

---

# Biotic interactions in species distribution modelling: retrieving fundamental niches from occupancy maps

---

Greg J. McInerny<sup>1 2</sup> & Drew W. Purves<sup>2</sup>.

<sup>1</sup> *Department of Computer Science, University of Oxford, Wolfson Building, Parks Road, Oxford OX1 3QD*

<sup>2</sup> *Computational Ecology and Environmental Sciences, Microsoft Research Limited, 7 JJ Thomson Avenue, Cambridge CB3 0FB, UK.*

\*Corresponding author: [gmcinerny@hotmail.com](mailto:gmcinerny@hotmail.com)

**Article Type:** *Research Paper*

**Running Title:** Biotic interactions in species distribution models (45 characters)

**Key words:** dimension reduction, realised distribution, climate change, range, likelihood, Bayesian, competition, potential distribution, facilitation, uncertainty.

**10** (10 max)

Word-count (excluding references, tables and figures): **5374** (5000 Max)

Abstract word-count: **193** (200 Max)

Number of References: **52** (50 Max)

Number of figures and tables: **6** (6 Max)

## 1 **ABSTRACT**

2 Biotic interactions are a long standing omission from Species Distribution Modelling (SDM). We  
3 define a model for a whole species assemblage whereby fundamental niches become modified via  
4 interactions to generate realized distributions. In a ‘virtual ecology’ test using presence / absence  
5 data simulated from a model with known parameters to generate artificial, we show that a Bayesian  
6 parameter estimation scheme can successfully retrieve the true fundamental niches and the true  
7 interaction matrix. This model correctly predicts the true effects of species removal on the remaining  
8 species (which other SDM approaches cannot). Our general method was enabled by an ecologically-  
9 inspired ‘dimension reduction’ model where interaction coefficients follow from differences  
10 between species’ use of resources along a resource axis. Crucially, the units of this axis need not be  
11 known beforehand. This allows unrecorded species to be included in analysis as hidden/latent  
12 variables (‘known unknowns’) and we can estimate their parameters in the same way as species with  
13 data. Future applications could yield both new ecological insights on species distributions, and new  
14 kinds of predictions regarding climate change responses, species extinctions, and other perturbations  
15 – especially when comparing this model with alternatives inspired by other ecological hypotheses.

## 17 **INTRODUCTION**

18 In a direct criticism of Species Distribution Modelling (SDM), Davis *et al.*, (1998) provide  
19 experimental evidence showing how biotic interactions might affect species’ occurrences along  
20 environmental gradients. Few SDM analyses have ever considered this central tenet of Ecology; that  
21 species interact and those interactions can affect where species occur (Darwin, 1859; McInerny &  
22 Etienne, 2012b).

23 Fourteen years after Davis *et al.* (1998) biotic interactions remain an unresolved issue in  
24 SDM (see, Kissling *et al.*, 2012; Wisz *et al.* 2013), often being listed as a caveat for studies that  
25 regress species distribution against climate data (e.g. Hof *et al.*, 2011). Including more ecology will  
26 always be challenging because so many processes could potentially determine species distributions  
27 (Gaston, 2009) and very little data is available beyond occupancy maps. Nonetheless, in order to  
28 estimate, understand, and predict species distributions, we must move beyond identifying biotic  
29 interactions as a problem (Sugden, 1992; Zurell *et al.*, 2009; Urban *et al.*, 2012), toward identifying  
30 and validating operational solutions that include biotic interactions in SDM analyses based only on  
31 occupancy data.

32 Some recent studies have included small sets of selected species interactions in SDM as  
33 additional independent predictor variables alongside climate and physical factors (Araújo & Luoto,  
34 2007; Heikkinen *et al.*, 2007; Preston *et al.*, 2010; Meier *et al.*, 2011; Schweiger *et al.*, 2012; Hof *et al.*  
35 *et al.*, 2012; also see Kissling *et al.*, 2012). To model a whole species assemblage using this approach

1 would require modelling each species separately as both a dependent variable (that is affected by  
2 other species) and as an independent predictor variable for the other species (that is unaffected by  
3 other species) (e.g. Meier *et al.*, 2011). This approach leads to very parameter rich models for any  
4 one species, and is difficult to apply in predictive mode since predictions for any focal species  
5 requires predictions of future distributions for all interacting species, which requires predictions for  
6 the focal species and all other species. And so on. There are some other related examples (e.g.  
7 Leathwick & Austin, 2001; Gotelli *et al.*, 2010), most notably for this paper the work of Ovaskainen  
8 *et al.* (2010) who modelled 22 fungal species using a Bayesian multiple regression approach. The  
9 strength of this research perhaps lies more in generating new hypotheses (Ovaskinen *et al.*, 2010),  
10 rather than in prediction (Wisz *et al.*, 2013)(also see Mutshinda & O'Hara, 2010).

11 A deeper treatment of biotic interactions in SDM would instead model *assemblages* of  
12 species within a single model, in terms of the 'fundamental niches' of each species (e.g. the  
13 underlying response to climate and physical factors) which then become modified through  
14 interspecific interactions (e.g. competition) to produce the *realized distributions* that we can observe  
15 (Hutchinson, 1957). For the vast majority of species we do not have detailed information on true  
16 fundamental niches nor the true interactions among species. In this paper we aim to demonstrate  
17 how such a community-based approach to SDM could infer all these model components using only  
18 presence-absence observations of the realized distribution. Fundamental niches may relate more  
19 closely to the true climatic tolerances of species and, alongside estimates of species' interactions,  
20 allow SDM to ask questions that are hard or impossible using correlative SDM methods; for  
21 example, by quantifying how differences in environmental dependence within communities or  
22 extinction of competitors might affect large scale biodiversity patterns.

23 We investigate our new method within a 'virtual ecology' study (Hilborn & Mangel, 1997;  
24 Mangel, 2006; Zurell *et al.*, 2010; McInerney & Purves, 2011). Using pseudo-data generated from a  
25 known model with known parameters, we test if our new method can retrieve the true fundamental  
26 niches and interactions parameters solely from the pseudo-data on species occupancy. This  
27 methodological verification is an essential prerequisite for applying any model to real data,  
28 especially if models might eventually inform policy (Clark *et al.*, 2001; Carter *et al.*, 2007;  
29 McMahon *et al.*, 2011). Testing SDM techniques against real data is also essential, but the virtual  
30 ecology approach focuses on the core principles of the technique - rather than on the specifics of  
31 particular regions, taxa, or datasets – which can avoid misunderstanding and uncritical usage of new  
32 methods (Renner & Warton, 2013).

## METHODS

Our approach defines a model that explicitly describes (1) the fundamental niche (defined below) for each species, (2) how species' interactions modify the fundamental niche to determine realized distributions, and (3) how those interaction coefficients arise from the overlap of species along a 'resource' axis. Importantly, this resource overlap model is general as the units of this resource do not have to be specified. We fit this model to the pseudo presence-absence data for all species, simultaneously, using a standard likelihood-based approach. This method is flexible because various functional forms could be used for 1, 2 & 3 above. We can also extract estimates and uncertainty on all model parameters which can then be propagated into model projections. Additionally, other SDM innovations could be included within the same overall likelihood-based framework, for example accounting for fine-scale environmental heterogeneity (McInerny & Purves, 2011).

### EFFECT OF PHYSICAL FACTORS: FUNDAMENTAL NICHES

We use the term 'fundamental niche' to define the distribution species  $j$  would adopt in the absence of other species, and 'realized distribution' to define the distribution that emerges when competitive interactions are included. This is similar to the notion of *pre-* and *post-interactive* distributions (Hutchinson 1957; McInerny & Etienne 2012 a, c) and a widespread understanding of this terminology. In our model, species distributions are determined by the species' responses to environmental variables in the absence of competition; hence we define the fundamental niche by these functions (see below).

Of course, this definition omits many other ecological and observation mechanisms that could affect observations of species' distributions (McInerny & Etienne 2012b). And so our model, as all models do, contains a variety of process and observation errors when applied to 'real' data (e.g. McInerny & Purves 2011; Marion et al. 2012). Models incorporating further ecological and observation mechanisms could potentially expand on this definition. However, the 'fundamental' and 'realised' niche terminology may be of reduced utility when ecological mechanisms are essential for any distribution to emerge (Godsoe, 2010; McInerny & Etienne 2012a, 2012c, 2013). Therefore, we selected this terminology and definition to communicate this approach to retrieving an observed environmental response function by accounting for biotic interactions in a SDM analysis (see McInerny & Etienne 2012c).

Thus our fundamental niche describes the response of a species to physical environmental factors (e.g. temperature and precipitation) which can be projected on to geographic space. The probability that species  $j$  would be recorded, in the absence of interactions, within a sample plot across a set of locations  $q$  (which is synonymous with the frequency of  $j$ , given a reference sample plot size, in the area immediately around  $q$ ). This probability is denoted  $P_{j,q}^F$  (where  $F$  denotes

fundamental niche) and depends on the environment at location  $q$  ( $V_{1,q}$ ,  $V_{2,q}$  etc.). To begin simply, we employ Gaussian responses for two arbitrary environmental variables:

$$P_{j,q}^F = K_j \cdot \exp\left(-1 \cdot \left[\frac{V_{1,q} - \phi_{1,j}}{\sigma_{1,j}}\right]^2\right) \cdot \exp\left(-1 \cdot \left[\frac{V_{2,q} - \phi_{2,j}}{\sigma_{2,j}}\right]^2\right) \quad (1)$$

Each species ( $j$ ) has an optimum  $\phi_{c,j}$  for each environmental variable  $c$ , and a tolerance parameter  $\sigma_{c,j}$  that modulates  $P_{j,q}^F$  away from this optimum (also see McNerny & Purves, 2011; Marion *et al.*, 2012). The maximum probability of occurrence for species  $j$  is set by the parameter  $K_j$ . Since this maximum must be between 0 and 1, we apply 0 and 1 as the bounds on the  $K_j$  parameters when they are estimated. Many functional forms could describe this response/fundamental niche (see Austin, 2002) and more environmental variables could be included without substantially increase the complexity of the analysis. We consider 2 environmental variables (equation 1) and estimate the parameter set  $\theta_{fund,j} = (\kappa_j, \phi_{1,j}, \sigma_{1,j}, \phi_{2,j}, \sigma_{2,j})$  for 10 species ( $j$ ).

## BIOTIC INTERACTIONS LEADING TO REALIZED DISTRIBUTIONS

The second step defines how realized distributions emerge from modification of the fundamental niches by biotic interactions. We restrict our analysis to competitive negative interactions and define a matrix of interaction coefficients ( $\alpha$ ), where each coefficient  $\alpha_{i,j}$  represents the effect of species  $i$  on species  $j$ . More precisely, we define  $\alpha$  as the reduction in frequency of species  $j$ , caused by a unit increase in the frequency of species  $i$  (see eqs 2-3). All intraspecific interactions ( $\alpha_{1,1}, \alpha_{2,2} \dots \alpha_{n,n}$ ) are equal to one but are accounted for in the formulation of the fundamental niche (Marion *et al.*, 2012) and so are ignored in the analysis. Our model combines interaction coefficients with the fundamental niches to produce the realized distribution ( $P_{j,q}^R$ ), defined as the equilibrium probability (i.e.  $\frac{dP_{j,q}^R}{dt} = 0$ ) of occurrence of species  $j$  at location  $q$  (the superscript  $R$  denoting realized niche):

$$P_{j,q}^R = \max(P_{j,q}^F - \Omega_{j,q}, 0) \quad (2)$$

Where the term  $\Omega_{j,q}$  is the summed effects of all other species ( $i \neq j$ ) on species  $j$  at location  $q$ . The effect of species  $i$  is the product of the relevant interaction coefficient ( $\alpha_{ij}$ ) and the realised niche (i.e. the frequency) for species  $i$  ( $P_{i,q}^R$ ):

$$\Omega_{j,q} = \sum_{i \neq j} \alpha_{ij} \cdot P_{i,q}^R \quad (3)$$

The  $\max()$  function in eq. 2 ensures that  $P_{j,q}^R \geq 0$ . From a computational and statistical standpoint this truncation is inconvenient, but it follows from the inclusion of additive species effects where a unit increase in the frequency of species  $i$  reduces the frequency of  $j$  by  $\alpha_{i,j}$ . Species are then affected by competitors' realized distributions rather than their fundamental niches.

From this straightforward model, simple fundamental niches (e.g. the Gaussians used here) can be converted into surprisingly complex, multinomial realized distributions (e.g. see Fig. 1). Nonetheless, the model does nothing more than specify a realized probability of occurrence for each species at each location  $q$ , given the model parameters ( $\theta_{fund}$  and the  $\alpha$  matrix) and the environmental variables at  $q$  ( $V_{1,q}, V_{2,q}$ ). The model could therefore be fitted to species presence-absence data using standard likelihood-based methods (see below).

There are however a couple of technical complications. Firstly, calculating the  $P_{j,q}^R$ , depends on knowing the realized distributions of all competing species ( $i \neq j$ ), and *vice versa*. This circularity in  $\Omega_{j,q}$  requires a numerical search for a set of realized distributions ( $P_{1,q}^R, P_{2,q}^R \dots P_{n,q}^R$ ) that is self-consistent as eq. 2 holds true for all species. This illustrates how modelling species interactions differs from a traditional approach of fitting each species separately (see introduction). Assemblage models such as this may be unlikely to have analytically tractable closed form solutions. Thus, we employed a simple numerical search for the self-consistent set of realized distributions for each location  $q$ , given the parameters (see *supplementary information*). In principle, such assemblage models could exhibit several alternate solutions (e.g. ‘either species a, or species b, but not both’). In most cases however, the species pool is relatively small because of the different environmental requirements. We did not find evidence for this in our simulations (see *supplementary information*). Moreover it would be unlikely to retrieve the parameter estimates if non-uniqueness of the solutions was a problem. Interestingly, with other models that exhibited stronger non-uniqueness, it may still be possible to utilize the models so long as the frequency of alternate solutions were estimated correctly.

## SCALABILITY & DIMENSION REDUCTION

As described above, each coefficient ( $\alpha_{i,j}$ ) could be estimated as a free parameter for small species assemblages. For large assemblages, however, the number of coefficients (parameters) increases explosively with the number of species  $n$  (number of free  $\alpha$  terms =  $(n^2 - n)/2$ , intra-specific effects removed); whereas the amount of data will increase at best linearly with  $n$  (i.e. each new species added has presence-absence observations). For example, the 158 species of European tree considered recently by Baselga & Araújo (2009) would require 12403  $\alpha$  coefficients in our analysis (the parameters for fundamental niches,  $\theta_{fund}$ , scale linearly with species number).

A practical method must reduce the number of free parameters (*‘dimension reduction’*). Dimension reduction could be approached by assigning species to groups and estimating coefficients for each group, or by creating a sparse matrix of coefficients by fixing most interaction terms to zero. Both these methods require considerable prior knowledge on the geometry of interactions. We explored an alternative dimension reduction approach that uses ecological ideas to reduce parameter

number. There are other potential models fitting this purpose including, but not limited to, decomposing interactions into effects and responses (Goldberg & Fleetwood, 1987; also see Kissling *et al.*, 2012), and Tilman's  $R^*$  and resource ratio theories (Tilman, 1987).

## DIMENSION REDUCTION VIA ECOLOGICAL THEORY

Our 'ecological' dimension reduction derives the  $\alpha$  matrix (which scales non-linearly with species number) from an ecological model, whose parameters scale linearly with species number. We denote these parameters  $\theta_{comp}$ . We implement a resource overlap model (inspired by MacArthur 1972, pp40-41) whereby interaction coefficients ( $\alpha_{i,j}$ ) follow from the overlap in resource use between species  $i$  and  $j$  on a single resource axis (MacArthur & Levins, 1967; MacArthur, 1972; also see Williams & Martinez, 2000):

$$\alpha_{i,j} = A \cdot \exp \left[ -1 \cdot \left[ \frac{x_j - x_i}{\omega} \right]^2 \right] \quad (4)$$

In our formulation, each species has a resource preference  $x_j$  and the closer species are on the resource axis the greater their competition coefficient. The axis represents resource partitioning between species co-occurring at a single location, e.g., rooting depth in plants, or preferred seed sizes for granivorous birds (MacArthur, 1972). The rate of decrease in competition is set by  $\omega$ , which applies equally to all species pairs (i.e. species use equivalent ranges of resources). Species with the same resource preference have a competition coefficient of  $A$  ( $x_j = x_i$ ,  $\alpha_{i,j} = A$ ) and compete as if they were conspecifics ( $\alpha_{i,i} = 1 = A$ ). Negative values of  $A$  (not considered here) could model facilitation, e.g. butterfly-host plant interactions (Araújo & Luoto, 2007) or some plant-plant interactions (see Brooker *et al.*, 2007). We limit ourselves to the 1-d case here, but see Williams & Purves (2011) for an analogous approach in food web modelling that using two axes. Competition coefficients could follow the species' positions in an n-dimensional resource-use space.

Crucially, this model can be fit to data without specifying what resources any such axes represent, whilst reducing the  $\alpha$  matrix to a single parameter ( $x_j$ ) per species and the parameter  $\omega$  (assuming we set  $A = 1$ ). We estimate this parameter set,  $\theta_{comp} = \{x_1, x_2 \dots x_n, \omega\}$ , from the data, at the same time as the fundamental niche parameters  $\theta_{fund}$ . If this model could describe competition between 158 European trees (as in Baselga & Araújo, 2009), we would need only 159 parameters to account for interactions between all species, compared to 12403 for a fully specified model.

## FITTING THE MODEL TO DATA

Because  $P_{j,q}^R$  is simply a probability, it is straightforward to state the associated likelihood function needed to estimate model parameters, given the presence-absence data, through Bayesian analysis enabled by MCMC sampling. Assuming independence among samples, the likelihood function of the whole data set ( $D_j^{observed}$ ) given the parameters  $\theta_{fund}$  and  $\theta_{comp}$ , is the sum of the logarithms of the probability for each datum (i.e. each presence-absence record) across all species  $j$  and locations  $q$ :

$$\mathcal{L}\{D_j^{observed}|\theta_{fund}, \theta_{comp}\} = \sum_q \sum_j \ln \left( Z(j, q|\theta_{fund}, \theta_{comp}) \right) \quad (5)$$

where  $Z(j, q|\theta_{fund}, \theta_{comp})$  denotes the probability of the data for species  $j$  and location  $q$ , given the parameters  $\theta_{fund}, \theta_{comp}$ :

$$Z(j, x|\theta) = \begin{cases} P_{j,q}^R & \text{if}(D_j^{observed} == 1) \\ 1 - P_{j,q}^R & \text{if}(D_j^{observed} == 0) \end{cases} \quad (6)$$

This likelihood (equations 5-6) fits all species together in one model.

A technical issue arises when fitting the ‘dimension reduced’ model. In its raw form, the model is ‘under constrained’ because multiple parameter combinations can give exactly the same fit to data. This implies that the data cannot fully specify all of the parameters. To see why, imagine reversing species positions along the resource axis, or adding any single value to each of their values. Reversal would have no effect as  $|x_j - x_i|$  is equivalent for  $x_j < x_i$  and  $x_j > x_i$ , and a translation along the axis again gives the same spacing between species (eq. 4). Similarly, adjusting the parameter  $\omega$  along whilst changing the spacing of the species can also give identical set of  $\alpha$  values. Thus species positions ( $x$ ) per se cannot be estimated from the data. Rather, there is an infinite family of equivalent solutions, defined by various stretches, translations and reversals on the resource axis, but all giving the same parameters, and the same uncertainty on parameters. This makes sense as the resource axis is not explicitly equated to particular units.

There are two solutions. Firstly, to estimate the parameters in the raw model form, then extract the unidentifiable parameters post-hoc (e.g. the species relative positions  $x$ ). This approach is quite involved and we did not pursue it further. Instead, after considerable experimentation, our solution fixes the positions ( $x$ ) for two randomly chosen species at an arbitrary distance – call these species  $a$  and  $b$ . The anchoring has no particular biological interpretation – it simply forces the Bayesian analysis to select one solution from the infinite family.

The parameter estimation can then only adjust  $\omega$  to give the correct value for  $\alpha_{a,b}$  as implied in the data. With  $x_a, x_b$  fixed and  $\omega$  strongly constrained by data, the remaining species positions are



constrained to positions estimated in relation to  $a$  and  $b$  (fig. 2). This also reduces the freedom of the analysis and we need to estimate two fewer parameters. Finally, we add a hierarchical distribution to encourage a normal distribution in the resource positions ( $x$ )(see *supplementary information*) and help constrain the least constrained species interactions, e.g. for species with limited overlapping of fundamental niches. With these two simple solutions in place, the model can be fit to data using a generic MCMC sampler (see below), which requires only the likelihood function (eqs 5-6) and a set of data providing, for a list of locations  $q$ , the presence or absence of each species, and the environmental variables at  $q$  (also see McInerny & Purves 2011).

## VERIFICATION USING VIRTUAL ECOLOGY

We verified this modelling approach in a ‘virtual ecology’ test by (a) generating pseudo-data from a the model using known parameters, and then (b) passing the pseudo-data to a separate implementation for the Bayesian analysis to test if that analysis could correctly retrieve the parameters underlying the pseudo-data set (Fig. 1, fig, S1). Importantly, step (b) proceeds exactly as it would for an analysis against real data because only occupancy map is passed to the Bayesian analysis and it uses no information on the true fundamental niches, or true interactions. The pseudo-data is subject to ‘noise’ derived from the sampling of the realised distribution (probabilities) to generate binary presence / absence records.

To ensure separation between the code generating the pseudo-data, and the code estimating parameters, the pseudo-data was generated in R (R Development Core Team, 2011). We randomly selecting variables from the distributions in table S1 for 10 species, then calculated the fundamental niches, simulated the realised distributions, and then sampled observations from realized distributions ( $P_{j,q}^R$ ). This data was then passed to the Bayesian analysis implemented in C++ using the ‘Filzbach’ library based on the Metropolis-Hastings algorithm. Filzbach enables reliable MCMC sampling over relatively complex, arbitrary models with large numbers of parameters (<http://research.microsoft.com/en-us/um/cambridge/groups/science/tools/filzbach/filzbach.htm>). In each step of the MCMC the realized distributions and observations are simulated using a C++ implementation of the model. The full workflow for this analysis is shown in fig S1 and details of the Bayesian analysis are given in *supplementary information*. For each model parameter the Bayesian analysis returns the posterior mean and credibility intervals. From these parameters, we can calculate fundamental niches, and realized distributions, and interaction coefficients; then compare these with their ‘true’ values, i.e., the values used to originally generate the pseudo-data.

## RESULTS

### REPRODUCTION OF REALIZED DISTRIBUTIONS AND FUNDAMENTAL NICHES

Comparing the true and estimated realised distributions shows that important qualitative features have been captured by the model (fig 3), even for those with complicated bimodal and/or highly skewed distributions (e.g. fig 3, species #2 & #9). However, the harder test is a comparison of the true and estimated fundamental niches which have not been directly observed (Figs. 3c). We found that the model inferred the qualitative and quantitative features of the fundamental niches very well (fig 3a, e). We observed some tendency to underestimate the fundamental niche ( $P_{j,q}^F$ ) but mostly in regions far from species' optimum environments where the true  $P_{j,q}^F$  is low (see fig S4). This means parameter estimates for the fundamental niches were conservative relative to the estimated realized distribution. The credible intervals on estimated fundamental niche optima parameters ( $\phi_{c,j}$ ) (fig 4a) contained the true parameter values for 18 of the 20 parameters estimated; and for the tolerance parameter ( $\sigma_{c,j}$ ) 17 of 20 parameters contained the true value (fig 4b).

The model also successfully extracted the species interactions, i.e., the  $\alpha$  matrix (Fig. 4c): with the credible intervals on  $\alpha$  values including the true values in most cases, even where species did not encounter each other because of the positions of their fundamental niches (fig 4c). There was a slight tendency to underestimate competition overall, which is consistent with the underestimation of fundamental niches (smaller fundamental niches require less 'trimming' to get to the same realized distributions). This effect was most marked for competition between species ID 6 and ID 2 with species ID 9 (fig 4c). We remind the reader that both in the true model, and in the model used in the parameter estimation scheme, the  $\alpha$  matrix were derived solely from the resource overlap dimension reduction scheme (fig. 2 & 3). The analysis correctly estimate the spacings of the species along the resource axis; but put the species on a different scale, and with the positions reversed (Fig. 4d). This is not a surprising, since the resource axis has no absolute units (see above) and the model recovers the true competition coefficients used to convert fundamental to realized niches.

### EFFECTS OF SPECIES REMOVAL

A more stringent model test is prediction beyond the conditions represented in the data. In one sense we have already demonstrated this by predicting fundamental niches from the binary presence-absence data (fig 4), equivalent to predicting a species' response to removal of all other species. In addition, we explored the effects of removing half of the species (fig 5) in the original model (that used to create the pseudo-data) and the estimated model (that based on parameters estimated from the pseudo-data). We found no obvious tendency to over- or under- predict the response to the species' extinctions (fig 5) although the predictive accuracy was dependent on exactly which species went extinct and which remained, largely because the accuracy of the estimates of the fundamental

1 niches (and so competition parameters) varied among species. Thus, changes in biotic context were  
2 reasonably robust (fig 5) because comparative estimates of  $\Delta P_{j,q}^R$  show unbiased predictions in the  
3 magnitude of change between pre- and post-extinction communities in the true and estimated  
4 models.

## 6 MODELLING THE EFFECTS OF UNRECORDED SPECIES

7 Unlike traditional SDM, our method fits a model for a whole assemblage of species to data from the  
8 whole assemblage (equation 5). Therefore, it could be possible to include a species in the model  
9 when we have no data for it if it has some effect on the recorded species, because the model  
10 functions (equations 1 - 4) cannot account for the observed realized distributions of those recorded  
11 species, without inferring the presence of, and so parameters for, the unrecorded species. This is  
12 analogous to the principle behind Le Verrier's inference that Neptune existed before it was observed  
13 directly, because of its effect on Uranus (Brookes, 1970).

14 Within SDM, including unrecorded species may be more valuable for accounting for their effects on  
15 the recorded species, rather than modelling their own distributions. For example, unobserved (and  
16 unnamed) boreal birds species that limit the northern boundaries of temperate birds could be  
17 included in a model, without any records of the boreal species. Without including the unobserved  
18 species, the distributional limit may be falsely attributed to the fundamental niche of the temperate  
19 birds.

20 To test this idea, we carried out exactly the same verification as above, but removed the  
21 presence-absence data for a single species  $j$ , and compared the outputs to the model including  
22 species  $j$  (Fig. 6). As shown in fig 6 a & b, a reasonable prediction of both fundamental niche and  
23 realised distribution can be retrieved for a species that was completely unrecorded. The 95% credible  
24 intervals on those estimates contain the true values, but importantly, they are larger than before (fig.  
25 4). Obviously, the retrieval of parameters for unrecorded species breaks down when the data  
26 contains little or no information on the effects of the unrecorded species on the recorded species.  
27 This can be seen in fig 6 e & f where the distribution is well constrained against variable 1 but is  
28 unconstrained against variable 2. Importantly, the Bayesian approach gives us a measure of that  
29 uncertainty (fig 6).

## 31 DISCUSSION

32 There is considerable, and long-standing, demand for techniques that advance ecological  
33 understanding by examining the effect of species interactions on their geographic distributions  
34 (Sugden, 1992; Davis *et al.*, 1998; Guisan & Thuiller, 2005; Thuiller *et al.*, 2008; Elith &  
35 Leathwick, 2009; Kissling *et al.*, 2012; Wisz *et al.*, 2013). We have presented a novel general

method for accounting for species interactions in Species Distribution Modelling (SDM) that proceeds by (a) defining an explicit model of the fundamental niche for each species (see niche definition above); (b) defining an explicit model of species interactions which augments these fundamental niches to create realized distributions; (c) reducing the dimensionality of the interactions model using an ecological model; (d) fitting the resultant model only to data sampled from the realized distributions but retrieving all model parameters. Importantly then, this methodology does not rely on new data resources, being based upon the same kind of presence-absence data (of realized distributions) that are routinely used in many Species Distribution Modelling studies already. Moreover, the analysis presented here is one case of a more general approach which can accommodate a wide variety of model formulations in specifying the fundamental niches, the interactions, and the dimensions reduction scheme. The virtual ecology test showed that, in principle, the approach can work. Parameterising a model on real data should always be followed by a virtual ecology test that ensures the estimated parameters can be again inferred from the data they simulate.

## NEW QUESTIONS

The principal benefit of modelling realized distributions from fundamental niches and species interactions is the opportunity to study new questions. Traditional SDM approaches are appropriate for estimating how the equilibrium distributions might be affected by climate change given equilibrium assumptions (Kissling *et al.* 2012). Whereas, the model presented here could also estimate changes in species distributions driven by interactions with other species. Regional changes in the abundance of one species (caused by climate change or other factors) could affect the geographical distribution of the other species in that region. We have shown here that, in principle, a model of the form presented here could correctly predict the effect of extinctions on the remaining species (Fig. 6) whereas a traditional SDM would necessarily predict no change.

Species interactions could also affect the response of species to climate change (and so alter the species assemblage) if different species shift at different rates (Urban *et al.*, 2012) or have different tolerances to the variables which are changing. At a given location, we are likely to find a mix of (1) fast-moving species which have arrived because the location has become newly suitable; (2) slow-moving species that remain, despite the reduction in unsuitability; and (3) species that are less affected by the climate change at that location. Our model could then, in principle, estimate how altered interactions generate transient changes across species' distributions. The model could of course be expanded to recognize that distributions can be in disequilibria with current climate where data allows (García-Valdés *et al.* 2013).

## MODEL EXTENSIONS

Using a ‘virtual ecology’ approach, we demonstrated that an explicit model of fundamental niches, species interactions, and realized distributions, could be parameterised from only data on realized distributions. In application to real data, at least two additional realities will need to be considered. First, in common with virtual ecology tests, we assumed that the model structure and dimension reduction scheme underlying true observations was known and the same model structure was used to generate the pseudo data, and in the parameter estimation (eqs. 1-4)(fig. S1). In application to real data the true underlying model cannot be known. Therefore, the model structure(s) used in the parameter estimation step are selected based on knowledge about the system, ecological theory, and computational plausibility. This patently calls for a variety of models to be developed and compared, and for caution in interpreting parameter estimates and model predictions from any single model without methodological validation. This limitation is true for any model-data analysis, including all SDM analyses, which specify a model structure(s) and apply those structure(s) to data.

Using a likelihood-based framework does mean we can propose several different model structures, then use widely-accepted rigorous methods to select from among them (e.g. AIC, BIC, or cross-validation techniques; Marion *et al.*, 2012). Selecting among competing models would give better model predictions and generate hypotheses about the ecology underlying distributions (Ovaskinen *et al.*, 2010). For example, it might be possible to conclude that interactions in one species assemblage are entirely competitive, or a mix of competitive or facilitative; that interactions are mostly strong, mostly weak, or a mix; that interactions tend to be asymmetric; and / or that species can be aggregated into functional groups.

A second limitation is that, in common with SDM analyses generally, our method makes a number of simplifying assumptions; such as assuming species distributions are at equilibrium with the climate, ignoring spatial interactions such as source-sink dynamic, assuming no dispersal limitation, and ignoring fine-scale variation in the environmental variables. Moreover, we assume no biases in sampling, or biases resulting from environmental variables not considered. However, the Bayesian methodology employed is a flexible format which could allow these issues to be addressed at the same time as interactions (e.g. Latimer *et al.*, 2006; McInerney & Purves, 2011; Marion *et al.*, 2012).

## DIMENSION REDUCTION

Modelling assemblages of interacting species can vastly increase the number of model parameters in general, not just within SDM. We addressed this problem using a ‘dimension reduction’ method inspired by ideas from theoretical ecology (niche overlap: see eqn 4, fig. 2, and MacArthur, 1972) which reduces the scaling between the number of parameters number and species. Arguably, all exercises in ecological modelling are dimension reductions compared to the full complexity of real

1 nature. For example, community ecology models often assume interactions depend only on species  
2 identities and ignore differences such as size, genotype, or sex. Our approach simply extends this  
3 philosophy, by assuming interactions between species depended on their separation in a one-  
4 dimensional resource-use space.

5 Thanks to fast computation, and computational statistics, and continuing increases in data  
6 availability, ecologists are now in a position to propose, select among, and parameterize a much  
7 wider variety of models, involving many more species, and more interesting and complex  
8 interactions among them. The SDM literature is a clear example, with recent analyses including  
9 hundreds of species (e.g. Baselga & Araújo, 2009). Therefore, we anticipate that the topic of  
10 dimension reduction may become more important in both ecology and SDM. Although generic  
11 approaches exist (e.g. Araújo & Luoto, 2007), it is worth considering novel methods (Wiens, 2011),  
12 especially when based on our ecological understanding. The model structures and associated  
13 parameter estimates used in dimension reduction may have great interest in themselves (Edwards &  
14 Smith, 2010).

## 16 **ACKNOWLEDGEMENTS**

17 We thank Bob O'Hara, Antoine Guisan and two anonymous referees for insightful comments on a  
18 previous version of this manuscript. The colours used for the graphics were selected using  
19 [www.colorbrewer2.org/](http://www.colorbrewer2.org/).

## 21 **REFERENCES**

- 22 1. Araújo, M.B. & Luoto, M. (2007) The importance of biotic interactions for modelling species  
23 distributions under climate change. *Global Ecology and Biogeography*, **16**, 743-753
- 24 2. Austin, M.P. (2002) Spatial prediction of species distribution: an interface between ecological  
25 theory and statistical modelling. *Ecological Modelling*, **157**, 101-118.
- 26 3. Baselga, A. & Araújo, M.B. (2009) Individualistic vs. community modelling of species  
27 distributions under climate change. *Ecography*, **32**, 55-65.
- 28 4. Brooker, R.W., Travis, J.M.J., Clark, E. & Dytham, C. (2007) Modelling species' range shifts in  
29 a changing climate: The impacts of biotic interactions, dispersal distance and the rate of climate  
30 change. *Journal of Theoretical Biology*, **245**, 59-65.
- 31 5. Brookes, C.J. (1970) On the Prediction of Neptune. *Celestial Mechanics*, **3**, 67-80.
- 32 6. Carter, T.R., Jones, R.N., Lu, X., Bhadwal, S., Conde, C., Mearns, L.O., O'Neill, B.C.,  
33 Rounsevell, M.D.A. & Zurek, M.B. (2007) New Assessment Methods and the Characterisation  
34 of Future Conditions. *Climate Change 2007: Impacts, Adaptation and Vulnerability*.  
35 Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental

- Panel on Climate Change, M.L. Parry, O.F. Canziani, J.P. Palutikof, P.J. van der Linden and C.E. Hanson, Eds., Cambridge University Press, Cambridge, UK, 133-171.
7. Clark, J.S., Carpenter, S.R., Barber, M., Collins, S.L., Dobson, A.P., Foley, J., Lodge, D.M., Pascual, M.R.A., Pielke, J., Pizer, W., Pringle, C.M., Reid, W., Rose, K., Sala, O., Schlesinger, W.H., Wall, D., Wear, D.N. (2001) Ecological Forecasts: An Emerging Imperative. *Science*, **293** (5530), 657-660.
  8. Darwin, C.R. (1859) *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. John Murray, London.
  9. Davis, A.J., Jenkinson, L.S., Lawton, J.H., Shorrocks, B. & Wood, S. (1998) Making mistakes when predicting shifts in species range in response to global warming. *Nature*, **391**, 783–786.
  10. Edwards, E.J. & Smith, S.A. (2010) Phylogenetic analyses reveal the shady history of C4 grasses. *Proceedings of the National Academy of Sciences, USA*, **107**, 2532–2537.
  11. Elith, J. & Leathwick, J.R. (2009) Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution and Systematics*, **40**, 677-697.
  12. García-Valdés, R., Zavala, M. A., Araújo, M. B., Purves, D. W. (2013) Chasing a moving target: projecting climate change-induced shifts in non-equilibrial tree species distributions. *Journal of Ecology*. doi: 10.1111/1365-2745.12049
  13. Gaston, K.J. (2009) Geographic range limits: achieving synthesis. *Proceedings of the Royal Society B*, **276**, 1395-1406.
  14. Goldberg, D.E. & Fleetwood, L. (1987) Competitive Effect and Response in Four Annual Plants. *Journal of Ecology*, **75**(4), 1131-1143.
  15. Godsoe, W. (2010) I can't define the niche but I know it when I see it: a formal link between statistical theory and the ecological niche. *Oikos*, **119**, 53-60.
  16. Gotelli, N.J., Graves, G.R. and Rahbek, C. (2010). Macroecological signals of species interactions in the Danish avifauna. *Proceedings of the National Academy of Sciences*, **107**, 5030-5035.
  17. Guisan, A. & Thuiller, W. (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, **8**, 993-1009.
  18. Heikkinen, R.K., Luoto, M., Virkkala, R., Pearson, R.G. & Korber, J.-H. (2007) Biotic interactions improve prediction of boreal bird distributions at macro-scales. *Global Ecology and Biogeography*, **16**, 754–763.
  19. Hilborn, R. & Mangel, M. (1997) *The Ecological Detective: Confronting Models with Data*. Princeton University Press, Princeton, NJ.
  20. Hof, C., Araújo, M.B., Jetz, W. & Rahbek, C. (2011) Additive threats from pathogens, climate and land-use change for global amphibian diversity. *Nature*, **480**, 516-521.

21. Hof, A. R., Jansson, R. and Nilsson, C. (2012), How biotic interactions may alter future predictions of species distributions: future threats to the persistence of the arctic fox in Fennoscandia. *Diversity and Distributions*. doi: 10.1111/j.1472-4642.2011.00876.x
22. Hutchinson, G.E. (1957) Concluding remarks. *Population Studies: Animal Ecology and Demography*. *Cold Spring Harbor Symposium on Quantitative Biology*, **22**, 415– 457.
23. Kissling, D W., Dormann, C., Groeneveld, J., Hickler, T., Kühn, I., McNerny, G J., Montoya, J M., Römermann, C., Schiffers, K., Schurr, F M., Singer, A., Svenning, J C., Zimmermann, N E., & O’Hara, B. (2012). Towards novel approaches to modelling biotic interactions in multispecies assemblages at large spatial extents. *Journal of Biogeography*. doi:10.1111/j.1365-2699.2011.02663.x
24. Latimer, A.M., Wu, S., Gelfand, A.E., & Silander, J A. (2006). Building statistical models to analyse species distributions. *Ecological Applications*, **16**, 33-50.
25. Leathwick, J.R. & Austin, M.P. (2001) Competitive interactions between tree species in New Zealand’s old-growth indigenous forests. *Ecology*, **82**, 2560–2573.
26. MacArthur, R. & Levins, R. (1967) The Limiting Similarity, Convergence, and Divergence of Coexisting Species. *The American Naturalist*. 101(921), 377-385.
27. MacArthur, R.H. (1972) *Geographical ecology*. Princeton University Press, Princeton.
28. Mangel, M. (2006). *The Theoretical Biologist's Toolbox - Quantitative Methods for Ecology and Evolutionary Biology*. Cambridge University Press. Cambridge.
29. Marion, G., McNerny, G J., Pagel, J., Catterall, S., Cook, A., Hartig, F., Schurr, F., & O’Hara, B. (2012). Parameter and uncertainty estimation for process-oriented models of species abundance and distribution: data, statistics and the niche. *Journal of Biogeography*. doi:10.1111/j.1365-2699.2012.02772.
30. McNerny, G J. & Purves, D W. (2011). Fine scale environmental variation in species distribution models: regression dilution, latent variables and neighbourly advice. *Methods in Ecology & Evolution*. **2** (3), 248-257. DOI: 10.1111/j.2041-210X.2010.00077.x
31. McNerny, G J. & Etienne, R S. (2012a). Ditch the niche - is the niche a useful concept in ecology or species distribution modelling? *Journal of Biogeography*, **39**, 2096-2102.
32. McNerny, G J. & Etienne, R S. (2012b). Stitch the niche - a practical philosophy and visual schematic for the niche concept. *Journal of Biogeography*, **39**, 2103-2111.
33. McNerny, G J. & Etienne, R S. (2012c). Enrich the Niche - taking responsibility for the concepts we use in ecology and species distribution modelling. *Journal of Biogeography*. **39**, 2112-2118.
34. McNerny, G J. & Etienne, R S. (2013). 'Niche' or 'Distribution' Modelling?: a response to Warren. *Trends in Ecology & Evolution*. DOI: 10.1016/j.tree.2013.01.007
35. McMahon, S., Harrison, S.P., Armbruster, W.S., Bartlein, P.J., Beale, C.M., Edwards, M.E., Kattge, J., Midgley, G., Morin, X., Prentice, I.C. (2011) Improving assessment and modelling of



- climate change impact on global terrestrial biodiversity. *Trends in Ecology and Evolution*, **26**(5), 249-259.
36. Meier, E.S., Edwards Jr, T.C., Kienast, F., Dobbertin, M. & Zimmermann, N.E. (2011) Co-occurrence patterns of trees along macro-climatic gradients and their potential influence on the present and future distribution of *Fagus sylvatica* l. *Journal of Biogeography*, **38**, 371-768.
37. Mutshinda, C. M., and R. B. O'Hara. 2010. Integrating the niche and neutral perspectives on community structure and dynamics. *Oecologia*, **166**, 241-251.
38. Ovaskainen, O., Hottola, J. & Siitonen, J. (2010) Modeling species co-occurrence by multivariate logistic regression generates new hypotheses on fungal interactions. *Ecology*, **91**, 2514–2521.
39. Preston, K.L., Rotenberry, J.T., Redak, R.A. & Allen, M.F. (2008) Habitat shifts of endangered species under altered climate conditions: importance of biotic interactions. *Global Change Biology*, **14**, 2501–2515.
40. R Development Core Team. 2011. R version 2.12.2. R Foundation for Statistical Computing, Vienna, Austria. (<http://cran.r-project.org/>)
41. Renner, I. W. and Warton, D. I. (2013), Equivalence of MAXENT and Poisson Point Process Models for Species Distribution Modeling in Ecology. *Biometrics*. doi: 10.1111/j.1541-0420.2012.01824.x
42. Schweiger, O., Heikkinen, R. K., Harpke, A., Hickler, T., Klotz, S., Kudrna, O., Kühn, I., Pöyry, J. and Settele, J. (2012) Increasing range mismatching of interacting species under global change is related to their ecological characteristics. *Global Ecology and Biogeography*, **21**, 88–99. doi: 10.1111/j.1466-8238.2010.00607.x
43. Sugden, A.M. (1992) Using biotic interactions to forecast the consequences of global climate change. *TREE*, **7**, 35–36
44. Thuiller, W., Albert, C.H., Araújo, M.B., Berry, P.M., Cabeza, M., Guisan, G., Hickler, T., Midgley, G.F., Paterson, J., Schurr, F.M., Sykes, M.T., Zimmermann, N.E. (2008) Predicting global change impacts on plant species distributions: future challenges. *Perspectives in Plant Ecology, Evolution and Systematics*, **9**, 137-152
45. Tilman, D. (1987) *Resource competition and community structure*. Princeton University Press, Princeton, NJ.
46. Urban, M., Tewksbury, J.J. & Sheldon, K.S. (2012) On a collision course: competition and dispersal differences create no-analogue communities and cause extinctions during climate change. *Proceedings of the Royal Society Series B* doi:10.1098/rspb.2011.2367
47. Wisz, M. S., Pottier, J., Kissling, W. D., Pellissier, L., Lenoir, J., Damgaard, C. F., Dormann, C. F., Forchhammer, M. C., Grytnes, J.-A., Guisan, A., Heikkinen, R. K., Høye, T. T., Kühn, I., Luoto, M., Maiorano, L., Nilsson, M.-C., Normand, S., Öckinger, E., Schmidt, N. M., Termansen, M., Timmermann, A., Wardle, D. A., Aastrup, P. & Svenning, J.-C. (2013) The

role of biotic interactions in shaping distributions and realised assemblages of species:  
implications for species distribution modelling. *Biological Reviews*, **88**, 15–30. doi:  
10.1111/j.1469-185X.2012.00235.x

48. Wiens, J. J. 2011. The niche, large-scale biogeography, and species interactions. *Philosophical Transactions of the Royal Society of London B*, 366, 2336–2350

49. Williams, R.J. & Martinez, N.D. (2000) Simple rules yield complex food webs. *Nature*, 404, 180–183.

50. Williams, R.J. & Purves, D.W. (2011) The probabilistic niche model reveals substantial variation in the niche structure of empirical food webs. *Ecology*, **92**, 1849–1857.

51. Zurell, D., Jeltsch, F., Dormann, C.F. and Schröder, B. (2009): Static species distribution models in dynamically changing systems: How good can predictions really be? *Ecography* 32(5): 733–744.

52. Zurell, D., Berger, U., Cabral, J.S., Jeltsch, F., Meynard, C.N., Münkemüller, T., Nehrbass, N., Pagel, J., Reineking, B., Schröder, B., Grimm, V. (2010): The virtual ecologist approach: simulating data and observers. *Oikos* 119(4): 622–635.

## **Biosketch**

Greg McInerny is a Senior Research Fellow in Information Visualisation working on new approaches to both spatial and visual analytics. Alongside his research in Species Distribution Modelling, Greg is interested in the feedbacks between the structure, generation and regulation of biodiversity, and the role of software in modern scientific practice.

Drew Purves is an ecologist, and head of the Computational Ecology and Environmental Science group at Microsoft Research Cambridge, UK. His primary research is in combining ecological theory with different sources of data, via computational statistics, to developing realistic, predictive models of ecological phenomena. Drew's research varies widely in scale and taxa, from the growth of individual annual plants over timescales of days, to the global dynamics of carbon and ecosystem function over timescales of centuries.

Author contributions: G.M. and D.P. conceived the ideas, carried out the analyses, and wrote the paper.

## FIGURE LEGENDS

Figure 1: We want to retrieve an estimate of the underlying response to climate variables, the *fundamental niche*  $P_{j,q}^F$ , defined here as the distribution a species would exhibit if there were no interactions with other species in the assemblage (shown for a focal species as the grey-scale shaded region of a, and as solid grey lines projected on to one or other environmental axis, c, d). The fundamental niche becomes modified by interactions to form the *realised distribution*  $P_{j,q}^R$  (shown for the focal species as orange contours, and for other species as grey contours, in a, c, d). In occupancy maps, we observe only presence / absence sample of the realized distribution (circles show presences, absences are all other grid cells). In panel a, contours show  $P_{j,q}^R = (0.1, 0.3, 0.5, 0.7, 0.9, 0.95)$ . The dotted lines in (a) show the position of the transects shown in (c) & (d). Note that in (d) the species does not occur, but we would like to estimate what its fundamental niche is. The method we present in this paper can do this by modeling the species interactions.

Figure 2: The dimensions of the full competition matrix of species coefficients  $\alpha$  (top) can be reduced by assuming a resource axis along which competition takes place (bottom) such that the coefficient for species  $i$  and  $j$ ,  $\alpha_{i,j}$ , depends only on their relative positions along the resource axis. Therefore, we need only estimate these positions from the data (a number of parameters that scales linearly with the number of species) rather than each separate  $\alpha$  parameters (the number of which scales with the square of the number of species).

Figure 3: Knowing the true model (that is, true parameter values) that produce the true fundamental  $P_{j,q}^F$  (a) and true realised  $P_{j,q}^R$  (b) distributions, allows us to generate pseudo-data consisting of presence / absence samples from the true model (c). These are passed to a parameter estimation scheme which estimates the model parameters. These model parameters can then be used to generate estimates of the realised (d) and fundamental (e) distributions, which can be compared to the true distributions (a, b). In an application of this model to ‘real’ data we would only have occupancy samples (c) of the distribution, and the fundamental niche would not be known. The fact that the fundamental niches are estimated well (compare a, e) shows that, in principle, it is possible to estimate the fundamental niche given only noisy data from the realized distribution (c) as are available in reality. All 10 species were included in the analysis, but we show 6 here. Contours for are shown for  $P_{j,q}^R = (0.1, 0.3, 0.5, 0.7, 0.9, 0.95)$ .

Figure 4: True vs. estimated parameters. In (a), for each species (different colours), the difference between true and estimated, with 95% credible intervals, is shown for the niche optimum (a) and niche breadth (b), which define the fundamental niche. These parameters are not directly observable from the data, but are estimated by the parameter estimation scheme, which is provided with presence / absence samples from the realized distributions only. At the same time, the analysis estimates the competition coefficient for each species pair (c: bars show 95% intervals on the difference between the estimated and true coefficient) by estimating the species positions ( $x$ ) along the resource axis (d: different species in different colours; bars are 95 credible intervals) (estimated value of  $\omega$ , mean = 0.953, 95% credible intervals,  $0.823 < \omega < 0.997$  (3dp)) In panel (d), species pairs are coloured according to the amount of overlap in their fundamental niches (darker shade implying more overlap). Note the underestimate of competition between species 9 and 2, and species 9 and 7 (see text) whereas for other species pairs the 95% intervals contain the true coefficient, showing that most competition coefficients were retrieved correctly.

Figure 5: Examination of the model's predictive ability. We begin with a full assemblage (a) then remove half of the species at random (grey contours in a). In response, the remaining species expand their realized distributions (b, c). This simulation can be carried out using the true model with known parameters (b) or with the estimated model (c). Panels (d) and (e) show the difference (after minus before) in the realized distributions of the remaining species. As the figure shows, the estimated model was able to predict the effect of extinctions correctly, even though it was parameterized only against presence / absence samples from the realized distributions prior to the extinctions. Contours for are shown for the species' realised distributions  $P_{j,q}^R = (0.1, 0.3, 0.5, 0.7, 0.9, 0.95)$ .

Figure 6: True vs. estimated distributions and parameters for two different unrecorded species, each of which had no data (see text). The top panels show the true and estimated fundamental and realised realized distributions of the unrecorded species (a, b, e, f). The bottom panels show the difference between estimated and true values for the fundamental niche estimates (c, d, g, h). Note the change in scales in the bottom panels.

FIGURES

Figure 1

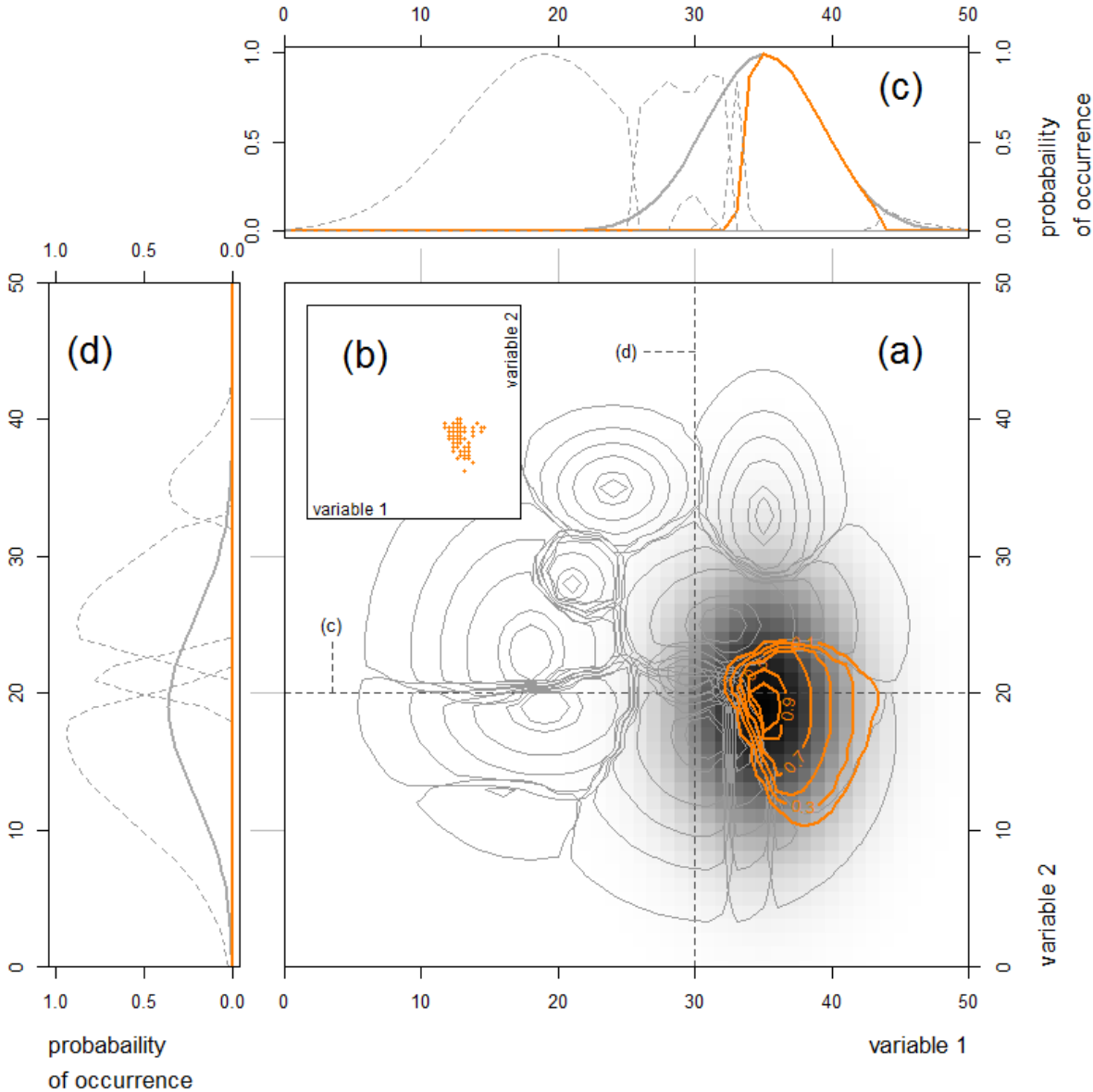


Figure 2

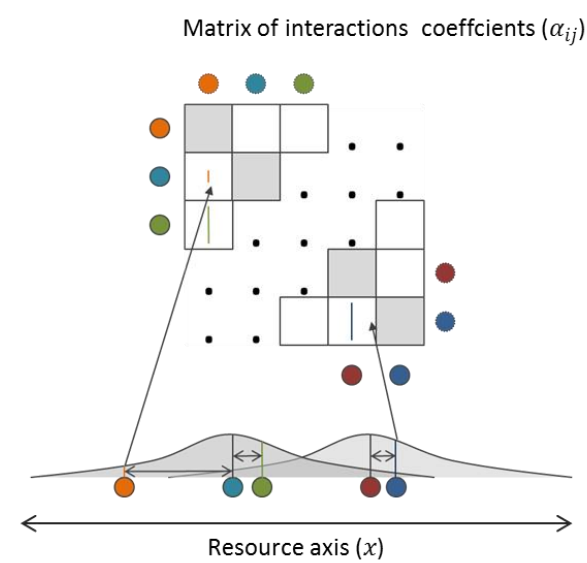
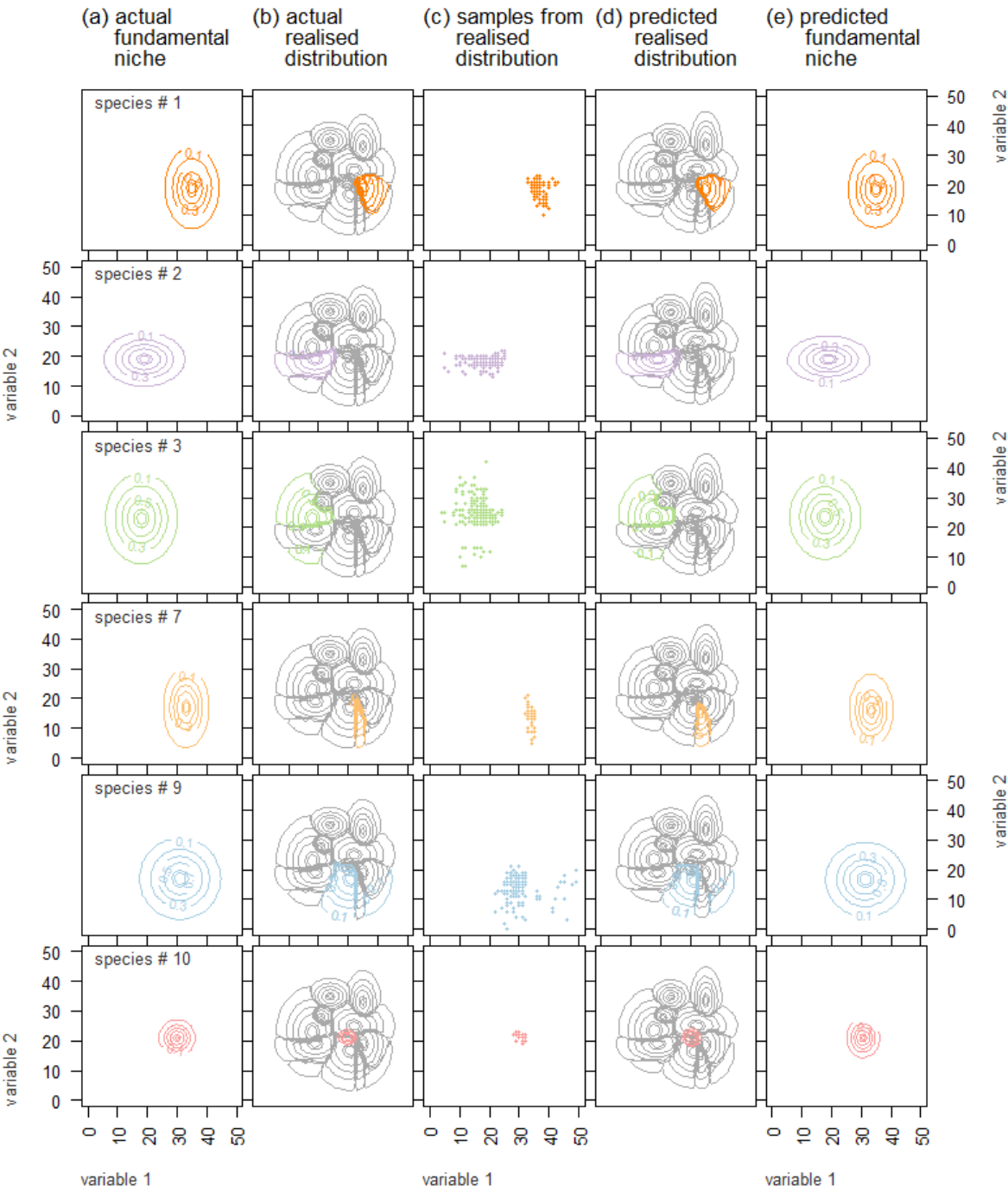
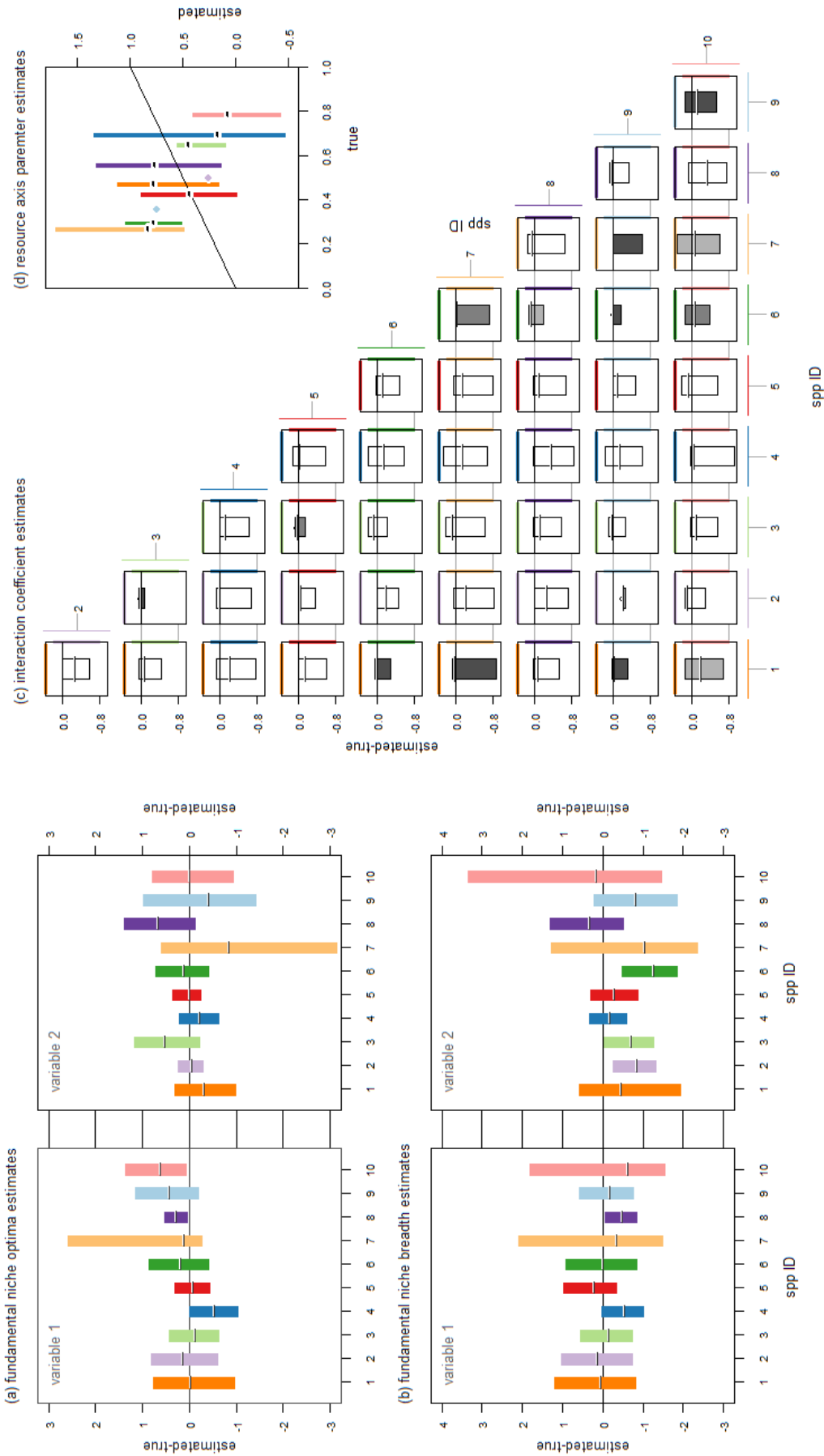


Figure 3



1 Figure 4



2  
3



1 Figure 5

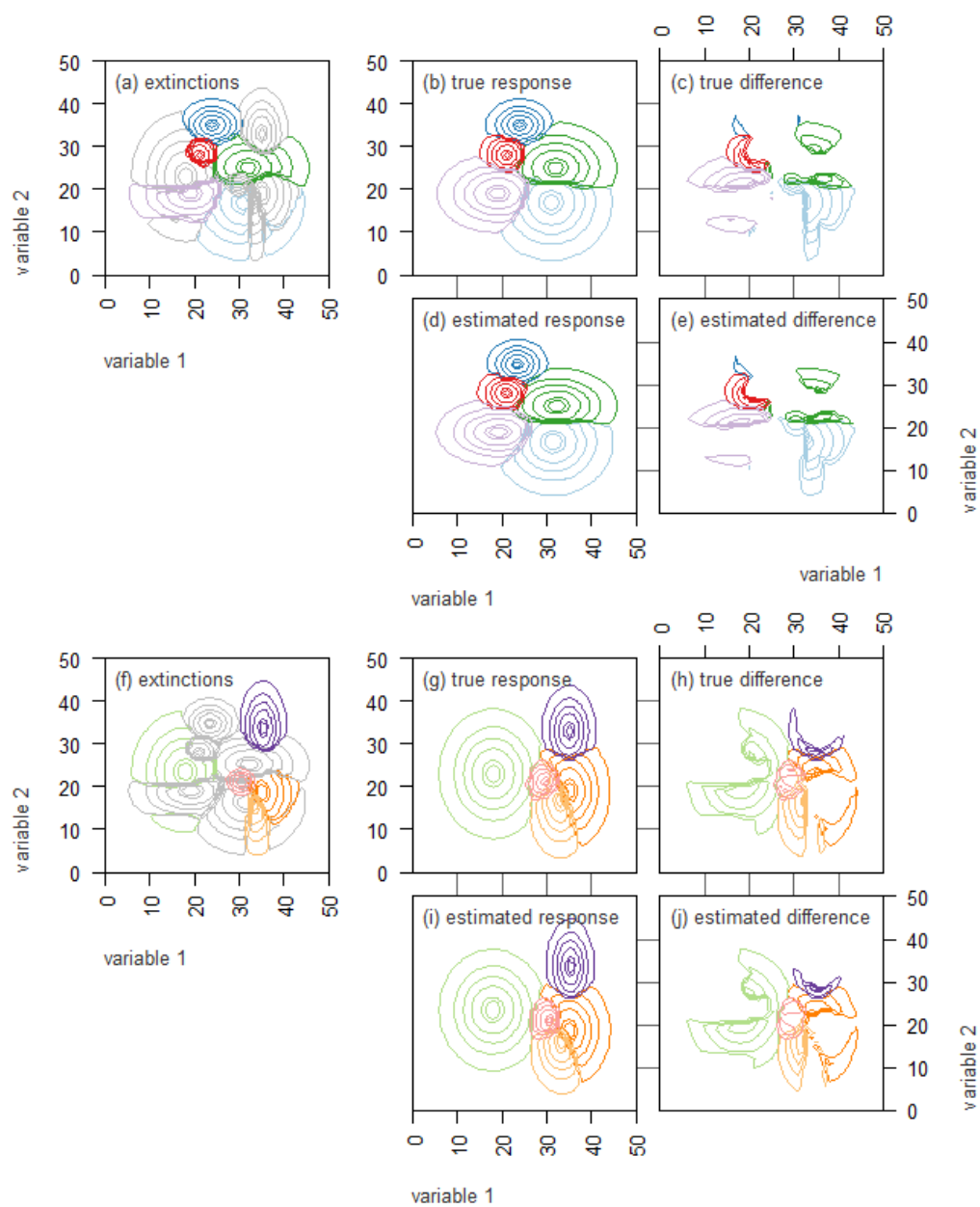


Figure 6

