scenarios of sampling bias. We used generalized dissimilarity modelling (GDM) to assess the role of geographical and environmental drivers in compositional variation within the predicted distributions.

Results: For the two community types, CDMs computed for the whole study area provided good performance when evaluated by random cross-validation and external validation. Geographical transferability provided lower but relatively good performance, while model extrapolation performed poorly when compared with interpolation. Generalized dissimilarity modelling showed a predominant effect of geographical distance on compositional variation, complemented with the environmental predictors that also influenced habitat suitability.

Main conclusions: Correlative approaches typically used for modelling the distribution of individual species are also useful for delineating the potential area of occupancy of community types at the continental scale, when using consistent definitions of the modelled entity and high data completeness. The combination of CDMs with GDM further improves the understanding of diversity patterns of plant communities, providing spatially explicit information for mapping vegetation diversity and related habitat types at large scales.

KEYWORDS
community distribution models, ecosystem properties, extent of occurrence, generalized dissimilarity modelling, habitat conservation, plant communities, vegetation

1 | INTRODUCTION

Understanding the spatial variation of vegetation across broad geographical extents is a major challenge in conservation biogeography (Prentice et al., 1992; Venevsky & Veneskaia, 2003). Although conservation initiatives need accurate information on broadscale vegetation patterns to evaluate the extent of occurrence and the area of occupancy of habitat types and related ecosystems (Janssen et al., 2016; Keith, Elith, & Simpson, 2014), detailed vegetation maps are generally not available for large areas, or they are based on expert interpretation at coarse spatial resolution (Bohn & Neuhäusl, 2003). This challenge can be addressed by modelling the relationships between vegetation and environment under functional or species-compositional approaches (Noss, 1990; Whittaker et al., 2005). The functional approach has been widely used for modelling the distribution of plant functional types across biomes (Box, 1981; Greve, Lykke, Blach-Overgaard, & Svenning, 2011; Prentice et al., 1992) but it reflects vegetation structure rather than species composition. In contrast, the compositional approach provides information about the constituent species of vegetation (Ferrier & Guisan, 2006) and it represents a promising tool for characterizing spatial patterns of vegetation diversity.

Under the species-compositional approach, broadscale vegetation patterns can be studied using community distribution models (CDMs), which comprise methods for modelling the occurrence of plant communities in relation to abiotic factors (Chapman & Purse, 2011; Potts, Hedderson, Franklin, & Cowling, 2013). Ferrier and Guisan (2006) suggested that modelling pre-defined plant community types (classifying species assemblages first and then predicting the distribution of the resulting groups) is a straightforward approach to achieve congruence with vegetation classifications. This emphasizes the Clementsian view of communities as discrete entities (Clements, 1936), providing a practical perspective for mapping community types at broad scales. This view of CDMs has been

TABLE 1 A selection of studies using community distribution models (CDMs) to assess the distribution of community types, vegetation types or ecosystems previously identified by plant species composition. The studies were selected to cover as much as possible the variation in modelled entities, geographical area, grid size, statistical methods and predictor variables

	Modelled entity	Geographical area	Grid size	Modelling method	Predictor variables
Brzeziecki, Kienast and Wildi (1995)	Forest community types	Switzerland	1 km	BM	Topography, climate, soil
Lechmere-Oertel and Cowling (1999)	Fynbos and karoo community types	Western Cape (South Africa)	1.8 km	GLM	Topography, climate, soil
Cairns (2001)	Treeline vegetation types	Glacier national Park (USA)	30 m	GLM, ANN, CT	Topography, geomorphology
Miller and Franklin (2002)	Vegetation alliances	Mojave desert (USA)	30 m	GLM, GAM, CT	Climate, topography, geomorphology
Vogiatzakis and Griffiths (2006)	Grassland community types	Lefka Ori, Crete (Greece)	10 m	CT	Topography, geomorphology
Marage and Gégout (2009)	Forest community types	France	50 m	LR	Climate, soil
Dlamini (2011)	Forest classes	Swaziland	0.7 km	EM	Climate, topography, soil
Essl et al. (2011)	Mire community types/habitats	Austria	250 m	Ensemble of methods	Present and future climate, topography, soil
Potts et al. (2013)	Thicket vegetation subtypes	Albany coast (South Africa)	~4 km	Ensemble of methods	Present and future climate
Keith et al. (2014)	Mire ecosystem	South-eastern Australia	250 m	BRT	Climate, topography, soil
Stenzel, Feilhauer, Mack, Metz and Schmidtlein (2014)	Natura 2000 habitat types	Munich region (Germany)	30 m	Maxent	RapidEye satellite imagery
Zhou et al. (2016)	Alpine vegetation groups	Qilian mountains (NW China)	30 m	DT, MLC, RF	Satellite images, climate
Janská et al. (2017)	Vegetation types	Siberia (Russia)	~4 km	Maxent	Climate, soil

ANN, artificial neural network; BM, Bayesian model; BRT, boosted regression tree; CT, classification tree; DT, decision tree; EM, expectation-maximization algorithm; GAM, generalized additive model; GLM, generalized linear model; LG, logistic regression; MLC, maximum-likelihood classification; RF, random forest.

mainly applied for modelling the distribution of vegetation types across small geographical extents, for example landscapes or regions (Table 1). However, modern vegetation classification uses comprehensive datasets collected across larger spatial extents (De Cáceres et al., 2015; Jennings, Faber-Langendoen, Loucks, Peet, & Roberts, 2009; Mucina et al., 2016), opening new possibilities for developing broadscale vegetation mapping products requested by conservation agencies (Álvarez-Martínez et al., 2017; Keith et al., 2014).

Community distribution models need to consider analogous assumptions to those employed in species distribution models (SDMs), especially equilibrium in time and space and stability of biotic interactions (Wiens, 2011), with the main difference that the response variable is a community type rather than a species (Franklin, 2013). A first question that arises is how to model the occurrence of community types as discrete entities (Mücher, Hennekens, Bunce, Schaminée, & Schaepman, 2009). This is related to taxonomic bias (i.e. inconsistent use of species' concepts across the data) but at the community level this can be less problematic because community types are usually defined by many co-existing species. Another issue is how discrete community types are defined (Ferrier & Guisan, 2006), and how consistent these definitions are at broad scales (De Cáceres et al., 2015). Uncertainties related to subjective judgement or ambiguity can occur when a community type is not consistently identified, that is when the entity to be modelled is not homogeneously described across its distribution range (Regan, Colyvan, & Burgman, 2008). This problem affects the classification of any biological entity (Keith et al., 2015) but it may be particularly problematic in CDMs.

In addition, geographical sampling bias (i.e. uneven sampling densities across the study area) is particularly relevant when it leads to reduced representation of environmental variability, which limits model generalization (Thuiller, Brotons, Araújo, & Lavorel, 2004). A crucial issue is whether predictions can be generalized to new geographical areas, considering spatial interpolation (applying predictions to data-deficient parts of the study area), transferability (applying predictions to areas not spatially overlapping with the calibration dataset, but with a similar range of predictor values) and extrapolation (applying predictions to new areas and different range of predictor values;

Peterson et al., 2011). Sampling bias affects model transferability and extrapolation in SDMs (Heikkinen, Marmion, & Luoto, 2012; Wenger & Olden, 2012). In the special case of CDMs, model generalization can be influenced by the compositional variation of the target communities across different regions, assuming the Gleasonian perspective that not all species may respond equally to environmental changes (Gleason, 1926). Therefore, a new question related to CDMs is how to assess the spatial complexity of species assemblages within the target vegetation. This question can be addressed by methods such as generalized dissimilarity modelling (GDM) for analysing the relative influence of geographical and environmental gradients on compositional variation (Ferrier, Manion, Elith, & Richardson, 2007). Generalized dissimilarity modelling might be useful for understanding the influence of sampling bias at broad scales, detecting to what extent the compositional variation of target communities is linked to environmental gradients (Brown, Cameron, Yoder, & Vences, 2014). Since GDM allows to project compositional variation into spatially explicit maps, they also can help in predicting vegetation variation within the extent of occurrence predicted by CDMs.

Here, we combine CDMs and GDM for modelling both the distribution and the compositional variation of community types at the continental scale. We hypothesized that, given a community type encompassing certain compositional consistency at the continental scale, environmental predictors will predict its occurrence under similar assumptions as in SDMs. We also hypothesized that geographical sampling bias influences model generalization (transferability and extrapolation) in relation to the internal compositional variation of the modelled entity across geographical and environmental gradients. We selected as study cases two plant community types that characterize acidophilous beech forests and base-rich fens, representing habitats of conservation concern which have been sampled comprehensively in Europe. We further discuss how both CDMs and GDM reveal spatial vegetation patterns to serve nature conservation at the continental scale.

2 | METHODS

2.1 | Plant community data

We used data stored in the European Vegetation Archive (EVA), a repository of vegetation-plot data containing full records of species co-occurring in relatively small areas (usually <1,000 m^2; Chytrý et al., 2016). Although some vegetation plots were assigned to vegetation types by the original authors, this classification is not consistent across the EVA database and many plots are unclassified. Therefore, the EVA database is currently being used to perform comprehensive classifications of major vegetation types at the level of phytosociological alliances (see www.euroveg.org/eva-database-eva-publications). Alliances are useful units in practical applications, because they are transferable to the European habitat classifications, which is one of the main uses of EVA for assisting nature conservation in Europe (Chytrý et al., 2016; Janssen et al., 2016).

We selected two vegetation datasets representing well-sampled communities in Europe. Our first dataset consists of vegetation plots interpreted as European acidophilous beech forests (hereafter, beech forests) in a classification project using compositional data (Willner et al., 2017). These data include 2,827 vegetation plots (Figure 1) sampled across surface areas of 100–500 m^2 and assigned to the alliance *Luzulo-Fagion sylvaticae*, which is characterized by oligotrophic and acidophilous species (Barbati, Piermaria, & Marchetti, 2007). This community type defines the habitat of conservation concern "9110 *Luzulo-Fagetum* beech forests" (Thauront & Stallegger, 2008), which is protected by the European Habitat Directive 92/43/ECC, and corresponds to the habitat "G1.6 *Fagus* woodland" of EUNIS classification (www.eunis.org).

The second dataset consists of 1,510 vegetation plots (Figure 1) sampled in areas from 1 to 100 m^2 assigned to mountain base-rich fens (hereafter, rich fens) in a pan-European classification (Jiménez-Alfaro et al., 2014). This study found this range of plot sizes appropriate for describing the alliance *Caricion davallianae* (Peterka et al., 2017), which is mainly associated with base-rich and waterlogged soils in cool areas with high precipitation (Essl, Dullinger, Moser, Rabitsch, & Kleinbauer, 2011; Jiménez-Alfaro et al., 2014). This vegetation characterizes the European protected habitat "7230 Alkaline fens" and the EUNIS type "D4.1—Rich fens, including eutrophic tall-herb fens and calcareous flushes and soaks" (Šefferová Stanová, Seffer, & Janák, 2008). The most frequent species of the two community types are presented in Appendix S1.

2.2 | Environmental data

We compiled the 18 bioclimatic variables of WorldClim (Hijmans, Cameron, Parra, Jones, & Jarvis, 2005) at the grid size of 2.5 arc minutes (c. 4.2 km). This grid resolution matches the geographical uncertainty estimated for the occurrence data, ranging in most cases from a few hundred metres to a few kilometres. In addition, we used the solar radiation toolset in ArcGIS 10.4 (ESRI, Redlands, California, USA) to model potential annual mean irradiation – reflecting the amount of energy incident on the earth surface for each grid cell); and obtained an estimate of potential evapotranspiration (PET) from the Global-PET Database (www.cgiar-csi.org) – reflecting the capacity for transpiration flow and primary production when water is not limiting (Fisher, Whittaker, & Malhi, 2011). PET is based on the temperature-radiation equation of Hargreaves, recommended for broadscale studies (Zomer, Trabucco, Straaten, & Bossio, 2006). We finally included a predictor estimating topsoil pH(H_2O) as provided by the ISRIC World Soil Information (http://www.isric.org/). This variable is useful for comparing the regional dominance of calcareous and non-calcareous bedrock. Other soil variables from the same source (e.g. fraction of silt and clay) were also explored, but they did not provide any contribution to the models and, consequently, they were not used.

We managed all spatial data with ArcGIS using the European ETRS89 (LAEA) projection system to minimize geographical distortion. Solar radiation, PET and pH were obtained at 1 km grid resolution and then aggregated to 4.2 km as the spatial resolution of the

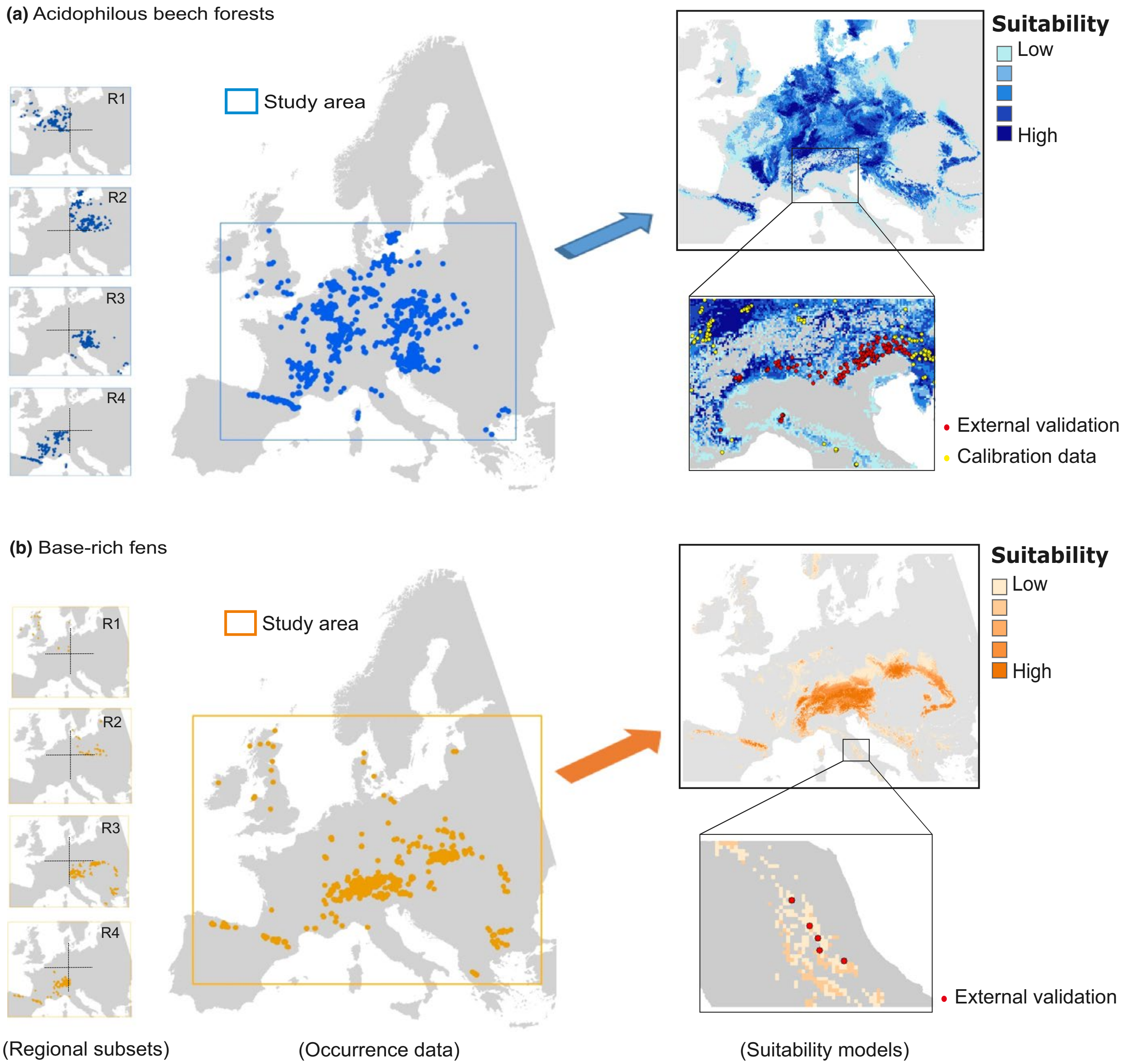

FIGURE 1 Occurrence data and habitat suitability predicted with community distribution models (CDMs) for (a) acidophilous beech forests and (b) base-rich fens in Europe. The study area was defined according to the known distribution range of the two community types and partitioned in four regional subsets based on geographical quadrants (R1–R4). Regions zoomed out from suitability maps show the distribution of localities used for external validation. [Colour figure can be viewed at wileyonlinelibrary.com]

study. Collinearity between pairs of variables was tested using the Pearson correlation coefficient *r*. Highly correlated predictors ($r > .7$) were excluded from each pair, choosing the variable with stronger ecological meaning for modelling plant diversity in temperate regions (e.g. Franklin et al., 2013). We finally kept eight predictors to reflect major environmental gradients that are supposed to influence the distribution and composition of the studied communities: temperature seasonality, mean temperature of the wettest quarter, annual precipitation, precipitation seasonality, precipitation of the warmest quarter, solar radiation, potential evapotranspiration and topsoil pH.

2.3 | Community distribution models

We used Maxent V.3.3.3k, a package implementing a presence-background method that combines machine-learning and statistical inference (Elith et al., 2011), to model the occurrence of community types. Maxent provides good performance in transferability to new climatic scenarios (Hijmans & Graham, 2006) and geographical regions (Heikkinen et al., 2012) and with low sample sizes when compared with presence-only or presence–absence methods (Elith et al., 2006; Gibson, Barrett, & Burbidge, 2007). Maxent is recommended for distribution data gathered from records sampled without a

TABLE 2 Averaged model performance (mean ± *SD*) of *n*-fold cross-validations performed for predicting the occurrence of beech forests and base-rich fens in Europe. Interpolation was based on random fourfold CV in the whole study area (cf. baseline models). Geographical transferability was computed with occurrences from three regions (3R) in the remaining one (R′). Geographical interpolation shows the predictive value of 75% of occurrences within one focal region ($1R_{75}$) to the remaining 25% of the same region ($1R_{25}$). Geographical extrapolation reflects the predictive value of 75% of occurrences within one focal region ($1R_{75}$) to the whole occurrences of the other three regions separately (3R′). AUC_{diff} reflects AUC for training data minus AUC for testing data

	Interpolation baseline model (all study area)	Geographical transferability (3R to 1R′)	Geographical interpolation ($1R_{75}$ to $1R_{25}$)	Geographical extrapolation ($1R_{75}$ to 3R′)
Cross-validation	Fourfold	Fourfold	Fourfold	12-fold
Acidophilous beech forests				
AUC	0.856 ± 0.013	0.697 ± 0.131	0.908 ± 0.030	0.616 ± 0.230
AUC_{diff}	0.018 ± 0.015	0.175 ± 0.127	0.007 ± 0.013	0.309 ± 0.230
Omission rate (min)	0.001 ± 0.002	0.002 ± 0.005	0.003 ± 0.006	0.178 ± 0.192
Omission rate (10th)	0.134 ± 0.037	0.494 ± 0.257	0.151 ± 0.043	0.693 ± 0.329
Base-rich fens				
AUC	0.907 ± 0.008	0.725 ± 0.208	0.781 ± 0.277	0.603 ± 0.214
AUC_{diff}	0.013 ± 0.009	0.189 ± 0.196	0.078 ± 0.125	0.256 ± 0.152
Omission rate (min)	0.004 ± 0.005	0.020 ± 0.040	0.073 ± 0.086	0.280 ± 0.258
Omission rate (10th)	0.117 ± 0.041	0.536 ± 0.443	0.145 ± 0.054	0.564 ± 0.354

unique sampling protocol that contain no reliable information about species absences (Phillips et al., 2009). Indeed, our data also rely on the compilation of regional and national datasets collected under different, usually subjective sampling schemes. Moreover, the spatial accuracy of the sampling units (up to a few kilometres) makes impossible to infer absences in a grid cell where another type has been recorded, since many community types occur within the same grid cell because of local habitat variation.

We first calibrated a "baseline model" using the default parameters of Maxent and 10,000 background points randomly selected within the area that has been sampled for each community type (Figure 1). We consider this model as one scenario with non-biased data and high data completeness. Model outputs were mapped using a minimum threshold of suitability based on the equate entropy in Maxent, because it provides an intermediate threshold between the most conservative (minimum predicted area) and the least conservative (equal sensitivity and specificity) options (Morán-Ordóñez, Suárez-Seoane, Elith, Calvo, & de Luis, 2012). Model discrimination was assessed with the area under the receiver-operating characteristic (ROC) curve (AUC), using the fraction of the study area predicted as suitable to calculate the commission error (1 – specificity) as implemented in Maxent. AUC was averaged from (1) model interpolation using a random fourfold cross-validation with 10 replicates (each replicate using 75% of occurrences for training and 25% for validation) and (2) external evaluation using a dataset provided by the Italian National Vegetation Database (BVN/ISPRA; Casella, Bianco, Angelini, & Morroni, 2012). These data consisted of plots sampled in beech forests from the Italian Alps (*n* = 255) and rich fens from the Apennines (*n* = 16) that were identified by the database as the same alliances analysed here.

We tested for geographical transferability by simulating scenarios in which the available dataset is sampled in ¾ of the study area, partitioning the data into subsets based on four geographical quadrants (Figure 1). The regions defined by these quadrants showed a clear climatic differentiation for the two vegetation types, as reflected by discriminant analyses computed with the environmental predictors (Appendix S2). We performed four regionally biased models using the occurrences of 3 of 4 geographical regions for calibration and the remaining one for evaluation, that is using a "geographical fourfold cross-validation" (Radosavljevic & Anderson, 2014). We also tested for extrapolation by simulating a more challenging scenario in which calibration data are taken from one region only. Since these models differ largely in the number of occurrences and regional features, we compared their performance with interpolation models computed for each region using the same calibration data. A random selection of 75% of occurrences from each region was used for calibration ($R1_{75}$, $R2_{75}$, $R3_{75}$ and $R4_{75}$) and the remaining 25% were kept for evaluation. Thus, a model calibrated with $R1_{75}$ was tested first for interpolation within the region R1 using $R1_{25}$ as evaluation data and, secondly, predictions were extrapolated separately to the other geographical regions (R2, R3 and R4). To deal with the sampling bias of calibration data, background points were masked in all models to the extent of the region/s used as data sources.

We quantified model overfitting for all model scenarios as the difference between training and testing AUC ($AUC_{DIFF} = AUC_{training} – AUC_{testing}$), where higher values indicate loss of performance (Warren & Seifert, 2011). To obtain a direct estimate of model performance using presence-only data, we also evaluated omission rates, which reflect the number of false-negative predictions based on a threshold of habitat suitability defined by the training data. We selected the minimum omission rate (minimum value of habitat suitability of training data) and the 10th percentile (minimum threshold for the 90% of occurrences with the highest suitability values) provided by Maxent. Omission rates range from 0 to 1, low values indicating good model performance and thus low overfitting. Spatial predictions of habitat suitability achieved for all modelling scenarios, including the baseline model, were compared with the Pearson correlation coefficient based on a random selection of 8,000 points. We also assessed potential differences in the contribution of ecological drivers to different models using the Jackknife evaluation method in Maxent.

2.4 | Generalized distribution modelling

We used generalized distribution modelling (GDM) as a method to predict spatial patterns of turnover in species composition (Ferrier et al., 2007) using the *gdm* package in R (Manion et al., 2017). Generalized distribution modelling uses a nonlinear matrix regression technique for analysing spatial patterns in compositional dissimilarity, providing fitted I-splines to describe the relationships between a dissimilarity metric (beta diversity) and predictors, then estimating the partial deviance explained by each predictor (Fitzpatrick et al., 2013). For each community type, we created a presence–absence matrix with the full species composition of each plot and generated site pairs described by their compositional (Bray–Curtis) distance, geographical coordinates and environmental variables. After running the GDM analysis, we quantified the importance of geographical distance and environmental predictors and computed predictions for the whole study area to visualize patterns of compositional dissimilarity. We created a series of maps to show the spatial variation for the first three axes of a principal component analysis (PCA) separately; and for the three PCA axes combined into a unique RGB (red-green-blue) layer.

3 | RESULTS

3.1 | Baseline model

The baseline models calibrated with the whole dataset of beech forests showed the highest habitat suitability in Central Europe, with a wide distribution range between Southern Scandinavia and the margins of the southernmost temperate mountains (Figure 1). The equate entropy threshold defined as non-suitable those areas with habitat suitability <0.18, distributed through the Mediterranean and Atlantic regions, as well as the easternmost regions of the study area. The most important predictors estimated from the Jackknife evaluation method were temperature seasonality (42% of total contribution), summer precipitation (21%), potential evapotranspiration (19%) and soil pH (10%) (Figure 2). For rich fens, the highest habitat suitability was predicted for the Central European mountain systems, especially for the Alps and the Carpathians, but also for the Pyrenees and Bulgarian mountains (Figure 1). The equate entropy threshold was 0.13, identifying as non-suitable the areas far from inland European mountains and continental islands. The main predictors were summer precipitation (74%), solar radiation (12%), potential evapotranspiration (5%) and temperature seasonality (4%) (Figure 2). Appendix S3 shows the variable contributions and response curves for both community types.

External validation suggested good model performance for the two vegetation types (Figure 1). In the beech forests, 100% of the occurrence data were predicted as suitable for the minimum training presence threshold and 70% for the 10th training presence threshold, indicating omission rates of 0 and 0.3, respectively. Mean habitat suitability was 0.38 (*SD* ± 0.13), AUC was 0.743, and AUC_{diff} was 0.120. For rich fens, 100% of the data were predicted as suitable using the minimum training presence threshold and 99% for the 10th training presence threshold. Mean habitat suitability was 0.22 (*SD* ± 0.05), AUC was 0.812, and AUC_{diff} was 0.099.

3.2 | Interpolation, transferability and extrapolation

Model interpolation of the baseline model showed better performance (higher average AUC, lower AUC_{diff} and lower omission rates) than models testing geographical transferability from three regions to one (Table 2). Nevertheless, many of the latter models showed fair performance (e.g. AUC ~ 0.7 or higher). We also found strong correlations between the spatial predictions computed with the baseline models and the subsets used for geographical transferability (Pearson's r, $n = 4$; mean ± *SD*; 0.92 ± 0.04 in forests; 0.96 ± 0.03 in fens). In the two community types, partial contribution of variables in transferability was very similar to that contributing in the baseline models (Figure 2). In contrast, extrapolation from a unique region to the others provided poor performance reflected by AUC (~0.6 or lower), high overfitting and high omission rates (Table 2). These results were worse than those achieved from the interpolation within each region. The correlations between models derived from extrapolation and the baseline model were significant, but the coefficients were lower than those from transferability models (0.52 ± 0.16 in forests; 0.56 ± 0.37 in fens). In the two community types, the importance of environmental variables differed strongly between the subsets and the baseline model computed for all regions (Figure 2).

3.3 | Compositional variation

The total deviance explained by GDM was 19% and 15% for beech forests and rich fens, respectively. In both cases, variable contribution was higher for geographical distance than for environmental distances (Table 3). In beech forests, the most important predictors (after geographical distance) were continental variation in solar radiation and temperature seasonality. According to the predicted dissimilarities, the first and second PCA axes represented main

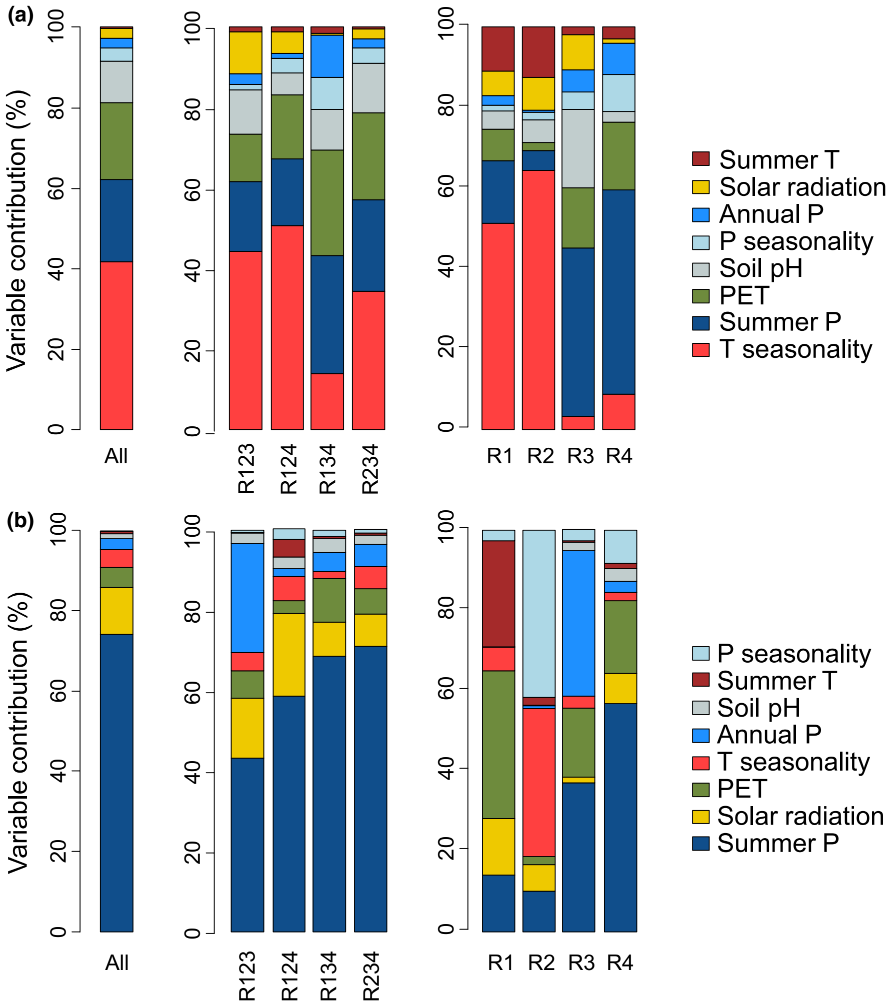

FIGURE 2 Variable contribution in community distribution models (CDMs) computed for plant community types of (a) acidophilous beech forests and (b) base-rich fens in Europe, using combinations of geographical regions (R1, R2, R3 and R4) as input data. Regions as in Figure 1. Bars in the central panel show results for models computed with three regions for testing transferability, and bars in the right panel show models computed for each region separately for testing extrapolation. [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 3 Variable importance (Var. Imp., in %) and relative contribution (Rel. Cont., summarizing coefficients in I-splines) of predictor variables in generalized dissimilarity modelling computed for the variation on plant species composition across acidophilous beech forests (total explained deviance = 19%) and base-rich fens (total explained deviance = 15%) in Europe

	Acidophilous beech forests		Base-rich fens	
	Var. Imp.	Rel. Cont.	Var. Imp.	Rel. Cont.
Geographical distance	17.2%	0.5	21.1%	0.5
Potential evapotranspiration	4.7%	0.3	1.1%	0.1
Regional soil pH	1.4%	0.1	0.6%	0.1
Summer precipitation	2.1%	0.3	8.2%	0.4
Solar radiation	7.8%	0.9	0.6%	0.1
Temperature seasonality	7.5%	0.8	0	0
Temperature of wettest quarter	1.5%	0.1	1.3%	0.2

variation across latitude and longitude, respectively, while the third axis differentiated lowland and mountain regions (Figure 3a). The combination of the three axes showed similar patterns of species composition (similar colours) in areas aggregated mainly in Central Europe, Eastern Europe, Atlantic regions and northern Europe. In rich fens, the best predictor after geographical distance was the continental variation in summer precipitation (Figure 3b). Predictions for the first and second PCA axes reflected longitudinal and latitudinal patterns, while the third PCA mainly pointed to the higher Central European mountain systems. The combination of the three PCA axes predicted the strongest differences along the longitudinal gradient.

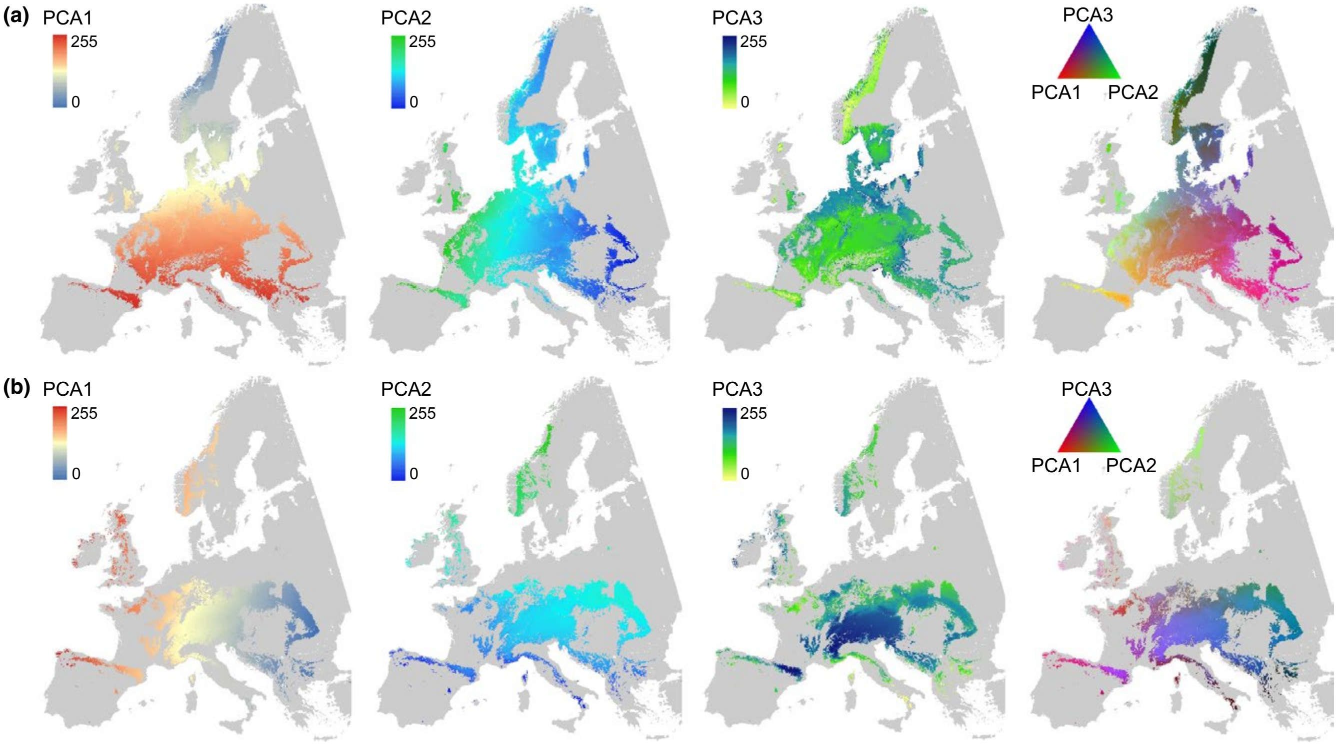

FIGURE 3 Spatial predictions obtained from generalized dissimilarity modelling for the compositional variation of (a) acidophilous beech forests and (b) base-rich fens in Europe. Maps show variation in community dissimilarity according to the values predicted for three axes of a principal component analysis (PCA), and the combination of the three axes into a unique model output using an RGB (red-green-blue) colour plate. The strongest differences in colour reflect the strongest predicted differences in species composition. All maps are masked to suitable regions predicted for each community type as in Figure 1. [Colour figure can be viewed at wileyonlinelibrary.com]

4 | DISCUSSION

4.1 | Modelling the distribution of discrete community types

This study demonstrates how correlative CDMs predict the distribution of discrete community types at the continental scale. The spatial patterns predicted here are in accordance with country-level maps of European habitats related to beech forests (Thauront & Stallegger, 2008) and mountain alkaline fens (Šefferová Stanová et al., 2008); and also with the maps provided for the European Red list of Habitats (Janssen et al., 2016). Nevertheless, our predictions offer a significant improvement to these initiatives by providing continental maps at relatively high resolution, covering non-sampled areas and providing further information about environmental drivers. We also were able to predict the two community types in non-sampled regions, supporting the role of deterministic environmental factors for explaining the distributions of discrete vegetation entities.

The predicted areas of suitability for acidophilous beech forests were mainly correlated with the interaction of precipitation and temperature, which is consistent with the ecological preferences known for these communities (Leuschner, Ellenberg, & Sutcliffe, 2017) and their predicted distribution at regional extents (e.g. *Luzulo-Fagion* in France, Marage & Gégout, 2009). Interestingly, this community type was not predicted to occur in the European Atlantic region, where another beech-dominated community type influenced by oceanic climate occurs on acidic soils (*Ilici-Fagion* alliance in northern Spain and Western France, Marage & Gégout, 2009). Similarly, the suitable areas for base-rich fens were associated with the main climatic drivers influencing these habitats in European mountains, namely summer precipitation and solar radiation (Essl et al., 2011). As it was shown in our models, mountain base-rich fens hardly occur in the Mediterranean region, where summer drought limits water availability (Jiménez-Alfaro et al., 2014).

We note that both community types also depend on soil factors, since they only occur on acidic soils (acidophilous beech forests) and in calcareous hard-water habitats (base-rich fens). However, the expected influence of soil pH was only confirmed in the case of the beech forests. This likely reflects their occurrence in zonal habitats of regions with predominant acidic bedrocks. In contrast, base-rich fens are azonal habitats usually associated with springs, while not all base-rich bedrocks support the existence of springs (Grootjans et al., 2005). This makes it difficult to predict the distribution of base-rich fens using coarse-grain variables related to soil conditions. Although the spatial accuracy of our occurrence data (~1–2 km) is good enough to estimate the extent of occurrence of the modelled entities, it seems also limited by high commission error, especially in azonal habitats. This issue, caused by the fact that broadscale predictors do not account for variation in local conditions, is actually a well-known limitation of SDMs (Guisan & Thuiller, 2005) or biodiversity models based on remote sensing (Rocchini et al., 2010).

4.2 | Model generalization and compositional dissimilarity

Our results support the hypothesis that CDMs strongly depend on data completeness at broad scales, as indicated by lower model performance under increasing geographical sampling bias. The models perform better when more regions are used for calibration and when used for interpolation, in agreement with the known effect of sampling bias in SDMs (Anderson & Gonzalez, 2011; Heikkinen et al., 2012; Phillips et al., 2009). In general, we found geographical transferability good enough to be applied from a large proportion of the study region to non-sampled areas. However, extrapolation seems too challenging for obtaining good models at the continental scale, as it has been found in similar scenarios using SDMs (Wenger & Olden, 2012). Under a scenario of strong sampling bias, our extrapolation exercise showed very low performance and low correlation with the baseline model. This was expected given the reduced geographical extent of the calibration data in the context of the whole study area, producing truncated responses when models are computed with data from one region only (Heikkinen et al., 2012; Thuiller et al., 2004). This confirms our hypothesis that increasing geographical sampling bias significantly decreases model performance.

The performance of spatial transferability and extrapolation was also dependent on the characteristics of the regions selected for model calibration, as it has been shown in SDMs (Suárez-Seoane, Virgós, Terroba, Pardavila, & Barea-Azcón, 2014). Some of these regions are more representative of the environmental variation available also outside the focal region, as indicated by the variance of AUC values and omission rates. This explains the lower model performance found in base-rich fens, for which some of the geographical partitions represented marginal regions with a low proportion of the environmental variation captured therein. The performance of model generalization was related to the compositional variation of the community types and the influence of environmental drivers observed with GDM. Indeed, habitat suitability and species dissimilarity were influenced by similar predictors in beech forests (e.g. temperature seasonality) and rich fens (e.g. summer precipitation). The influence of geographical distance was, however, more important than environmental variation in both community types, indicating that compositional variation is also related to assembly processes such as biogeographical history or dispersal limitation (HilleRisLambers, Adler, Harpole, Levine, & Mayfield, 2012). These results suggest that model transferability and extrapolation are limited by the truncation of the environmental space in the training data, but also by the internal variation in species composition across geographical gradients. This challenges the assumption of equilibrium in space and time, especially when increasing regional sampling bias. Thus, extrapolation from a restricted geographical region to the whole continent seems to be a risky business in terms of statistical performance and reliability of spatial predictions, except for the hypothetical case that a focal region contains the full gradient of environmental conditions that is available at the continental extent.

4.3 | Applications for broadscale vegetation mapping

In contrast with the geographical extent traditionally used in CDMs, this is one of few studies modelling the distribution of plant community types at the continental scale. Similar models may be useful for estimating the extent of occurrence and the area of occupancy of vegetation types in ecosystem assessment (Rodríguez et al., 2015) and for supporting projections under past (Potts et al., 2013) and future (Keith et al., 2014) environmental conditions. The distribution maps produced with our CDMs can be also used to estimate the potential area of occupancy of specific habitats related to forests and mires, which is a current requirement for conservation in Europe (Álvarez-Martínez et al., 2017). We note that, in analogy with the spatial hierarchy of species beta-niches in SDMs (Ackerly, Schwilk, & Webb, 2006), predictions in grid cells of a few km represent heterogeneous landscapes where different community types may co-occur. The potential area of occupancy therefore assumes that different community types can occur in the same grid cell, which is an important difference to vegetation mapping based on high-resolution data (Miller & Franklin, 2002), or CDMs using remote sensing (Álvarez-Martínez et al., 2017). In addition, GDM offers complementary information about the compositional variation within the potential area of occupancy. GDM has been used to address biodiversity patterns and spatial regionalization of different organisms (Brown et al., 2014; Fitzpatrick et al., 2013; Lasram, Hattab, Halouani, Romdhane, & Le Loc'h, 2015), while similar approaches for modelling species turnover have also been suggested with remote sensing applications (Rocchini et al., 2010). Nevertheless, these approaches are still poorly integrated in the analysis of broadscale vegetation patterns.

Although we modelled vegetation alliances as entities that are known to respond to broadscale environmental factors in Europe (Jiménez-Alfaro et al., 2014; Mucina et al., 2016), we conclude that any community type at any hierarchical level may be modelled at continental extent, provided it is consistently defined by species composition and constrained by environmental factors. A key assumption of this approach is that community types defined by species composition are discrete entities that respond to abiotic factors, thus adopting a Clementsian view. This community resolution encapsulates, to a certain extent, the effect of assembly processes such as environmental filtering and biotic interactions (HilleRisLambers et al., 2012). A major limitation attributed to SDMs – the lack of information about biotic interactions (Wisz et al., 2013) – is thus expected to have a minor impact on models computed for plant community types that involve interspecific interactions implicitly (Lortie et al., 2004). However, community types defined by similar species assemblages and co-occurring niches are not floristically identical across their distribution range, and complementary methods such as GDM allow to integrate a Gleasonian perspective, evaluating the extent of internal variation in species responses. The combination of both CDMs and

GDM is therefore a promising tool for mapping large-scale vegetation patterns, by considering that continental vegetation classifications are nowadays on the agenda of biodiversity research, and similar data will be accessible due to the increasing availability of international databases (Franklin, Serra-Diaz, Syphard, & Regan, 2017).

ACKNOWLEDGEMENTS

We thank all scientists who collected vegetation-plot data in the field and integrated these data in European databases. We also thank the European Environment Agency (project EEA/NSV/14/006) for funding; and P. de Ruffray, Y. Ferrez and G. Corriol for data gathering. BJ-A was funded by the European Social Fund and the State Budget of the Czech Republic (project no. CZ.1.07/2.3.00/30.0037), and MC and MH by the Czech Science Foundation (project no. 14-36079G, Centre of Excellence PLADIAS).

ORCID

Borja Jiménez-Alfaro http://orcid.org/0000-0001-6601-9597
Susana Suárez-Seoane http://orcid.org/0000-0001-7656-4214
Milan Chytrý http://orcid.org/0000-0002-8122-3075
Michal Hájek http://orcid.org/0000-0002-5201-2682
Emiliano Agrillo http://orcid.org/0000-0003-2346-8346
Jose M. Álvarez-Martínez http://orcid.org/0000-0002-8150-0802
Ariel Bergamini http://orcid.org/0000-0001-8816-1420
Jörg Brunet http://orcid.org/0000-0003-2667-4575
Laura Casella http://orcid.org/0000-0003-2550-3010
Xavier Font https://orcid.org/0000-0002-7253-8905
François Gillet http://orcid.org/0000-0002-3334-1069
Florian Jansen http://orcid.org/0000-0002-0331-5185
Zygmunt Kącki http://orcid.org/0000-0002-2241-1631
Jonathan Lenoir http://orcid.org/0000-0003-0638-9582
Jozef Šibík http://orcid.org/0000-0002-5949-862X

REFERENCES

Ackerly, D. D., Schwilk, D. W., & Webb, C. O. (2006). Niche evolution and adaptive radiation: Testing the order of trait divergence. *Ecology*, *87*, S50–S61. https://doi.org/10.1890/0012-9658(2006)87[50:NEAART]2.0.CO;2

Álvarez-Martínez, J., Jiménez-Alfaro, B., Barquín, J., Ondivieta, B., Silió-Calzada, A., & Juanes, A. (2017). Modelling the area of occupancy of habitat types with remote sensing. *Methods in Ecology and Evolution*, https://doi.org/10.1111/2041-210X.12925

Anderson, R. P., & Gonzalez, I. (2011). Species-specific tuning increases robustness to sampling bias in models of species distributions: An implementation with Maxent. *Ecological Modelling*, *222*, 2796–2811. https://doi.org/10.1016/j.ecolmodel.2011.04.011

Barbati, A., Piermaria, C., & Marchetti, M. (2007). *European forest types – Categories and types for sustainable forest management reporting and policy*. Technical report No 9/2006, Copenhagen, Denmark: European Environmental Agency.

Bohn, U., & Neuhäusl, R. (Eds.) (2003). *Map of the natural vegetation of Europe. Explanatory text with CD-ROM*. Bonn, Germany: German Federal Agency for Nature Conservation.

Box, E. E. O. (1981). *Macroclimate and plant forms – An introduction to predictive modeling in phytogeography*. The Hague, the Netherlands: Junk.

Brown, J. L., Cameron, A., Yoder, A. D., & Vences, M. (2014). A necessarily complex model to explain the biogeography of the amphibians and reptiles of Madagascar. *Nature Communications*, *5*, 5046. https://doi.org/10.1038/ncomms6046

Brzeziecki, B., Kienast, F., & Wildi, O. (1995). Modelling potential impacts of climate change on the spatial distribution of zonal forest communities in Switzerland. *Journal of Vegetation Science*, *6*, 257–268. https://doi.org/10.2307/3236221

Cairns, D. M. (2001). A comparison of methods for predicting vegetation type. *Plant Ecology*, *156*, 3–18. https://doi.org/10.1023/A:1011975321668

Casella, L., Bianco, P., Angelini, P., & Morroni, M. (2012). Italian national vegetation database (BVN/ISPRA). *Biodiversity and Ecology*, *4*, 404. https://doi.org/10.7809/b-e

Chapman, D. S., & Purse, B. V. (2011). Community versus single-species distribution models for British plants. *Journal of Biogeography*, *38*, 1524–1535. https://doi.org/10.1111/j.1365-2699.2011.02517.x

Chytrý, M., Hennekens, S. M., Jiménez-Alfaro, B., Knollová, I., Dengler, J., Jansen, F., … Yamalov, S. (2016). European vegetation archive (EVA): An integrated database of European vegetation plots. *Applied Vegetation Science*, *19*, 173–180. https://doi.org/10.1111/avsc.12191

Clements, F. E. (1936). Nature and structure of the climax. *The Journal of Ecology*, *24*, 252. https://doi.org/10.2307/2256278

De Cáceres, M., Chytrý, M., Agrillo, E., Attorre, F., Botta-Dukát, Z., Capelo, J., … Willner, W. (2015). A comparative framework for broad-scale plot-based vegetation classification. *Applied Vegetation Science*, *18*, 543–560. https://doi.org/10.1111/avsc.12179

Dlamini, W. M. (2011). A data mining approach to predictive vegetation mapping using probabilistic graphical models. *Ecological Informatics*, *6*, 111–124. https://doi.org/10.1016/j.ecoinf.2010.12.005

Elith, J., H. Graham, C., P. Anderson, R., Dudiík, M., Ferrier, S., Guisan, A., … Zimmerman, N. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, *29*, 129–151. https://doi.org/10.1111/j.2006.0906-7590.04596.x

Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., & Yates, C. J. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, *17*, 43–57. https://doi.org/10.1111/j.1472-4642.2010.00725.x

Essl, F., Dullinger, S., Moser, D., Rabitsch, W., & Kleinbauer, I. (2011). Vulnerability of mires under climate change: Implications for nature conservation and climate change adaptation. *Biodiversity and Conservation*, *21*, 655–669.

Ferrier, S., & Guisan, A. (2006). Spatial modelling of biodiversity at the community level. *Journal of Applied Ecology*, *43*, 393–404. https://doi.org/10.1111/j.1365-2664.2006.01149.x

Ferrier, S., Manion, G., Elith, J., & Richardson, K. (2007). Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Diversity and Distributions*, *13*, 252–264. https://doi.org/10.1111/j.1472-4642.2007.00341.x

Fisher, J. B., Whittaker, R. J., & Malhi, Y. (2011). ET come home: Potential evapotranspiration in geographical ecology. *Global Ecology and Biogeography*, *20*, 1–18. https://doi.org/10.1111/j.1466-8238.2010.00578.x

Fitzpatrick, M. C., Sanders, N. J., Normand, S., Svenning, J.-C., Ferrier, S., Gove, A. D., & Dunn, R. R. (2013). Environmental and historical imprints on beta diversity: Insights from variation in rates of species turnover along gradients. *Proceeding of the Royal Society B*, *280*, 20131201. https://doi.org/10.1098/rspb.2013.1201

Franklin, J. (2013). Mapping Vegetation from Landscape to Regional Scales. In E. van der Maarel & J. Franklin (Ed.), *Vegetation Ecology* (pp. 486–508). Oxford, UK: John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118452592.ch16

Franklin, J., Davis, F. W., Ikegami, M., Syphard, A. D., Flint, L. E., Flint, A. L., & Hannah, L. (2013). Modeling plant species distributions under future climates: How fine scale do climate projections need to be? *Global Change Biology*, *19*, 473–483. https://doi.org/10.1111/gcb.12051

Franklin, J., Serra-Diaz, J. M., Syphard, A. D., & Regan, H. M. (2017). Big data for forecasting the impacts of global change on plant communities. *Global Ecology and Biogeography*, *26*, 6–17. https://doi.org/10.1111/geb.12501

Gibson, L., Barrett, B., & Burbidge, A. (2007). Dealing with uncertain absences in habitat modelling: A case study of a rare ground-dwelling parrot. *Diversity and Distributions*, *13*, 704–713. https://doi.org/10.1111/j.1472-4642.2007.00365.x

Gleason, H. A. (1926). The individualistic concept of the plant association. *Bulletin of the Torrey Botanical Club*, *53*, 7. https://doi.org/10.2307/2479933

Greve, M., Lykke, A. M., Blach-Overgaard, A., & Svenning, J.-C. (2011). Environmental and anthropogenic determinants of vegetation distribution across Africa. *Global Ecology and Biogeography*, *20*, 661–674. https://doi.org/10.1111/j.1466-8238.2011.00666.x

Grootjans, A., Alserda, A., Bekker, R., Janáková, M., Kemmers, R., Madaras, M., … Wolejko, L. (2005). Calcareous spring mires in Slovakia: Jewels in the crown of the Mire Kingdom. *Mires*, *From Siberia to Tierra del Fuego*. *Stapfia*, *85*, 97–115.

Guisan, A., & Thuiller, W. (2005). Predicting species distribution: Offering more than simple habitat models. *Ecology Letters*, *8*, 993–1009. https://doi.org/10.1111/j.1461-0248.2005.00792.x

Heikkinen, R. K., Marmion, M., & Luoto, M. (2012). Does the interpolation accuracy of species distribution models come at the expense of transferability? *Ecography*, *35*, 276–288. https://doi.org/10.1111/j.1600-0587.2011.06999.x

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, *25*, 1965–1978. https://doi.org/10.1002/(ISSN)1097-0088

Hijmans, R. J., & Graham, C. H. (2006). The ability of climate envelope models to predict the effect of climate change on species distributions. *Global Change Biology*, *12*, 2272–2281. https://doi.org/10.1111/j.1365-2486.2006.01256.x

HilleRisLambers, J., Adler, P. B., Harpole, W. S., Levine, J. M., & Mayfield, M. M. (2012). Rethinking community assembly through the lens of coexistence theory. *Annual Review of Ecology, Evolution, and Systematics*, *43*, 227–248. https://doi.org/10.1146/annurev-ecolsys-110411-160411

Janská, V., Jiménez-Alfaro, B., Chytrý, M., Divíšek, J., Anenkhonov, O., Korolyuk, A., … Culek, M. (2017). Palaeodistribution modelling of European vegetation types at the last glacial maximum using modern analogues from Siberia: Prospects and limitations. *Quaternary Science Reviews*, *159*, 103–115. https://doi.org/10.1016/j.quascirev.2017.01.011

Janssen, J. A. M., Rodwell, J. S., García Criado, M., Gubbay, S., Haynes, T., Nieto, A., … Valachovič, M. (2016). *European red list of habitats – Part 2. Terrestrial and freshwater habitats*. Luxembourg: Publications Office of the European Union.

Jennings, M. D., Faber-Langendoen, D., Loucks, O. L., Peet, R. K., & Roberts, D. (2009). Standards for associations and alliances of the U.S. National Vegetation Classification. *Ecological Monographs*, *79*, 173–199. https://doi.org/10.1890/07-1804.1

Jiménez-Alfaro, B., Hájek, M., Ejrnaes, R., Rodwell, J., Pawlikowski, P., Weeda, E. J., … Díaz, T. E. (2014). Biogeographic patterns of base-rich fen vegetation across Europe. *Applied Vegetation Science*, *17*, 367–380. https://doi.org/10.1111/avsc.12065

Keith, D. A., Elith, J., & Simpson, C. C. (2014). Predicting distribution changes of a mire ecosystem under future climates. *Diversity and Distributions*, *20*, 440–454. https://doi.org/10.1111/ddi.12173

Keith, D. A., Rodríguez, J. P., Brooks, T. M., Burgman, M. A., Barrow, E. G., Bland, L., & Spalding, M. D. (2015). The IUCN red list of ecosystems: Motivations, challenges, and applications. *Conservation Letters*, *8*, 214–226. https://doi.org/10.1111/conl.12167

Lasram, F. B. R., Hattab, T., Halouani, G., Romdhane, M. S., & Le Loc'h, F. (2015). Modeling of beta diversity in tunisian waters: Predictions using generalized dissimilarity modeling and bioregionalisation using fuzzy clustering. *PLoS One*, *10*, e0131728. https://doi.org/10.1371/journal.pone.0131728

Lechmere-Oertel, R. G., & Cowling, R. M. (1999). Predicting the distribution of fynbos and succulent karoo biome boundaries and plant communities using generalised linear models and geographic information systems. *South African Journal of Botany*, *65*, 89–96. https://doi.org/10.1016/S0254-6299(15)30945-5

Leuschner, C., Ellenberg, H., & Sutcliffe, L. (2017). *Vegetation ecology of Central Europe*. Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-319-43042-3

Lortie, C. J., Brooker, R. W., Choler, P., Kikvidze, Z., Michalet, R., Pugnaire, F. I., & Callaway, R. M. (2004). Rethinking plant community theory. *Oikos*, *107*, 433–438. https://doi.org/10.1111/j.0030-1299.2004.13250.x

Manion, G., Lisk, M., Ferrier, S., Nieto-Lugilde, D., Mokany, K., & Fitzpatrick, M. (2017). gdm R package Retrieved from https://cran.r-project.org/package=gdm

Marage, D., & Gégout, J.-C. (2009). Importance of soil nutrients in the distribution of forest communities on a large geographical scale. *Global Ecology and Biogeography*, *18*, 88–97. https://doi.org/10.1111/j.1466-8238.2008.00428.x

Miller, J., & Franklin, J. (2002). Modeling the distribution of four vegetation alliances using generalized linear models and classification trees with spatial dependence. *Ecological Modelling*, *157*, 227–247. https://doi.org/10.1016/S0304-3800(02)00196-5

Morán-Ordóñez, A., Suárez-Seoane, S., Elith, J., Calvo, L., & de Luis, E. (2012). Satellite surface reflectance improves habitat distribution mapping: A case study on heath and shrub formations in the Cantabrian Mountains (NW Spain). *Diversity and Distributions*, *18*, 588–602. https://doi.org/10.1111/j.1472-4642.2011.00855.x

Mücher, C. A., Hennekens, S. M., Bunce, R. G. H., Schaminée, J. H. J., & Schaepman, M. E. (2009). Modelling the spatial distribution of Natura 2000 habitats across Europe. *Landscape and Urban Planning*, *92*, 148–159. https://doi.org/10.1016/j.landurbplan.2009.04.003

Mucina, L., Bültmann, H., Dierßen, K., Theurillat, J.-P., Raus, T., Čarni, A., … Tichý, L. (2016). Vegetation of Europe: Hierarchical floristic classification system of vascular plant, bryophyte, lichen, and algal communities. *Applied Vegetation Science*, *19*, 3–264. https://doi.org/10.1111/avsc.12257

Noss, R. F. (1990). Indicators for monitoring biodiversity: A hierarchical approach. *Conservation Biology*, *4*, 355–364. https://doi.org/10.1111/j.1523-1739.1990.tb00309.x

Peterka, T., Hájek, M., Jiroušek, M., Jiménez-Alfaro, B., Aunina, L., Bergamini, A., … Chytrý, M. (2017). Formalized classification of European fen vegetation at the alliance level. *Applied Vegetation Science*, *20*, 124–142. https://doi.org/10.1111/avsc.12271

Peterson, A., Soberón, J., Pearson, R., Anderson, R., Martínez-Meyer, E., Nakamura, M., & Araújo, M. (2011). *Ecological niches and geographic distributions*. Princeton, NJ: Princeton University Press.

Phillips, S. J., Dudík, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J., & Ferrier, S. (2009). Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. *Ecological Applications*, *19*, 181–197. https://doi.org/10.1890/07-2153.1

Potts, A. J., Hedderson, T. A., Franklin, J., & Cowling, R. M. (2013). The last glacial maximum distribution of South African subtropical thicket inferred from community distribution modelling. *Journal of Biogeography*, *40*, 310–322. https://doi.org/10.1111/j.1365-2699.2012.02788.x

Prentice, I. C., Cramer, W., Harrison, S. P., Leemans, R., Monserud, R. A., Solomon, A. M., … Mar, N. (1992). A global biome model based on plant physiology and dominance, soil properties and climate. *Plant Physiology*, *19*, 117–134.

Radosavljevic, A., & Anderson, R. P. (2014). Making better Maxent models of species distributions: Complexity, overfitting and evaluation. *Journal of Biogeography*, *41*, 629–643. https://doi.org/10.1111/jbi.12227

Regan, H. M., Colyvan, M., & Burgman, M. A. (2008). A taxonomy and treatment of uncertainty for ecology and conservation biology. *Ecological Applications*, *12*, 618–628.

Rocchini, D., Balkenhol, N., Carter, G. A., Foody, G. M., Gillespie, T. W., He, K. S., … Neteler, M. (2010). Remotely sensed spectral heterogeneity as a proxy of species diversity: Recent advances and open challenges. *Ecological Informatics*, *5*, 318–329. https://doi.org/10.1016/j.ecoinf.2010.06.001

Rodríguez, J. P., Keith, D. A., Rodríguez-Clark, K. M., Murray, N. J., Nicholson, E., Regan, T. J., … Wit, P. (2015). A practical guide to the application of the IUCN red list of Ecosystems criteria. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *370*, 20140003.

Šefferová Stanová, V., Seffer, J., & Janák, M. (2008). *Management of Natura 2000 habitats 7230 Alkaline fens*. Bruxelles: European Commission.

Stenzel, S., Feilhauer, H., Mack, B., Metz, A., & Schmidtlein, S. (2014). Remote sensing of scattered natura 2000 habitats using a one-class classifier. *International Journal of Applied Earth Observation and Geoinformation*, *33*, 211–217. https://doi.org/10.1016/j.jag.2014.05.012

Suárez-Seoane, S., Virgós, E., Terroba, O., Pardavila, X., & Barea-Azcón, J. M. (2014). Scaling of species distribution models across spatial resolutions and extents along a biogeographic gradient. The case of the Iberian mole *Talpa occidentalis*. *Ecography*, *37*, 279–292. https://doi.org/10.1111/j.1600-0587.2013.00077.x

Thauront, M., & Stallegger, M. (2008). *Management of Natura 2000 habitats. 9110 Luzulo-Fagetum beech forests*. Bruxelles: European Commission.

Thuiller, W., Brotons, L., Araújo, M. B., & Lavorel, S. (2004). Effects of restricting environmental range of data to project current and future species distributions. *Ecography*, *27*, 165–172. https://doi.org/10.1111/j.0906-7590.2004.03673.x

Venevsky, S., & Veneskaia, I. (2003). Large-scale energetic and landscape factors of vegetation diversity. *Ecology Letters*, *6*, 1004–1016. https://doi.org/10.1046/j.1461-0248.2003.00527.x

Vogiatzakis, I. N., & Griffiths, G. H. (2006). A GIS-based empirical model for vegetation prediction in Lefka Ori, Crete. *Plant Ecology*, *184*, 311–323. https://doi.org/10.1007/s11258-005-9075-2

Warren, D. L., & Seifert, S. N. (2011). Ecological niche modeling in Maxent: The importance of model complexity and the performance of model selection criteria. *Ecological Applications*, *21*, 335–342. https://doi.org/10.1890/10-1171.1

Wenger, S. J., & Olden, J. D. (2012). Assessing transferability of ecological models: An underappreciated aspect of statistical validation. *Methods in Ecology and Evolution*, *3*, 260–267. https://doi.org/10.1111/j.2041-210X.2011.00170.x

Whittaker, R. J., Araújo, M. B., Jepson, P., Ladle, R. J., Watson, J. E. M., & Willis, K. J. (2005). Conservation Biogeography: Assessment and prospect. *Diversity and Distributions*, *11*, 3–23. https://doi.org/10.1111/j.1366-9516.2005.00143.x

Wiens, J. J. (2011). The niche, biogeography and species interactions. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *366*, 2336–2350. https://doi.org/10.1098/rstb.2011.0059

Willner, W., Jiménez-Alfaro, B., Agrillo, E., Biurrun, I., Campos, J. A., Čarni, A., … Chytrý, M. (2017). Classification of European beech forests: A Gordian Knot? *Applied Vegetation Science*, *20*, 494–512. https://doi.org/10.1111/avsc.12299

Wisz, M. S., Pottier, J., Kissling, W. D., Pelissier, L., Lenoir, J., Damgaard, C. F., … Svenning, J.-C. (2013). The role of biotic interactions in shaping distributions and realised assemblages of species: Implications for species distribution modelling. *Biological Reviews of the Cambridge Philosophical Society*, *88*, 15–30. https://doi.org/10.1111/j.1469-185X.2012.00235.x

Zhou, J., Lai, L., Guan, T., Cai, W., Gao, N., Zhang, X., … Zheng, Y. (2016). Comparison modeling for alpine vegetation distribution in an arid area. *Environmental Monitoring and Assessment*, *188*, 408. https://doi.org/10.1007/s10661-016-5417-x

Zomer, R., Trabucco, A., van Straaten, O., & Bossio, D. (2006). Carbon, land and water: A global analysis of the hydrologic dimensions of climate change mitigation through Afforestation/Reforestation. http://dx.doi.org/10.3910/2009.101

BIOSKETCH

Borja Jiménez-Alfaro is a plant ecologist interested in community ecology and biogeography, with expertise in the analysis of big vegetation databases. This study was developed with the European Vegetation Archive (www.euroveg.org/eva-database) to support the European Environmental Agency for mapping European habitats.

Author contributions: B.J.-A. designed the study with support of SSS and MC. WW conducted the classification of vegetation plots of beech forests. BJA lead the writing and conducted the analyses. All authors provided their own data and participated actively in the project, interpreting the results and commenting on the manuscript.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Jiménez-Alfaro B, Suárez-Seoane S, Chytrý M, et al. Modelling the distribution and compositional variation of plant communities at the continental scale. *Divers Distrib*. 2018;24:978–990. https://doi.org/10.1111/ddi.12736