Individualistic vs community modelling of species distributions under climate change

Andrés Baselga and Miguel B. Araújo

A. Baselga, Depto de Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales, CSIC, C/Gutierrez Abascal, 2, ES-28006 Madrid, Spain, and Depto de Zoología, Facultad de Biología, Univ. de Santiago de Compostela, Rúa Lope Gómez de Marzoa, ES-15782 Santiago de Compostela, Spain. – M. B. Araújo (maraujo@mncn.csic.es), Depto de Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales, CSIC, C/Gutierrez Abascal, 2, ES-28006 Madrid, Spain, Laboratorio Internacional de Cambio Global, UC-CSIC, Depto de Ecología, Facultad de Ciencias Biológicas, PUC, Alameda 340, PC 6513677, Santiago, Chile, and Rui Nabeiro Biodiversity Chair, Univ. de Évora, Largo dos Colegiais, PT-7000 Évora, Portugal.

Studies investigating the consequences of future climate changes on species distributions usually start with the assumption that species respond to climate changes in an individualistic fashion. This assumption has led researchers to use bioclimate envelope models that use present climate-range relationships to characterize species' limits of tolerance to climate, and then apply climate-change scenarios to enable projections of altered species distributions. However, there are techniques that combine climate variables together with information on the composition of assemblages to enable projections that are expected to mimic community dynamics. Here, we compare, for the first time, the performance of GLM (generalized linear model) and CQO (canonical quadratic ordination; a type of community-based GLM) for projecting distributions of species under climate change scenarios. We found that projections from these two methods varied both in terms of accuracy (GLM providing generally more accurate projections than CQO) and in the broad diversity patterns yielded (higher species richness values projected with CQO). Model outputs were also affected by species-specific traits, such as species range size and species geographical positions, supporting the view that methods are sensitive to different degrees of equilibrium of species distributions with climate. This study reveals differences in projections between individual- and community-based approaches that require further scrutiny, but it does not find support for unsupervised use community-based models for investigating climate change impacts on species distributions. Reasons for this lack of support are discussed.

Studies using models to investigate the consequences of future climate changes on species distributions often start with the assumption that species respond to climate changes in an individualistic fashion. In other words, that complex community dynamics, or simpler biotic dependencies between pairs of species, are not important at the grand scale in which global environmental changes operate. The assumption that species respond individualistically to climate changes has received support from analysis of the fossil record (reviewed by Graham and Grimm 1990). For example, analysis of pollen cores shows that even when species composition remains relatively stable through time, abundances change significantly and non-analogue communities can emerge (Williams et al. 2001, Simakova 2006). There is also evidence of individualistic responses of species from short-term experimental manipulations of climatic variables (McGeoch et al. 2006). Despite an abundant literature supporting the predominance of individualistic species responses to climate changes, the topic remains controversial (Callaway 2007). Proponents of community responses of species to climate changes

highlight that synchronous dynamics of organisms in the fossil record have been reported. For example, Labandeira and co-workers argued that major extinctions of flowering plants were followed by the decline in the diversity of insects during the middle to late Pennsylvanian extinctions, during the Permian event and at the Cretaceous/Tertiary boundary (Labandeira 2002, Labandeira et al. 2002). Others have reported cases of stability in the composition of assemblages over time (Lyons 2003). Developments in community ecology have also demonstrated that biotic interactions (namely positive interactions) can affect species' ability to adapt to changes in their environment. For example, Jordano (2000) suggests that >90% of tropical plant species rely on animals for the dispersal of their seeds. Should these animals be removed from the system, plants would find it extremely difficult to track climate changes through dispersal. Simulation (Koh et al. 2004, Travis et al. 2005) and modelling studies (Araújo and Luoto 2007) provided additional support for the idea that simple biotic interactions have an important role

in either facilitating or preventing adaptation of species to climate change.

Modifying individual species distributions models to account for complex biotic interactions is far from trivial (Araújo and Luoto 2007, Heikkinen et al. 2007). Indeed, it requires information on the biology of organisms that is either unavailable or available for specific case studies alone. An alternative approach involves the statistical analysis of species co-occurrence in environmental space. This approach has been termed community-based modelling of species distributions (Ferrier et al. 2002). Community modelling relates distributions of multiple species to sets of environmental variables and produces an analysis of the collective properties of species distributions (Ferrier and Guisan 2006). It is implicit to these methods that statistical patterns of co-occurrence (and co-exclusion) among species capture meaningful biotic interactions among species, thereby providing a useful tool for modelling community dynamics.

Although community-based models have been previously weighted against individual species distribution models, comparisons have been performed for models projecting distributions in the same region (and time) where data were sampled (Leathwick et al. 2005, Elith et al. 2006, Chatfield 2008). As widely acknowledged, though, climate change poses challenges of generality (making useful predictions beyond the training data) rather than precision (making accurate predictions within the training data) (Guisan and Zimmermann 2000), which is the criterion usually used for benchmarking of the species distribution models (Araújo and Rahbek 2006). Therefore, it is possible that models thought to be robust for a particular application (Segurado and Araújo 2004, Elith et al. 2006, Meynard and Quinn 2007, Tsoar et al. 2007) might provide poor results for others (see Araújo et al. 2005, or the recent debate between Peterson et al. 2007, and Phillips 2008). Here, we provide a first comparison of community and individual-based models, using techniques based in generalized linear models (GLM), for forecasting changes in the potential distributions of species under climate change. In particular we examine: 1) whether projections made with a particular implementation of community-based models differ, in terms of accuracy, from equivalent individualistic models. 2) Whether differences in accuracy of community vs individualistic models can be associated with species-specific traits, such as range size and geographical position. 3) Whether differences in projected species richness and assemblage composition exist and, if so, whether strong biological explanations for these divergent patterns can be inferred.

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 is relatively well known compared with 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) data set (Araújo and Williams 2000); and 3) they are long-lived organisms and their distribution is relatively stable in comparison with some other groups. The species and subspecies presence–absence data are a subset of AFE (Jalas and Suominen 1972–1996), which was digitized by Lahti and 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 <25 grid cells were excluded from analyses to avoid problems of modelling species with small sample sizes (Stockwell and Peterson 2002); the reduced dataset comprised 119 taxa (Supplementary material Appendix 1) that are referred as ''species'' throughout the text for simplicity.

For this study, we were limited to the use of two environmental predictor variables (see explanation below). In order to select the two variables, we started with the analysis of a larger set of variables that are often considered in studies modelling distributions of trees and other plant species (Thuiller et al. 2003): GDD (growing degree days); TANN (mean annual temperature); MTC (mean temperature of the coldest month); MWC (mean temperature of the warmest month); A2P (mean ratio of annual actual evapotranspiration over annual potential evapotranspiration); PANN (mean annual precipitation sum); PSPR (mean annual spring precipitation); PSUM (mean annual summer precipitation); PAUT (mean annual autumn precipitation); PWIN (mean annual winter precipitation); and RANN (annual solar long-wave radiation). The eleven predictor variables were analysed with 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 correlated with first two PCA components: GDD (component 1 loading = −0.97) and PANN (component 2 loading = 0.95). Therefore, GDD and PANN were used to fit the models and to project species distributions under present (high resolution climatic data for 10' quadrats; New et al. 2002), and future climatic conditions (scenario A1 of the HadCM3 GCM for 2050 and 2080; Schröter et al. 2005).

Selection of the community-based model used

There are several methods that allow exploring assemblage-level interactions in species distribution models. Approaches include GDM (generalized dissimilarity modelling; Ferrier et al. 2007), MARS-community (multiple adaptive regression splines adapted for community modelling; Leathwick et al. 2005), CQO (canonical quadratic ordination; Yee 2004), and CAO (canonical additive ordination; Yee 2006). GDM models compositional dissimilarity in assemblages and have been successfully used as a pre-processing step for modelling species distributions (Ferrier et al. 2007). However, GDM, itself,

was not conceived to make projections for individual species distributions. MARS-community uses non-linear relationships and identifies the combination of environmental variables best able to project the occurrence of the component species. MARS has been shown to make precise projections within a training set (Moisen and Frescino 2002), but its ability to make useful projections under climate change remains untested. CQO and CAO explicitly account for co-occurrence and exclusion patterns as other ordination techniques, while enabling projections of the distribution of each species. CQO is fitted with GLM (generalized linear model) and assumes quadratic responses of species to predictor environmental variables. CAO does not make a priori assumptions about the shape of the species responses to the environmental predictors and is fitted with GAM (generalized additive model). After carefully considering the alternative options, we decided to use CQO for three reasons. First, results are easily interpretable. Second, CQO uses quadratic responses of species to environmental predictors, which are well rooted in the ecological theory (Austin 2002). Third, the current implementation of the CQO allows fitting two latent predictor variables to models (Rank-2 CQO), whereas CAO only allows fitting one latent variable (Rank-1 CAO). Since species distributions are best modelled with respect to, at least, two variables (one reflecting the available energy and the other the available water), CQO was preferred over CAO.

Given that our objective was to compare community-based versus individual distribution models, we were careful to avoid differences in model outputs that could be attributed to slightly different combinations of environmental predictors in models. This would happen, for example, if we used latent variables summarised with principal component analysis, as the structure of the correlation matrix would naturally change when climate variables were projected into the future. To avoid this problem, we identified two orthogonal variables (with PCA, see above), thought to be ecologically meaningful, and fitted these variables to 1) individual distribution models for each species (referred as GLM throughout the text) and 2) a community-based model simultaneously including all the species in a Rank-2 CQO model (referred as CQO throughout the text). CQO identifies a set of orthogonal latent variables from a combination of several variables. By using only two orthogonal variables we ensured that the latent variables were equivalent to the individual variables entering the model. With this procedure we ensured that differences between the Rank-2 CQO and GLM models could only be attributable to the co-occurrence/exclusion patterns (Fig. 1).

GLM model

Species distributions were modelled using GLM with binomial errors, logit link and quadratic functions. Response variables were species occurrence records and predictor variables were GDD and PANN (for more details see R script in Supplementary material Appendix 1). No variable selection was implemented and quadratic and linear terms of GDD and PANN were automatically included in models for all species in order to allow full comparability with CQO. The functions fitted using the complete dataset were used to project the species distributions under current and future (2050 and 2080) climates.

CQO model

We first tested how well the quadratic functions described the species responses to the environmental predictors, by fitting both Rank-1 CAO and CQO and visually comparing the response curves (Fig. 2). We found that quadratic functions described reasonably well species responses to environmental predictors (most species have a bell-shaped response curve), thus lending support for our choice of using CQO rather than CAO. A Rank-2 CQO was fitted to the occurrence of the 119 species, using binomial errors, logit link and GDD and PANN as predictor variables (for more details see R script in Supplementary material Appendix 2). As with GLM, the functions fitted with CQO were used to project the species distributions under current and future (2050 and 2008) climates.

Model verification

In most studies investigating climate change impacts on species distributions, independent data for evaluation of the models are difficult to obtain (Araújo et al. 2005). When this is the case, alternative approaches are required for assessing models performance. One approach is to internally evaluate models (i.e. verify them), by measuring how well predictions fit the calibration data or, preferably, a subset of the calibration data withheld for evaluation of the models (Araújo and Guisan 2006). When data are split into calibration and evaluation, the measured accuracy of the models may be affected by how data were initially split. To account for such sensitivity in model outcomes, models were cross-validated 10 times. Splitting of the data was done randomly and the size of the random splits was determined by application of a commonly used heuristic for identifying the ratio of calibration and cross-validation sets in presence and absence models: $[1 + (p - 1)^{1/2}]^{-1}$, where p is the number of predictors (Fielding and Bell 1997). Since species prevalence is highly correlated with the cut-off value that minimizes the difference between sensitivity (the probability that the model will correctly classify a presence) and specificity (probability that the model will correctly classify an absence) (Liu et al. 2005, Jiménez-Valverde and Lobo 2006), we used the prevalence of each species in the calibration set as a cut-off for converting the projected probabilities into presence-absence scores (see also Araújo and Luoto 2007). Model sensitivity, specificity and Kappa statistic were calculated for all species and for each method (i.e. GLM and CQO, for more details see R script in Supplementary material Appendix 2) and for each one of the 10 cross-validation datasets using the PresenceAbsence package (Freeman 2007). Comparison of Kappa statistic values across species is problematic because this measure is affected by prevalence (Allouche et al. 2006). However, it is an appropriate method to test for

differences across modelling methods for the same species. The t statistic for dependent samples was used to assess for significant differences in sensitivity, specificity and Kappa between GLM and CQO, for each species. The t values were then regressed against range size (logarithm of the area in km^2) and mean latitude of each species to assess for species-specific trends in the performance of each modelling method.
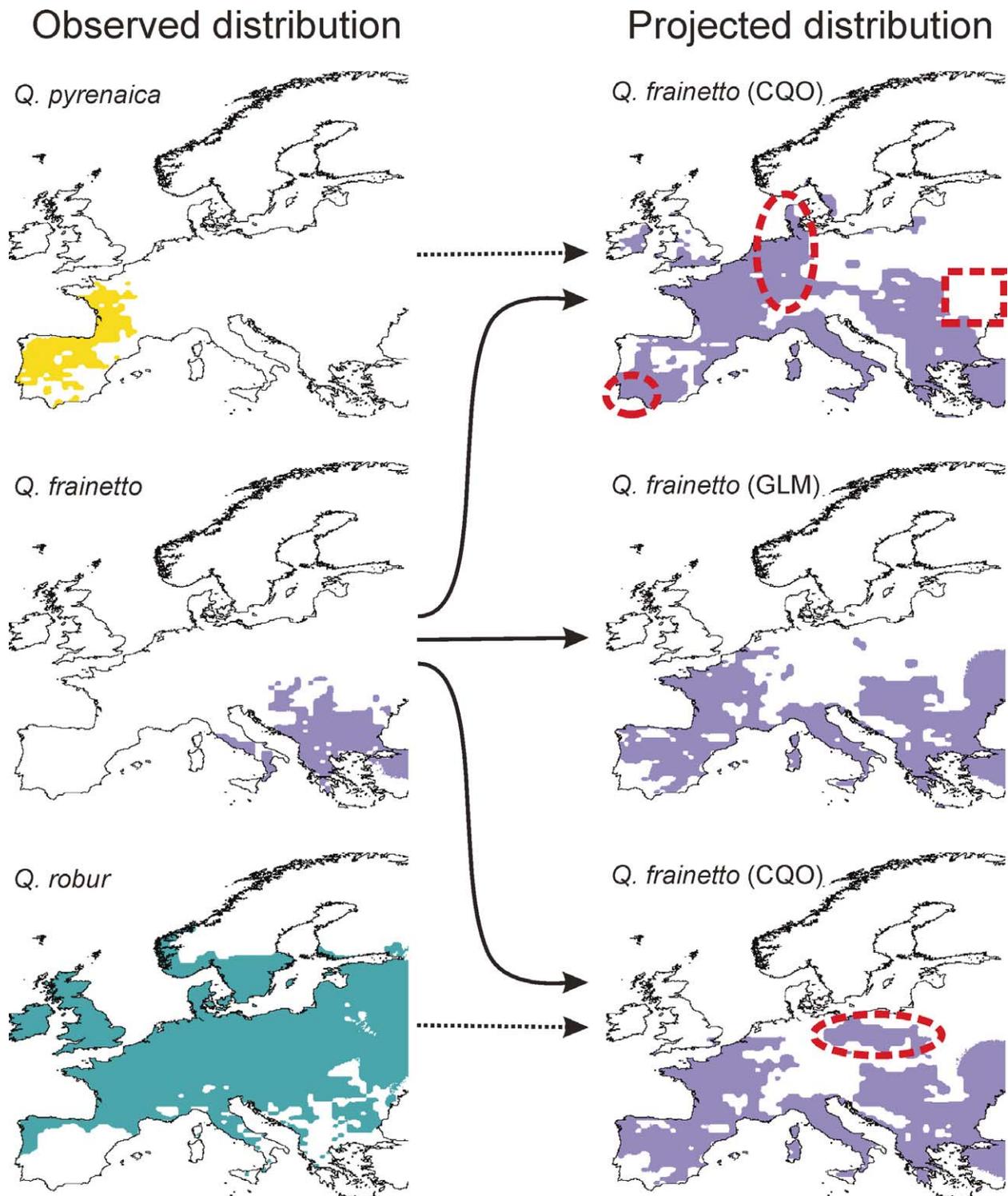

Figure 1. Projections of *Quercus frainetto* using individualistic (GLM) and community-based (CQO) response models. Red ovals indicate regions where CQO predicts the presence of *Q. frainetto* and GLM predicts its absence; red squares indicate regions where GLM predicts the presence of *Q. frainetto* and CQO predicts its absence. Two alternative projections with CQO are provided: one with *Q. pyrenaica* as covariate in the models and the other with *Q. robur* as a covariate.

Comparison between model projections

The GLM and CQO models fitted to the complete dataset were used to project the probability of occurrence of each species under present and future (2050 and 2080) climatic conditions (see above). The projected probabilities were converted into presence and absence using the prevalence of each species as the threshold for conversion. Differences between model projections were investigated at both the individual species level (comparing measures of model accuracy across methods using cross-validated samples) and the assemblage level. Assemblage variation can be examined both in terms of the numbers of species (species richness) and the identity of species (species composition) present in any location (Harrison et al. 1992, Baselga et al. 2007). Thus, we compared differences in modelled patterns of species richness and assemblage composition (dissimilarity) obtained with community and individualistic methods.

Species richness was computed for each method (S_{GLM} and S_{CQO}) as the sum of all presences projected in each cell. The difference between both values ($\Delta S = S_{CQO} - S_{GLM}$) was mapped and regressed against geographical coordinates (longitude and latitude) and environmental predictors (GDD and PANN) to assess for geographical and environmental trends in models. To examine differences in species composition, we computed the Simpson's index of dissimilarity (Koleff et al. 2003) between the assemblages projected with GLM and CQO in each cell. The Simpson's index of dissimilarity (β_{sim}) was preferred because it is independent of differences in richness between samples (Baselga 2007); ΔS and β_{sim} are two methodologically independent measures capturing complementary information on the differences between assemblages projected by GLMs and CQO.

Results

Model accuracy

The fitted models were moderately accurate, with sensitivity ranging from 0.67 to 1.00 with GLMs (mean $\pm$ SD: 0.86 ± 0.06) and 0.57 to 1.00 with CQO (mean $\pm$ SD: 0.83 ± 0.08). Specificity ranged from 0.53 to 0.96 in GLMs (mean $\pm$ SD: 0.76 ± 0.09) and 0.57 to 0.94 in CQO (mean $\pm$ SD: 0.75 ± 0.09). The tendency for GLM to have higher sensitivity values than CQO, and for CQO to have higher specificity than GLM was significant for the majority of species, though differences in performance between models were greater for sensitivity than for specificity ($p < 0.05$, Table 1). A tendency for improved performance of GLM over CQO was maintained when kappa values, that weight sensitivity and specificity equally, were used for benchmarking of the models (Table 1).

Differences in the relative performance of methods were significantly related to geographical properties of species ranges. Using t statistics accounting for differences between the two methods, we found that species' range size was negatively related with the t static accounting for differences in sensitivity between methods (Pearson $r = -0.37$, $p < 0.001$). In other words, sensitivity obtained with CQO models was higher among restricted range species than GLM, whereas GLM had higher sensitivity values than CQO among wide-ranging species. When the same test was used to examine differences in specificity and kappa between methods, we found a significant negative relationship between latitude of the centroid of species distributions and the t statistic accounting for differences in specificity and kappa ($r = -0.40$, $p < 0.001$ and $r = -0.43$, $p < 0.001$, respectively). This shows that specificity and kappa values were higher for CQO than for GLM in southerly distributed species, whereas the reverse pattern was recorded for northerly distributed species. All other assessed relationships between t statistics and geographical characteristics of species were not significant (i.e. $p > 0.05$).

Model projections

Current species richness (S), as projected with CQO, was significantly higher than projected richness with GLM
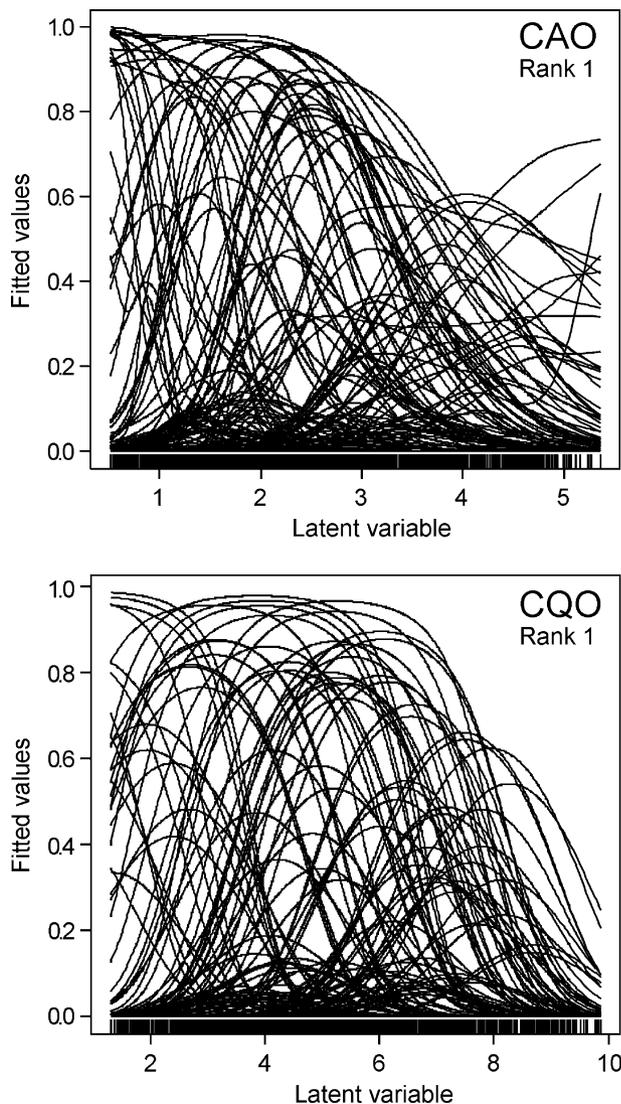

Figure 2. Response curves of 119 tree species against a latent climatic variable. Response curves were estimated with Rank-1 CAO (canonical additive ordination) and CQO (canonical quadratic ordination).

Table 1. Relative accuracy of GLM and CQO models of European tree species distributions, assessed through t tests for dependent samples after 10-fold cross validation. Scores are the number of species (n =119) for which accuracy is significantly higher for any of the methods considered (p <0.05), or for which there are not significant differences.

	GLM >CQO	GLM <CQO	n.s.
Sensitivity	71	16	32
Specificity	44	47	28
Kappa	60	15	44

(mean $\Delta S = 3.00$; SD $= 4.60$; t $= 112.91$, p <0.001), although this relationship was not constant across geographical and environmental space (Fig. 3). ΔS was positively and significantly related to latitude ($R^2 = 0.16$, p <0.001) and longitude ($R^2 = 0.12$, p <0.001), as well as to quadratic functions of GDD ($R^2 = 0.09$, p <0.001) and PANN ($R^2 = 0.20$, p <0.001). A complete environmental model composed by the second order polynomial for GDD and PANN accounted for near a half of the variation in ΔS ($R^2 = 0.46$, p <0.001). With projections for 2050, areas where GLM projected higher species richness than CQO were more frequent (Fig. 3), although species richness projected by CQO was still significantly higher than that projected by GLM (mean $\Delta S = 3.46$; SD $= 5.58$; t $= 106.74$, p <0.001). For this period, ΔS was also positively and significantly related to latitude ($R^2 = 0.29$, p <0.001) and longitude ($R^2 = 0.10$, p <0.001), as well as to quadratic functions of GDD ($R^2 = 0.32$, p <0.001) and PANN ($R^2 = 0.03$, p <0.001). A complete environmental model composed by the second order polynomial for GDD and PANN accounted for more than half of the variation in ΔS ($R^2 = 0.54$, p <0.001). For 2080, areas where GLM projected more species than CQO became even more widespread (Fig. 3) and the general difference between richness projected by CQO and GLM was still significant but lower than for previous periods (mean $\Delta S = 1.22$; SD $= 5.99$; t $= 35.07$, p <0.001). Latitudinal and environmental patterns in ΔS were more marked, with the GLM increasing the tendency for projecting higher richness in lower latitudes. Indeed, for 2080, projected ΔS was positively and significantly related to latitude ($R^2 = 0.46$, p <0.001) and longitude ($R^2 = 0.03$, p <0.001), negatively related to GDD ($R^2 = 0.51$, p <0.001) and positively related to PANN ($R^2 = 0.05$, p <0.001). A complete environmental model composed by GDD, PANN and their interaction accounted for more than half of the variation in ΔS ($R^2 = 0.54$, p <0.001).

Differences in projected species composition (β_{sim}) with GLM and CQO did not present a marked geographical structure (Fig. 4). Although β_{sim} was significantly related with latitude and longitude (probably due to the extremely large number of cases) the amount of explained variance was negligible ($R^2 = 0.01$, p <0.001 and $R^2 = 0.03$, p < 0.001, respectively). The same weak relationship was found with GDD ($R^2 = 0.001$, p <0.001) but, conversely, β_{sim} was strongly related to the quadratic function of PANN ($R^2 = 0.69$, p <0.001). A complete environmental model composed by GDD, the quadratic function of PANN, and their interaction accounted for almost three quarters of the variation in β_{sim} ($R^2 = 0.72$, p <0.001). Similar results were found for 2050 and 2080, for which all the tested relationships were negligible (all $R^2 < 0.05$, p <0.001) with exception of that of PANN ($R^2 = 0.61$, p <0.001 and $R^2 = 0.30$, p <0.001, respectively). A complete environmental model for 2050 composed by GDD, the second order polynomial of PANN and their interaction accounted for almost three quarters of the variation in β_{sim} ($R^2 = 0.72$, p <0.001). A complete environmental model for 2050 composed by the second order polynomial for GDD and PANN accounted for nearly two thirds of the variation in β_{sim} ($R^2 = 0.63$ p <0.001), whereas the model for 2080 accounted for a third of the variation in β_{sim} ($R^2 = 0.34$, p <0.001).

Discussion

Would community-based models improve forecasts of species distributional change under climate change? In our study, an approach that modelled individual species responses to climate provided more accurate projections than modelling assemblages. Such improved accuracy shows that GLM was able to fit the data better than CQO, but it does not necessarily clarify if GLM would be able to make more useful predictions of range changes under climate change than CQO (Araújo et al. 2005, Randin et al. 2006, Peterson et al. 2007). Answering this question would require independent validation data that were unavailable to us (Araújo and Rahbek 2006). However, our results also showed that models including community interactions differed markedly from models that do not include them and that differences are spatially structured. This was true for the three measures of model accuracy used and, more importantly, for projected species richness and assemblage composition in the present and in the future. Generally, it was found that GLM provided more accurate projections than CQO and that differences in accuracy between the two methods were associated with species-specific traits, such as geographical location and range size. When individual species projections were combined, in order to estimate the potential species richness and assemblage composition of areas, the obtained patterns varied. For example, overall, CQO tended to predict higher species richness than GLM. This increase of species richness with community-based models invites the interpretation that the identification of patterns of co-existence among species might lead to an increase in estimated realized niche of species. In contrast to species richness, projected assemblage composition was similar between the two methods, with the exception of areas with cold and wet climates in the northern-west of Europe where the two approaches projected very different assemblages.

Accuracy of individual vs community-based models

Few studies have compared individualistic vs community models. Those that did, provided mixed results (Elith et al. 2006, Leathwick et al. 2006, Chatfield 2008). In our study, individualistic GLM models had significantly higher
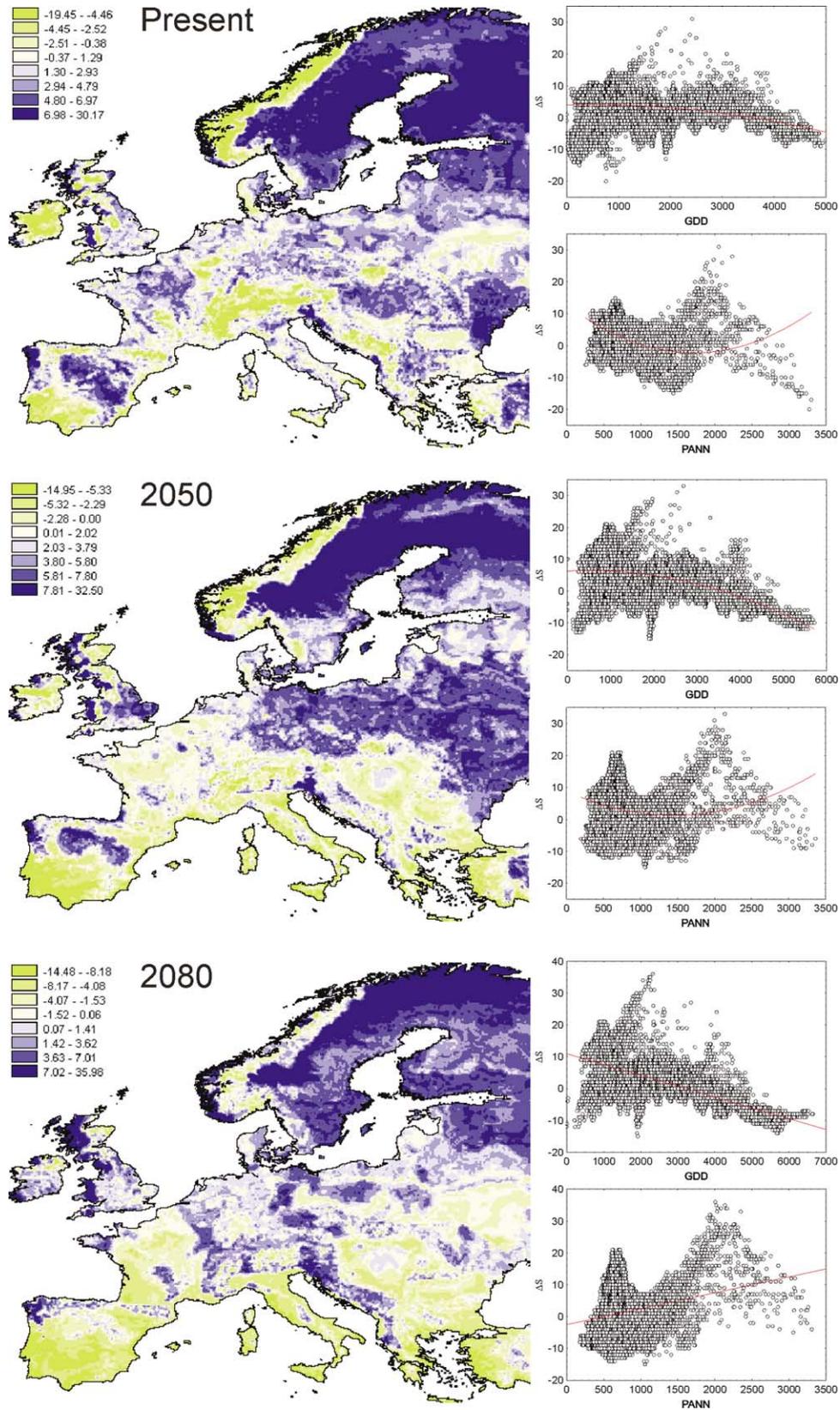
Figure 3. Differences in species richness (ΔS) as projected by CQO and GLM for the present and for the future (2050 and 2080), and environmental correlates of these differences. Negative values (yellow) on the map correspond to cells where GLM project higher species richness than CQO, whereas positive values (blue) correspond to cells where CQO projects higher species richness than GLM.

accuracy (with kappa) than CQO for about half of the species. The improved performance of GLM was related with its superiority in predicting presences (thus increasing sensitivity). In contrast, both methods projected absences accurately (high specificity) for a similar number of species. Our analysis also revealed that sensitivity among restricted range species was relatively higher with CQO than with GLM. Analogous results were found by Leathwick et al. (2006) for New Zealand freshwater fishes. In their study, community-based modelling with MARS provided more accurate projections (with AUC) than individualistic models for the least prevalent species. Conversely, a study of demersal fish species in Western Australia (Chatfield 2008), found the opposite pattern, i.e. of relatively lower AUC and sensitivity values among restricted range species modelled with community-based approaches.

When investigating the geographical distribution of accuracy values we also found that specificity values obtained with CQO were higher than GLM among low-latitude species. In other words, community-based models were better at predicting absences in the south of Europe, where many restricted range species occur. In contrast, CQO was less effective than GLM in predicting absences (lower specificity) among widespread species in central and northern Europe. Spatial patterns of overall model accuracy followed the pattern of specificity; in other words, kappa values were higher with CQO among southern species and GLM provided more accurate projections in central and northern Europe.

Projected distribution patterns with individual and community-based models

Community-based models (CQO) consistently projected higher local species richness than individual (GLM) models. In other words, by weighting patterns of co-existence among species, CQO effectively emulates positive relationships among species, "forcing" coexistence in projections of species niches, and thus leading to higher local species richness. No clear relationships between simulated increases of positive interactions (i.e. positive differences in richness between CQO and GLM, Fig. 3) and environmental gradients were found. However, both empirical evidence (Callaway et al. 2002) and simulation experiments (Travis et al. 2005) have demonstrated a tendency for preponderance of positive interactions in extreme environmental gradients. If this was true at biogeographical scales, increased species richness would be expected to occur in the extreme temperature and precipitation gradients. However, this expectation was not supported by our analysis. Even though there was slight tendency for increased species richness in the higher (colder) latitudes, higher species richness with CQO was found in intermediate levels of GDD (proxy for energy) and annual precipitation (proxy for water availability) (Fig. 3).

In contrast with species richness, where marked geographical differences between models were recorded, there was high similarity between assemblages projected with individualistic and community-based models. Since we measured assemblage similarity with an index independent of richness, this means that, for a given cell, the assemblage

projected by one of the methods is almost a subset of the assemblage projected by the alternate method. In other words, differences in species richness between individualistic and community-based models are the result of the inclusion of additional species by any of the two methods, but the projected assemblages projected in a given cell rarely include sets of mutually exclusive species.

Lessons for the future – are community-based models more useful for climate change studies than individual-based models?

As shown by several authors, model accuracy, as measured with non-independent evaluation procedures, does not necessarily equate to predictive accuracy in an independent setting (Araújo et al. 2005, Randin et al. 2006, Peterson et al. 2007). Thus one needs to be cautious when extrapolating conclusions about robustness of models with non-independent evaluations (Araújo and Rahbek 2006). For example, it would be unwarranted to conclude that individualistic GLM is superior to community-based implementations of GLM (i.e. CQO), simply because GLM tends to fit the data better than CQO. It is the ability to predict novel situations (model generality), not the ability to predict the data used for calibration of the models (model accuracy), or non-independent evaluation data, that matters in studies examining the potential impacts of climate change on species distributions. Nevertheless, the mismatch with projections obtained with community and individualistic models does lend support to the view that patterns of species co-occurrence and co-exclusion can affect the predictive power of species distribution models under climate change (Araújo and Luoto 2007, Heikkinen et al. 2007). The critical question is whether differences accrued from using community models lead to unrealistic projections or whether they add useful information for improving projections of species distributional change under climate change?

These are difficult questions that our analyses cannot fully clarify. Notwithstanding, let us consider the following. Firstly, CQO (as well as other community models) deals with statistical patterns of co-occurrence and co-exclusion among modelled species, but 1) the biological significance of these interactions is not explored; and 2) biotic interactions with important interacting taxa (e.g. pollinators, seeds dispersers, and other plant eating animals) are usually overlooked because studies are commonly performed with similar taxa. Secondly and more importantly, interactions in community-based models are explored in the multidimensional space composed of environmental predictors. In other words, spatial co-occurrence and co-exclusion patterns are not explicitly taken into account. This is a serious shortcoming of an approach that is supposedly modelling communities, because species interactions occur primarily in geographical rather than in environmental space. One consequence of this feature is that spatial patterns derived from vicariant speciation, competition, or other spatial representations of biotic interactions are incorporated in community models only indirectly or partially. For example, although distributions of *Quercus frainetto* and *Q. pyrenaica* are spatially
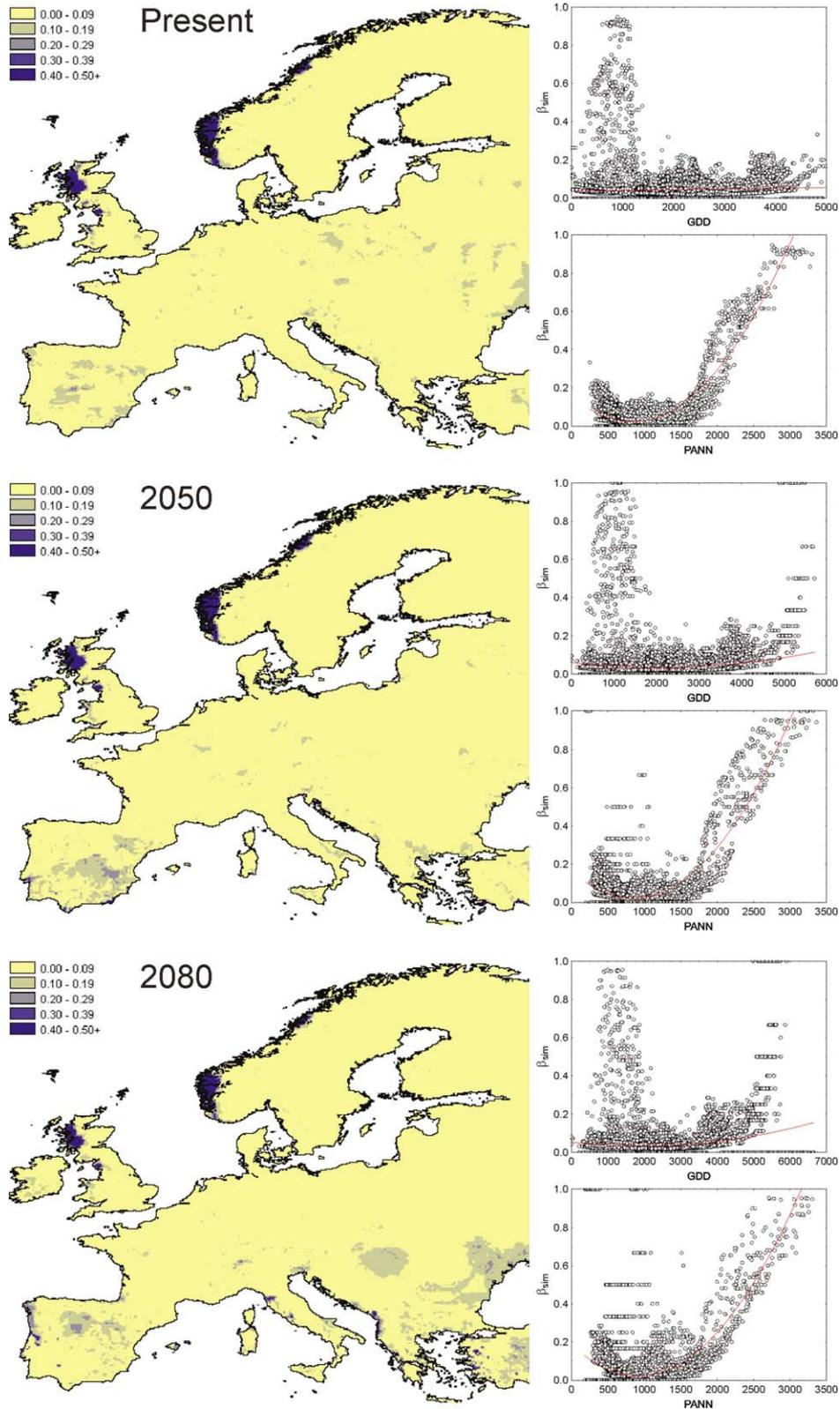
Figure 4. Differences in species composition (β_{sim}) projected by CQO and GLM for the present and for the future (2050 and 2080), and environmental correlates of these differences. Zero values (yellow) correspond to cells where GLM and CQO project identical communities (measured as Simpson dissimilarity index), whereas positive values (shades of blue) represent cells where both models project different community composition.

disjunct in Europe (Fig. 1), the CQO model would not consider this distribution to represent a pattern of co-exclusion because the two species co-occur in similar environments (standard GLM would be no different). To put it differently, for CQO (and presumably for other existing community-based model) to represent species interactions in geographical space, assumptions of equilibrium of species distributions with climate (i.e. species occur in areas with suitable environments while being absent from areas with unsuitable ones) need to be verified. This is of course a problem, shared with all models that derive species-environment association from an analysis of the statistical relationships between variables (as shown by, Araújo and Pearson 2005, Midgley et al. 2006, Morin et al. 2008, Svenning et al. 2008).

The benefits of community-based models have been described as including "faster processing of species distributions data, increased power to detect shared patterns of environmental response across rarely recorded species, and enhanced capacity to synthesise complex data into a form more readily interpretable by scientists and decision makers" (Ferrier and Guisan 2006). However, the properties of these models are not yet fully understood. Their ability to make useful projections into independent settings, i.e. different times, regions, or resolutions, has not been assessed; and differences between alternative community modelling techniques have not been thoroughly tested. Our results show that overall accuracy of models can be reduced with community-based models, and that projections under climate change can vary markedly between individual species response models and models that are expected to account for community interactions. Results are, of course, contingent on the studied taxa, types of interactions considered (e.g. ignoring pollinators, herbivores, seed dispersers, etc.), and the historical contingencies of Europe, but evidence that community-based models improve the usefulness of predictions of species range shift under climate change with regards to individualistic response models is still lacking. We suggest that improving current projections of future species responses to climate requires modellers to include more ecology into the models (Austin 2002, Thuiller et al. 2003), to develop new tools for integration of biotic interactions into models (Dos Santos et al. 2008), and to measure uncertainties from models more explicitly (for review see Heikkinen et al. 2006, Araújo and New 2007). Promising research is now being undertaken aiming at increasing the realism of individual species distributions models by explicitly accounting for population dynamics (Keith et al. 2008) and complex physiological feedbacks (Rickebusch et al. 2008). The next step will be require an integration of complex community dynamics into models that already account for demographic processes as well as physiological feedbacks.

References

Allouche, O. et al. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). – J. Appl. Ecol. 43: 1223–1232.

Araújo, M. B. and Williams, P. H. 2000. Selecting areas for species persistence using occurrence data. – Biol. Conserv. 96: 331–345.

Araújo, M. B. and Pearson, R. G. 2005. Equilibrium of species' distributions with climate. – Ecography 28: 693–695.

Araújo, M. B. and Guisan, A. 2006. Five (or so) challenges for species distribution modelling. – J. Biogeogr. 33: 1677–1688.

Araújo, M. B. and Rahbek, C. 2006. How does climate change affect biodiversity? – Science 313: 1396–1397.

Araújo, M. B. and Luoto, M. 2007. The importance of biotic interactions for modelling species distributions under climate change. – Global Ecol. Biogeogr. 16: 743–753.

Araújo, M. B. and New, M. 2007. Ensemble forecasting of species distributions. – Trends Ecol. Evol. 22: 42–47.

Araújo, M. B. et al. 2005. Validation of species-climate impact models under climate change. – Global Change Biol. 11: 1504–1513.

Austin, M. P. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. – Ecol. Model. 157: 101–118.

Baselga, A. 2007. Disentangling distance decay of similarity from richness gradients: response to Soininen et al. 2007. – Ecography 30: 838–841.

Baselga, A. et al. 2007. A multiple-site similarity measure independent of richness. – Biol. Lett. 3: 642–645.

Callaway, R. M. 2007. Positive interactions and interdependence in plant communities. – Springer.

Callaway, R. M. et al. 2002. Positive interactions among alpine plants increase with stress. – Nature 417: 844–848.

Chatfield, B. S. 2008. How to find the one that got away. Predicting the distribution of temperate demersal fish from environmental variables. – Ph. D. thesis, School of Earth and Geographical Sciences, Univ. of Western Australia.

Dos Santos, D. A. et al. 2008. Sympatry inference and network analysis in biogeography. – Syst. Biol. 57: 432–448.

Elith, J. et al. 2006. Novel methods improve prediction of species' distributions from occurrence data. – Ecography 29: 129–151.

Ferrier, S. and Guisan, A. 2006. Spatial modelling of biodiversity at the community level. – J. Appl. Ecol. 43: 393–404.

Ferrier, S. et al. 2002. Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. II. Community-level modelling. – Biodivers. Conserv. 11: 2309–2338.

Ferrier, S. et al. 2007. Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. – Divers. Distrib. 13: 252–264.

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

Freeman, E. 2007. PresenceAbsence: an R Package for presence-absence model evaluation. – <cran.r-project.org/>.

Graham, R. W. and Grimm, E. C. 1990. Effects of global climate change on the patterns of terrestrial biological communities. – Trends Ecol. Evol. 5: 289–292.

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

Harrison, S. et al. 1992. Beta-diversity on geographic gradients in britain. – J. Anim. Ecol. 61: 151–158.

Heikkinen, R. K. et al. 2006. Methods and uncertainties in bioclimatic envelope modelling under climate change. – Prog. Phys. Geogr. 30: 751–777.

Heikkinen, R. K. et al. 2007. Biotic interactions improve prediction of boreal bird distributions at macro-scales. – Global Ecol. Biogeogr. 16: 754–763.

Humphries, C. H. et al. 1999. Plant diversity in Europe: Atlas Florae Europaea and Worldmap. – Acta Bot. Fenn. 162: 11–21.

Jalas, J. and Suominen, J. (eds) 1972–1996. Atlas Florae Europaeae. – The Committee for Mapping the Flora of Europe and Societas Biologica Fennica Vanamo.

Jiménez-Valverde, A. and Lobo, J. M. 2006. The ghost of unbalanced species distribution data in geographical model predictions. – Divers. Distrib. 12: 521–524.

Jordano, P. 2000. Fruits and frugivory. – In: Fenner, M. (ed.), Seeds: the ecology of regeneration in natural plant communities. Commonw. Agric. Bur. Int., pp. 125–166.

Keith, D. A. et al. 2008. Predicting extinction risks under climate change: coupling stochastic population models with dynamic bioclimatic habitat models. – Biol. Lett. 4: 560–563.

Koh, L. P. et al. 2004. Species coextinctions and the biodiversity crisis. – Science 305: 1632–1634.

Koleff, P. et al. 2003. Measuring beta diversity for presence-absence data. – J. Anim. Ecol. 72: 367–382.

Labandeira, C. 2002. The history of associations between plants and animals. – In: Herrera, C. and Pellmyr, O. (eds), Plant–animal interactions. An evolutionary approach. Blackwell, pp. 26–74.

Labandeira, C. C. et al. 2002. Impact of the terminal Cretaceous event on plant–insect associations. – Proc. Nat. Acad. Sci. USA 99: 2061–2066.

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

Leathwick, J. R. et al. 2005. Using multivariate adaptive regression splines to predict the distributions of New Zealand's freshwater diadromous fish. – Freshwater Biol. 50: 2034–2052.

Leathwick, J. R. et al. 2006. Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions. – Ecol. Model. 199: 188–196.

Liu, C. R. et al. 2005. Selecting thresholds of occurrence in the prediction of species distributions. – Ecography 28: 385–393.

Lyons, S. K. 2003. A quantitative assessment of the range shifts of Pleistocene mammals. – J. Mammal. 84: 385–402.

McGeoch, M. A. et al. 2006. Species and community responses to short-term climate manipulation: microarthropods in the sub-Antarctic. – Austral Ecol. 31: 719–731.

Meynard, C. N. and Quinn, J. F. 2007. Predicting species distributions: a critical comparison of the most common statistical models using artificial species. – J. Biogeogr. 34: 1455–1469.

Midgley, G. F. et al. 2006. Migration rate limitations on climate change-induced range shifts in Cape Proteaceae. – Divers. Distrib. 12: 555–562.

Moisen, G. G. and Frescino, T. S. 2002. Comparing five modelling techniques for predicting forest characteristics. – Ecol. Model. 157: 209–225.

Morin, X. et al. 2008. Tree species range shifts at a continental scale: new predictive insights from a process-based model. – J. Ecol. 96: 784–794.

New, M. et al. 2002. A high-resolution data set of surface climate over global land areas. – Clim. Res. 21: 1–25.

Peterson, A. T. et al. 2007. Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. – Ecography 30: 550–560.

Phillips, S. J. 2008. Transferability, sample selection bias and background data in presence-only modelling: a response to Peterson et al. (2007). – Ecography 31: 272–278.

Randin, C. F. et al. 2006. Are niche-based species distribution models transferable in space? – J. Biogeogr. 33: 1689–1703.

Rickebusch, S. et al. 2008. Incorporating the effects of changes in vegetation functioning and CO_2 on water availability in plant habitat models. – Biol. Lett. 4: 556–559.

Schröter, D. et al. 2005. Ecosystem service supply and vulnerability to global change in Europe. – Science 310: 1333–1337.

Segurado, P. and Araújo, M. B. 2004. An evaluation of methods for modelling species distributions. – J. Biogeogr. 31: 1555–1568.

Simakova, A. N. 2006. The vegetation of the Russian Plain during the second part of the Late Pleistocene (33–18 ka). – Quat. Int. 149: 110–114.

Stockwell, D. R. B. and Peterson, A. T. 2002. Effects of sample size on accuracy of species distribution models. – Ecol. Model. 148: 1–13.

Svenning, J. C. et al. 2008. Postglacial dispersal limitation of widespread forest plant species in nemoral Europe. – Ecography 31: 316–326.

Thuiller, W. et al. 2003. Generalized models vs classification tree analysis: predicting spatial distributions of plant species at different scales. – J. Veg. Sci. 14: 669–680.

Travis, J. M. J. et al. 2005. The interplay of positive and negative species interactions across an environmental gradient: insights from an individual-based simulation model. – Biol. Lett. 1: 5–8.

Tsoar, A. et al. 2007. A comparative evaluation of presence-only methods for modelling species distribution. – Divers. Distrib. 13: 397–405.

Williams, J. W. et al. 2001. Dissimilarity analyses of late-Quaternary vegetation and climate in eastern North America. – Ecology 82: 3346–3362.

Williams, P. H. et al. 2000. Endemism and important areas for conserving European biodiversity: a preliminary exploration of Atlas data for plants and terrestrial vertebrates. – Belgian J. Entomol. 2: 21–46.

Yee, T. W. 2004. A new technique for maximum-likelihood canonical Gaussian ordination. – Ecol. Monogr. 74: 685–701.

Yee, T. W. 2006. Constrained additive ordination. – Ecology 87: 203–213.