



ORIGINAL
ARTICLE



Do community-level models describe community variation effectively?

Andrés Baselga^{1,2*} and Miguel B. Araújo^{1,3,4}

¹Departamento de Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales, CSIC, C/José Gutiérrez Abascal, 2, 28006 Madrid, Spain, ²Departamento de Zoología, Facultad de Biología, Universidad de Santiago de Compostela, Rúa Lope Gómez de Marzoa s/n, 15782 Santiago de Compostela, Spain, ³Laboratorio Internacional de Cambio Global, UC-CSIC, Departamento de Ecología, Facultad de Ciencias Biológicas, PUC, Alameda 340, PC 6513677, Santiago, Chile, ⁴Rui Nabeiro Biodiversity Chair, CIBIO, Universidade de Évora, Largo dos Colegiais, 7000 Évora, Portugal

ABSTRACT

Aim The aim of community-level modelling is to improve the performance of species distributional models by taking patterns of co-occurrence among species into account. Here, we test this expectation by examining how well three community-level modelling strategies ('assemble first, predict later', 'predict first, assemble later', and 'assemble and predict together') spatially project the observed composition of species assemblages.

Location Europe.

Methods Variation in the composition of European tree assemblages and its spatial and environmental correlates were examined with cluster analysis and constrained analysis of principal coordinates. Results were used to benchmark spatial projections from three community-based strategies: (1) assemble first, predict later (cluster analysis first, then generalized linear models, GLMs); (2) predict first, assemble later (GLMs first, then cluster analysis); and (3) assemble and predict together (constrained quadratic ordination).

Results None of the community-level modelling strategies was able to accurately model the observed distribution of tree assemblages in Europe. Uncertainty was particularly high in southern Europe, where modelled assemblages were markedly different from observed ones. Assembling first and predicting later led to distribution models with the simultaneous occurrence of several types of assemblages in southern Europe that do not co-occur, and the remaining strategies yielded models with the presence of non-analogue assemblages that presently do not exist and that are much more strongly correlated with environmental gradients than with the real assemblages.

Main conclusions Community-level models were unable to characterize the distribution of European tree assemblages effectively. Models accounting for co-occurrence patterns along environmental gradients did not outperform methods that assume individual responses of species to climate. Unrealistic assemblages were generated because of the models' inability to capture fundamental processes causing patterns of covariation among species. The usefulness of these forms of community-based models thus remains uncertain and further research is required to demonstrate their utility.

Keywords

Bioclimatic envelope models, biotic interactions, community-level modelling, ecological niches, Europe, species distribution modelling, trees.

*Correspondence: Andrés Baselga, Departamento de Zoología, Facultad de Biología, Universidad de Santiago de Compostela, Rúa Lope Gómez de Marzoa s/n, 15782 Santiago de Compostela, Spain.
E-mail: andres.baselga@usc.es

INTRODUCTION

Community-level modelling combines distributions from several species to produce synthetic representations of the

'spatial pattern in the distribution of biodiversity at a collective community level' (Ferrier & Guisan, 2006). Three community-level modelling approaches have been proposed as alternatives to the familiar individual species distribution

models (Ferrier & Guisan, 2006). (1) 'Assemble first, predict later' is an approach whereby species distributions are first combined with classification or ordination methods and the resulting assemblages are then modelled using machine-learning or regression-based approaches. (2) 'Predict first, assemble later' is an approach whereby individual species distributions are modelled first and the resulting potential species distributions are then combined (i.e. the result is in fact the summation of individualistic models). (3) 'Assemble and predict together' is an approach whereby species distribution models are fitted using both environmental predictors and information on species co-occurrence. The potential usefulness of community-level modelling has been discussed by Ferrier *et al.* (2002) and Ferrier & Guisan (2006) on theoretical grounds. The authors suggest that the virtue of the 'assemble first, predict later' strategy is that it enforces congruence of spatially projected and observed assemblages, whereas the 'predict first, assemble later' and the 'assemble and predict together' strategies are expected to extrapolate beyond known assemblages. The authors also clarify that the 'predict first, assemble later' strategy does not consider patterns of species co-occurrence in the modelling process, whereas the 'assemble and predict together' strategy is the only one that allows individual species response curves to environmental variables to be combined with inter-specific covariation in species ranges (Ferrier & Guisan, 2006), thereby providing 'scope to address interactions between the distributions of different species, such as those resulting from competition or predation'.

Despite the suggestion that community-level modelling strategies are useful in a number of model applications, few studies have assessed the implications of the three proposed approaches (Ferrier *et al.*, 2002; Ferrier & Guisan, 2006). In fact, the three approaches are not alternative modelling strategies for the same problem. They are deeply rooted in different concepts – the Clementsian and Gleasonian concepts in community ecology – and therefore represent different hypotheses on the mechanisms driving variation in the composition of assemblages. The Clementsian concept views communities as rigid combinations of co-occurring species (Clements, 1916), and thus underlies the 'assemble first, predict later' approach, in which the distribution of assemblages is modelled as if the communities were stable and fixed entities. In contrast, the 'predict first, assemble later' approach can be interpreted as a formalization of the Gleasonian concept, which views assemblages as the result of collective individualistic responses of species to abiotic factors (Gleason, 1939). Finally, the 'assemble and predict together' strategy assumes the existence of interactions between species, but it avoids the extreme Clementsian view of communities being completely fixed and rigid entities (Callaway, 1997).

There are many examples illustrating that species are often sorted along environmental gradients in a seemingly individualistic fashion, but interactions among species have been shown to constrain these responses (e.g. Labandeira *et al.*,

2002; Koh *et al.*, 2004; Travis *et al.*, 2005; Araújo & Luoto, 2007; Thrush *et al.*, 2008). The 'assemble and predict together' strategy may thus be seen as providing a tool to reconcile the evidence of individual sorting of species along environmental gradients with the evidence of the existence of interactions among co-occurring species.

If the concepts underlying each of the alternative community-based models are different, should the outcomes of the models be different? What exactly are they representing? Are they effective tools with which to model community dynamics at varying spatial scales? Surprisingly, empirical tests assessing the merits of the three community-based modelling strategies are scarce and the results are inconclusive. For example, Ferrier *et al.* (2002) found no major differences between the 'assemble first, predict later' and the 'predict first, assemble later' strategies. Olden *et al.* (2006) found that a particular 'assemble and predict together' strategy (implemented with a multi-response artificial neural network, MANN) outperformed the predictive capacity of two alternative community-level modelling strategies (implemented with logistic regression and multiple discriminant analysis, respectively). However, because different methods were used in this latter study (e.g. neural networks versus logistic regression) it is difficult to know whether differences between model outputs arose because different algorithms were used or because of differences in the conceptual underpinning of the models. Ideally, if the goal is to assess the conceptual implications of alternative modelling strategies, the algorithms should be standardized to ensure comparability. When such standardization was carried out, the overall accuracy of community-based strategies (i.e. 'assemble and predict together') was reduced compared with that of familiar individualistic models ('predict first, assemble later') (Baselga & Araújo, 2009).

Comparisons of community-based models with individual species models can improve our understanding of the strengths and weaknesses of each approach (Leathwick *et al.*, 2005; Elith *et al.*, 2006; Olden *et al.*, 2006; Chatfield, 2008; Baselga & Araújo, 2009). However, appropriate tests for benchmarking community-based models should ideally examine how well they recover observed patterns of assemblage composition. Here, we start with the premise that it is a reasonable expectation that models accounting for patterns of co-occurrence along the environmental space should interpolate patterns of compositional variation of assemblages better than individual species models do. To examine this expectation, we use a well-known dataset on the distribution of European trees. We begin by characterizing the observed patterns of assemblage variation across geographical and environmental space. Then we model assemblage composition with the three modelling strategies discussed above. Finally, correlates of species composition in the observed assemblages are compared with the correlates among projected assemblages with the three community-based strategies.

MATERIALS AND METHODS

Biological data and environmental predictors

For this study, 158 native tree species and subspecies distributed across Europe were considered. This covers most of the important timber taxa of Europe, including most gymnosperm softwoods (Pinales and Taxales) and some hardwoods (Myricales, Malpighiales, Rosales, Juglandales and Fagales) (Humphries *et al.*, 1999). Trees were chosen because: (1) their distribution and ecology are relatively well known compared to other plant taxa; (2) their richness is correlated (Spearman correlation $\rho = 0.80$, $P < 0.001$) with the overall richness of the Atlas Flora Europaeae (AFE) dataset (Araújo & Williams, 2000); and (3) they are long-lived organisms and their distribution is relatively stable compared to some other groups. Furthermore, they have been used as a test case for studies of species distribution models for more than 10 years (Araújo & Williams, 2000; Thuiller *et al.*, 2003; Svenning & Skov, 2005; Rieckebusch *et al.*, 2008; Baselga & Araújo, 2009). The presence-absence data of species and subspecies constitute a subset of Atlas Flora Europaeae (Jalas & Suominen, 1972–1996), which was digitized by Lahti & Lampinen (1999). Data are located in 4419 UTM (Universal Transverse Mercator) 50×50 km grid cells. We used only 2130 grid cells, excluding most of the eastern European countries (except for the Baltic States) because of low recording efforts in these areas (Williams *et al.*, 2000). Taxa occurring in fewer than 25 grid cells were excluded from the analyses to avoid problems associated with modelling species with small sample sizes (Stockwell & Peterson, 2002): the reduced dataset comprised 119 taxa (see supplementary material in Baselga & Araújo, 2009), which are hereafter referred to as ‘species’ for simplicity.

For this study, we were limited to the use of two climatic predictor variables (high-resolution climatic data for 10' quadrats; New *et al.*, 2002) owing to the methodological constraints of the community-based model used herein (for details see Baselga & Araújo, 2009). For this reason, two climatic variables, GDD (mean growing degree days, > 5 °C) and P_{ann} (mean annual precipitation sum, mm), were selected from eleven predictor variables using PCA (principal components analysis). The first two components accounted for 86% of the variance. Examining the component loadings of the environmental variables, we selected the two variables most strongly correlated with first two PCA components: GDD (Component 1 loading = -0.97) and P_{ann} (Component 2 loading = 0.95). Therefore, GDD and P_{ann} were used to fit the models and to project species distributions (for details see Baselga & Araújo, 2009).

Description of observed patterns of assemblage variation

First, we characterized the patterns of variation in species composition among European tree assemblages and investigated how variation in species composition is correlated with

geographical and environmental factors. This characterization of compositional variation of assemblages across geographical space was then used as a benchmark against which to compare patterns of assemblage variation as projected by the three modelling strategies used. For each of the three approaches, we analysed the variation in species composition of the modelled assemblages using the same approach as for the observed assemblages. The specific characteristics of the models and how they were selected from similar alternative approaches are explained below.

Variation in assemblage composition between all pairs of cells was measured using Simpson's index of dissimilarity (β_{sim}). This index was preferred to other alternatives because it is independent of species richness gradients (Koleff *et al.*, 2003; Baselga, 2007). Simpson's dissimilarities were computed in R (R Development Core Team, 2006) using the function provided by Baselga (2010). This dissimilarity matrix was then used to aggregate data into clusters using the R cluster package (Maechler *et al.*, 2005). Clusters were built with the average linkage method. In order to visualize the spatial patterns of species composition, an arbitrary cut-off of 10 clusters was set using the `MAPTREE` package (White, 2007). The significance of these 10 groups was assessed by means of analysis of similarity (ANOSIM) tests (Clarke, 1993) using the `VEGAN` package (Oksanen *et al.*, 2007). Thereafter, the geographical distribution of the clusters was mapped using `IDRISI` (Clark Labs, 2000). This mapping allowed for a visual inspection of the geographical structure of European tree assemblages. The spatial and environmental correlates of assemblage composition were then assessed using constrained analyses of principal coordinates (CAP) (Oksanen *et al.*, 2007). This analysis allowed the relationship between variability in the table of species occurrences and in the tables of the two sets of predictor variables (environmental factors: GDD and P_{ann} ; and spatial position: longitude and latitude) to be examined. CAP was selected because it can be computed with any dissimilarity index and, therefore, Simpson dissimilarity was preserved in the constrained ordination. Owing to the large size of our matrices and the computational limitations of R, the significance of variables could not be computed with the permutation tests (`VEGAN` command `permutest`; Oksanen *et al.*, 2007). For this reason, we have not tested the inclusion of further variables, or polynomial terms, in order to avoid the inclusion of non-significant terms that could inflate the amount of explained variation. GDD, P_{ann} , longitude and latitude are likely to be good predictors of variation in tree species composition, so the amount of variation explained in our results should be considered conservative, enabling the comparison of both sets of predictors (environment and geographical position) with the same number of variables. Finally, variation in species composition was partitioned among environmental and geographical predictors, subtracting the variation explained by each set from the variation explained by a complete model (Legendre & Legendre, 1998), yielding estimates of fractions independently and jointly explained by environment and spatial position (Borcard *et al.*, 1992).

Selection of modelling procedures

There are several methods enabling the implementation of each of the community-level model strategies investigated herein. We used generalized linear models (GLMs) in conjunction with clustering methodologies for the ‘predict first, assemble later’, and for the ‘assemble first, predict later’ strategies. Constrained quadratic ordination (CQO) (Yee, 2004) was used for the ‘assemble and predict together’ strategy. GLMs and CQO are analogous and therefore should provide comparable results, except for the fact that the latter explicitly accounts for the co-occurrence and co-exclusion of species along environmental gradients. Therefore, straightforward comparisons between GLMs and CQO methods are possible, and differences are directly attributable to changes in patterns of range overlap between species. For an extended discussion of the two approaches, see Baselga & Araújo (2009).

Assemble first, predict later

The distributions of the 10 major clusters found in the observed assemblages were modelled using GLMs with binomial errors, logit link and quadratic functions. Response variables were the occurrence of each one of the 10 cluster classes, and predictor variables included GDD and P_{ann} . No variable selection was implemented, and quadratic and linear terms of GDD and P_{ann} were automatically included in models for all clusters in order to allow full comparability with the other strategies. The functions fitted using the complete dataset were used to project the cluster distributions under the current climate. Because each cluster (assemblage type) is considered a rigid entity, several clusters can be predicted to be present in the same cell. The projected presence of each one of the 10 clusters or several clusters together (up to 30 combinations in our results), as well as the number of clusters predicted to be present in each cell were mapped using IDRISI (Clark Labs, 2000). This allowed visual comparison of the observed and projected geographical distribution of European tree assemblages. Because the ‘assemble first, predict later’ strategy does not preserve the identities of species projected for each cell, no further analysis of the geographical and environmental structure of the assemblages could be conducted.

Predict first, assemble later

Species distributions were modelled using GLMs with binomial errors, logit link and quadratic functions. Response variables were species occurrence records, and predictor variables included GDD and P_{ann} (for further details see ‘individualistic models’ in Baselga & Araújo, 2009). The functions fitted using the complete dataset were used to interpolate the species distributions under the current climate. We built a table of predicted presences by cells that was subject to the same analyses as the observed assemblages. Differences between this strategy and the observed patterns, and those derived from other strategies, were assessed by means of two

analyses. First, we measured the correlation between dissimilarity matrices (Mantel tests conducted in R with the package *vegan*; Oksanen *et al.*, 2007). Second, we compared the fractions of variation in species composition explained by environmental and spatial factors in the CAP analyses.

Assemble and predict together

A rank-2 CQO (two latent variables) was fitted to the occurrence of the 119 species, using binomial errors, logit link, and GDD and P_{ann} as predictor variables (for more details see R script in Baselga & Araújo, 2009). As with GLMs, the functions fitted with CQO were used to interpolate species distributions under the current climate, and a table of the predicted species compositions was built. This table was subject to all analyses applied to the observed communities and the communities predicted by GLMs, as explained above.

RESULTS

Compositional variation in the observed assemblages

The 10 biogeographical clusters summarizing different groups of tree species were significantly different from each other (ANOSIM $R = 0.77$, $P < 0.01$). The observed clusters (Fig. 1a) match the widely recognized discontinuity between Mediterranean and Eurosiberian biotas (Rivas-Martínez, 1990). Mediterranean tree assemblages were divided into several major groups, corresponding to the three southern peninsulas, as well as other more restricted clusters, whereas Eurosiberian tree assemblages were much more uniform in composition and were only further subdivided into a boreal and a temperate cluster. The analysis of the environmental and spatial correlates of the observed patterns of variation in species composition showed that a complete model including both the environmental (GDD and P_{ann}) and spatial (longitude and latitude) set of predictors explained only 26% of the variation in species composition. Partitioning of the explained variation showed that fractions exclusively explained by the environment (4%) and geography (7%) were small compared with the fraction explained by the collinear effects of both sets of variables (16%).

Compositional variation from the ‘assemble first, predict later’ strategy

The interpolation of the 10 observed clusters using the function fitted by GLMs showed a clear lack of environmental structure in Mediterranean assemblages (Fig. 1b). This is evident from the high number of clusters that are predicted to occur simultaneously in southern cells (Fig. 1c). Only boreal and temperate Eurosiberian clusters are projected with relative accuracy, whereas Mediterranean clusters are extrapolated to regions where they are not actually present, generating up to 29 combinations of cluster occurrences. No further analyses could be conducted owing to methodological constraints imposed by

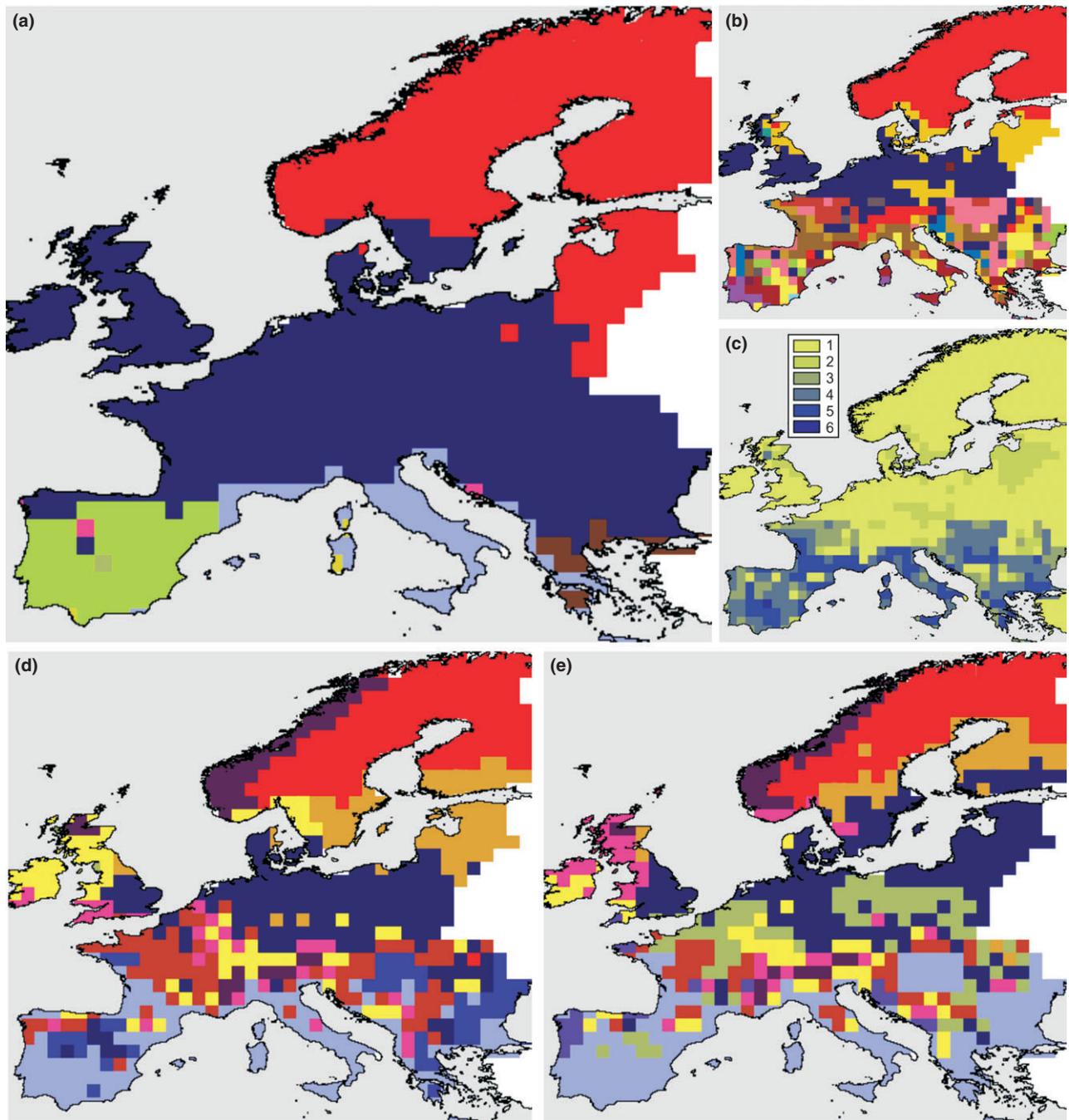


Figure 1 Distribution of 10 clusters summarizing the geographical structure of European tree assemblages. (a) Observed assemblages; (b) projection yielded by the 'assemble first, predict later' strategy; (c) number of different clusters projected in the same cell by the 'assemble first, predict later' strategy; (d) projection by the 'predict first, assemble later' strategy; and (e) projection by the 'assemble and predict together' strategy.

this strategy, mainly derived from the loss of species identities when they are assembled.

Compositional variation from the 'predict first, assemble later' strategy

The cluster analysis of species composition interpolated by GLMs showed clear differences from the observed pattern

of species composition of European trees (Fig. 1d). The assemblages interpolated by GLMs were mostly structured in latitudinal bands (i.e. the whole Mediterranean region was predicted to harbour similar assemblages that were clustered together). The 10 clusters projected by GLMs were even more clearly structured than observed ones (ANOSIM $R = 0.93$, $P < 0.01$). Mantel tests revealed a moderate correlation between the observed dissimilarity matrix and that yielded

by GLMs ($r = 0.67$, $P < 0.01$). Considering assemblages projected by GLMs, the amount of variation in species composition explained by spatial and climatic factors was markedly higher (59%). Partitioning of the variation in species composition interpolated by GLMs yielded a fraction exclusively explained by environmental factors (17%) that was much larger than that exclusively explained by spatial position (5%); 37% of this variation was explained by the collinear effect of the two sets of variables.

Compositional variation from the 'assemble and predict together' strategy

Results of the cluster analysis for species composition interpolated with CQO (Fig. 1e) were very similar to those projected with GLMs but different from the observed ones; assemblages projected by CQO were mostly structured in latitudinal bands. As for GLMs, the 10 clusters projected by CQO were more clearly structured than the observed ones (ANOSIM $R = 0.93$, $P < 0.01$). Mantel tests yielded moderate correlations between the observed dissimilarity matrix and those interpolated by the CQO model ($r = 0.65$, $P < 0.01$). In contrast, dissimilarity matrices derived from assemblages interpolated by GLM and CQO models were strongly correlated ($r = 0.97$, $P < 0.01$). For the assemblages interpolated by CQO, the amount of variation in species composition explained by spatial and environmental factors was similar to that of GLM assemblages (57%). Partitioning of the variation in species composition interpolated by CQO models yielded purely environmental (18%), spatial (4%) and shared (35%) fractions that were also similar to those derived from the GLMs.

DISCUSSION

Previous studies have assessed whether community-based approaches can improve projections of individual species distributions (Leathwick *et al.*, 2005; Elith *et al.*, 2006; Olden *et al.*, 2006; Chatfield, 2008; Baselga & Araújo, 2009). Here, we address a slightly different question. We ask whether community-level models characterize community variation effectively. The results show that the three community-level model strategies tested herein yield patterns of assemblage composition that are markedly different from the observed ones. This mismatch between modelled and observed assemblages is interpreted as an indication that the models failed to capture the underlying mechanisms generating co-occurrence (and co-exclusion) of species distributions. Revealingly, community-based models that take statistical patterns of co-occurrence among species distributions into account did not reproduce observed assemblages more closely than approaches that simply combined the results of individualistic models. In fact, both approaches yielded equally unrealistic patterns of assemblage variation.

The starting premise of this study was that models accounting for patterns of co-occurrence among species distributions were likely to characterize the observed compo-

sitional variation of assemblages more closely than individual species models. The rationale was that models taking into account the co-occurrence of species along environmental gradients are more likely to reflect, at least partially, underlying patterns of species interactions than models ignoring patterns of co-occurrence among species. Of course, not all patterns of co-occurrence or co-exclusion reflect functional interactions between species (Araújo & Guisan, 2006). Patterns of co-occurrence may also be caused by shared physiological requirements. When this is the case, species may respond to climatic factors in a similar fashion without necessarily interacting. Likewise, co-exclusion among species can arise because of species having different environmental requirements and thus occupying different parts of environmental gradients. In other cases, species with similar environmental requirements may have disjoint distributions because of dispersal constraints. Regardless of whether biotic interactions or biogeographical contingencies are responsible for generating the patterns of co-occurrence in the data, it is expected that the mechanisms generating them are somehow reflected in the outputs of the 'assemble and predict together' and the 'assemble first, predict later' strategies. The degree to which mechanisms generating co-occurrence and co-exclusion among species distributions are reflected in the statistical summaries of these two community-based models is unknown, but it is reasonable to expect that they should convey more information than the 'predict first, assemble later' approach, which assumes unconstrained responses of species to environmental gradients. However, the strong discrepancies between the observed and modelled assemblages suggest that the statistical signals associated with co-occurrence among European tree distributions are ineffectively characterized by community-based models. In fact, the 'predict first, assemble later' and 'assemble and predict together' strategies yielded patterns of variation in species composition that were highly correlated, supporting the view that the inclusion of species co-occurrences along environmental gradients as input for the models did not improve their ability to fit real patterns of assemblage variation.

In addition to comparing observed and modelled assemblages, we analysed how spatial and environmental correlates changed between them. Specifically, we treated observed correlations as a reference with which to compare the correlations obtained with the three community-level model strategies. The correlation analysis showed that the observed composition of assemblages of European trees is poorly explained by geographical or environmental gradients. In contrast, the patterns of interpolated assemblage variation with any of the community-level models are strongly correlated with climatic and geographical factors. The reason for such differences is that, while observed assemblages may be weakly correlated with climatic gradients, community-level models force this correlation to be strong. In fact, models take into account existing climate gradients and patterns of covariation of species in climate space and interpolate 'potential' assemblages that may have no bearing on observed ones. The degree

of matching between observed and potential assemblages is possibly related to the degree of equilibrium of individual species distributions with the predictor variables used to fit the models. Low equilibrium, that is, species being absent from many climatically suitable areas, may arise for several reasons, such as species dispersal limitation, inappropriate combinations of biotic interactions, or unsuitable land use (Araújo & Pearson, 2005).

Global analyses of equilibrium of species distributions and climate are still lacking, but it is expected that areas with marked historical signatures will harbour species with lower degrees of equilibrium with climate (Araújo & Pearson, 2005). When this is the case, community-level model strategies are more likely to fail to describe community variation effectively. Indeed, our results show that modelled assemblages in Europe moderately match observed ones in northern regions, where tree species have been shown to be at greater equilibrium with climate (Svenning & Skov, 2004). In contrast, matching of observed and modelled assemblages is reduced in southern regions, where species have lower degrees of equilibrium with climate (Svenning & Skov, 2004) (Fig. 1). A contrasting performance of models in southern and northern Europe possibly reflects the historical effects of glacial–interglacial periods (Hewitt, 1999). Species that today occur at higher latitudes are likely either to have persisted in ‘cryptic’ refugia (Bhagwat & Willis, 2008) or to have dispersed northwards from southern refugia (Taberlet *et al.*, 1998; Hewitt, 1999), thus showing improved ability to track climate changes and reach equilibrium with the current climate. In contrast, Mediterranean assemblages are the result of the long-term persistence of isolated populations affected by dispersal limitation, as suggested in the context of species distribution modelling studies (Svenning & Skov, 2004) and supported later by phylogeographic data (Petit *et al.*, 2005). At the assemblage level, in addition to the latitudinal richness gradient (Svenning & Skov, 2007), the lower degree of equilibrium of southern species with climate could be related to the fact that the distribution of assemblages in southern Europe is not structured in latitudinal bands following climatic gradients. This distribution of assemblages leads to the low amount of variation explained by climatic correlates, as previously found by Svenning & Skov (2005), and for other groups, such as reptiles and amphibians (Araújo *et al.*, 2008), and longhorn beetles (Baselga, 2008). The post-glacial recolonization process yields a wide-ranging uniform assemblage in the Eurosiberian region, in which only two large temperate and boreal regions are defined. As shown in several studies, wide-ranging biogeographical units (resulting from wide-ranging species) are more likely to be related to contemporary climatic constraints than narrow-ranging units, which are more commonly associated with historical factors (Jetz & Rahbek, 2002; Svenning & Skov, 2005; Araújo *et al.*, 2008).

Mismatches between observed and potential assemblages could be reduced if spatial predictors were added to community-level models. The same is true for individual species distribution models, but, in both cases, the price of including

spatial predictors for improving the fit of the model with the observations is to hinder their predictive ability (Dormann *et al.*, 2007). As such, spatial predictors are justifiable outside the realm of prediction, or when it is not essential to understand the mechanisms driving the distributions.

CONCLUSIONS

Our implementation of community-level models was unable to accurately characterize the distribution of European assemblages of trees. Community-based models accounting for co-occurrence patterns along environmental space did not match observed assemblages better than familiar species distribution models that assume individualistic responses of species to environmental gradients. It seems clear that many factors interact to shape compositional variation among assemblages of trees, and that without approaches that account mechanistically for such interactions it is difficult to represent existing community complexities, let alone future ones. Another question is whether community-based models help to model the distributions of individual species by providing information on shared responses of species to environmental variation. In situations where biological sampling is insufficient, such prospects look promising, but in a companion paper (Baselga & Araújo, 2009) we did not find unequivocal evidence in support of this assertion. The usefulness of community-based models thus remains uncertain, and further research is required to demonstrate their utility.

ACKNOWLEDGEMENTS

Digital species distribution data were kindly supplied by Raino Lampinen and pre-processed by the late Chris Humphries and Paul Williams. We thank Simon Ferrier and one anonymous referee for valuable comments and suggestions. Research by A.B. and M.B.A. was originally funded by the EC FP6 MACIS (Minimisation of and Adaptation to Climate Change: Impacts on Biodiversity, contract no. 044399) project. A.B. is currently funded by the Spanish Ministry of Science and Innovation (project CGL2009-10111/BOS). M.B.A. is currently funded by the EC FP6 ECOCHANGE project (Challenges in Assessing and Forecasting Biodiversity and Ecosystem Changes in Europe, contract no. 036866-GOCE) and by the Spanish Ministry of Science and Innovation (complementary action no. CGL2008-01198-E/BOS).

REFERENCES

- Araújo, M.B. & Guisan, A. (2006) Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, **33**, 1677–1688.
- Araújo, M.B. & Luoto, M. (2007) The importance of biotic interactions for modelling species distributions under climate change. *Global Ecology and Biogeography*, **16**, 743–753.

- Araújo, M.B. & Pearson, R.G. (2005) Equilibrium of species' distributions with climate. *Ecography*, **28**, 693–695.
- Araújo, M.B. & Williams, P.H. (2000) Selecting areas for species persistence using occurrence data. *Biological Conservation*, **96**, 331–345.
- Araújo, M.B., Nogués-Bravo, D., Diniz-Filho, J.A.F., Haywood, A.M., Valdes, P.J. & Rahbek, C. (2008) Quaternary climate changes explain diversity among reptiles and amphibians. *Ecography*, **31**, 8–15.
- Baselga, A. (2007) Disentangling distance decay of similarity from richness gradients: response to Soininen *et al.* 2007. *Ecography*, **30**, 838–841.
- Baselga, A. (2008) Determinants of species richness, endemism and turnover in European longhorn beetles. *Ecography*, **31**, 263–271.
- Baselga, A. (2010) Partitioning the turnover and nestedness components of beta diversity. *Global Ecology and Biogeography*, **19**, 134–143.
- Baselga, A. & Araújo, M.B. (2009) Individualistic vs community modelling of species distributions under climate change. *Ecography*, **35**, 55–65.
- Bhagwat, S.A. & Willis, K.J. (2008) Species persistence in northerly glacial refugia of Europe: a matter of chance or biogeographical traits? *Journal of Biogeography*, **35**, 464–482.
- Borcard, D., Legendre, P. & Drapeau, P. (1992) Partialling out the spatial component of ecological variation. *Ecology*, **73**, 1045–1055.
- Callaway, R.M. (1997) Positive interactions in plant communities and the individualistic-continuum concept. *Oecologia*, **112**, 143–149.
- Chatfield, B.S. (2008) *How to find the one that got away. Predicting the distribution of temperate demersal fish from environmental variables*. PhD Thesis, School of Earth and Geographical Sciences, University of Western Australia, Perth.
- Clark Labs (2000) *Idrisi 32.02. GIS software package*. Clark University, Worcester, MA.
- Clarke, K.R. (1993) Non-parametric multivariate analysis of changes in community structure. *Australian Journal of Ecology*, **18**, 117–143.
- Clements, F.E. (1916) *Plant succession: an analysis of the development of vegetation*. Carnegie Institution of Washington, Washington, DC.
- Dormann, C.F., McPherson, J.M., Araújo, M.B., Bivand, R., Bolliger, J., Carl, G., Davies, R.G., Hirzel, A., Jetz, W., Kissling, W.D., Kuhn, I., Ohlemuller, R., Peres-Neto, P.R., Reineking, B., Schroder, B., Schurr, F.M. & Wilson, R. (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, **30**, 609–628.
- Elith, J., Graham, C.H., Anderson, R.P. *et al.* (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.
- Ferrier, S. & Guisan, A. (2006) Spatial modelling of biodiversity at the community level. *Journal of Applied Ecology*, **43**, 393–404.
- Ferrier, S., Drielsma, M., Manion, G. & Watson, G. (2002) Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. II. Community-level modelling. *Biodiversity and Conservation*, **11**, 2309–2338.
- Gleason, H.A. (1939) The individualistic concept of the plant association. *American Midland Naturalist*, **21**, 92–110.
- Hewitt, G.M. (1999) Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society*, **68**, 87–112.
- Humphries, C.H., Araújo, M.B., Williams, P.H., Lampinen, R., Lahti, T. & Uotila, P. (1999) Plant diversity in Europe: Atlas Florae Europaeae and WORLDMAP. *Acta Botanica Fennica*, **162**, 11–21.
- Jalas, J. & Suominen, J. (1972–1996) *Atlas Florae Europaeae*. The Committee for Mapping the Flora of Europe and Societas Biologica Fennica Vanamo, Helsinki.
- Jetz, W. & Rahbek, C. (2002) Geographic range size and determinants of avian species richness. *Science*, **297**, 1548–1551.
- Koh, L.P., Dunn, R.R., Sodhi, N.S., Colwell, R.K., Proctor, H.C. & Smith, V.S. (2004) Species coextinctions and the biodiversity crisis. *Science*, **305**, 1632–1634.
- Koleff, P., Gaston, K.J. & Lennon, J.K. (2003) Measuring beta diversity for presence-absence data. *Journal of Animal Ecology*, **72**, 367–382.
- Labandeira, C.C., Johnson, K.R. & Wilf, P. (2002) Impact of the terminal Cretaceous event on plant–insect associations. *Proceedings of the National Academy of Sciences USA*, **99**, 2061–2066.
- Lahti, T. & Lampinen, R. (1999) From dot maps to bitmaps – Atlas Florae Europaeae goes digital. *Acta Botanica Fennica*, **162**, 5–9.
- Leathwick, J.R., Rowe, D., Richardson, J., Elith, J. & Hastie, T. (2005) Using multivariate adaptive regression splines to predict the distributions of New Zealand's freshwater diadromous fish. *Freshwater Biology*, **50**, 2034–2052.
- Legendre, P. & Legendre, L. (1998) *Numerical ecology*, 2nd edn. Elsevier, Amsterdam.
- Maechler, M., Rousseeuw, P., Struyf, A. & Hubert, M. (2005) *Cluster analysis basics and extensions*. R package. Available at: <http://cran.r-project.org/>.
- New, M., Lister, D., Hulme, M. & Makin, I. (2002) A high-resolution data set of surface climate over global land areas. *Climate Research*, **21**, 1–25.
- Oksanen, J., Kindt, R., Legendre, P. & O'Hara, B. (2007) *vegan: Community Ecology Package*. R package version 1.8-5. Available at: <http://cran.r-project.org/>.
- Olden, J.D., Joy, M.K. & Death, R.G. (2006) Rediscovering the species in community-wide predictive modelling. *Ecological Applications*, **16**, 1449–1460.
- Petit, R.J., Hampe, A. & Cheddadi, R. (2005) Climate changes and tree phylogeography in the Mediterranean. *Taxon*, **54**, 877–885.
- R Development Core Team (2006) *R: a language and environment for statistical computing*. R Foundation for

- Statistical Computing, Vienna. Available at: <http://www.r-project.org>.
- Rickebusch, S., Thuiller, W., Hickler, T., Araújo, M.B., Sykes, M.T., Schweiger, O. & Laffourcade, B. (2008) Incorporating the effects of changes in vegetation functioning and CO₂ on water availability in plant habitat models. *Biology Letters*, **4**, 556–559.
- Rivas-Martínez, S. (1990) Bioclimatology and biogeography of West Europe. *Climate and global change. Proceedings of the European School of Climatology and Natural Hazards* (ed. by V.C. Duplessey, A. Poms and R. Fantecli), pp. 225–246. European Commission, Brussels.
- Stockwell, D.R.B. & Peterson, A.T. (2002) Effects of sample size on accuracy of species distribution models. *Ecological Modelling*, **148**, 1–13.
- Svenning, J.C. & Skov, F. (2004) Limited filling of the potential range in European tree species. *Ecology Letters*, **7**, 565–573.
- Svenning, J.C. & Skov, F. (2005) The relative roles of environment and history as controls of tree species composition and richness in Europe. *Journal of Biogeography*, **32**, 1019–1033.
- Svenning, J.C. & Skov, F. (2007) Could the tree diversity pattern in Europe be generated by postglacial dispersal limitation? *Ecology Letters*, **10**, 453–460.
- Taberlet, P., Fumagalli, L., Wust-Saucy, A.G. & Cosson, J.F. (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, **7**, 453–464.
- Thrush, S.F., Coco, G. & Hewitt, J.E. (2008) Complex positive connections between functional groups are revealed by Neural Network Analysis of ecological time series. *The American Naturalist*, **171**, 669–677.
- Thuiller, W., Araújo, M.B. & Lavorel, S. (2003) Generalized models vs. classification tree analysis: predicting spatial distributions of plant species at different scales. *Journal of Vegetation Science*, **14**, 669–680.
- Travis, J.M.J., Brooker, R.W. & Dytham, C. (2005) The interplay of positive and negative species interactions across an environmental gradient: insights from an individual-based simulation model. *Biology Letters*, **1**, 5–8.
- White, D. (2007) *maptree: mapping, pruning, and graphing tree models*. R package version 1.4-4. Available at: <http://www.epa.gov/wed/pages/staff/white/>.
- Williams, P.H., Humphries, C.H., Araújo, M.B., Lampinen, R., Hagemeyer, W., Gasc, J.-P. & Mitchell-Jones, A. (2000) Endemism and important areas for conserving European biodiversity: a preliminary exploration of atlas data for plants and terrestrial vertebrates. *Belgian Journal of Entomology*, **2**, 21–46.
- Yee, T.W. (2004) A new technique for maximum-likelihood canonical Gaussian ordination. *Ecological Monographs*, **74**, 685–701.

BIOSKETCHES

Andrés Baselga (PhD, University of Santiago de Compostela) is interested in the integration of several biodiversity-related disciplines (including phylogenetics and macroecology) as a way to search for robust hypotheses for the causes of biodiversity. He is especially interested in the integration of beta diversity patterns in the central debate about large-scale gradients of biodiversity.

Miguel B. Araújo (PhD, University of London) is a senior researcher of the Spanish Research Council (CISC) at the National Museum of Natural Sciences in Madrid (<http://www.biochange-lab.eu>), and 'Rui Nabeiro' Biodiversity Chair (visiting full professor) at the University of Évora (<http://www.catedra.uevora.pt/rui-nabeiro/>). His research interests span various topics of conservation biogeography, global-change biology and macroecology.

Editors: Kate Parr and Robert Whittaker