Improving species distribution models using biotic interactions: a case study of parasites, pollinators and plants

Tereza Cristina Giannini, Daniel S. Chapman, Antonio Mauro Saraiva, Isabel Alves-dos-Santos and Jacobus C. Biesmeijer

T. C. Giannini (giannini@usp.br), Bioscience Inst., Univ. of São Paulo, R. do Matão, 321, BR-05508-900, São Paulo, Brazil. – D. S. Chapman, Centre for Ecology and Hydrology, Bush Estate, Edinburgh EH26 0QB, UK. – A. M. Saraiva, School of Engineering, Univ. of São Paulo, Av. Prof. Luciano Gualberto, 380, BR-05508-970, São Paulo, Brazil. – I. Alves-dos-Santos, Univ. of Sao Paulo, R. do Matão, 321, BR-05508-900, São Paulo, Brazil. – J. C. Biesmeijer, Naturalis Biodiversity Center, PO Box 9517, NL-2300 RA Leiden, Netherlands.

Biotic interactions have been considered as an important factor to be included in species distribution modelling, but little is known about how different types of interaction or different strategies for modelling affect model performance. This study compares different methods for including interspecific interactions in distribution models for bees, their brood parasites, and the plants they pollinate. Host–parasite interactions among bumble bees (genus *Bombus*: generalist pollinators and brood parasites) and specialist plant–pollinator interactions between *Centris* bees and *Krameria* flowers were used as case studies. We used 7 different modelling algorithms available in the BIOMOD R package. For *Bombus*, the inclusion of interacting species distributions generally increased model predictive accuracy. The improvement was better when the interacting species was included with its raw distribution rather than with its modeled suitability. However, incorporating the distributions of non-interacting species sometimes resulted in similarly increased model accuracy despite their being no significance of any interaction for the distribution. For the *Centris-Krameria* system the best strategy for modelling biotic interactions was to include the interacting species model-predicted values. However, the results were less consistent than those for *Bombus* species, and most models including biotic interactions showed no significant improvement over abiotic models. Our results are consistent with previous studies showing that biotic interactions can be important in structuring species distributions at regional scales. However, correlations between species distributions are not necessarily indicative of interactions. Therefore, choosing the correct biotic information, based on biological and ecological knowledge, is critical to improve the accuracy of species distribution models and forecast distribution change.

Species distribution modelling has been used as a computational tool with multiple objectives, including predicting the impacts of global change (Franklin 2009). This technique combines occurrence data with environmental variables based on a correlative approach to build a representation of a species' ecological requirements (Anderson et al. 2003). Several algorithms have been applied to create such models, and results show areas that are similar to those where the species is known to occur thus representing potential areas of occurrence (Stockwell and Peters 1999, Phillips et al. 2006). Future scenarios can be projected assuming that current climatic requirements of a species remain unchanged under climate change (Thuiller et al. 2005).

Recently, the importance of considering biotic information when analyzing species distribution models (SDMs) has been advocated as a way of representing effects of interspecific interactions (e.g. competition, predation, parasitism, mutualism) on species distributions (Guisan and Thuiller 2005, Soberón and Peterson 2005) and when analyzing impacts of climatic change (Brooker et al. 2007, Hegland et al. 2009, Van der Putten et al. 2010).

In previous studies, some interacting species were modelled separately and their distributions were compared, emphasizing the role of interactions in shaping the geographical ranges. For example, two competitor species of pocket mice (*Heteromys australis* and *H. anomalus*) where used to illustrate the concept of competitive exclusion (Anderson et al. 2002). Considering specialised interactions and the use of SDM, Godsoe et al. (2009) and Warren et al. (2010) discussed the evolutionary process that shapes the geographical distribution of *Ficus* tree species and their obligatory wasp pollinators, and Giannini et al. (2011) analyzed the case of *Peponapis* bees and the *Cucurbita* plant species that they pollinate.

In other studies, the biotic information was explicitly included in the modelling process. The inclusion of biotic predictors generally improved the models fit (Heikkinen et al. 2007, Meier et al. 2010, Pellissier et al. 2010) and altered the projections of its future distribution (Araújo and Luoto 2007, Hof et al. 2012).

Once the biotic predictors have been identified and selected, they can be incorporated in the models in different

ways. Firstly, the occurrence records of the interacting species can be added as a predictor variable for the focal species. This might be particularly useful when the biotic component has an obligate or a highly specialized relationship with the focal species. Secondly, the estimated occurrence or habitat suitability of the interacting species (i.e. its SDM prediction) can be used as input layer in the focal species models. This might be more useful if the distribution of the interacting species is incompletely known. To our knowledge, a comparative analysis considering both options has not been performed.

The potential value of biotic information to species distribution models should depend on the relationship between the focal species and the biotic component. The occurrence of an obligate or highly specialized parasite, for example, can be taken as good evidence for the presence of its host. While the presence of a single food plant species of a generalist forager merely points to the potential occurrence of that forager species. It is even possible that information on a non-interacting species can improve the distribution model if both species happen to be associated with similar environmental features (e.g. soil, climate, micro habitat) or share a similar spatial structure (Chapman 2010). As such, it is not clear whether the inclusion of interacting species as predictors improves the model because it represents mechanisms with predictive power or because of chance.

In this paper, we examined the extent to which the modelling methods, trophic position (here plant/pollinator/brood parasite), and level of ecological dependence change the SDM projections of the focal species. We chose two different groups of interacting organisms: one group of host–parasite pairs and one group of specialized plant–pollinator species. The host–parasite group was comprised of species of bumblebees and their congeneric parasites (both *Bombus* spp.) that are common throughout Europe and Asia and tend to have a broad floral diet (Benton 2006). Brood parasite *Bombus* species enter the nests of host bees, kill the host larva and lay their own eggs in the provisioned cells (Michener 2000). The specialized plant–pollinator group was comprised of species of *Krameria* plants (Krameriaceae) and *Centris* bee pollinators – two genera with a Neotropical distribution. *Krameria* plants are pollinated only by *Centris* bees that collect floral oil to raise their offspring (Vogel 1974, Simpson 1989).

The main objectives of this work were 1) to compare the fit of species distribution models with biotic predictors included in different ways or without them; 2) to compare the effect of including biotic predictors pertaining to different trophic positions or ecological dependence; 3) to analyze the extent to which projections under climate change depend on modelling methods and the strategy used to include biotic interactions in the models.

Material and methods

Details of the data sources used in the models can be found in the Supplementary material Appendix 1. The selected species of *Bombus*, representing host–parasite pairings, were: 1) *B. terrestris* (host) and *B. vestalis* (parasite), and 2) *B. lucorum*, *B. cryptarum* and *B. magnus* (all hosts) and

B. bohemicus (parasite of all three hosts) (Loken 1984). The species of the *Bombus lucorum* host complex (*B. cryptarum*, *B. magnus* and *B. lucorum*) are very similar and not easily identifiable which can result in misidentification. In addition, the knowledge of host–parasite relationships in this genus is still incomplete. In particular, it is unclear how commonly *B. bohemicus* parasitizes *B. cryptarum* and *B. magnus* (P. Rasmont pers. comm.). Therefore, although we modelled different linkages between these three hosts and the parasite *B. bohemicus*, we acknowledge that this is based on current consensus and may (slightly) change in the future. Even though these species occur across large parts of Europe and Asia, our analyses were focused on Great Britain for which good distribution records were available. We selected occurrences reported since 1990 and considered the cells with no reported occurrence as true absence points.

The selected species of bees and plants representing pollinator–plant interactions were: 1) *Krameria erecta* (plant); *Centris atripes* and *C. rhodopus* (pollinating bees) all occurring in western areas of Mexico and USA, and 2) *K. tomentosa* (plant) and *C. fuscata* and *C. tarsata* (pollinating bees), occurring in northeastern and central areas of Brazil. We did not filter out older records, as was the case for *Bombus*, because of more limited data availability. As such, almost 30% of records were from before 1990. Occurrences were projected onto a 10×10 km grid with each cell classified as a presence (at least one record of the focal species) or pseudo-absence (no record) (Godsoe et al. 2009).

The BIOMOD package (Thuiller et al. 2009b) for R 2.11.1 (The R Foundation for Statistical Computing) was applied to model present day and future species distributions. We created pseudo-absence points for each Neotropical species using the BIOMOD 'circles' strategy. With this strategy, a circle is drawn individually around each presence point with a radius of 5° and pseudo-absence points randomly selected from this region (Thuiller et al. 2009b). This ensures that the model is fitted using data from a region where the species are likely to be able to reach by dispersal. The total number of pseudo-absences per species was set so that the number of occurrences and pseudo-absences summed to 10 000 (Van der Wal et al. 2009, Lobo and Tognelli 2011). The original databases were randomly split in calibration (80%) and evaluation subsets (20%) (Fielding and Bell 1997).

Using BIOMOD, we applied 7 model classes: artificial neural networks (ANN) (Ripley 1996), classification tree analysis (CTA) (Breiman et al. 1984), generalised linear models (GLM) (McCullagh and Nelder 1989), generalised boosted models (GBM) (Ridgeway 1999), flexible discriminant analysis (FDA) (Hastie et al. 1994), multivariate adaptive regression splines (MARS) (Friedman 1991) and random forest (RF) (Breiman 2001). To evaluate the models, we used the cross-validated area under the receiver operating characteristic curve (AUC), kappa statistic, true skill statistic (TSS), sensitivity (proportion of presences correctly predicted), and specificity (proportion of absences correctly predicted) for the 20% evaluation subset (Thuiller et al. 2009b). Other than AUC, these require conversion of model predictions into predicted presence/absence data.

To do this we used threshold classification, with thresholds selected to maximise the sum of sensitivity and specificity (Thuiller et al. 2009a). We also evaluated variable importance as the correlation scores between the original prediction and the prediction made with a permuted variable, normalized on the scale 0–100% (Thuiller et al. 2009a).

We used 19 bioclimatic variables with a grid resolution of 5 arc-minutes to model the species' geographical distributions. These represent gradients in average temperature and precipitation, over diurnal, seasonal and annual timescales (Hijmans et al. 2005). In modelling the present day species distributions, we also used land use information for the year 2000 (Global Land Cover 2000) that was converted to 10 km resolution. Because many of the environmental variables were highly correlated we applied principal component analysis (PCA) using R 2.11.1 (The R Foundation for Statistical Computing) (Powney et al. 2010). We identified six independent axes of variables representing 80% of the total environmental variation in Britain, 97% in North America and 92% in South America.

We added the information about biotic interactions as a supplementary layer to the six PCA variables following different strategies. First, we performed modelling with only the abiotic variables to obtain the model-predicted value of occurrence for each species. Second, we created new models including the biotic information in two different ways: A) we included the actual presence-absence of the biotic component as a supplementary explanatory variable (predictor) to modelling the focal species; B) we included the model-predicted value of occurrence for interacting species from abiotic SDMs as a supplementary predictor. Where the modeled species had more than one interacting species (e.g. B. bohemicus with its hosts and Krameria with its pollinators), we fitted individual models for each interacting species, models in which all interacting species were included individually, and models based on the richness of the interacting species. Where models were based on occurrence values, a richness index was derived as the sum of values for the interacting species. We restricted the models of Bombus to Great Britain and the models of Krameria and Centris were restricted to areas where both species of Krameria are known to occur (Simpson 1989), using the strategy of pseudo-absences described above.

In addition to modelling the effects of interacting species, we repeated the modelling exercise specifying non-relevant interactions in the SDMs, i.e. we included interactions between species that do not strongly interact via specialist host–parasite or plant–pollinator relationships. With such strategy, we aimed to test whether the AUC variation was due to the inclusion of biotic interactions or simply due to increased model complexity or statistical artifacts of spatially structured predictors. It was not possible to apply this procedure to Krameria species because the different interacting species groups occupied distinct geographic regions. Thus, we considered non-relevant interactions between parasitic Bombus species and hosts they did not use, and interactions between regionally co-occurring Centris bees. Differences in the resulting model evaluation statistics of the abiotic and different biotic models were tested using paired t tests.

We also analyzed the effect of including biotic data on ensemble model projections with current and future climate conditions (Araújo and New 2007, Thuiller et al. 2009b). For future projections, downscaled bioclimatic variables were obtained for 2050, assuming a moderate climatic change scenario (Special Report on Emission Scenarios B2) (Ramirez and Jarvis 2008) and the HADCM3 model (Gordon et al. 2000). Future land use scenarios were not available to us, and so for this exercise we developed present day models based on just the climate and interspecific interactions. For brevity, we only forecasted the distribution of the species for which biotic interactions had the largest effect on the SDMs for present day conditions (B. bohemicus). With BIOMOD ensemble forecasting (Araújo and New 2007, Thuiller et al. 2009b), the final model is a consensual result of all models used (in our case, the same seven models described before), weighted on the basis of their predictive accuracy on the test data. Using this function, different measures of accuracy were available and we have chosen AUC to calculate the weights.

Results

For Bombus, the best method for including biotic interactions was to use the interacting species' actual distribution as a predictor variable, as shown by significantly higher AUC, TSS and kappa (Table 1). However, for the Centris-Krameria system the best method was to include model-predicted occurrence values, as shown by significantly higher TSS, and non-significantly higher values of all the other accuracy measures (Table 1). In both cases the biotic component of the model was the single most important predictor variable, though this was more pronounced for Bombus than Krameria-Centris (Fig. 1).

Including relevant interactions nearly always increased predictive accuracy over the abiotic models for Bombus (Table 2). More specifically, including information on the parasite species had stronger effects on model performance

Table 1. Mean values of seven cross-validated model accuracy measures for biotic models that use the actual distribution of interacting species as predictors (presence/absence) or use model-predicted values for the interacting species as predictors. P-values are from paired t-tests between the results for both modelling strategy. Only relevant interactions were used.

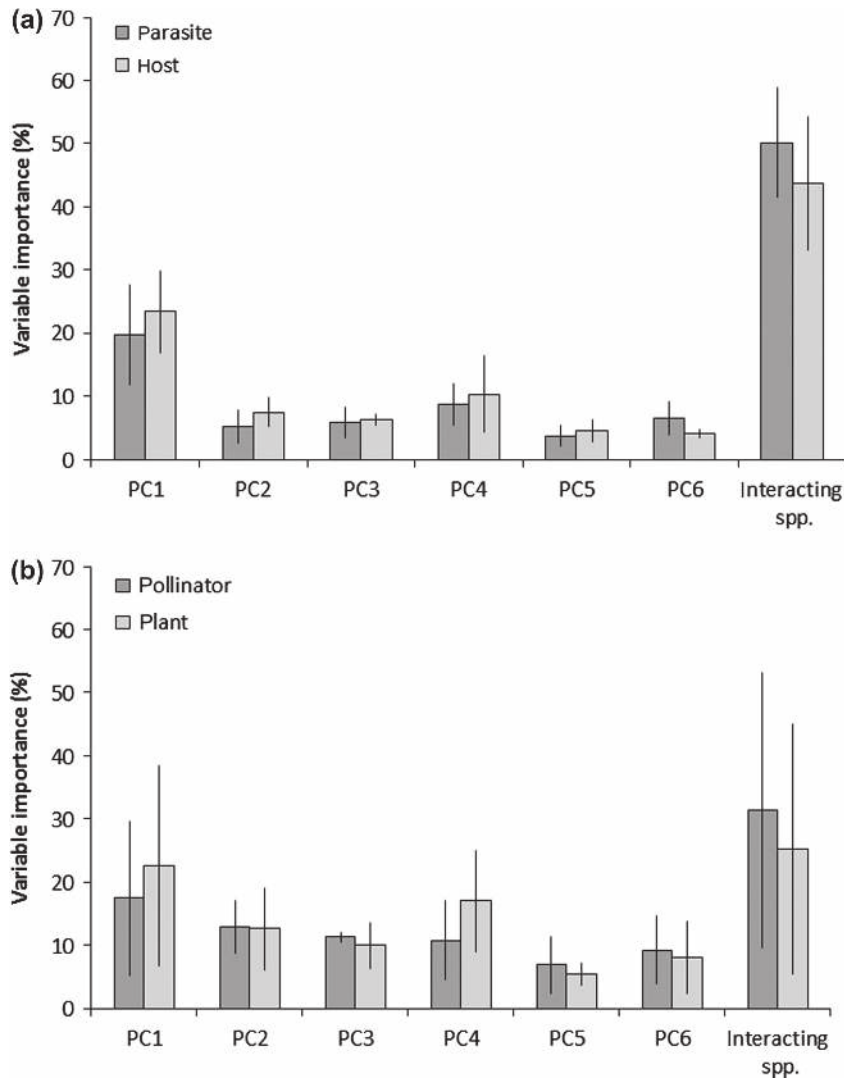| Species	Method of inclusion	AUC	TSS	Kappa	Sensitivity	Specificity
Bombus species	presence/absence	0.786	0.352	0.489	64.056	87.821
	predicted	0.761	0.311	0.442	62.867	86.721
	p	0.003	0.002	0.008	0.484	0.117
Centris and *Krameria*	presence/absence	0.884	0.495	0.667	57.504	98.888
	predicted	0.903	0.563	0.710	64.474	99.002
	p	0.087	<0.001	0.174	0.118	0.496

Figure 1. Mean variable importance (with standard deviation) of each principal component (abiotic features) and interacting species (biotic feature) in models for (a) *Bombus* species and (b) *Krameria-Centris*. Variable importance was evaluated as the mean correlation score between model predictions and those made with permuted variable, normalized on the scale 0–100%.

than including the host. For the parasite with multiple hosts (*B. bohemicus*), there was little difference between models that used all three hosts individually, used host richness, or only used the most widespread host (*B. lucorum*). However, inclusion of the rarer hosts resulted in slight model improvement. When the *Bombus* models featured non-relevant interactions, unexpected significant increases in model accuracy were detected in four out of eight cases, while only two cases gave significant decreases in model accuracy (Table 2).

By contrast, the inclusion of biotic interactions in models for the Neotropical plant and pollinator species resulted in almost no significant variation in model accuracy measures (Table 2). Significant improvement in at least one model accuracy measure was observed in only three out of 12 cases when including relevant interactions, while in three cases a relevant interaction caused a significant decrease in model accuracy. When non-relevant interactions were modeled, significantly increased model accuracy was observed in two out of four cases.

Despite variation in accuracy measures, ensemble projections of current day distributions for *B. bohemicus* (a species showing a large improvement in model accuracy for biotic models) were similar for abiotic models and biotic models with appropriate and inappropriate interactions (Fig. 2a–d). However, ensemble projections for 2050 under moderate climate change differed markedly. The abiotic model and the biotic model based on a non-relevant interaction both predicted that *B. bohemicus* would retreat into its present-day range core in Scotland (northwest Britain) (Fig. 2e, h). By contrast, both biotic models suggested the species would maintain or expand its Scottish distribution and also expand into Wales (southwest Britain) (Fig. 2f–g).

Discussion

This study compared species distribution models fitted with abiotic variables versus models that also included biotic variables. In some cases the biotic variables were the actual

Table 2. Effects of including biotic variables (R = relevant and NR = non-relevant) in distribution models. Cells report the mean difference in seven cross-validated model accuracy statistics between biotic and abiotic models. Positive values (+) indicate average increase, and negative (−) indicate decrease. Bold labels indicate statistically significant differences (p < 0.05 in paired t-tests). Following Table 1, biotic models for *Bombus* included the actual interacting species distribution, while biotic models for *Centris* and *Krameria* included predicted values for the interacting species.

Niche	Modelled species	Interacting species	Interaction type	Variation				
				AUC	TSS	Kappa	Sensitivity	Specificity
Parasite	*B. bohemicus*	All hosts	R	**+0.12**	**+0.17**	**+0.12**	+7.61	+3.14
	B. bohemicus	Host richness	R	**+0.13**	**+0.20**	**+0.12**	+5.98	+3.58
	B. bohemicus	*B. lucorum*	R	**+0.12**	**+0.17**	**+0.11**	+7.45	+2.41
	B. bohemicus	*B. magnus*	R	**+0.02**	+0.02	**+0.04**	−0.62	+3.46
	B. bohemicus	*B. cryptarum*	R	**+0.03**	+0.02	**+0.05**	−0.97	+2.31
	B. vestalis	*B. terrestris*	R	**+0.07**	**+0.22**	**+0.17**	+2.96	+7.31
	B. bohemicus	*B. terrestris*	NR	**+0.03**	+0.02	+0.01	+0.35	+0.78
	B. vestalis	*B. lucorum*	NR	**+0.07**	**+0.08**	**+0.09**	+0.55	+5.27
	B. vestalis	*B. magnus*	NR	+0.01	+0.02	+0.01	−1.58	+1.23
	B. vestalis	*B. cryptarum*	NR	+0.01	0	0	−4.41	+1.90
Host	*B. lucorum*	*B. bohemicus*	R	**+0.10**	**+0.12**	**+0.12**	+0.05	+5.62
	B. magnus	*B. bohemicus*	R	**+0.01**	−0.02	+0.02	+0.49	+0.75
	B. cryptarum	*B. bohemicus*	R	0	0	+0.03	−1.37	+2.17
	B. terrestris	*B. vestalis*	R	**+0.03**	**+0.03**	**+0.06**	−2.23	+5.80
	B. lucorum	*B. vestalis*	NR	**+0.08**	**+0.09**	**+0.09**	+3.51	+0.75
	B. magnus	*B. vestalis*	NR	**−0.03**	**−0.07**	**−0.04**	−1.51	+0.39
	B. cryptarum	*B. vestalis*	NR	**−0.09**	**−0.13**	**−0.08**	+0.59	−0.09
	B. terrestris	*B. bohemicus*	NR	0	**+0.02**	**+0.02**	−1.27	+2.46
Pollinator	*C. atripes*	*K. erecta*	R	0	0	−0.03	+1.17	+0.01
	C. fuscata	*K. tomentosa*	R	**+0.05**	−0.01	+0.08	+2.56	+0.08
	C. rhodopus	*K. erecta*	R	+0.01	+0.13	**+0.13**	**+14.35**	−0.56
	C. tarsata	*K. tomentosa*	R	−0.01	**−0.08**	−0.05	+8.04	−0.49
	C. atripes	*C. rhodopus*	NR	+0.03	+0.07	+0.02	+2.81	−0.11
	C. fuscata	*C. tarsata*	NR	**+0.10**	**+0.10**	+0.08	+7.68	+0.82
	C. rhodopus	*C. atripes*	NR	0	+0.03	**+0.08**	+3.26	+0.04
	C. tarsata	*C. fuscata*	NR	+0.01	−0.05	−0.11	+9.46	−0.53
Plant	*K. erecta*	All pollinators	R	0	+0.05	+0.03	**+6.75**	+0.21
	K. erecta	Pollinator richness	R	−0.02	+0.03	+0.06	+8.23	+0.06
	K. erecta	*C. atripes*	R	**+0.09**	**+0.39**	**+0.55**	**+49.91**	**+1.26**
	K. erecta	*C. rhodopus*	R	0	+0.01	+0.07	**+8.37**	−0.08
	K. tomentosa	All pollinators	R	0	−0.02	**−0.03**	−4.01	**+0.24**
	K. tomentosa	Pollinator richness	R	−0.01	−0.03	−0.03	−1.40	+0.07
	K. tomentosa	*C. fuscata*	R	−0.02	−0.01	−0.01	+1.34	−0.12
	K. tomentosa	*C. tarsata*	R	−0.01	0	**−0.03**	−2.34	**+0.15**

distributions of the interacting species, while in other cases it included their predicted suitability. Species added as covariates in the models had known interactions with the modeled species. In order to examine whether correlations between distributions are indicative of biotic interactions or represent statistical artifacts we also fitted models including species known not to interact strongly with the target bee species. We found that the different strategies for including biotic information resulted in different model accuracies, as measured by various cross-validated statistics. Including the actual distributions of interacting species generally resulted in significant higher accuracy for *Bombus* species. However, in many cases, including the distributions of non-interacting *Bombus* species also improved model accuracy. For the *Krameria-Centris* system, there was little overall difference between abiotic and biotic models, though the best results were obtained by adding the interacting species as its abiotic-model predicted values rather than its actual distribution.

There are a number of possible reasons for the differences between our findings for British *Bombus* and those for the neotropical *Krameria* and *Centris*. Great Britain has a more accurate distribution data than the Neotropics, so data quality might have affected the *Krameria-Centris* analysis. For example, while we can be reasonably confident of absences in Britain, those in the tropics are much less certain and this may affect model accuracy statistics designed for presence/absence data (Peterson et al. 2011). Also, inaccurate geo-referencing and our use of older records may make current climate and land use less useful as predictors (Johnson and Gillingham 2008). Alternatively, the interactions among *Bombus* species might be more specialized than those between *Krameria* and *Centris*. *Krameria* plants depend more on the *Centris* bees (their only pollinator) than the bees depend on *Krameria* oil rewards, since these bees can obtain their resource from other oil plant producers, such as Malpighiaceae species (Simpson 1989). *Centris* is also a widespread genus and other congeneric species could potentially pollinate the *Krameria* species analyzed here. Thus, our results support the view that specialized biotic interactions are more important than weak
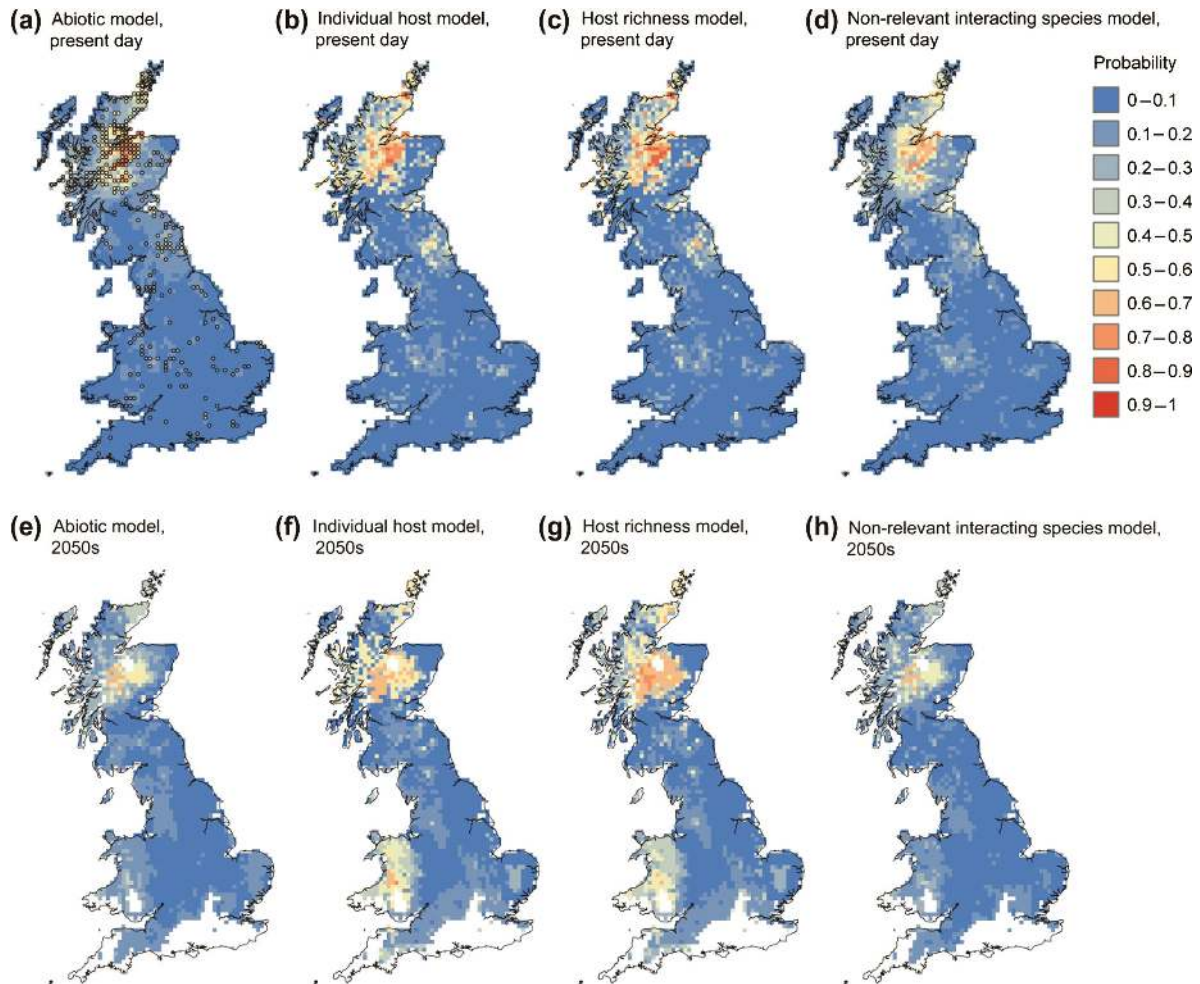
(a) Abiotic model, present day
(b) Individual host model, present day
(c) Host richness model, present day
(d) Non-relevant interacting species model, present day

Probability
0 − 0.1
0.1 − 0.2
0.2 − 0.3
0.3 − 0.4
0.4 − 0.5
0.5 − 0.6
0.6 − 0.7
0.7 − 0.8
0.8 − 0.9
0.9 − 1

(e) Abiotic model, 2050s
(f) Individual host model, 2050s
(g) Host richness model, 2050s
(h) Non-relevant interacting species model, 2050s

Figure 2. Projections of *Bombus bohemicus* occurrence areas. Top: present climatic conditions using (a) abiotic information only; (b) the most widespread host *B. lucorum*; (c) richness of its three host species (*B. lucorum*, *B. magnus* and *B. cryptarum*); (d) non-relevant interacting species (*B. terrestris*). Bottom: equivalent projections under expected 2050 climatic conditions for a moderate climate change scenario. White areas in the future projection maps show the regions where one or more of the climatic PCA gradients are outside the present day range, and on which projections may be unreliable.

and generalized interactions in determining regional-scale distribution patterns.

The 10 km resolution of environmental layers used could also have played a role since *Bombus* is a genus that contains large species with wide foraging ranges (Spaethe and Weidenmüller 2002). *Centris* is likely to be more sedentary and so the effects of biotic interactions on its regional-scale distribution may be masked by those of broad climatic gradients. However, this proposition is difficult to evaluate. Soberón (2010) argued that models including biotic interactions should use environmental layers with higher resolution since interactions are expected to occur at finer spatial scales. However, others have also proposed that the geographical signature of interactions can occasionally be expressed at higher geographical resolutions and extents (Gotelli et al. 2010, Araújo et al. 2011).

Our analysis considered two alternative ways of incorporating information on interacting species into distribution models – the use of the actual distribution and the use of 'habitat' suitability values from the abiotic models. An alternative approach would have been to use the actual or

predicted distribution of the interacting species to 'clip' the extent over which modelling was performed – i.e. to model the focal species only within areas where its host, parasite, pollinator or food plant are found. In effect, this approach was implicit to our modelling strategy since the BIOMOD system includes sophisticated machine learning algorithms that automatically detect and model interactions among variables (Thuiller et al. 2009b). As such they will model a variable interaction rule such as 'if the interacting species is (predicted) absent then the focal species is absent, regardless of the climate' provided the data support it.

Our results revealed an improvement in *Bombus*' model accuracy when including known interspecific interactions in species distribution models. This is consistent with previous works that used similar strategies. For example, including the relative abundance (Meier et al. 2010), frequency of occurrence (Heikkinen et al. 2007, Pellissier et al. 2010) and the current and future distribution of the interacting species (Hof et al. 2012) also enhanced the accuracy of the focal species model. However, our results also showed that in some cases adding information

on non-relevant species also improved the models. Reasons for this possible artefactual improvement of model accuracy include correlated bias in recording effort across species (Hortal et al. 2008), equal or opposite responses to environmental gradients (Baselga and Araújo 2010), influence of unmeasured environmental variables (Bateman et al. 2012), or correlation among spatially autocorrelated distribution patterns due to chance or shared evolutionary history (Chapman 2010). Therefore, our results suggest that increased species distribution model accuracy when proposed biotic interactions are included may not always provide evidence for the existence of such interactions. Based on our findings, we suggest that previous knowledge about species interactions are required prior to adding variables reflecting biotic interactions to species distribution models and for correct interpretation of the results (i.e. what is the underlying ecological or biological mechanism).

Our results suggest also that, for *Bombus*, the higher trophic level (parasite) might benefit more from including interactions than the lower trophic level (host), as shown by increases in model accuracy scores. This could reflect a bottom-up control of host–parasite communities, as suggested by Vazquez et al. (2007) who analyzed trophic networks. One would expect hosts to be able to exist in areas unsuitable for their parasites, but not the other way around. Recent reviews also emphasize that higher trophic levels are often disproportionately affected by drivers such as climate change, competition from invasive species, and habitat modification (Tylianakis et al. 2008, Gilman et al. 2010).

Consistent with other studies using ensembles of forecasts with climatic variables (Araújo et al. 2005, Pearson et al. 2006, Garcia et al. 2012), we found that although ensemble projections from the model variants were similar for the current day distribution, projected distributions under climate change were more sensitive to the inclusion of biotic interactions (Araújo and Luoto 2007, Meier et al. 2010, Hof et al. 2012). However, the analyses addressed here highlight the importance of considering different methods to include this information, and also the importance of reliable knowledge about focal species biology.

In this study we considered species pairs with specialized, presumably strong interactions. However, weak interactions dominate communities (Bascompte and Jordano 2007), and it is not clear whether a species' response to climate change would be influenced more by a small number of strong interactions or by a large number of weak ones. Such weak interactions might be better represented by multivariate distribution model approaches that represent interspecific associations or shared responses to environmental gradients and predict the distributions of multiple species (Elith and Leathwick 2007, Baselga and Araújo 2009). According to Tylianakis et al. (2008) the presence of weak interactions promotes web and meta-community stability at landscape scales. On the other hand, specialized pollinators, particularly ones with narrow diets, are vulnerable to disruption in the plant community (Memmott et al. 2007). This is particularly important since, as pointed out by Gilman et al. (2010), the key question is not the resulting effects of climate change on individual species, but the stability of the system as a whole.

Conclusions

We found that the distributions of species that have strong and specialized interactions with a focal species are useful predictors in species distribution models. However, including non-relevant interactions (i.e. other species that do not interact with the modeled species) can also yield a model improvement, showing that correlations between species distribution patterns are not necessarily indicative of interspecific interactions. Furthermore, we showed how projections of climate change impacts on species distributions are very sensitive to the specification of biotic interactions in the model. Thus, we suggest that a better understanding of species interactions, and how species assemblages face global changes as an integrated system, is needed to understand their resilience to these impacts.

References

Anderson, R. P. et al. 2002. Using niche-based GIS modeling to test geographic predictions of competitive exclusion and competitive release in South American pocket mice. – Oikos 98: 3–16.

Anderson, R. P. et al. 2003. Evaluating predictive models of species' distributions: criteria for selecting optimal models. – Ecol. Model. 162: 211–232.

Araújo, M. 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. Reducing uncertainty in projections of extinction risk from climate change. – Global Ecol. Biogeogr. 14: 529–538.

Araújo, M. B. et al. 2011. Using species co-occurrence networks to assess the impacts of climate change. – Ecography 34: 897–908.

Bascompte, J. and Jordano, P. 2007. Plant–animal mutualistic networks: the architecture of biodiversity. – Annu. Rev. Ecol. Evol. Syst. 38: 567–593.

Baselga, A. and Araújo, M. B. 2009. Individualistic vs community modelling of species distributions under climate change. – Ecography 32: 55–65.

Baselga, A. and Araújo, M. B. 2010. Do community-level models describe community variation effectively? – J. Biogeogr. 37: 1842–1850.

Bateman, B. L. et al. 2012. Nice weather for bettongs: using weather events, not climate means, in species distribution models. – Ecography 35: 306–314.

Benton, T. 2006. Bumblebees: the natural history and identification of the species found in Britain. – Collins.

Breiman, L. 2001. Random forests. – Mach. Learn. 45: 5–32.

Breiman, L. et al. 1984. Classification and regression trees. – Chapman and Hall.

Brooker, R. W. et al. 2007. Modelling species' range shifts in a changing climate: the impacts of biotic interactions, dispersal distance and the rate of climate change. – J. Theor. Biol. 245: 59–65.

Chapman, D. S. 2010. Weak climatic associations among British plant distributions. – Global Ecol. Biogeogr. 19: 831–841.

Elith, J. and Leathwick, J. 2007. Predicting species distributions from museum and herbarium records using multiresponse models fitted with multivariate adaptive regression splines. – Divers. Distrib. 13: 265–275.

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.

Franklin, J. 2009. Mapping species distribution. Spatial inference and prediction. – Cambridge Univ. Press.

Friedman, J. 1991. Multivariate adaptive regression splines. – Ann. Stat. 19: 1–141.

Garcia, R. A. et al. 2012. Exploring consensus in 21st century projections of climatically suitable areas for African vertebrates. – Global Change Biol. 18: 1253–1269.

Giannini, T. C. et al. 2011. Ecological niche similarities of *Peponapis* bees and non-domesticated *Cucurbita* species. – Ecol. Model. 222: 2011–2018.

Gilman, S. E. et al. 2010. A framework for community interactions under climate change. – Trends Ecol. Evol. 25: 325–331.

Global Land Cover 2000. Global Land Cover database. – European Commission, Joint Research Centre, 2003, < http: //bioval.jrc.ec.europa.eu/glc2000/glc2000.php >.

Godsoe, W. et al. 2009. Divergence in an obligate mutualism is not explained by divergent climatic factors. – New Phytol. 183: 589–599.

Gordon, C. et al. 2000. The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. – Clim. Dyn. 16: 147–168.

Gotelli, N. J. et al. 2010. Macroecological signals of species interactions in the Danish avifauna. – Proc. Natl Acad. Sci. USA 107: 5030–5035.

Guisan, A. and Thuiller, W. 2005. Predicting species distribution: offering more than simple habitat models. – Ecol. Lett. 8: 993–1009.

Hastie, T. et al. 1994. Flexible discriminant analysis by optimal scoring. – J. Am. Stat. Assoc. 89: 1255–1270.

Hegland, S. J. et al. 2009. How does climate warming affect plant–pollinator interactions? – Ecol. Lett. 12: 184–195.

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

Hijmans, J. R. et al. 2005. Very high resolution interpolated climate surfaces for global land areas. – Int. J. Climatol. 25: 1965–1978.

Hof, A. R. et al. 2012. How biotic interactions may alter future predictions of species distributions: future threats to the persistence of the arctic fox in Fennoscandia. – Divers. Distrib. 18: 554–562.

Hortal, J. et al. 2008. Historical bias in biodiversity inventories affects the observed environmental niche of the species. – Oikos 117: 847–858.

Johnson, C. J. and Gillingham, M. P. 2008. Sensitivity of species-distribution models to error, bias, and model design: an application of resource selection functions for woodland caribou. – Ecol. Model. 213: 143–155.

Lobo, J. M. and Tognelli, M. F. 2011. Exploring the effects of quantity and location of pseudo-absences and sampling biases on the performance of distribution models with limited point occurrence data. – J. Nat. Conserv. 19: 1–7.

Loken, A. 1984. Scandinavian species of the genus *Psithyrus* Lepeletier (Hymenoptera: Apidae). – Entomol. Scand. 23: 1–45.

McCullagh, P. and Nelder, J. A. 1989. Generalized linear models. – Chapman and Hall.

Meier, E. S. et al. 2010. Co-occurrence patterns of trees along macro-climatic gradients and their potential influence on the present and future distribution of *Fagus sylvatica* L. – J. Biogeogr. 38: 371–382.

Memmott, J. et al. 2007. Global warming and the disruption of plant–pollinator interactions. – Ecol. Lett. 10: 710–717.

Michener, C. D. 2000. The bees of the World. – Johns Hopkins Univ. Press.

Pearson, R. G. et al. 2006. Model-based uncertainty in species' range prediction. – J. Biogeogr. 33: 1704–1711.

Pellissier, L. et al. 2010. Species distribution models reveal apparent competitive and facilitative effects of a dominant species on the distribution of tundra plants. – Ecography 33: 1004–1014.

Peterson, A. T. et al. 2011. Ecological niches and geographical distributions. – Princeton Univ. Press.

Phillips, S. J. et al. 2006. Maximum entropy modelling of species geographic distributions. – Ecol. Model. 190: 231–259.

Powney, G. D. et al. 2010. Synchrony of butterfly populations across species' geographic ranges. – Oikos 119: 1690–1696.

Ramirez, J. and Jarvis, A. 2008. High resolution statistically downscaled future climate surfaces. – International Centre for tropical Agriculture, CIAT, < http://gisweb.ciat.cgiar.org/GCMPage >.

Ridgeway, G. 1999. The state of boosting. – Comput. Sci. Stat. 31: 172–181.

Ripley, B. D. 1996. Pattern recognition and neural networks. – Cambridge Univ. Press.

Simpson, B. B. 1989. Krameriaceae. – Flora Neotrop. Monogr. 49: 1–109.

Soberón, J. and Peterson, A. T. 2005. Interpretation of models of fundamental ecological niches and species' distributional areas. – Biodivers. Inform. 2: 1–10.

Soberón, J. M. 2010. Niche and area of distribution modeling: a population ecology perspective. – Ecography 33: 159–167.

Spaethe, J. and Weidenmüller, A. 2002. Size variation and foraging rate in bumblebees (*Bombus terrestris*). – Insect Soc. 49: 142–146.

Stockwell, D. and Peters, D. 1999. The GARP modeling system: problems and solutions to automated spatial prediction. – Int. J. Geogr. Inform. Sci. 13: 143–158.

Thuiller, W. et al. 2005. Climate change threats to plant diversity in Europe. – Proc. Natl Acad. Sci. USA 102: 8245–8250.

Thuiller, W. et al. 2009a. ModOperating manual for BIOMOD. – Laboratoire d'Écologie Alpine and Univ. Joseph Fourier.

Thuiller, W. et al. 2009b. BIOMOD: a platform for ensemble forecasting of species distributions. – Ecography 32: 369–373.

Tylianakis, J. M. et al. 2008. Global change and species interactions in terrestrial ecosystems. – Ecol. Lett. 11: 1351–1363.

Van der Putten, W. H. et al. 2010. Predicting species distribution and abundance responses to climate change: why it is essential to include biotic interactions across trophic levels. – Phil. Trans. R. Soc. B 365: 2025–2034.

Van der Wal, J. et al. 2009. Selecting pseudo-absence data for presence-only distribution modeling: how far should you stray from what you know? – Ecol. Model. 220: 589–594.

Vazquez, D. P. et al. 2007. Species abundance and asymmetric interaction strength in ecological networks. – Oikos 116: 1120–1127.

Vogel, S. 1974. Ölblumen und ölsammelnde Bienen. – Tropische Subtropische Pflanzenwelt 7: 285–547.

Warren, M. et al. 2010. A comparative approach to understanding factors limiting abundance patterns and distributions in a fig tree–fig wasp mutualism. – Ecography 33: 148–158.