

LETTER

Accounting for dispersal and biotic interactions to disentangle the drivers of species distributions and their abundances

Isabelle Boulangeat,^{1*} Dominique Gravel² and Wilfried Thuiller¹

¹Laboratoire d'Ecologie Alpine, UMR CNRS 5553, Université Joseph Fourier, BP 53, 38041 Grenoble Cedex 9, France

²Université du Québec à Rimouski, Département de Biologie, Chimie et Géographie, 300 Allée des Ursulines, Québec G5L 3A1, Canada

*Correspondence: E-mail: isabelle.boulangeat@gmail.com

Abstract

Although abiotic factors, together with dispersal and biotic interactions, are often suggested to explain the distribution of species and their abundances, species distribution models usually focus on abiotic factors only. We propose an integrative framework linking ecological theory, empirical data and statistical models to understand the distribution of species and their abundances together with the underlying community assembly dynamics. We illustrate our approach with 21 plant species in the French Alps. We show that a spatially nested modelling framework significantly improves the model's performance and that the spatial variations of species presence–absence and abundances are predominantly explained by different factors. We also show that incorporating abiotic, dispersal and biotic factors into the same model bring new insights to our understanding of community assembly. This approach, at the crossroads between community ecology and biogeography, is a promising avenue for a better understanding of species co-existence and biodiversity distribution.

Keywords

Abiotic niche, co-occurrence index, community assembly rules, dispersal mechanisms, fundamental niche, niche overlap, species distribution model, species pool.

Ecology Letters (2012) 15: 584–593

INTRODUCTION

Understanding the factors driving the distribution of species and their abundances is an important research area in ecology since it refers to species coexistence and the maintenance of species diversity (Chesson 2000). Soberon (2007) proposed a conceptual framework distinguishing three main drivers. Abiotic constraints delimit the species' fundamental niche within which the species could establish and maintain itself given its intrinsic physiological limits (Chase & Leibold 2003). Dispersal limitations may then restrict a species' range by preventing it from reaching a suitable site. Dispersal limitation is inherently linked to the species' dispersal capability, but may also be influenced by historical legacies (Vellend *et al.* 2007). Some species may reach high abundance in unsuitable sites due to frequent immigration from neighbouring suitable sites, as predicted by source-sink dynamics or mass effect (Pulliam 2000). Finally, biotic interactions may modify either the resources availability or the local abiotic environment with potentially contrasting consequences on abundance (e.g. competition and facilitation, Lortie *et al.* 2004). These three main drivers could interact together and influence the observed spatial distribution of the environmental conditions suitable for a given species (i.e. species' realised niche).

These drivers of species distributions may act at different spatial scales (Kneitel & Chase 2004). For instance, frequent consideration is given to abiotic variables acting at large spatial scales (Davies *et al.* 2004). Climate and soil variables have been shown to be the most relevant variables when predicting continental to regional-scale plant species presence–absence distributions (Thuiller *et al.* 2004). Other factors, such as land cover, are also important to understand species distribution at regional scales (Randin *et al.* 2009). On the other hand, biotic interactions and consumer–resource dynamics are assumed to occur at smaller spatial scales (Soberon 2007). Dispersal limitations

are most likely to occur at a larger extent than species dispersal distances, whereas source-sink dynamics occur at the same scale as dispersal.

So far, species distribution has been modelled using species distribution models (SDMs), which, most of the time, ignore the effects of dispersal and biotic interactions (VanDerWal *et al.* 2009). Even if SDMs provide satisfactory predictions for presence–absence, they typically fail to explain and predict species' abundances across sites (Pearce & Ferrier 2001; Sagarin *et al.* 2006). Here we argue that this failure is due to the omission of the above-mentioned dynamic drivers (i.e. dispersal, biotic interactions) and that all processes acting at different spatial scales are fitted in a unique response model. The abundance of a given species is obviously driven to some extent by the same processes as the presence–absence, but additional processes might come into play to determine the abundance when the species is present. They operate at finer resolutions where community composition, population dynamics and the microenvironment interact. For explaining and modelling species abundances, the use of a spatially nested approach, which model separately the processes underlying presence–absence and abundance variation, seems therefore particularly relevant.

We propose to account explicitly for the spatial nestedness of processes in a comprehensive modelling framework where the species undergoes successive filters. The first filter, related to larger scale factors, determines the presence or absence of the species. This filter is expected to be primarily driven by abiotic variables (ecophysiological limitations) and by dispersal when site accessibility is limited. Once a species is able to reach a suitable site, the second filter, expected to be mainly determined by local-scale factors such as biotic interactions and dispersal in case source-sink dynamics occur, will determine the local abundance and other absences (Fig. 1, Theoretical model). Our objective here is to test these expectations using an

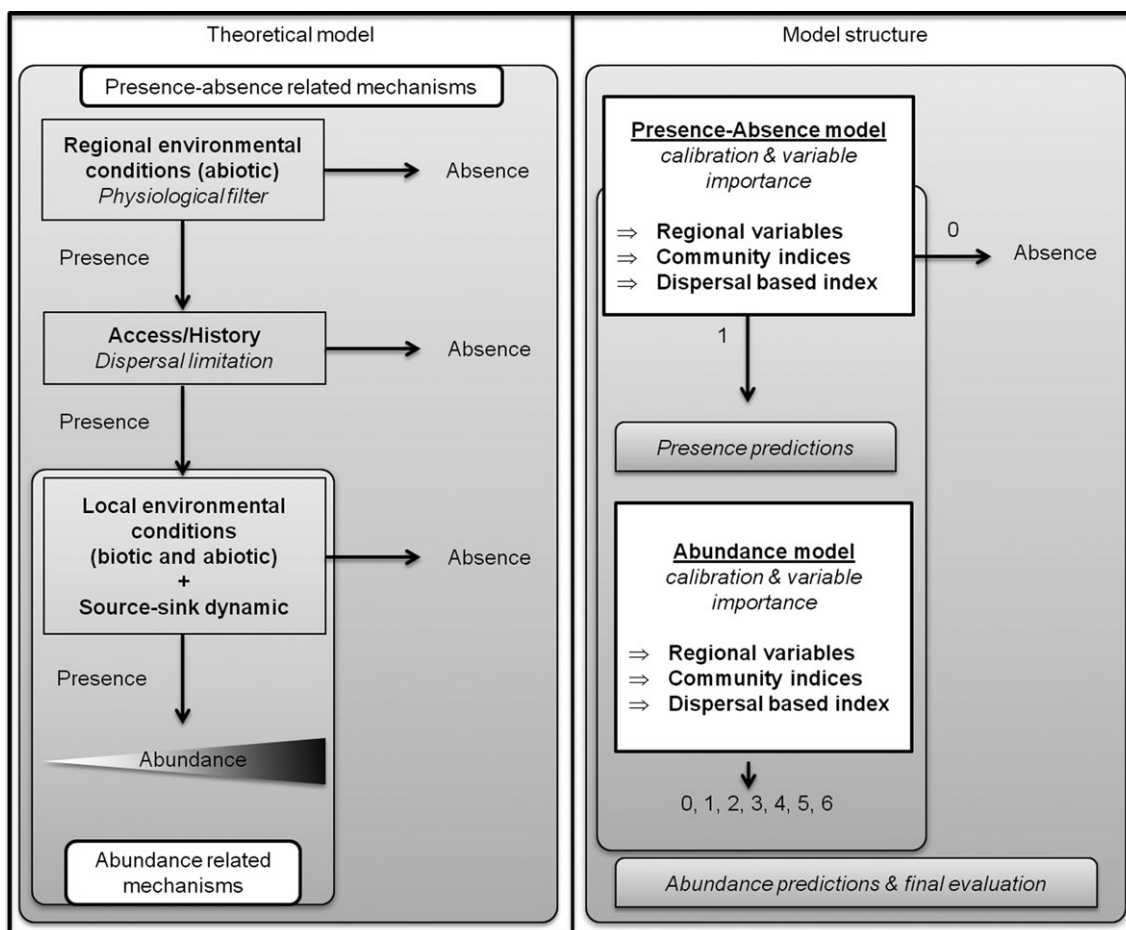


Figure 1 Theoretical framework and model structure. The aim of the nested model structure is to represent the theoretical framework. The first model focuses on presence–absence only and is expected to primarily involve abiotic drivers due to physiological filtering and dispersal mechanisms due to dispersal limitation. The second model focuses on abundance when presence has been assessed and is expected to involve local-scale mechanisms, including abiotic and biotic community-scale effects and source-sink dynamics.

operational statistical framework (Fig. 1, Model structure). First, we introduce a set of key innovations to account for biotic interactions, using co-occurrence based indices, and for dispersal by proposing a measurement based on species dispersal capabilities and observed spatial distributions. Second, to separate presence–absence from abundance underlying processes, we propose a two-step modelling approach: (1) modelling species presence and absence and (2) modelling abundances once presence has been determined. We highlight the innovative nature and outputs of our framework with the modelling of 21 plant species in the French Alps. We show how our framework separates the drivers of presence–absence and abundances. We then demonstrate how consideration of biotic interactions and dispersal not only improves predictions of species distributions, but also provides better knowledge of the elusive fundamental niche and explores the key drivers of abundances within their niches.

MATERIAL AND METHODS

We applied our framework independently to 21 plant species representative of the French Alps. We selected species with a reasonable number of observations in different abundance classes and represent varying life forms and dispersal abilities (Table S1 and S2). The idea was to test our framework and underlying hypotheses with a

set of species that were sufficiently different to permit robust conclusions. Sampling spread over a region of 30 000 km², from lowlands to alpine summits (Figure S1).

Vegetation database

We used a database of vegetation surveys provided by the National Alpine Botanical Conservatory (CBNA), including 8160 community-plots sampled in natural or semi-natural areas from 1980 to 2009 (Figure S1 left) and with a total of 2170 plant species (Boulangeat *et al.* 2012). Plot size information was not systematically available but was approximately 10 × 10 m. Within each community-plot, species abundances were recorded using a cover scheme with six classes (1: less than 1%; 2: from 1 to 5%; 3: from 5 to 25%; 4: from 25 to 50%; 5: from 50 to 75%; 6: up to 75%) (Braun-Blanquet 1946). Species abundance classes were converted to relative abundances for co-occurrence based indices (see below). We first converted cover classes to their mean percentages (0.5, 3, 15, 37.5, 62.5 and 87.5%) and then normalised them between 0 and 1 to obtain the relative abundance of each species.

We used a second dataset from the same source (CBNA) to build the dispersal-based index and some of the co-occurrence based indices (see below). This dataset contains approximately two million spatially

localised single occurrences (i.e. presence-only data, Figure S1 right) recorded from 1980 to 2009. Each of the 2170 plant species has been recorded at least 20 times.

Abiotic variables

We considered climate, topography and soil (see Table S3 and Figure S2). Four climatic variables (temperature, precipitation, moisture index and radiation) were extracted from the Aurelhy meteorological model (Bénichou and Le Breton, 1987), based on interpolated measurements at a resolution of 100×100 m. We used a 50×50 m digital elevation model to generate slope angle ($^\circ$), topography (elevation range within the pixels) and the topographic wetness index. Soil carbon and available water capacity were extracted from the European soil database (1×1 km). We used land cover maps (100×100 m) extracted from the 2006 Europe-wide Corine Land Cover (artificial surfaces, agricultural areas, forest, scrub, open spaces, wetlands and water bodies) (Table S3) to build the dispersal index and the expected community assembly (see below). For community plots, we assigned a land cover type from the plot description provided by the botanists (i.e. wetland, open land, shrubland and wood), pairing the European Corine land cover classification (Bossard *et al.* 2000).

Dispersal based index

We developed an index modelling the potential effects of dispersal on species distribution and abundance. It aimed to account explicitly for the spatial auto-correlation of species distributions with the inclusion of the biological hypotheses underlying species dispersal abilities. This index was based on a seed rain model and the integration of the species presence-absence distribution in the neighbourhood of the community (Fig. 2).

We first approximated the actual spatial patterns of species presences and absences at the regional scale using traditional SDMs because a complete sampling of the entire region at a reasonable resolution is practically impossible. We assumed that there would be a strong match between the first environmental filter and the presence-absence distribution at a coarse resolution (100×100 m). We used random forest models as they have been shown to be of very good

predictive accuracy (Peters *et al.* 2007). Our objective was to closely match the observed distribution and only detect potential missing presences (not sampled). The calibrations were made using all available species observations (i.e. using both datasets, with a minimum of 500 presences per focal species) and all gridded abiotic variables (Figure S2).

Based on these estimated presence-absence distributions (Figure S4), we modelled the potential seed rain received each year in a plot using a spatially explicit dispersal model (adapted from Engler *et al.* 2009). The expected seed rain of a species in a plot was modelled as an integration of the presences in the neighbourhood weighted by a species-specific seed dispersal kernel. The neighbouring pixels (from 0 m to a threshold distance d_k) contributed to the seed rain as:

$$\frac{\exp\left((r - R) \cdot \frac{\ln(1 - k)}{d_k}\right) - \exp\left(r \cdot \frac{\ln(1 - k)}{d_k}\right)}{\frac{2 \cdot \pi \cdot r}{R}} \quad (1)$$

where r is the distance from the pixel to the community, k is the proportion of seeds that fall before the distance d_k and R is the pixel size (here 100 m). The long distance pixels, from distance d_k to maximal dispersal distance ldd , contributed to the seed rain as:

$$\frac{1 - k}{\pi \cdot (ldd + d_k) \cdot (ldd - d_k)} \quad (2)$$

The parameter k was fixed at 0.99, and ldd and d_{99} were assigned from independent literature information in accordance with Vittoz & Engler (2007) and Engler *et al.* (2009) (Table S2). The species-specific dispersal index for a site was the overall potential seed rain received in a community-plot (eqn 1 and 2) integrated over all neighbouring pixels (Fig. 2). Despite the effect of the abiotic environment on species distributions and thus on the dispersal index, the correlations between dispersal index and abiotic variables were lower than 0.7 for the 21 focal species and all variables (Table S4).

Co-occurrence based indices

We computed four site-specific co-occurrence-based indices to estimate the effects of biotic interactions on species distribution.

Neighbourhood



Pixels contributions

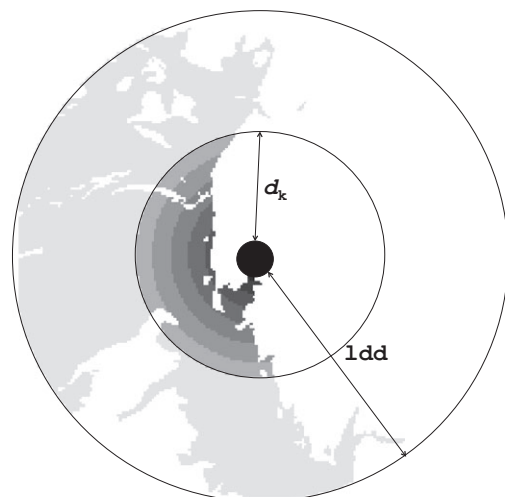
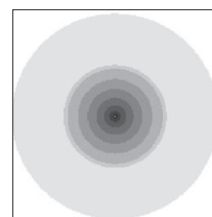


Figure 2 Contribution of neighbouring cells to the dispersal-based index. A kernel function is applied to weight species presences around each focal cell (sampled point, black dot). Pixels under distance d_{99} contribute by short distance dispersal and pixels between d_{99} and ldd contribute by long distance dispersal (see equations). The neighbourhood map displays the species presence (black) and absence (white) around the focal community. The pixel contribution map shows the weights of each pixel. The resulting map is then added up to obtain the potential seed rain, known as the dispersal-based index.

The indices aggregated pair-wise species interaction strength measurements at community level. We took into account the interactions between each of our 21 species with all species occurring in the French Alps dataset. We used three different strategies based on (1) the observed co-occurrences, (2) the expected community composition according to abiotic environmental filters and (3) a comparison between observations and expectations. The rationale was to depict resource competition and coexistence resulting from species interactions. Observed species co-occurrences (1) were a proxy for actual species interactions. Expected species co-occurrences (2) described the potential competitive pressure. The comparison between the two and (3) measured the resulting species interactions when controlling for the effect of the abiotic environment.

Observed co-occurrences (community aggregated C-score)

Pair-wise species association strengths between the focal species j and the co-occurring species k were calculated using the C-score measure (Stone & Roberts 1990). Pair-wise C-score represents the average number of checkerboard units (i.e. when the two species occur in distinct sites) and was calculated as $(N_j - N_{jk}) \cdot (N_k - N_{jk})$, where N_j and N_k are the regional number of observations of species j and k respectively, and N_{jk} the number of co-occurrences between j and k . Pair-wise C-scores were then aggregated at the community level, weighted by the relative abundance of species k (p_k).

$$I_{jx} = \sum_k I_{jk} \cdot p_k \quad (3)$$

where I_{jk} was the pair-wise C-score between the focal species j and another species k and I_{jx} the community aggregated C-score at site x . A value close to zero indicates independence between the focal species and the local community. Conversely, a high value suggests a strong repulsion between the focal species and the local community.

Expected co-occurrences (niche overlap index)

Expected communities (i.e. expected species co-occurrences) were constructed by relating observations to abiotic variables for all species involved in the sampled communities (2170 species). A species was expected to be present in a community-plot if the local conditions fell into the observed species' environmental range, delimited by minimum and maximum observed values for all abiotic variables simultaneously. The rationale behind this was to identify those species that were not observed in locations potentially suitable for them. Species found in the community-plots could be viewed as the 'winners' that had already withstood the prevailing biotic interactions within their communities. By accounting for all species for which the environment of the observed community-plots were suitable, we took into account the 'absent' part of biodiversity or 'dark diversity' that theoretically could inhabit a particular site after accounting for the physiological constraints (Pärtel *et al.* 2011).

The niche overlap index was evaluated in each plot by accounting for all species expected to be present in the site and their niche overlap with the focal species. The abiotic niche space was defined by the first two axes of a principal component analysis (PCA, the first two axes accounted for 73% of the total variation) of all pixels of the study region (at a resolution of 100 × 100 m) and all abiotic variables. Species niche overlaps were calculated in the abiotic space using the D statistic (Schoener 1968). It ranges from zero to one

(method adapted from Broennimann *et al.* 2012). The niche overlap index was calculated as the mean D between all species expected to be present in a given site and the focal species. It aims at describing the competitive pressure of the expected species pool on the focal species.

Observed and expected co-occurrence comparison (attraction and repulsion indices)

We derived two indices based on a comparison between the observed and expected species co-occurrences. The rationale for this comparison was to partition the effect of positive vs. negative biotic interactions (or other local factors). Both indices (for repulsion and attraction) were estimated by comparing the two community matrices (observed and expected, Table 1). Negative co-distribution (fewer co-occurrences than expected, repulsion) should occur if species compete intensively with each other (Table 1). Positive co-distribution (higher co-occurrences than expected, attraction) should occur if they positively interact with each other, or alternatively the presence of one species changes the local environmental conditions in such a way that a suitable habitat is created for the non-expected species (Gilpin & Diamond 1982). Two types of pair-wise interactions were calculated to quantify repulsion, ($I_{jk} = (C_{01} + C_{10}) / (C_{01} + C_{10} + C_N + C_0)$) or attraction ($I_{jk} = (F_{01} + F_{10}) / (F_{01} + F_{10} + F_N + F_0)$) (Table 1) and aggregated at the community level (from eqn 3). Resulting indices quantified total negative (repulsion) or positive (attraction) interaction between the community and the focal species.

Species abundance model

We followed the nested modelling framework presented in the Introduction (Fig. 1). Inspired by the general idea of Zero Inflated Poisson models for count data (Welsh *et al.* 1996), we first modelled species presence and absence and then, in a second step, we did the same with abundance cover classes for locations with predicted

Table 1 Classification of the different cases arising from the comparison between expected and observed co-occurrences. Comparison of the observed (realised) co-occurrences with the expected ones (potential) for each species pair. Situation 1: both species are expected (1/1). If one species is not observed (C_{10} or C_{01}), it is the consequence of competitive exclusion or of an unsuitable local environment. Other configurations are neutral (C_N) or are the result of unconsidered factors (C_0). Situation 2: only one species is expected (1/0 or 0/1). If both species are observed (F_{01} or F_{10}), it is the consequence of a facilitation effect or a common local suitable environment. Other situations are neutral (F_N) or are the result of unconsidered factors (F_0)

Expected co-occurrence	Observed co-occurrence	Interpretation	Name
1/1	1/0	Competition or specific environment	C_{10}
	0/1		C_{01}
	1/1	Neutral	C_N
	0/0	Other factors	C_0
1/0 0/1	1/1	Facilitation	F_{10} F_{01}
	1/0 0/1	Neutral	F_N
1/0 0/1	0/1	Other factors	F_0
	1/0		
	1/0		
	0/1		

presence. To evaluate the models, final predictions were given by absences predicted from the first model and abundance classes (potentially including other absences) predicted using the second model (see Figure S3 for a detailed workflow of the method). We compared this nested model to a non-nested model (i.e. without the first step) to evaluate its performance.

We used random forests (RF) to model both presence–absence and abundance data. This choice was driven by the ability of RF to deal with both binary and multilevel data and to estimate the importance of predictor variables in high dimensional settings. They provide estimates of the independent contribution of each predictor (Strobl *et al.* 2009). We performed nine repetitions, following a cross-validation procedure (see Figure S3).

We calibrated and evaluated four models using different sets of predictor variables to evaluate the performance of our framework and estimate the relevance of the proposed indices: abiotic variables (A), abiotic variables and co-occurrence based indices (AC), abiotic variables and dispersal index (AD) and all predictors (ACD). Predictions about final abundance classes were evaluated using the Hanssen-Kuipers discriminant statistic (Gandin & Murphy 1992) (see supporting information eqn S1 for the formula). We computed variable importance with a re-sampling method implemented into random forest models (Broennimann *et al.* 2012) (see supporting information eqn S2 for details and formula). We averaged the importance values across the nine repetitions and then normalised the results for each species by calculating the relative importance of the different variables. A variable was considered as not significant when the confidence interval included zero.

All analyses were carried out in R (R Development Core Team 2010) with the ‘*raster*’ package for spatial data management, the ‘*ade4*’ package for PCAs, and the ‘*party*’ package for random forest models and variable importance.

Biotic interactions and dispersal effects on the abiotic niche

We illustrated how community-scale and dispersal mechanisms affected the abiotic niche for the four species for which the addition of the proposed indices increased performance the most. We derived different predictions from the complete model (ACD) and projected them onto the abiotic niche space. This space was defined by the first two axes of a PCA involving the abiotic variables of the sampled plots. These two axes together explained 53% of the total inertia. In this space, we drew the density of presences derived from different model predictions, computed with observed values of explicative variables or with dispersal and/or biotic interactions set to zero. We grouped all repetitions into a single prediction. First, absences were determined where predicted absences had a majority amongst repetitions and then for presences, abundance classes were given by the majority amongst repetitions’ predicted classes. Following Soberon (2007) we defined sinks as the sites where the model predicted presences with observed values but absences with null dispersal and we defined sources as the sites where the model predicted presences with observed values as with null dispersal. Then, to locate competition and facilitation areas, we compared the current predictions with predictions where all co-occurrence indices were set to zero. A predicted absence that became a presence with a release of biotic interactions indicated a negative impact of biotic interactions and conversely, a presence

changed to absence indicated a positive impact of biotic interactions. We also compared the two response curves (current predictions and those with co-occurrence indices equalling zero) as a function of the temperature, which is the most important climatic gradient in the region (explaining 70% of the first axis of the PCA). We estimated these curves using local least square regressions between probabilities of presences and temperature. We used the ‘loess’ function (Cleveland *et al.* 1992).

RESULTS

Nested vs. non-nested models of species’ abundances across sites

The nested framework improved our ability to predict abundance irrespective of the set of variables used and the focal species (see Supporting Information Figure S5). The performance of the nested model was better than the direct model (without the nested framework) in 99% of the cases among all models, repetitions and focal species, with an average increase for the Hanssen-Kuipers discriminant equal to 0.26.

Is it useful to integrate biotic interactions and dispersal into SDMS?

The comparison of the performance of the three models (AC, AD and ACD) with the model that only uses regional environmental filters (A) demonstrated the importance of biotic interactions and dispersal to explain the species’ abundances across sites (Fig. 3). In general, the inclusion of co-occurrence indices (i.e. biotic interactions) was sufficient to significantly improve the model performance. For some species, neither dispersal nor biotic interactions were relevant, whereas their interaction greatly improved the performance of models (e.g. *Ranunculus glacialis*, *Phragmites australis*, Fig. 3). Finally, only one species showed no improvement using the proposed indices (*Festuca paniculata*, Fig. 3).

Disentangling species abundance and species presence–absence drivers

The separate analysis of variable importance at each model step highlighted major differences in the main drivers of presence–absence or abundance (Fig. 4). In general, dispersal was relatively important in predicting presence or absence but less for the abundance class where presence was expected (Fig. 4). The most important biotic interaction index was the repulsion one, followed by the community aggregated C-score, and the attraction index for some species. The niche overlap index was a poor predictor of both presence–absence and abundance. The biotic interactions indices were more important (for all species) in the second step of the model in predicting the abundance class where presence was expected. For instance, *Bromus erectus* well illustrated these results (Fig. 4). Dispersal was more important for presence-absence, and the importance of biotic interactions increased for abundance. Abiotic variables, although they varied in relative importance between species, were relevant to both modelling steps (see Fig. S6 for details).

Biotic interactions and dispersal effects on species’ abiotic niches

Our modelling framework allowed us to visualise the effects of the different indices on the abiotic niche space (e.g. *B. erectus* in Fig. 5, other species in Figure S7). For *B. erectus*, we found that the realised

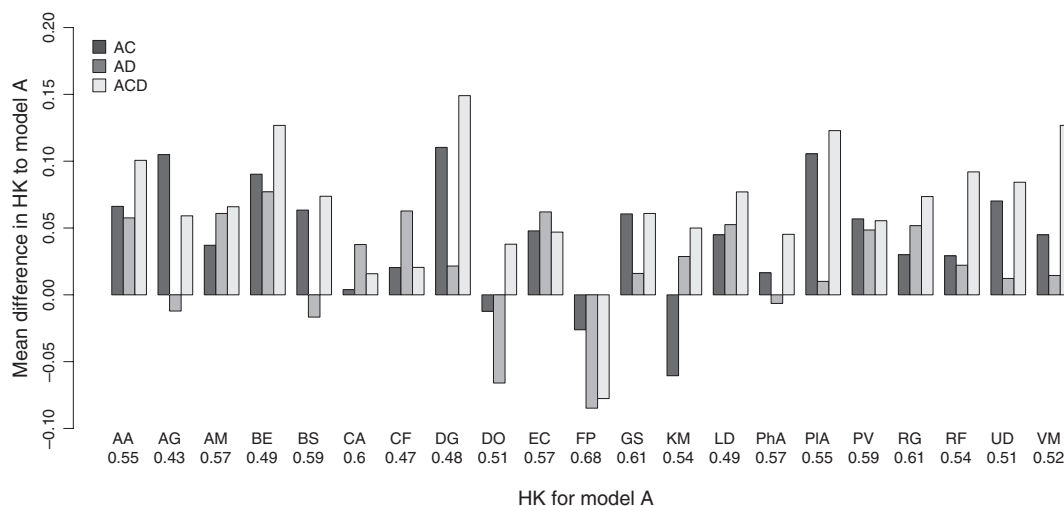


Figure 3 Comparison between model A and models AC, AD and ACD. Each bar represents the average difference across all repetitions between the predictive accuracy of model A and the models AC, AD and ACD. Accuracy was measured using the Hanssen-Kuipers discriminant (HK), which varies from 0 to 1 for perfect fit. The numeric values on the x-axis are the mean predictive accuracy of model A. The following abbreviations are used to name the species: AA = *Abies alba*, AG = *Alnus glutinosa*, AM = *Arnica montana*, BE = *Bromus erectus*, BS = *Buxus sempervirens*, CA = *Cacalia alliardae*, CF = *Carex ferruginae*, DG = *Dactylis glomerata*, DO = *Dryas octopetala*, EC = *Euphorbia cyparissias*, FP = *Festuca paniculata*, GS = *Geranium sempervirens*, KM = *Kobresia myosuroides*, LD = *Larix decidua*, PhA = *Phragmites australis*, PIA = *Plantago alpina*, PV = *Polygonum viviparum*, RG = *Ranunculus glacialis*, RF = *Rhododendron ferrugineum*, UD = *Urtica dioica*, VM = *Vaccinium myrtillus*.

abiotic niche lies in drier and warmer places than the average of the sampled plots (Fig. 5a). Using a comparison between the realised niche and a prediction with no dispersal, we identified abiotic conditions of sources and sinks (Fig. 5b). Surprisingly, sources were located where the density of presences was relatively low (Fig. 5a,b). Interestingly, sources corresponded to high abundance and sinks to lower abundance (Fig. 5a,b). With a release of biotic interactions, the abiotic niche did not expand significantly but became denser, particularly in the warmer part of the gradient (Fig. 5c), suggesting there is competition in these areas (red, Fig. 5c, right panel). At the colder edge of the niche, the density decreased, pointing to positive effects of biotic interactions (green, Fig. 5c, right panel). Some sinks were potentially due to biotic interactions, as they occurred in the same abiotic region as negative biotic interactions (Fig. 5b,c).

DISCUSSION

We proposed and tested a comprehensive framework for investigating mechanisms underlying species distributions and their abundances. First, we demonstrated that the nested modelling structure greatly improves our understanding of distribution and abundance, increasing model performance for all species studied here. Second, our nested framework allowed us to show that the ranking in variable importance was reversed between presence-absence and abundance when presence was expected, supporting the hypothesis that different processes were intertwined. Third, we showed that including the three groups of drivers (abiotic, biotic and dispersal) in the same model could offer insights into the mechanisms of community assembly and revealed how they ultimately shape the realised niche from the abiotic niche of species.

Biotic interactions

We showed that community-scale effects of biotic interactions were more important for the second modelling step in explaining

abundance when presence was established. This result supports the hypothesis that mechanisms underlying abundance variations occur at community scale, which is where species interact. In our example (*B. erectus*), the interaction with the community is essentially negative, particularly at the warm edge of the abiotic niche (Fig. 5d) where the environment is suitable for a large number of species. Moreover, co-occurrence indices associated to repulsion are relatively important for the abundance model step (Fig. 4b). These results suggest that co-occurrence indices are strongly related to negative biotic interactions, such as competition for resources (Chase & Leibold 2003; Soberon 2007). At the same time, we observed positive effects of biotic interactions at the colder edge of the abiotic niche. This may be the consequence of facilitation, which has been suggested to be an important factor of species coexistence in harsh conditions (Choler *et al.* 2001; Callaway *et al.* 2002).

Observed species co-occurrence has long been used to infer community assembly rules (Gilpin & Diamond 1982; Gotelli & McCabe 2002) and several co-occurrence based indices have been used in SDMs as a proxy for unmeasured abiotic conditions and species interactions (Leathwick 2002; Cottenie 2005; Araujo & Luoto 2007; Meier *et al.* 2010). The main limitation of these approaches is that non co-occurring species pairs could be interpreted either as the result of competitive exclusion or different species' environmental niches (Gilpin & Diamond 1982). Here, we propose to use a comparison between observed and expected community composition to tease apart the effects of environment from those of competition (Chesson 2000; Chase & Leibold 2003; Silvertown 2004). Moreover, these new metrics allow distinguishing between positive (attraction) and negative (repulsion) effects. In further studies, it would also be interesting to separate the cases where a species undergoes competition (or facilitation) from the situation where the species has the strongest competitive ability (or facilitate other species) (Table 1). These indices still cannot distinguish biotic interactions from micro-environmental conditions effects that may have a prominent role (e.g. topographical heterogeneity inducing thermal differences,

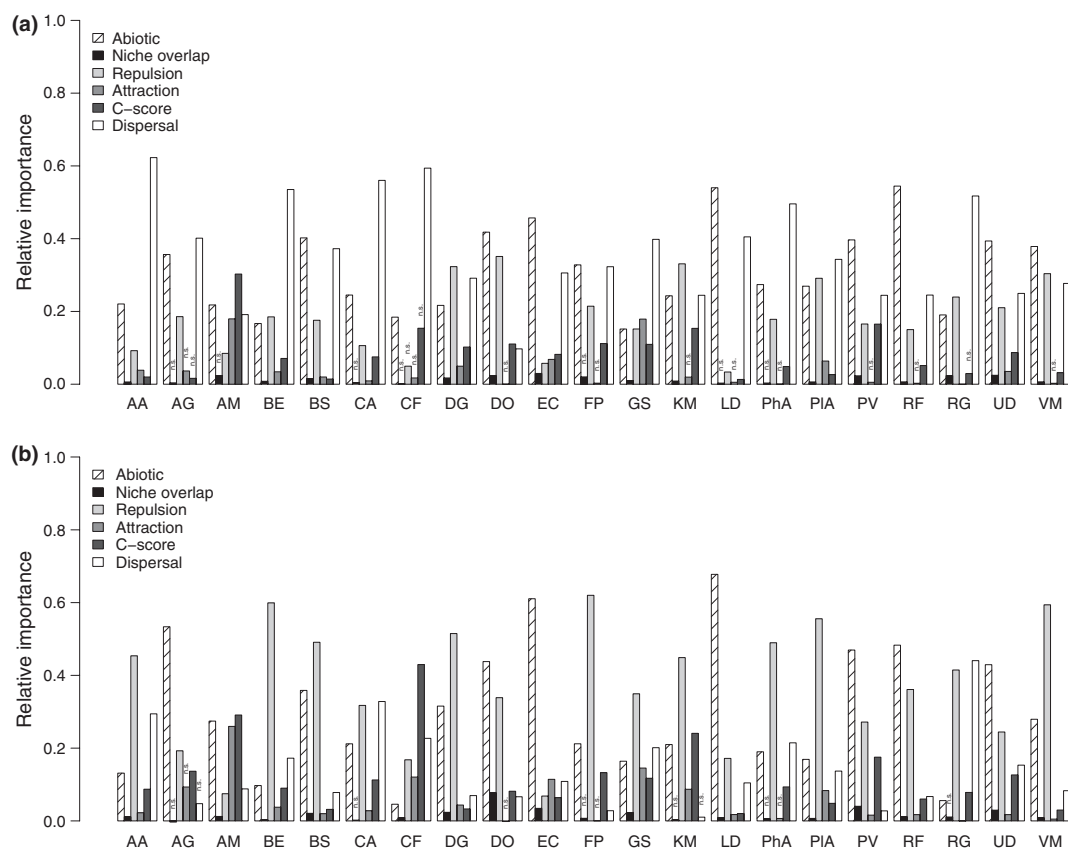


Figure 4 Importance of variables. Each barplot represents the relative importance of each variable or group of variables. The relative importance of each abiotic variable was added together. Species abbreviations are the same as for Fig. 3. Non-significant variables are marked with n.s. (a) Importance of variables to explain presence–absence distribution (modelling step 1). (b) Importance of variables to explain abundance distribution (modelling step 2).

Scherrer & Körner 2011). For instance, the co-occurrence indices for *P. australis* possibly relate to specific local abiotic conditions (small wetlands) whereas the repulsion index for *Dactylis glomerata*, a common and widely distributed grass, probably reveals its competitive strategy. Because we will never be able to measure every abiotic variable at community scale, the attraction and repulsion indices provide an interesting proxy for describing local abiotic and biotic environments. Their influence on species distributions are stronger than the niche overlap index and are relatively easy to interpret in respect to the aggregated C-score.

The construction of the expected species pool is crucial for most of the proposed indices and requires particular attention. As we assume no dispersal limitation when calculating the expected species pool, we may overestimate expected species richness at inaccessible sites. For some species, important limiting variables might also be missing or imprecise (e.g. soil data). These species would thus be wrongly added to the species pool. However, this bias cannot explain alone the poor predictive power of the niche overlap index because the expected species pool is also used to build the repulsion index. This pattern might suggest that competition does occur in few plots only, while in the others plots, spatial and temporal environmental heterogeneity allow species coexistence (Silvertown 2004). Finally, although competition would preferentially occur between functionally similar species, we did not consider species identity. The addition of weights based on functional traits to our indices might be worth considering,

as they may be directly linked to resource acquisition or exploitation (Lavorel & Garnier 2002).

Dispersal

There are now several studies supporting the importance of dispersal on species distributions (e.g. Bahn & McGill 2007). This is, however, usually approximated by spatial autocorrelation functions (e.g. Borcard *et al.* 1992), generally built with little or no attention to the true spatial processes that drive biogeographical patterns. Although the fraction of variance explained by these variables is often interpreted as the spatial signature of dispersal limitations (Beale *et al.* 2008), this spatial structure could also result from unaccounted spatially auto-correlated environmental factors (Gravel *et al.* 2008; Araujo *et al.* 2009). Here, we had an underlying hypothesis with an explicit formulation based on species-specific dispersal distances, which allows us to progress in our understanding of the processes driving spatial distribution. Confusion may however still arise when missing spatially autocorrelated abiotic drivers operate at the same spatial scale as dispersal mechanisms (Cottenie 2005).

Our study showed that dispersal mechanisms were more important for determining presence–absence patterns, suggesting strong dispersal limitations. Far from its geographical range, a species is systematically predicted to be absent because it cannot reach the site, even if the environment is suitable. It can introduce a bias in the

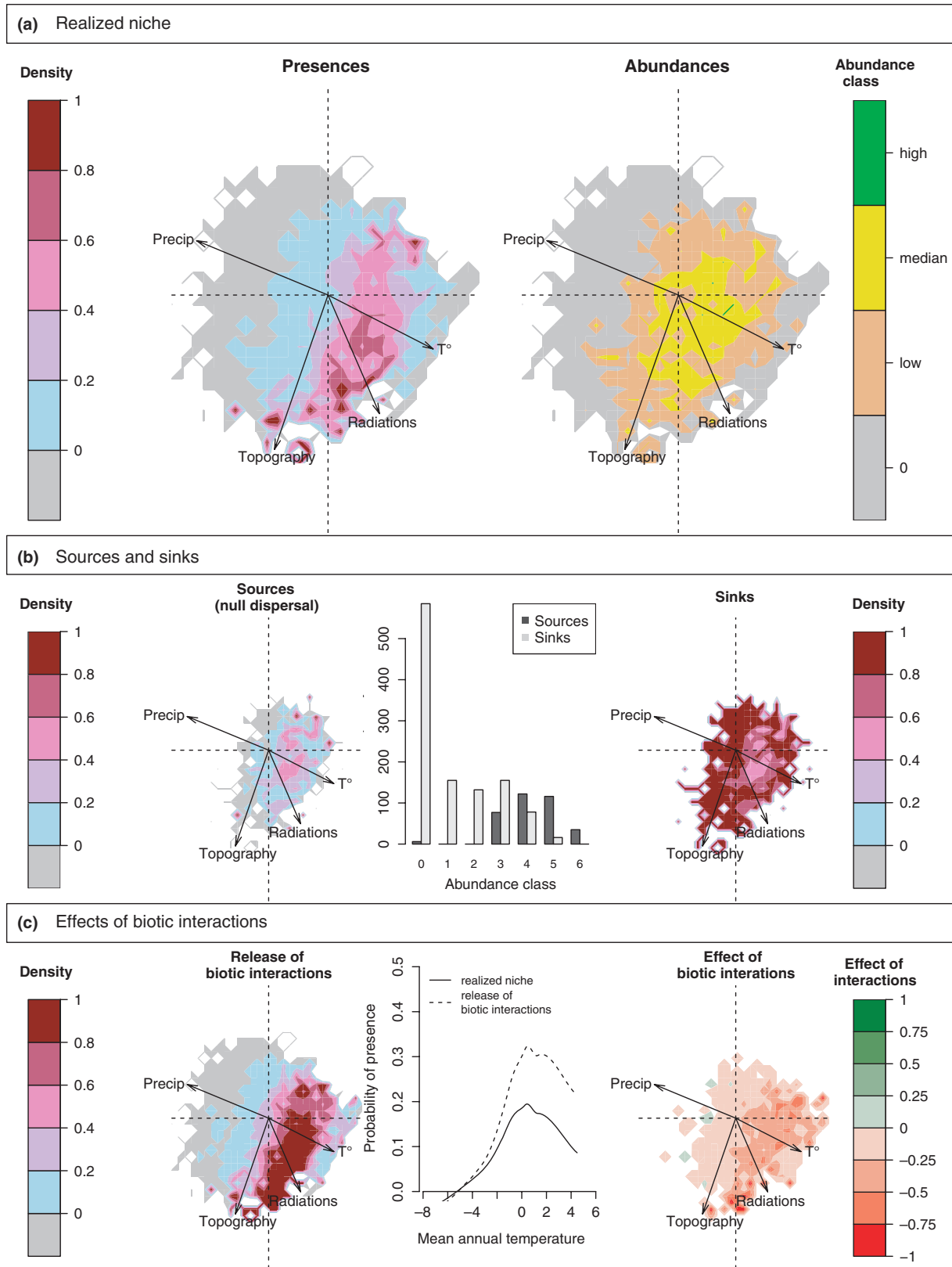


Figure 5 Effects of the different drivers on the abiotic niche for *Bromus erectus*. The abiotic niche space is represented by the first two axes (53% of inertia) of a PCA of the abiotic variables. (a) Realized niche. Predictions of model ACD. Left: density of predicted presences normalised by the number of sample plots within each grid cell. Right: third quartile of predicted abundance class within each grid cell. Low: < 5% cover; Medium: 5–25% cover; High: > 25% cover. (b) Left/right: Proportion of sources/sinks among predicted presences. Middle: abundances in source and sink plots. (c) Effect of biotic interactions. Left: density of predicted presences with co-occurrence indices equalling zero. Right: negative and positive effects of the biotic interactions.

relationship between species distribution and abiotic variables, because in these sites, abiotic constraints might be neglected. It is not possible however to reveal the real cause of such absences if all sampled plots with similar abiotic conditions are systematically far from the species geographical distribution. Ideally, to estimate the true relationship between species distribution and abiotic variables (i.e. the fundamental abiotic niche), the sampling should be homogeneous not only in relation to the abiotic variables but also spatially.

The importance of the dispersal index in the second model step highlights the implication of dispersal mechanisms at local scale, such as source-sink dynamics. Interestingly, the sources identified in our case study (and three other examples, Figure S7) are not related to the density of presence but to species' abundance (Fig. 5). These results support the source-sink theory, which predicts species occurrences in unsuitable habitat if immigration from surrounding source populations is sufficiently large (Pulliam 2000).

Species distribution and abiotic niches

Species distribution models generally use topoclimatic predictors only (Guisan & Zimmermann 2000). Consequently, the relationship they model between abiotic variables and the distribution includes at the same time abiotic constraints, dispersal mechanisms and interaction between species (Guisan & Zimmermann 2000; Soberon 2007). Here, because we introduced other explicative variables to describe dispersal and species interactions mechanisms, the modelled relationship between abiotic variables and the species distribution is refined and should be closer to the fundamental niche. Although the approximation of the fundamental niche is obviously constrained by the area considered and should be valid only for the gradient sampled, our method refines the relationship between topoclimatic variables and species distribution for our region. Whilst we agree that only experiments can define the fundamental niche (e.g. Vetaas 2002; Kearney & Porter 2009) whereas we only approximate it here, our model allows us to explore the effects of dispersal and biotic interactions on the abiotic niche, including the identification of sources and sinks (Fig. 5b) and is able to point out potential sites where facilitation or competition have a strong impact (Fig. 5c).

Perspectives

Our framework extends the boundaries of SDMs and should allow important ecological questions to be addressed. It offers an innovative way to improve our understanding of community assembly processes for large spatial scales and for many species at once, based on largely available data: coarse scale environmental variables, community surveys at the scale of species interactions, and species-specific dispersal abilities. For instance, understanding and predicting species invasion might benefit from this framework, as the potential distribution of the invader species in a region is often unknown because the realised distribution of the species usually reflects the place of introduction, dispersal capability of the species and biotic resistance of the native communities (Gallien *et al.* 2010). More fundamentally, not all species are influenced by the same factors. Applying our framework to a large set of species could help providing general rules or patterns for groups of species. For instance, dominant and generalist species are in theory less likely to be influenced by negative biotic interactions than subordinated species (Boulangeat *et al.* 2012). These relatively old but generally unresolved questions in

ecology and biogeography might be now tackled or addressed from a different angle.

ACKNOWLEDGEMENTS

The research leading to these results had received funding from the European Research Council under the European Community's Seven Framework Programme FP7/2007-2013 Grant Agreement no. 281422. IB was funded by the French "Agence Nationale de la Recherche" with the project SCION (ANR-08-PEXT-03) and initially by the European Commission's FP6 ECOCHANGE project (GOCE-CT-2007-036866). DG acknowledged support from the NSERC Discovery. The authors wish to thank the EMABIO team for insightful comments on the preliminary ideas for the article, and Pascal Vittoz for the dispersal data. Most of the computations presented in this article were performed using the CIMENT infrastructure (<https://ciment.ujf-grenoble.fr>), which is supported by the Rhône-Alpes region (GRANT CPER07 13 CIRA: <http://www.ci-ra.org>) and France-Grilles (<http://www.france-grilles.fr>). Thanks also to Version Original for checking and correcting the English in this article.

AUTHORSHIP

WT, DG and IB designed the study. IB performed the analyses and led the writing. All the authors contributed substantially to the writing.

REFERENCES

- Araujo, M.B. & Luoto, M. (2007). The importance of biotic interactions for modelling species distributions under climate change. *Glob. Ecol. Biogeogr.*, 16, 743–753.
- Araujo, M.B., Thuiller, W. & Yoccoz, N.G. (2009). Reopening the climate envelope reveals macroscale associations with climate in European birds. *Proc. Natl. Acad. Sci. USA*, 106, E45–E46.
- Bahn, V. & McGill, B.J. (2007). Can niche-based distribution models outperform spatial interpolation? *Glob. Ecol. Biogeogr.*, 16, 733–742.
- Beale, C.M., Lennon, J.J. & Gimona, A. (2008). Opening the climate envelope reveals no macroscale associations with climate in European birds. *Proc. Natl. Acad. Sci. USA*, 105, 14908–14912.
- Bénichou, P. & Le breton, O. (1987). Prise en compte de la topographie pour la cartographie des champs pluviométriques statistiques. *La Météorologie*, 7.
- Borcard, D., Legendre, P. & Drapeau, P. (1992). Partialling out the spatial component of ecological variation. *Ecology*, 73, 1045–1055.
- Bossard, M., Feranec, J. & Otahel, J. (2000). CORINE land cover technical guide – addendum 2000. European Environment Agency, Copenhagen, Denmark.
- Boulangeat, I., Lavergne, S., Van Es, J., Garraud, L. & Thuiller, W. (2012). Niche breadth, rarity & ecological characteristics within a regional flora spanning large environmental gradients. *J. Biogeogr.*, 39, 204–214.
- Braun-Blanquet, J. (1946). Über den Deckungswert der Arten in den Pflanzengesellschaften der Ordnung Vaccinio-Piceetalia. *Jahresber. Naturforsch. Ges. Graubündens.*, 130, 115–119.
- Broennimann, O., Fitzpatrick, M.C., Pearman, P.B., Petitpierre, B., Pellissier, L., Yoccoz, N.G. *et al.* (2012). Measuring ecological niche overlap from occurrence & spatial environmental data. *Glob. Ecol. Biogeogr.*, 21, 481–487.
- Callaway, R.M., Brooker, R.W., Choler, P., Kikvidze, Z., Lortie, C.J., Michalet, R. *et al.* (2002). Positive interactions among alpine plants increase with stress. *Nature*, 417, 844–848.
- Chase, J.M. & Leibold, M.A. (2003). *Ecological Niches*. Chicago University Press, Chicago.
- Chesson, P. (2000). Mechanisms of maintenance of species diversity. *Annu. Rev. Ecol. Syst.*, 31, 343–366.
- Choler, P., Michalet, R. & Callaway, R.M. (2001). Facilitation & competition on gradients in alpine plant communities. *Ecology*, 82, 3295–3308.

- Cleveland, W.C., Grosse, E. & Shyu, W.M. (1992). Local regression models. In: *Statistical Models* (ed. Hastie, JMCaTJ). Chapman and Hall, New York, 309–376.
- Cottenie, K. (2005). Integrating environmental and spatial processes in ecological community dynamics. *Ecol. Lett.*, 8, 1175–1182.
- Davies, T.J., Barraclough, T.G., Savolainen, V. & Chase, M.W. (2004). Environmental causes for plant biodiversity gradients. *Phil. Trans. R. Soc. Lond. B*, 359, 1645–1656.
- Engler, R., Randin, C.F., Vittoz, P., Czaka, T., Beniston, M., Zimmermann, N.E. *et al.* (2009). Predicting future distributions of mountain plants under climate change: does dispersal capacity matter? *Ecography*, 32, 34–45.
- Gallien, L., Munkemüller, T., Albert, C.H., Boulangeat, I. & Thuiller, W. (2010). Predicting potential distributions of invasive species: where to go from here? *Divers. Distrib.*, 16, 331–342.
- Gandin, L.S. & Murphy, A.H. (1992). Equitable skill scores for categorical forecasts. *Mon. Weather Rev.*, 120, 361–370.
- Gilpin, M.E. & Diamond, J.M. (1982). Factors contributing to non-randomness in species co-occurrences on islands. *Oecologia*, 52, 75–84.
- Gotelli, N.J. & McCabe, D.J. (2002). Species co-occurrence: a meta-analysis of Diamond's, J. M. Diamond's assembly rules model. *Ecology*, 83, 2091–2096.
- Gravel, D., Beaudet, M. & Messier, C. (2008). Partitioning the factors of spatial variation in regeneration density of shade-tolerant tree species. *Ecology*, 89, 2879–2888.
- Guisan, A. & Zimmermann, N.E. (2000). Predictive habitat distribution models in ecology. *Ecol. Model.*, 135, 147–186.
- Kearney, M. & Porter, W. (2009). Mechanistic niche modelling: combining physiological & spatial data to predict species' ranges. *Ecol. Lett.*, 12, 334–350.
- Kneitel, J.M. & Chase, J.M. (2004). Trade-offs in community ecology: linking spatial scale & species coexistence. *Ecol. Lett.*, 7, 69–80.
- Lavorel, S. & Garnier, E. (2002). Predicting changes in community composition & ecosystem functioning from plant trait: revisiting the Holy Grail. *Funct. Ecol.*, 16, 545–556.
- Leathwick, J.R. (2002). Intra-generic competition among *Nothofagus* in Zealand's, New primary indigenous forests. *Biodivers. Conserv.*, 11, 2117–2187.
- Lortie, C.J., Brooker, R.W., Callaway, R.M., Choler, P., Kikvidze, Z., Michalet, R. *et al.* (2004). Rethinking plant community theory. *Oikos*, 107, 433–438.
- Meier, E.S., Kienast, F., Pearman, P.B., Svenning, J.C., Thuiller, W., Araujo, M.B. *et al.* (2010). Biotic & abiotic variables show little redundancy in explaining tree species distributions. *Ecography*, 33, 1038–1048.
- Pärtel, M., Szava-Kovats, R. & Zobel, M. (2011). Dark diversity: shedding light on absent species. *Trends Ecol. Evol.*, 26, 124–128.
- Pearce, J. & Ferrier, S. (2001). The practical value of modelling relative abundance of species for regional conservation planning: a case study. *Biol. Conserv.*, 98, 33–43.
- Peters, J., De Baets, B., Verhoest, N.E.C., Samson, R., Degroev, S., De Becker, P. *et al.* (2007). Random forests as a tool for ecohydrological distribution modelling. *Ecol. Model.*, 207, 304–318.
- Pulliam, H.R. (2000). On the relationship between niche & distribution. *Ecol. Lett.*, 3, 349–361.
- R Development Core Team (2010). *R: A Language & Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Available online at <http://www.R-project.org>.
- Randin, C.F., Jaccard, H., Vittoz, P., Yoccoz, N.G. & Guisan, A. (2009). Land use improves spatial predictions of mountain plant abundance but not presence-absence. *J. Veg. Sci.*, 20, 996–1008.
- Sagarin, R.D., Gaines, S.D. & Gaylord, B. (2006). Moving beyond assumptions to understand abundance distributions across the ranges of species. *Trends Ecol. Evol.*, 21, 524–530.
- Scherrer, D. & Körner, C. (2011). Topographically controlled thermal-habitat differentiation buffers alpine plant diversity against climate warming. *J. Biogeogr.*, 38, 406–416.
- Schoener, T.W. (1968). *Anolis* lizards of Bimini – resource partitioning in a complex fauna. *Ecology*, 49, 704.
- Silvertown, J. (2004). Plant coexistence & the niche. *Trends Ecol. Evol.*, 19, 605–611.
- Soberon, J. (2007). Grinnellian & Eltonian niches & geographic distributions of species. *Ecol. Lett.*, 10, 1115–1123.
- Stone, L. & Roberts, A. (1990). The checkerboard score & species distributions. *Oecologia*, 85, 74–79.
- Strobl, C., Malley, J. & Tutz, G. (2009). Introduction, An to Partitioning, Recursive: rationale, Application & Characteristics of Classification & Trees, Regression, Bagging & Forests, Random. *Psychol. Methods*, 14, 323–348.
- Thuiller, W., Araújo, M.B. & Lavorel, S. (2004). Do we need land-cover data to model species distributions in Europe? *J. Biogeogr.*, 31, 353–361.
- VanDerWal, J., Shoo, L.P., Johnson, C.N. & Williams, S.E. (2009). Abundance & the Environmental Niche: environmental Suitability Estimated from Niche Models Predicts the Upper Limit of Local Abundance. *Am. Natur.*, 174, 282–291.
- Vellend, M., Verheyen, K., Flinn, K.M., Jacquemyn, H., Kolb, A., Van Calster, H. *et al.* (2007). Homogenization of forest plant communities & weakening of species–environment relationships via agricultural land use. *J. Ecol.*, 95, 565–573.
- Vetaas, O.R. (2002). Realized & potential climate niches: a comparison of four *Rhododendron* tree species. *J. Biogeogr.*, 29, 545–554.
- Vittoz, P. & Engler, R. (2007). Seed dispersal distances: a typology based on dispersal modes & plant traits. *Bot. Helv.*, 117, 109–124.
- Welsh, A.H., Cunningham, R.B., Donnelly, C.F. & Lindenmayer, D.B. (1996). Modelling the abundance of rare species: statistical models for counts with extra zeros. *Ecol. Model.*, 88, 297–308.

SUPPORTING INFORMATION

Additional Supporting Information may be downloaded via the online version of this article at Wiley Online Library (www.ecologyletters.com).

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organised for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

Editor, David Storch

Manuscript received 3 January 2012

First decision made 7 February 2012

Manuscript accepted 7 March 2012