Towards novel approaches to modelling biotic interactions in multispecies assemblages at large spatial extents

W. D. Kissling1, Carsten F. Dormann2,3, Jürgen Groeneveld4,5, Thomas Hickler6, Ingolf Kühn7, Greg J. McInerney8, José M. Montoya9, Christine Römermann10,11, Katja Schifflers12, Frank M. Schurr10,13, Alexander Singer4, Jens-Christian Svenning1, Niklaus E. Zimmermann14 and Robert B. O’Hara6

ABSTRACT

Aim Biotic interactions – within guilds or across trophic levels – have widely been ignored in species distribution models (SDMs). This synthesis outlines the development of ‘species interaction distribution models’ (SIDMs), which aim to incorporate multispecies interactions at large spatial extents using interaction matrices.

Location Local to global.

Methods We review recent approaches for extending classical SDMs to incorporate biotic interactions, and identify some methodological and conceptual limitations. To illustrate possible directions for conceptual advancement we explore three principal ways of modelling multispecies interactions using interaction matrices: simple qualitative linkages between species, quantitative interaction coefficients reflecting interaction strengths, and interactions mediated by interaction currencies. We explain methodological advancements for static interaction data and multispecies time series, and outline methods to reduce complexity when modelling multispecies interactions.

Results Classical SDMs ignore biotic interactions and recent SDM extensions only include the unidirectional influence of one or a few species. However, novel methods using error matrices in multivariate regression models allow interactions between multiple species to be modelled explicitly with spatial co-occurrence data. If time series are available, multivariate versions of population dynamic models can be applied that account for the effects and relative importance of species interactions and environmental drivers. These methods need to be extended by incorporating the non-stationarity in interaction coefficients across space and time, and are challenged by the limited empirical knowledge on spatio-temporal variation in the existence and strength of species interactions. Model complexity may be reduced by: (1) using prior ecological knowledge to set a subset of interaction coefficients to zero, (2) modelling guilds and functional groups rather than individual species, and (3) modelling interaction currencies and species’ effect and response traits.

Main conclusions There is great potential for developing novel approaches that incorporate multispecies interactions into the projection of species distributions and community structure at large spatial extents. Progress can be made by: (1) developing statistical models with interaction matrices for multispecies co-occurrence datasets across large-scale environmental gradients, (2) testing the potential and limitations of methods for complexity reduction, and (3) sampling and monitoring comprehensive spatio-temporal data on biotic interactions in multispecies communities.
INTRODUCTION

Whether the impacts of climate change on biodiversity and species distributions can be accurately predicted remains unclear. Theoretical and empirical studies suggest that biotic interactions play an important role in determining the response of species and communities to changing environments (Araújo & Luoto, 2007; Heikkinen et al., 2007; Kissling et al., 2007; Tylianakis et al., 2008; Gilman et al., 2010). The lack or limited incorporation of biotic interactions within forecast models might therefore be one reason for their failure to predict the consequences of global change on species and ecosystems (Gilman et al., 2010; Kissling et al., 2010). Thus far, little emphasis has been given to incorporating biotic interactions into species distribution models at large spatial scales (Pearson & Dawson, 2003; Araújo & Luoto, 2007; Heikkinen et al., 2007; Schweiger et al., 2008).

A species’ niche may be defined as encompassing the combinations of conditions that allow a species to survive in a region (i.e. a population growth rate \( r > 0 \); Hutchinson, 1957, 1978). These conditions include abiotic (i.e. physical–chemical) environment factors, and also abiotic and biotic consumable resources (e.g. soil nutrients, plants, animal prey) and the whole biotic interaction milieu (e.g. competitors, predators or prey, pathogens or hosts etc.) (Hutchinson, 1957, 1978). Quantifying species’ niches may also require models that account for the impacts of species on the ecological communities they inhabit (Chase & Leibold, 2003). This includes ecosystem engineers and apex consumers, which may exert massive impacts on their biotic and abiotic environment (Linder et al., in review). However, translating species’ niches into geographic distributions is a complex and difficult task (Holt, 2009), and the modelling of spatio-temporal dynamics in multispecies communities remains a significant challenge in ecology (Solé & Bascompte, 2006).

In this paper, we provide a synthesis of how biotic interactions can be modelled. Our goal is to stimulate the development of novel models for the projection of large-scale, multispecies distribution patterns. We focus on species that interact locally and where those interactions scale-up to coarse spatial extents and resolutions. Such models should allow interactions between multiple species to be estimated simultaneously, enabling projections of spatial and temporal patterns of species assemblages in changing environments. We start by summarizing how classical species distribution models (SDMs; Elith & Leathwick, 2009) have recently been extended to incorporate biotic interactions to a limited extent (‘SDM extensions’). We then review principal ways of modelling multispecies interactions and illustrate promising new tools for incorporating multispecies interactions via interaction matrices. Collectively, we refer to these novel approaches as ‘species interaction distribution models’ (SIDMs). SIDMs offer novel avenues for modelling multispecies assemblages across large spatial extents and for describing the structure of species assemblages that could arise under global environmental change.

BIOTIC INTERACTIONS IN SPECIES DISTRIBUTION MODELS

Biotic interactions have been modelled and quantified using a diversity of methods that cover a range of spatial extents (Table 1). In contrast to the many traditional models that mainly focus on local spatial extents (Table 1) and fine resolutions (e.g. 10 × 10 m vegetation plots), SDMs aim to model species distributions across large extents by linking observation records (usually presence-only or presence/absence data) to abiotic environmental variables (Elith & Leathwick, 2009). These classical SDMs usually neglect biotic interactions. A range of different statistical methods is used to fit response surfaces (Elith et al., 2006), which can also be extrapolated to assess the effects of climate change on species distributions by using climatic surfaces from a range of regional or global circulation models.

Numerous assumptions are made during model fitting and subsequent application of classical SDMs (Guisan & Thuiller, 2005), including the assumption that biotic interactions are equally strong and important across the entire species’ range and unimportant for range dynamics at large spatial scales. In most cases, the coarse resolutions (e.g. 10 × 10 km) of SDM applications are used to justify the omission of biotic interactions (Pearson & Dawson, 2003). However, it remains unclear how biotic interactions scale-up from local to global scales and whether coarse grain sizes of sampling units are unsuitable for detecting and representing biotic interactions. Additionally, the projection of SDMs into the future assumes that the strength and direction of biotic interactions remain stable over space and time even though a wide range of empirical studies show that species interactions are altered by global environmental change (Tylianakis et al., 2008).

An increasing awareness of the importance of biotic interactions has stimulated attempts to incorporate species interactions into the classical SDM framework. These SDM extensions include the following.
Table 1 Examples of how biotic interactions have been quantified and modelled. The list is intended to exemplify various approaches but is not intended to be exhaustive.

<table>
<thead>
<tr>
<th>Model type</th>
<th>Biotic interaction</th>
<th>Spatial extent</th>
<th>Implementation of interaction</th>
<th>Data</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Species distribution models</td>
<td>Competition, facilitation, trophic interactions</td>
<td>Regional to global</td>
<td>Additional predictors in statistical models or constraining model predictions to the presence of interacting species</td>
<td>Frequency or count data, basal area, proportional abundance, presence/absence, presence-only (all georeferenced)</td>
<td>Leathwick &amp; Austin (2001), Rouget et al. (2001), Araújo &amp; Luoto (2007), Heikkinen et al. (2007), Schweiger et al. (2008, 2012), Meier et al. (2010, 2011a)</td>
</tr>
<tr>
<td>Zone of influence models</td>
<td>Plant competition</td>
<td>Local</td>
<td>Modelling overlap of individuals’ areas of influence</td>
<td>Physiological data of resource uptake</td>
<td>Gates &amp; Westcott (1978), Schifflers et al. (2011)</td>
</tr>
<tr>
<td>Population dynamic models</td>
<td>Competition, host–parasitoid</td>
<td>Local to regional</td>
<td>Discrete-time stochastic Gompertz model, coupled-map lattice model</td>
<td>Abundance data, e.g. time-series from trapping, surveys and counts</td>
<td>Mutshinda et al. (2009, 2011)</td>
</tr>
<tr>
<td>Epidemiological models</td>
<td>Infectious diseases</td>
<td>Local to global</td>
<td>Wavelet time series analysis, individual-based models, stage-based (susceptible, infected, recovered, SIR) models</td>
<td>Frequency of outbreaks, abundance, transmission rates, demographic rates, behavioural data, etc.</td>
<td>Grenfell et al. (2001), Keeling et al. (2001)</td>
</tr>
<tr>
<td>Multivariate regression models</td>
<td>Competition, facilitation, mutualism</td>
<td>Local to regional</td>
<td>Spatial associations in residuals of regression models (cross-covariance matrices)</td>
<td>Presence/absence, abundance</td>
<td>Banerjee et al. (2008), Latimer et al. (2009), Ovaskainen et al. (2010), Sebastián-González et al. (2010)</td>
</tr>
<tr>
<td>Dynamic vegetation models</td>
<td>Plant competition</td>
<td>Landscape to global</td>
<td>Simulation of individuals, cohorts or populations competing for resources (e.g. space, light or water)</td>
<td>Mathematical functions for growth, mortality, establishment etc.</td>
<td>Lischke et al. (2006), Prentice et al. (2007), Rüger et al. (2008)</td>
</tr>
<tr>
<td>Ecological networks</td>
<td>Mutualistic interactions, food webs, host–parasitoid</td>
<td>Local</td>
<td>Analysis of linkages between species, network properties</td>
<td>Presence/absence of feeding links, per capita interaction strength, visitation frequencies</td>
<td>Berlow et al. (2004), Wootton &amp; Emmerson (2005), Scotti et al. (2007), Ings et al. (2009)</td>
</tr>
</tbody>
</table>

1. Adding an interacting species as additional predictor. This is the most straightforward and widely used approach for SDM extensions and has been implemented for a variety of interaction types. For instance, competition between trees has been considered by adding plot-level predictors representing frequency or count data (Leathwick & Austin, 2001), basal area (Rouget et al., 2001), or proportion data (Meier et al., 2010) of co-occurring tree species. Similarly, the distribution of host plants has been added as a predictor to model butterfly distributions (Araújo & Luoto, 2007; Preston et al., 2008). Interactions between animals have also been attempted, e.g. by using occurrence data of woodpeckers (providing nesting cavities for secondary cavity nesters) to model their effects on owl distributions (Heikkinen et al., 2007).

2. Modelling the distributions of interacting species separately by using abiotic variables, and then representing interactions by restricting the distribution of one species to the modelled distribution of the other. This approach has been implemented for butterfly distributions that can be restricted to the distribution of their larval host plants (Schweiger et al., 2008, 2012). It differs from 1 (above) by minimizing potential problems of false absences of the butterfly due to missing host plants and collinearity within regression models when a predictor species responds to similar ecological factors as the response species.

3. Integrating classical SDMs with process-based models of biotic interactions. This has been predominantly applied to plant species distributions by modelling bioclimatic limits and physiological representations of competition for resources (e.g. light and water) as derived from process-based forest gap models (e.g. Lischke et al., 2006; Hickler et al., 2012). These models can be combined with classical SDM approaches (e.g. Meier et al., 2011b).
These examples illustrate how biotic interactions are increasingly represented within extensions of classical SDMs. However, in most cases only one or a few species that potentially interact with a target model species are included and data usually come from a single snapshot in time. In all cases, the interaction type is unidirectional, i.e. one species depends on one or several others (Fig. 1a), and reciprocal effects of interactions have usually been neglected. Most important, species interactions have not been explicitly (or directly) modelled and are usually assumed to be static in space and time. Hence, the classical SDM framework with its extensions needs to be complemented by SIDMs, which explicitly model interactions among multiple species.

**PRINCIPAL WAYS OF MODELLING INTERACTIONS BETWEEN MULTIPLE SPECIES**

In addition to simple unidirectional interaction effects that have occasionally been used in SDM extensions (Fig. 1a; see above), we distinguish three ways of describing biotic interactions for multispecies systems: (1) simple qualitative linkages between species (Fig. 1b), (2) quantitative interaction effects that have occasionally been used in SDM extensions (Fig. 1a), and (3) interactions mediated by currencies (e.g. resources). In (c) and (d) the thickness of arrows represents the strength of interactions (thickness is proportional to values in the interaction matrix). Column and row names (in bold) of the matrices refer to species (numbers 1–4) and currencies (letters A and B). In (d) the first matrix represents the impact of a species on currencies (e.g. resources) whereas the second matrix represents the currency requirements of species (i.e. the effect of the currency on the species). Pairwise species interactions (e.g. highlighted in grey for species pair 1–2) can thus be unidirectional and asymmetric (a), reciprocal and symmetric in strength (b), reciprocal and asymmetric in strength (c), and via mediating currencies (d).
coefficients that reflect interaction strengths (Fig. 1c), and (3) parameters describing how each species has an effect on, and responds to, interaction currencies that mediate interactions (i.e. resources or conditions, Fig. 1d). All three approaches can be described by interaction matrices (either between species, or between species and interaction currencies; Fig. 1b–d). These interaction matrices quantify the effects of many species upon each other for each species pair, either as pairwise interaction coefficients or as functions that describe how pairwise interactions depend on other factors (e.g. environmental variables). Pairwise interactions can also be retrieved from interaction currencies. An advantage of using interaction currencies is that pairwise interactions between species do not need to be quantified directly, but emerge indirectly from modelling the dependencies of species on currencies. In the following, we illustrate these principal ways of modelling multispecies interactions via interaction matrices.

### Simple linkages between species

Many ecological studies are interested in discovering whether interactions occur between species in a community (Fig. 1b). This approach typically investigates communities of a single locality. Interactions can be summarized with a matrix $Z$ of size $S^2$, where $S$ is the number of species. If species $i$ interacts with species $j$ this is indicated by $Z_{ij} = 1$, and $Z_{ij} = 0$ indicates no interaction. Such binary information ($0 = \text{no interaction, } 1 = \text{interaction occurs}$) has played a key role in ecological network studies (Montoya et al., 2006; Bascompte & Jordano, 2007) where simple linkages between species in a community context have been described for food webs (e.g. $Z_{ij} = 1$ indicating consumption), mutualistic networks (e.g. $Z_{ij} = 1$ indicating flower visitation in a plant–pollinator network), or host–parasitoid networks (e.g. $Z_{ij} = 1$ indicating parasitism) (Ingts et al., 2009). Recent network studies have discovered new structural patterns in community-level interactions (Montoya et al., 2006; Bascompte & Jordano, 2007) that can help to simplify the parameterization of SIDMs (see section 'Ways to reduce complexity', below). However, no study has yet extended the essentially non-spatial network approach to actual species distribution data or modelled species' distributions or species' niches along large-scale environmental gradients. The review of Dale & Fortin (2011) considers some aspects of this challenge. For instance, network theory could be extended to analyse 'graph of graphs' where each node of a spatial graph contains a network of multispecies interactions (Dale & Fortin, 2011). Another starting point for extending the network approach to a spatial domain could be the 'niche model' of Williams & Martinez (2000) where diet breadth and food web interactions are modelled using simple assumptions along a one-dimensional trophic niche space (a resource axis). This model has been developed further to describe binary (0/1) interactions in a probabilistic way, which enables likelihood and Bayesian approaches to parameterize the model (Williams et al., 2010). Network models such as the niche model are usually parameterized with data from local food webs. However, this approach could be combined with data across environmental gradients and provides a route to predicting how species interact across large spatial extents.

### Interaction coefficients defined by interaction strengths

Moving beyond binary information on linkages between species, interaction coefficients can describe the strength of the effect of one species on another, and vice versa (Fig. 1c). Pairwise interactions have traditionally been represented as interaction coefficients describing the per capita effects of one species on the growth rate of a population of the other species, e.g. for the dynamics of two species competing by direct interference (Lotka, 1925; Volterra, 1926; May, 1973; Solé & Bascompte, 2006) or for modelling host–parasitoid interactions using discrete (rather than continuous) generation life cycles (Nicholson & Bailey, 1935). Interaction strengths for competitive interactions can also be specified by quantifying species’ relative positions along a resource axis and then calculating the overlap of their utilization curves (MacArthur & Levins, 1967; MacArthur, 1972, pp. 40–41). In ecological networks such as food webs, numerous measures of predator–prey interaction strength have been applied, ranging from simple linear Lotka–Volterra interaction coefficients to more complex (i.e. nonlinear) prey density-dependent coefficients and even multiple predator interference (see reviews by Berlow et al., 2004; Wootton & Emmerson, 2005). In mutualistic interactions such as plant–frugivore and plant–pollinator networks, interaction frequencies are often considered to be surrogates for interaction strengths (Vázquez et al., 2005; Scotti et al., 2007).

### Interactions mediated by interaction currencies

A third conceptual approach for modelling species interactions is to describe how interactions are mediated by interaction currencies (Fig. 1d). We use the term ‘interaction currency’ to include not only resources (which can be consumed, i.e. bionomic variables sensu Hutchinson, 1978), but also other non-consumable environmental conditions that mediate interactions (e.g. modulated environmental variables sensu Linder et al., in review). For instance, plants can alter temperature conditions beneath their canopies and planktonic organisms can alter pH, but none of these interaction currencies is actually consumed. Ideally, currencies should be chosen to capture the most important/limiting factors for the interacting species. In the classical resource–consumer framework, interaction currencies are resources and species interactions are modelled via their impacts on and requirements for these resources (May, 1973; Chase & Leibold, 2003). A major advantage is that the explicit description of species’ effects on and responses to interaction currencies can yield more realistic multispecies models. For instance, food web models that represent energetic and chemical elements as interaction currencies can ensure that thermodynamic and chemical
principles are obeyed, which is generally not the case for models based on pairwise interaction coefficients (Woodward et al., 2005). Moreover, interactions mediated via interaction currencies have also been widely implemented in forest models (Bugmann, 2001; Lischke et al., 2006; Rüger et al., 2008) and dynamic vegetation models (Hickler et al., 2004; Prentice et al., 2007), where inter- or intra-specific competition for light, space and water (and in some cases nitrogen) are modelled via empirical or physiology-based mathematical functions describing resource competition, growth and demographic processes of establishment and mortality.

EXAMPLES OF MODELLING MULTISPECIES INTERACTIONS WITH INTERACTION MATRICES USING SPECIES DISTRIBUTION DATA

We now explore how these interaction matrices may be applied to species distribution data. Possibly the simplest way to incorporate biotic interactions into classical SDMs is to use the presence of one species as a predictor for the focal species or restricting the model of one species to a model of the other (see above). However, this only models a one-way interaction (i.e. unidirectional rather than reciprocal interactions) (Fig. 1a) and does not generally estimate interaction coefficients between species (Fig. 1b–d). The modelled interaction will further be biased towards zero if the presence/absence of the predictor species is not known with certainty because extra noise added by misclassifications will reduce the observed correlation between the presences of the two species (Gustafson, 2004, Chapter 3). This approach suffers from at least two additional methodological problems: when a species is used as a predictor variable (e.g. prey or host plants) and a responding species (e.g. predators or herbivores) shares similar ecological or environmental niches, multicollinearity between predictor species and environmental variables can lead to problems. Then it is unclear whether the presence of the chosen predictor species or specific ecological conditions determine the focal species’ occurrence. Additionally, the range of suitable environmental conditions for the responding species may be outside the range of the predictor species. Classical SDMs would recognize such empty cells as ‘unsuitable environmental (climatic) conditions’ and not as ‘missing predictor species’. At least partially, such problems may be circumvented by explicitly including an interaction matrix in a simultaneous analysis of several species (Fig. 2). The precise way in which this is done will depend on the details of the ecological system, the model, and the data available. We outline two basic examples (Fig. 2). The first is based on static distribution data (Fig. 2a) and the second on temporal abundance dynamics (Fig. 2b). We also discuss challenges for modelling non-stationarity in interaction strengths across space and time.

**Static distributions modelled in the error matrix of multiple regression models**

Using a static snapshot of presence/absence patterns, multivariate logistic regressions can model species’ co-occurrences via residual co-variation in the error matrices (Fig. 2a; Ovaskainen et al., 2010; O’Hara & Zimmermann, in prep.). The easiest way to understand these models is to see them as threshold models: there is an unobserved latent variable, $y_i$, that represents the propensity for species $i$ to be present. If this latent variable has a value that exceeds a threshold, $t$, then the species is present (i.e. if $y_i > t$). The latent variable can then be modelled in the same way as in regression or generalized linear modelling (using a matrix $X$ with environmental variables and a vector $\beta$ with regression coefficients):

$$y_i = X_i \beta + \epsilon_i.$$

This is identical to a logistic regression if the residual error ($\epsilon_i$) follows a logistic distribution. When assuming that $\epsilon_i$ is normally distributed, then this model is identical to a generalized linear model with a probit link function.

![Figure 2](https://example.com/figure2.png)  
**Figure 2** Schematic representation of two methods for modelling multispecies co-occurrences using interaction matrices. (a) Static distributions (e.g. spatial presence/absence data) can be used to model species’ co-occurrences in the error matrices of multivariate logistic regressions, whereas (b) temporal dynamics (e.g. time-series of multispecies abundances) can be modelled with multivariate versions of population dynamic models. Both approaches rely on quantification of the interaction matrix, which describes pairwise interactions for multiple species pairs. See text for more details.
The advantage of the normal error approach is that it can be extended to more than one species (e.g. Ovaskainen et al., 2010; O’Hara & Zimmermann, in prep.). In this case, each species has its own latent variable and their errors, $e_i$, are correlated. For example, errors can be drawn from a multivariate normal distribution (MVN): $e_i \sim \text{MVN}(0, R)$, with $R$ being a correlation matrix with non-zero covariance (the scale of the latent variable is arbitrary, so its residual variance can be set to 1, and hence this covariance matrix is also the correlation matrix; Chib & Greenberg, 1998). Naively, this correlation is the interaction strength mentioned in the previous section (measured on the scale of the latent variable): a positive covariance means that if one species has a higher latent variable, the other is more likely to have a higher latent variable, so their joint probability of presence is higher.

Of course, a correlation between two species’ presences might not necessarily be the effect of a direct interaction. It may also be caused by a common response to an environmental variable (e.g. both species prefer a warm climate). If the responses of the latent variable for species 1 and 2 to an environmental gradient, $X$, are linear with slopes $\beta_1$ and $\beta_2$, then the covariance between the latent variables is $\beta_1\beta_2\text{Var}(X) + \rho_{12}$, where $\rho_{12}$ is the residual correlation of the latent variable (i.e. the correlation term in the matrix $R$). Thus, we can test whether apparent interactions between species are due to measured covariates (e.g. environmental variables) which allow us to estimate the effects of interactions through the correlation matrix $R$ while accounting for environmental variables (Fig. 2a). Yet it must always be remembered that there could be additional unobserved factors that jointly influence several species which would change the estimate of $\rho_{12}$.

This idea of modelling correlations can be extended to cases where the driver variable is unknown. Conceptually, this is the same approach as factor analysis (Manly, 2004; Chapter 7), which assumes that correlations between species result from a small number of unobserved latent variables, i.e. like covariates in a regression where the covariate is unobserved. This approach forms the basis of the joint modelling of densities of occurrences of species by Latimer et al. (2009) where single species models are built with the spatial variation in incidence quantified as a smooth surface representing the decay in correlation between pairs of points with distance. The co-occurrences between species can then be modelled by quantifying such surfaces for each species and by relating those to the other species’ presences through a spatial weight matrix of cross-covariance parameters (Latimer et al., 2009).

Statistical methods can also be used to model interactions mediated by resources or other interaction currencies (as described in the previous section) or for describing the effects of ecosystem engineers (see Linder et al., in review). If the interaction currency is known and measured, then it can be used as a covariate in the analysis, similar or additional to environmental covariates. However, even if it is not, a latent (i.e. unknown) interaction currency can be modelled: if species 1 produces (or requires) $\gamma$ of the interaction currency per individual and the latent variable of species 2 is increased (or decreased) by an amount $\delta$ per unit interaction currency. The covariance between species is then $\gamma \delta \sigma_i^2 + \rho$ (noting again that the latent variables have a residual variance of 1), where $\sigma_i^2$ is the total variance in the population size of species 1. If the interaction currency has not been measured, this model should be interpreted with caution as a supposed ‘currency’ may just represent the effect of an unmeasured environmental driver.

**Temporal dynamics modelled with multispecies time series**

Correlations between two species’ presences or abundances can be caused by several factors, which cannot be disentangled from spatial occurrence data alone, unless all drivers that affect species’ distributions are known a priori. However, if time series of multispecies co-occurrences (e.g. abundances) are available, the effects and relative importance of environmental stochasticity, environmental drivers, and inter-/intraspecific interactions can be estimated (Fig. 2b). Temporal fluctuations of species abundances have been modelled with multivariate versions of population dynamic models for single communities (Mutshinda et al., 2009, 2011), but could in principle be extended to whole range dynamics [for a discussion of range dynamics of a single species see Pagel & Schurr (2011) and Schurr et al. (in review)].

If the log-abundances $x$ of two species $i$ at time $t$ are $x_i(t)$ for $i = 1, 2$ and for simplicity we assume Gompertz dynamics (Mutshinda et al., 2009), then the interactions based on abundance data can be modelled as

$$
\begin{align*}
  x_1(t + 1) &= x_1(t) + r_1 \left( 1 - \frac{x_1(t) + \alpha_{12} x_2(t)}{K_1} \right) + e_1(t) \\
  x_2(t + 1) &= x_2(t) + r_2 \left( 1 - \frac{x_2(t) + \alpha_{21} x_1(t)}{K_2} \right) + e_2(t)
\end{align*}
$$

where $r_i$ and $K_i$ are the intrinsic growth rate and the carrying capacity of species $i$, respectively, and $e_i(t)$ is the environmental noise (Mutshinda et al., 2009, 2011). The interaction matrix described previously enters the model through the $\alpha$ terms. Interaction coefficients are usually asymmetric, i.e. in general $\alpha_{12} \neq \alpha_{21}$. Responses to a common environment will enter through effects on $e_1(t)$ and $e_2(t)$, which can be decomposed into the specific effects of these environmental factors plus any residual effects (Mutshinda et al., 2011).

An interaction currency (e.g. a common resource) can be included in the model, but it depends on the time scale of the interaction of the currency use and the species’ responses. If, for instance, the production and use of a resource is fast compared to the time scale of the population dynamics, then $\alpha_{12}$ will require a term $\gamma \delta$ (using the same notation and argument as above). In contrast, if resource production and use is slow, then a dynamic notation should be added for the resource, $c(t)$:

$$
c(t + 1) = c(t) + \gamma x_1(t) - \delta x_2(t) + e_3(t)
$$

which is equivalent to treating the resource as a species, but with an infinite carrying capacity. In general, the model...
outlined here captures the dynamics of a single multispecies
assemblage through effects on and responses to interaction
currencies. The model is essentially non-spatial and needs
extension to many sites in order to model species distributions
across large spatial extents.

With presence/absence data the model can be combined
with the latent variable approach above to model correlated
changes in species presences, i.e. \( x_j(t) \) is a latent variable. This
approach has been used to model the presence of waterbird
species breeding in artificial irrigation ponds (Sebastián-
González et al., 2010) where the presence/absence of each
species is attributed to a set of (environmental and spatial)
covariates and also to the presence of con- and heterospecifics
in the previous year and the presence of heterospecifics in the
same year (Fig. 2b). To describe meta-community dynamics
across large spatial extents, such descriptions of local community
dynamics have to be integrated with estimates of long-
distance dispersal (Schurr et al., in review). In this context, a
mechanistic understanding of dispersal is particularly valuable
currently most developed for plants, Nathan et al., 2008
which quantifies how biotic exchange between local communities
will depend on interspecific variation in functional traits
(Schurr et al., 2007) and environmental variation (Kuparinen
et al., 2009).

Varying interactions over space and time

Most models described above assume that the parameters (i.e.
interactions) are constant over the spatial and temporal scales
considered. However, it is reasonable to expect that interaction
effects vary in space (Schemske et al., 2009) and time (Olesen
et al., 2008; Petanidou et al., 2008). For instance, butterflies and
herbivorous insects can change food plants in space and time, competition between plants may vary across environ-
mental gradients, and pollinators might only be active during
particular seasons. The spatial and temporal resolution of data
thus becomes an important issue. In classical SDMs with fine
resolution data (e.g. vegetation plots), statistical parameters
have been added to describe the spatial and environmental
dependencies of unidirectional species interactions (e.g. Leath-
wick & Austin, 2001). Coarse resolutions (e.g. 10 \times 10 \text{ km grid}
cells) will introduce higher within-cell heterogeneity (both in
time and space), which adds noise to the biotic interaction
signal. Multispecies distribution data at fine spatial and temporal resolutions covering large spatial extents are there-
fore best suited for developing and testing SIDMs. With the
availability of sophisticated methods for rigorous statistical
inference the modelling of such data sets becomes increasingly
manageable (Hartig et al., in review; Marion et al., in review).
Alternatively, spatial scaling functions may be incorporated
that link interaction parameters and their environmental
dependencies between resolutions.

For regression models a number of methods have been
developed to describe spatial non-stationarity in parameter
estimates by the modelling of one or more smooth surfaces
over space (e.g. Fotheringham et al., 2002; Banerjee et al.,
2008; Finley, 2011; Hothorn et al., 2011). So far, these
approaches have only been used to incorporate the spatial
non-stationarity of environmental predictor variables (e.g.
Finley et al., 2009; Finley, 2011). In contrast, forest growth
and dynamic vegetation models explicitly simulate spatial
variation in the outcomes of competition, but these models are
not generally applicable to a large number of species from
different organism groups or to large spatial extents with fine
resolutions. Spatial and temporal non-stationarity of parameter
estimates describing species interactions will be an
important component for the future development of SIDMs
(Table 2).

WAYS TO REDUCE COMPLEXITY

Modelling large spatial and temporal data sets with multiple
interacting species and spatially (and maybe temporally)
varying parameter estimates and non-stationary covariance
structures will obviously produce considerable challenges for
statistical model fitting. For instance, the number of param-
eters in a pairwise interaction matrix increases as the square of
the number of species which can make it impractical to
estimate all parameters. A number of statistical methods exist
to reduce the complexity in the modelling (Table 2), including
approximations for spatial processes (e.g. Banerjee et al.,
2008), model selection procedures (Burnham & Anderson,
2002; O’Hara & Sillanpää, 2009), or shrinkage methods, which
simultaneously estimate parameters whilst shrinking the
unimportant ones towards zero (Tibshirani, 1996; Reineking
& Schröder, 2006). However, statistical methods are not
guaranteed to find a correct or biologically sensible answer. We
thus advocate making use of our ecological knowledge to
reduce the complexity of interaction matrices. We highlight a
number of ways (Fig. 3) how this can be achieved through
applying ecological knowledge and developing ecological
theory.

Using prior ecological knowledge to set interaction
coefficients to zero

One way to reduce the size of a full interaction matrix (Fig. 3a)
is to set some interaction coefficients to zero by using a priori
ecological knowledge about species interactions (Fig. 3b). Such
information may come from natural history knowledge (often
being published in the grey literature rather than the primary
literature) or from experiments, e.g. if some species are known
not to interact or if interactions are judged irrelevant for
determining a species’ distribution. For ecosystem engineers,
the massive impact of one or several species (‘modulators’
\textit{sensu} Linder et al., in review) on a range of other species may
be used to reduce the complexity of the full interaction matrix
by only including the unidirectional influence of ecosystem
engineers or modulators in the model (reciprocal effects can
usually be ignored). Another common approach is to use
community modules (Gilman et al., 2010) – groups of few
species known to interact strongly (e.g. a predator–prey
Table 2 Important components for the future development of approaches to modelling biotic interactions in multispecies assemblages across large spatial extents.

<table>
<thead>
<tr>
<th>Component</th>
<th>Related references (as a starting point)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical models</td>
<td>Latimer et al. (2009), Ovaskainen et al. (2010), O’Hara &amp; Zimmermann, in prep.</td>
</tr>
<tr>
<td>Models applying interaction matrices to temporal patterns of multispecies co-occurrences</td>
<td>Fotheringham et al. (2002), Banerjee et al. (2008), Fink et al. (2010), Hothorn et al. (2011)</td>
</tr>
<tr>
<td>Exploratory and predictive models for spatio-temporal data</td>
<td>Banerjee et al. (2008), Finley et al. (2009), Finley (2011), Hothorn et al. (2011)</td>
</tr>
<tr>
<td>Models incorporating spatial non-stationarity of parameters</td>
<td></td>
</tr>
<tr>
<td>Statistical methods (approximation of spatial processes, model selection, shrinkage methods, etc.)</td>
<td>Bascompte &amp; Jordano (2007), Olesen et al. (2007), Gilman et al. (2010), Linder et al., in review</td>
</tr>
<tr>
<td>A priori ecological knowledge (ecological networks, modules, apex consumers, ecosystem engineers etc.)</td>
<td>Simberloff &amp; Dayan (1991), Smith et al. (1997), Rüger et al. (2008), Blaum et al. (2011)</td>
</tr>
<tr>
<td>Functional group classifications (guilds, functional types, etc.)</td>
<td>Lavorel &amp; Garnier (2002), Eviner &amp; Chapin (2003), McGill et al. (2006), Suding et al. (2008)</td>
</tr>
<tr>
<td>Data collection and quantification of interaction patterns</td>
<td></td>
</tr>
<tr>
<td>Collecting long-term time-series of multispecies distribution and abundance patterns</td>
<td>Berlow et al. (2004), Wootton &amp; Emmerson (2005), Olesen et al. (2008), Schemske et al. (2009)</td>
</tr>
<tr>
<td>Quantifying how interaction strength and effectiveness vary with environment, space and time</td>
<td></td>
</tr>
</tbody>
</table>

Any inference must account for the fact that links might only be observed at certain time periods (e.g. seasonal dynamics; Olesen et al., 2008; Petanidou et al., 2008).

Modelling guilds and functional groups

A large body of ecological literature has focused on defining groups of species with similar ecological, physiological or morphological characteristics (Simberloff & Dayan, 1991; Smith et al., 1997). Often termed guilds, functional groups or functional types, such groups of species are a potentially useful way to reduce model complexity (Fig. 3c). This might be particularly helpful when investigating global change effects on biotic communities (Rüger et al., 2008; Küster et al., 2011). In animal studies, the guild concept has been applied to define groups of species that exploit the same class of environmental resources (e.g. food or suitable habitat) in a similar way (Simberloff & Dayan, 1991), e.g. for modelling environmental responses of bird species with similar dietary preferences across large spatial extents (Kissling et al., 2011). Recently, there have been suggestions for animal ecology to develop more comprehensive grouping schemes that consider traits other than those involved in resource acquisition (Blaum et al., 2011). In plant ecology, global scale vegetation patterns (at equilibrium) can be predicted from climate and soil characteristics assuming simple competition rules between such functional types (Prentice et al., 1992). Dynamic global vegetation models (DGVMs) and landscape-scale process-based forest models use...
a large number of physiological and demographic parameters to determine the outcome of competition for resources among plant functional types (Prentice et al., 2007; Rüger et al., 2008). Overall, functional group-based approaches to complexity reduction (Fig. 3c) seem to be particularly useful for modelling biotic interactions when data on responses of individual species are scarce, for example in species-rich tropical ecosystems or for studies with a global spatial extent. In most cases, these approaches are used within a trophic level but applications across trophic levels might also be possible. For instance, in plant–pollinator interactions flowers are classified as being pollinated by specific groups (guilds) of insects according to their morphology (‘pollination syndromes’, e.g. Faegri & van der Pijl, 1979). This could reduce the complexity of the interaction matrix, but often this does not mean that the pollinators of particular plant species can be predicted (Ollerton et al., 2009). An interesting avenue for future research could be to use functional groups defined by species’ position within interaction networks, e.g. by ordering species by their topological importance within the network (e.g. important connectors or generalist consumers, e.g. Montoya et al., 2009; Guimerá et al., 2010), by interaction strengths, or by trait-related network parameters (e.g. degree of size matching; Stang et al., 2009).

Modelling interaction currencies and effect and response traits

The classification of species into discrete categories (e.g. guilds, functional groups or plant functional types) is sensible if the traits, which are relevant for interactions, vary more between categories than within. However, it has been shown that plant species classified into the same functional group may show large variation in interaction-relevant traits (Wardle et al., 1998) and many plant species can have unique trait combinations (Evimer & Chapin, 2003). Discrete categorizations may then be of limited use if they miss a large proportion of interspecific variation or require an excessive number of categories. It has been suggested that categorical classification of communities should be replaced by continuous trait-based descriptions whenever possible (Evimer & Chapin, 2003; McGill et al., 2006). Continuous trait-based descriptions could...
reduce the dimensionality of the interaction matrix if species interactions can be described via their impacts on, and requirements for, shared currencies (Fig. 3d).

The development of such trait-based community models involves two conceptual steps. The first step simplifies the full interaction matrix into a matrix quantifying how each species affects interaction currencies, and a second matrix describing how each species responds to these currencies (Figs 1d & 3d). These interaction currencies are not only consumable resources (e.g. nutrients, prey) but also microclimatic conditions (light, temperature) or pollination and seed dispersal services (i.e. non-consumable factors that mediate interactions). The second step replaces species-specific effects and responses by quantitative traits, which describe how trait values affect a given interaction currency (‘effect traits’) and how in turn they respond to that currency (‘response traits’) (Fig. 3e; see also Lavorel & Garnier, 2002; McGill et al., 2006; Suding et al., 2008). For instance, plant competition might be mediated through seed mass, relative growth rate, leaf characteristics and plant height (Lavorel & Garnier, 2002), frugivory and seed dispersal might be affected by fruit size, energy content and body size (Schurr et al., 2009; Peters et al., 2010), and body size and stoichiometry may determine the strength of trophic interactions (Brown et al., 2004; Woodward et al., 2005). The complexity will be reduced when the number of interaction currencies (step 1) and the number of traits (step 2) are smaller than the number of species. Such trait-based approaches typically move the focus from the responses of single species to community-level changes in ecosystem structure and functioning, an approach that might not always be desirable (McGill et al., 2006).

CONCLUSIONS

Our synthesis suggests that there are promising new ways to model multispecies interactions using interaction matrices.
These novel approaches – collectively referred to as SIDMs – provide a general framework for modelling multispecies assemblages (summarized in Fig. 4). Although no single SIDM will fit all biotic interaction types across all spatial and temporal extents and grain sizes, a number of features emerge that are particularly important for the future development of such models (Table 2). We suggest the following areas as research priorities.

1. **Statistical modelling.** There is a general paucity of modelling studies estimating interaction coefficients for multispecies assemblages (interaction matrices) across multiple sites and/or multiple time-steps. The recent increase in ecological network studies is promising but most are non-spatial (i.e. come from a single location) and/or have no temporal dimension. An important step will be the development and testing of statistical models that allow spatial, temporal, and spatio-temporal multispecies co-occurrence data to be modelled. Estimating and quantifying the non-stationarity of interaction coefficients across space and time will become important here. Models explicitly including interaction currencies are largely lacking and their reliability and applicability needs to be tested.

2. **Methods for complexity reduction.** Ecological and statistical methods for complexity reduction are available and might be useful in particular circumstances (see above). However, it is unclear which ecological aspects of community-level dynamics and species interactions can be adequately captured if the complexity of the interaction matrix is reduced. For instance, the variability of effect and response traits and their relationships to interaction currencies across large-scale environmental and spatial gradients is largely unexplored. Additionally, complexity reduction through functional groupings based on species’ positions within ecological networks is a promising research avenue. Overall, complexity reduction approaches (both ecological and statistical) have not yet been widely tested or embedded into a statistical framework suitable for projecting large-scale multispecies assemblages under environmental change.

3. **Collecting comprehensive spatio-temporal data on multispecies systems.** There are few time-series of multispecies abundance dynamics available across large-scale environmental gradients, which limits our ability to infer biotic interactions. Most existing datasets are either purely spatial (e.g. no temporal but a large geographic extent) or purely temporal (e.g. time-series of multispecies abundances, but only from few local sites), and interaction types, strengths or currencies are usually not measured. Our review reinforces the importance of establishing and maintaining long-term and large-scale ecological monitoring programmes that measure abundances, interaction strengths, and interaction currencies for multiple species, locations and time steps. If possible, these monitoring programmes should be complemented by manipulative experiments that directly measure the same interactions. For instance, there is an urgent need to collect data on how interaction parameters of multiple species (e.g. presence/absence, strength, and effectiveness of interactions) vary with environmental conditions and along geographic gradients. Particularly needed are multispecies distribution data at fine spatial and temporal resolutions covering large spatial extents.

We are far from being able to accurately predict the consequences of global change for ecological communities and human well-being. Given the central role of biotic interactions in the origin and maintenance of biodiversity, we see a pressing need to develop novel models, which incorporate the effects of multispecies interactions into the projection of species distributions and community structure across large spatial extents. The methods and studies illustrated in our synthesis suggest that realistic progress in the development of SIDMs can be made in the coming decade and we have highlighted where important steps forward can be made.

**ACKNOWLEDGEMENTS**

This work was initiated by two workshops on ‘The ecological niche as a window to biodiversity’, organized by Steven Higgins, Bob O’Hara and Christine Römermann and funded by the LOEWE initiative for scientific and economic excellence of the German federal state of Hesse. We thank Daniel Montoya, Catherine Graham, Peter Linder, Boris Schröder, two anonymous referees and workshop participants for comments and discussion. Our work is supported by The Danish Council for Independent Research | Natural Sciences (Steno stipend to W.D.K.), the Villum K. Rasmussen Foundation (grantVKR09b-141 to J.-C.S.), the European Union (IEF Marie Curie Fellowship 252811 to K.S., Marie Curie Outgoing International Fellowship MOIF-CT-2006-40571 to J.G., and GOCE-CT-2007-036866 and ENV-CT-2009-226544 to N.E.Z.), the German Research Foundation (DFG (grants RO 3842/1-1 to C.R. and SCHU 2259/3-1 to F.M.S.), and the research programme ‘LOEWE – Landes-Offensive zur Entwicklung Wissenschaftlich-ökonomischer Exzellenz’ of Hesse’s Ministry of Higher Education, Research, and the Arts, Germany (R.B.O’H. and C.R.). J.M.M. is supported by the McyI (Ramon y Cajal Fellowship RYC-2008-03664) and the Generalitat de Catalunya.

**REFERENCES**


Lotka, A.J. (1925) *Elements of physical biology*. Williams and Wilkins, Baltimore, MD.


Pearson, R.G. & Dawson, T.P. (2003) Predicting the impacts of climate change on the distribution of species: are bioclimate


**BIOSKETCH**

W. Daniel Kissling is interested in biotic interactions, animal habitat and space use, macroecology and biogeography, and global change. Research interests of the co-authors include global change biology, community ecology, plant ecology, ecological networks, conservation biogeography, dynamic vegetation models, ecological modelling, and statistical and computational methods for modelling species distributions.

Author contributions: A core team (W.D.K., B.O’H., G.J.M., A.S., K.S., J.G., F.M.S. and N.E.Z.) conceived the principal ideas and the main structure of the manuscript, W.D.K. led the writing, and all authors commented on the manuscript and provided additional ideas and input.

Editor: Peter Linder

The papers in this Special Issue arose from two workshops entitled 'The ecological niche as a window to biodiversity' held on 26–30 July 2010 and 24–27 January 2011 in Arnoldshain near Frankfurt, Germany. The workshops combined recent advances in our empirical and theoretical understanding of the niche with advances in statistical modelling, with the aim of developing a more mechanistic theory of the niche. Funding for the workshops was provided by the Biodiversity and Climate Research Centre (BiK-F), which is part of the LOEWE programme 'Landes-Offensive zur Entwicklung Wissenschaftlich-ökonomischer Exzellenz' of Hesse’s Ministry of Higher Education, Research and the Arts.