SPECIAL ISSUE

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

W. D. Kissling[1]*, Carsten F. Dormann[2,3], Jürgen Groeneveld[4,5], Thomas Hickler[6], Ingolf Kühn[7], Greg J. McInerny[8], José M. Montoya[9], Christine Römermann[10,11], Katja Schiffers[12], Frank M. Schurr[10,13], Alexander Singer[4], Jens-Christian Svenning[1], Niklaus E. Zimmermann[14] and Robert B. O'Hara[6]

[1]*Ecoinformatics & Biodiversity Group, Department of Bioscience, Aarhus University, DK-8000 Aarhus C, Denmark,* [2]*Biometry and Environmental System Analysis, Faculty of Forest and Environmental Sciences, University of Freiburg, 79106 Freiburg, Germany,* [3]*Helmholtz Centre for Environmental Research – UFZ, Department of Computational Landscape Ecology, 04318 Leipzig, Germany,* [4]*Helmholtz Centre for Environmental Research – UFZ, Department of Ecological Modelling, 04318 Leipzig, Germany,* [5]*School of Environment, The University of Auckland, Auckland, New Zealand,* [6]*Biodiversity and Climate Research Centre (BiK-F), 60325 Frankfurt am Main, Germany,* [7]*Helmholtz Centre for Environmental Research – UFZ, Department of Community Ecology, 06120 Halle, Germany,* [8]*Computational Ecology and Environmental Science Group, Computational Science Laboratory, Microsoft Research, Cambridge CB3 0FB, UK,* [9]*Instituto de Ciencias del Mar, Consejo Superior de Investigaciones Científicas, E-08003 Barcelona, Spain,* [10]*Institute for Physical Geography, Goethe University Frankfurt, 60438 Frankfurt am Main, Germany,* [11]*Theoretical Ecology, Faculty of Biology and Preclinical Medicine, University of Regensburg, 93040 Regensburg, Germany,* [12]*Laboratoire d'Ecologie Alpine, UMR-CNRS 5553, Université J. Fourier, 38041 Grenoble Cedex 9, France,* [13]*Plant Ecology and Nature Conservation, Institute of Biochemistry and Biology, University of Potsdam, 14469 Potsdam, Germany,* [14]*Landscape Dynamics, Swiss Federal Research Institute WSL, CH-8903 Birmensdorf, Switzerland*

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.

Journal of Biogeography

W. D. Kissling *et al.*

*Correspondence: W. Daniel Kissling,
Ecoinformatics & Biodiversity Group,
Department of Bioscience, Aarhus University,
Ny Munkegade 114, DK-8000 Aarhus C,
Denmark.
E-mail: danielkissling@web.de.

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.

Model type	Biotic interaction	Spatial extent	Implementation of interaction	Data	Reference(s)
Species distribution models	Competition, facilitation, trophic interactions	Regional to global	Additional predictors in statistical models or constraining model predictions to the presence of interacting species	Frequency or count data, basal area, proportional abundance, presence/absence, presence-only (all georeferenced)	Leathwick & Austin (2001), Rouget et al. (2001), Araújo & Luoto (2007), Heikkinen et al. (2007), Schweiger et al. (2008, 2012), Meier et al. (2010, 2011a)
Lotka–Volterra models	Competition, predator–prey, host–parasitoid	Local	Differential equations with interaction coefficients	Density data, simulations	Lotka (1925), Volterra (1926), Nicholson & Bailey (1935), May (1973), Solé & Bascompte (2006)
Zone of influence models	Plant competition	Local	Modelling overlap of individuals' areas of influence	Physiological data of resource uptake	Gates & Westcott (1978), Schiffers et al. (2011)
Population dynamic models	Competition, host–parasitoid	Local to regional	Discrete-time stochastic Gompertz model, coupled-map lattice model	Abundance data, e.g. time-series from trapping, surveys and counts	Mutshinda et al. (2009, 2011)
Epidemiological models	Infectious diseases	Local to global	Wavelet time series analysis, individual-based models, stage-based (susceptible, infected, recovered, SIR) models	Frequency of outbreaks, abundance, transmission rates, demographic rates, behavioural data, etc.	Grenfell et al. (2001), Keeling et al. (2001)
Multivariate regression models	Competition, facilitation, mutualism	Local to regional	Spatial associations in residuals of regression models (cross-covariance matrices)	Presence/absence, abundance	Banerjee et al. (2008), Latimer et al. (2009), Ovaskainen et al. (2010), Sebastián-González et al. (2010)
Dynamic vegetation models	Plant competition	Landscape to global	Simulation of individuals, cohorts or populations competing for resources (e.g. space, light or water)	Mathematical functions for growth, mortality, establishment etc.	Lischke et al. (2006), Prentice et al. (2007), Rüger et al. (2008)
Ecological networks	Mutualistic interactions, food webs, host–parasitoid	Local	Analysis of linkages between species, network properties	Presence/absence of feeding links, per capita interaction strengths, visitation frequencies	Berlow et al. (2004), Wootton & Emmerson (2005), Scotti et al. (2007), Ings et al. (2009)

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
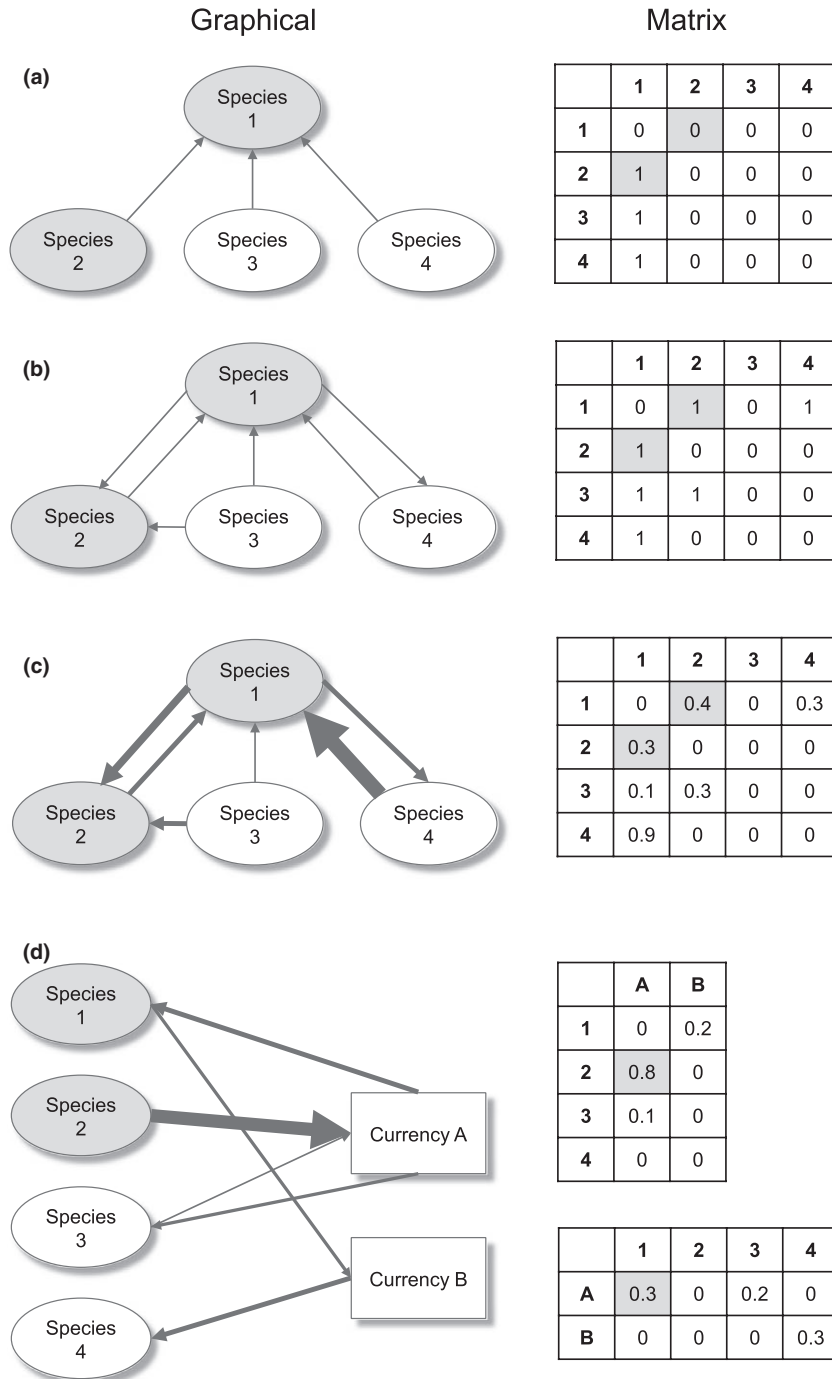
Figure 1 Graphical representations (left) and matrix representation (right) of interactions between species. (a) Unidirectional, i.e. one species is affected by others (as typically implemented in extensions of classical species distribution models), (b) multispecies interactions quantified by simple linkages between species pairs, (c) multispecies interactions where the strength of interaction is included (interaction coefficients), and (d) 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 $\mathbf{Z}$ of size S^2, where S is the number of species. If species i interacts with species j this is indicated by $\mathbf{Z}_{ij} = 1$, and $\mathbf{Z}_{ij} = 0$ indicates no interaction. Such binary information (0 = no interaction, 1 = 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. $\mathbf{Z}_{ij} = 1$ indicating consumption), mutualistic networks (e.g. $\mathbf{Z}_{ij} = 1$ indicating flower visitation in a plant–pollinator network), or host–parasitoid networks (e.g. $\mathbf{Z}_{ij} = 1$ indicating parasitism) (Ings 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 $\mathbf{X}$ with environmental variables and a vector β with regression coefficients):

$$y_i = \mathbf{X}_i\beta + \varepsilon_i.$$

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

(a) Static distributions

(b) Temporal dynamics
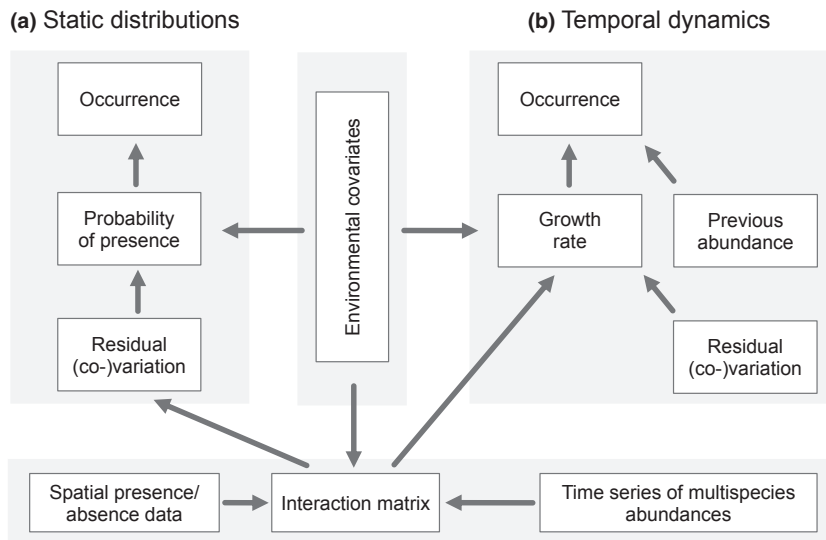
Figure 2 Schematic representation of two methods for modelling multispecies co-occurrences using interaction matrices. (a) Static distributions (e.g. spatial presence/absence data) can be used to model species' co-occurrences in the error matrices of multivariate logistic regressions, whereas (b) temporal 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, ε_i, are correlated. For example, errors can be drawn from a multivariate normal distribution (MVN): $\varepsilon_i \sim \mathrm{MVN}(0, \mathbf{R})$, with $\mathbf{R}$ 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). Naïvely, 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 driver 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 β_1 and β_2, then the covariance between the latent variables is $\beta_1 \beta_2 \mathrm{Var}(X) + \rho_{12}$, where ρ_{12} is the residual correlation of the latent variable (i.e. the correlation term in the matrix $\mathbf{R}$). 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 $\mathbf{R}$ 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 ρ_{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) γ of the interaction currency per individual and the latent variable of species 2 is increased (or decreased) by an amount δ per unit interaction currency. The covariance between species is then $\gamma\delta\sigma_1^2 + \rho$ (noting again that the latent variables have a residual variance of 1), where σ_1^2 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

$$x_1(t+1) = x_1(t) + r_1\left(1 - \frac{x_1(t) + \alpha_{12}x_2(t)}{K_1}\right) + \varepsilon_1(t)$$

$$x_2(t+1) = x_2(t) + r_2\left(1 - \frac{x_2(t) + \alpha_{21}x_1(t)}{K_2}\right) + \varepsilon_2(t)$$

where r_i and K_i are the intrinsic growth rate and the carrying capacity of species i, respectively, and $\varepsilon_i(t)$ is the environmental noise (Mutshinda *et al.*, 2009, 2011). The interaction matrix described previously enters the model through the α 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 $\varepsilon_1(t)$ and $\varepsilon_2(t)$, 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 α_{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, $c(t)$:

$$c(t+1) = c(t) + \gamma x_1(t) - \delta x_2(t) + \varepsilon_3(t)$$

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. $x_i(t)$ 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 environmental 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. Leathwick & Austin, 2001). Coarse resolutions (e.g. 10×10 km grid cells) will introduce higher within-cell heterogeneity (both in space and time), which adds noise to the biotic interaction signal. Multispecies distribution data at fine spatial and temporal resolutions covering large spatial extents are therefore 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 parameters 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' *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.

Component	Related references (as a starting point)
Statistical models	
Models for spatial multispecies co-occurrence patterns	Latimer *et al.* (2009), Ovaskainen *et al.* (2010), O'Hara & Zimmermann, in prep.
Models applying interaction matrices to temporal patterns of multispecies co-occurrences	Mutshinda *et al.* (2009, 2011), Sebastián-González *et al.* (2010)
Exploratory and predictive models for spatio-temporal data	Fotheringham *et al.* (2002), Banerjee *et al.* (2008), Fink *et al.* (2010), Hothorn *et al.* (2011)
Models incorporating spatial non-stationarity of parameters	Banerjee *et al.* (2008), Finley *et al.* (2009), Finley (2011), Hothorn *et al.* (2011)
Methods for complexity reduction	
Statistical methods (approximation of spatial processes, model selection, shrinkage methods, etc.)	Tibshirani (1996), Burnham & Anderson (2002), Reineking & Schröder (2006), Banerjee *et al.* (2008), O'Hara & Sillanpää (2009)
A priori ecological knowledge (ecological networks, modules, apex consumers, ecosystem engineers etc.)	Bascompte & Jordano (2007), Olesen *et al.* (2007), Gilman *et al.* (2010), Linder *et al.*, in review
Functional group classifications (guilds, functional types, etc.)	Simberloff & Dayan (1991), Smith *et al.* (1997), Rüger *et al.* (2008), Blaum *et al.* (2011)
Interaction currencies and effect and response traits (trait-based community models)	Lavorel & Garnier (2002), Eviner & Chapin (2003), McGill *et al.* (2006), Suding *et al.* (2008)
Data collection and quantification of interaction patterns	
Collecting long-term time-series of multispecies distribution and abundance patterns	Woiwod & Harrington (1994), Sauer *et al.* (2007)
Quantifying how interaction strength and effectiveness vary with environment, space and time	Berlow *et al.* (2004), Wootton & Emmerson (2005), Olesen *et al.* (2008), Schemske *et al.* (2009)

module composed of a lion and medium- to large-sized ungulates). Some community modules are particularly well studied, for instance in food webs (Amarasekare, 2008). Ecological networks may also hold useful information that could reduce the complexity of the interaction matrix. For instance, many networks are: (1) very heterogeneous (the bulk of the species have few interactions, but a few species are much more connected than expected by chance), (2) nested (specialists have a subset of generalists' interactions) and compartmentalized (some groups of species interact more among them than with the rest of the species), and (3) built on weak and asymmetric links among species (Montoya *et al.*, 2006; Bascompte & Jordano, 2007). This knowledge may be useful to reduce the number of interactions that have to be considered in models. For instance, the connectance of empirical food webs (i.e. the number of realized links, L, divided by the total number of possible links, S^2) decreases hyperbolically with the number of species S (Montoya & Solé, 2003). Thus, for 100 species (for example) the connectance is around 0.1 and only 10% of possible interactions are realized. Other examples include searching for forbidden links (i.e. potential links among species that are not realized; Olesen *et al.*, 2011), asymmetrical links (strong symmetrical links are more or less absent in food webs and mutualistic networks, except where the interaction is very intimate, e.g. plant–ant domatia relationships; Bascompte *et al.*, 2006), or modularity (modules consist of strongly connected species which are weakly interlinked with other modules; Olesen *et al.*, 2007).

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
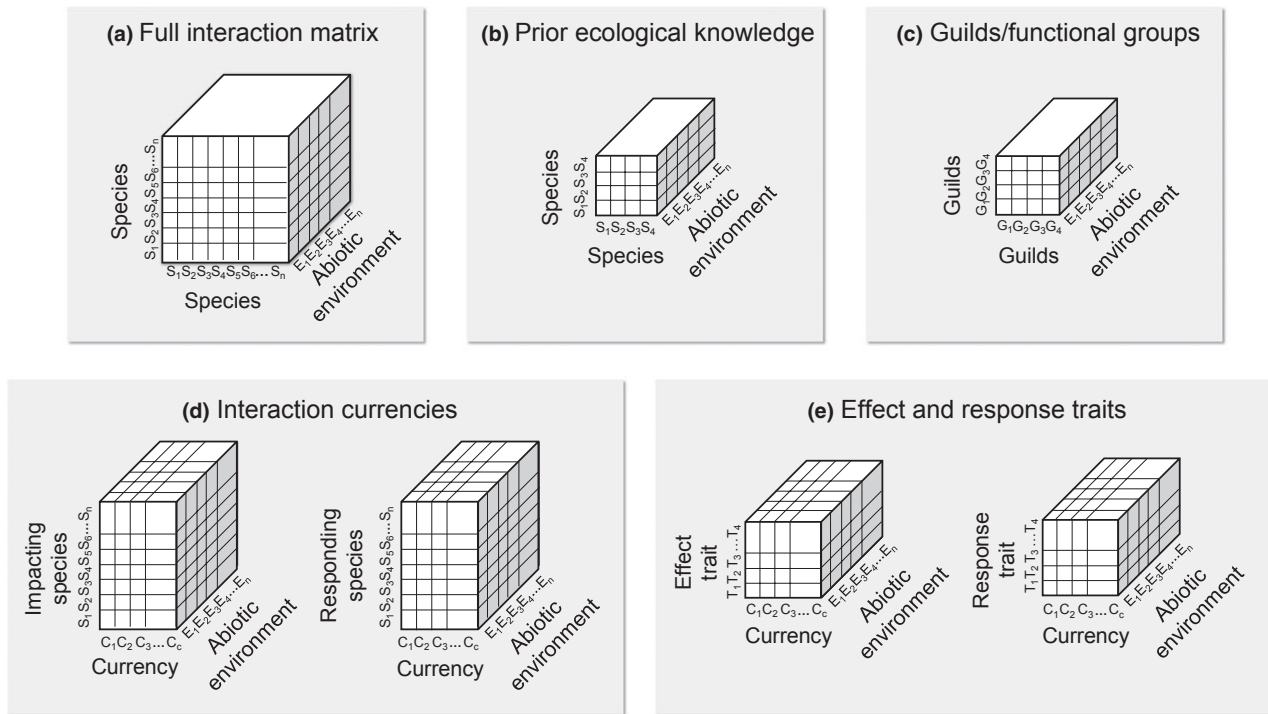
Figure 3 Simplified schematic examples for reducing the complexity and dimensionality when modelling biotic interactions in variable environments. (a) A full interaction matrix with S^2 species quantifies all possible species-by-species interaction coefficients and how they vary as a function of e environmental conditions. (b) Based on a priori ecological knowledge the number of interacting species is smaller than in the full interaction matrix because some species are known not to interact (from field studies, experiments, ecological networks, or natural history information). (c) Aggregating species into groups with similar biological, ecological, physiological or morphological characteristics allows the modelling of guilds, functional groups, or functional types. (d) Modelling species' impacts and responses to interaction currencies (e.g. consumable resources, microclimatic conditions, or pollination and seed dispersal services) can reduce the complexity if the number of interaction currencies is smaller than the number of species. (e) The impacts and responses of species as derived from modelling continuous species' effect and response traits. Here, quantitative traits (rather than species) are used to describe the responses to, and impacts on, specific currencies that mediate interactions.

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 (Eviner & 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 (Eviner & 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.
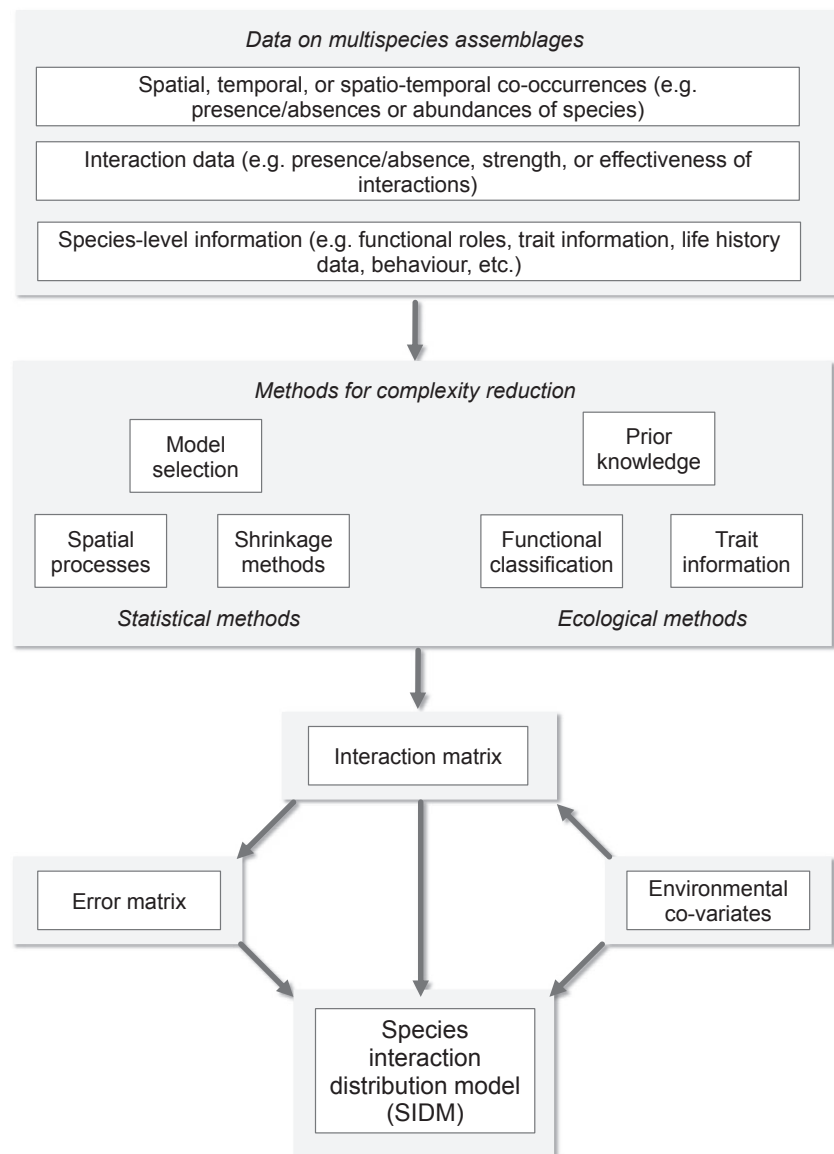
Figure 4 A general framework for implementing species interaction distribution models (SIDMs). Data on multispecies assemblages are combined with methods for complexity reduction to estimate the interaction matrix. The interaction matrix together with environmental covariates and the error matrix are used to specify the final SIDM.

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 nor embedded into a statistical framework suitable for projecting large-scale multispecies assemblages under environmental change.

3. *Collating 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

REFERENCES

Amarasekare, P. (2008) Spatial dynamics of foodwebs. *Annual Review of Ecology, Evolution, and Systematics*, **39,** 479–500.

Araújo, M.B. & Luoto, M. (2007) The importance of biotic interactions for modelling species distributions under climate change. *Global Ecology and Biogeography*, **16,** 743–753.

Banerjee, S., Gelfand, A.E., Finley, A.O. & Sang, H. (2008) Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **70,** 825–848.

Bascompte, J. & Jordano, P. (2007) Plant–animal mutualistic networks: the architecture of biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, **38,** 567–593.

Bascompte, J., Jordano, P. & Olesen, J.M. (2006) Asymmetric coevolutionary networks facilitate biodiversity maintenance. *Science*, **312,** 431–433.

Berlow, E.L., Neutel, A.-M., Cohen, J.E., de Ruiter, P.C., Ebenman, B., Emmerson, M., Fox, J.W., Jansen, V.A.A., Jones, J.I., Kokkoris, G.D., Logofet, D.O., McKane, A.J., Montoya, J.M. & Petchey, O. (2004) Interaction strengths in food webs: issues and opportunities. *Journal of Animal Ecology*, **73**, 585–598.

Blaum, N., Mosner, E., Schwager, M. & Jeltsch, F. (2011) How functional is functional? Ecological groupings in terrestrial animal ecology: towards an animal functional type approach. *Biodiversity and Conservation*, **20**, 2333–2345.

Brown, J.H., Gillooly, J.F., Allen, A.P., Savage, V.M. & West, G.B. (2004) Toward a metabolic theory of ecology. *Ecology*, **85**, 1771–1789.

Bugmann, H. (2001) A review of forest gap models. *Climatic Change*, **51**, 259–305.

Burnham, K.P. & Anderson, D.R. (2002) *Model selection and multimodel inference: a practical information-theoretic approach*, 2nd edn. Springer, New York.

Chase, J.M. & Leibold, M.A. (2003) *Ecological niches: linking classical and contemporary approaches*. University of Chicago Press, Chicago, IL.

Chib, S. & Greenberg, E. (1998) Analysis of multivariate probit models. *Biometrika*, **85**, 347–361.

Dale, M.R.T. & Fortin, M.-J. (2011) From graphs to spatial graphs. *Annual Review of Ecology, Evolution, and Systematics*, **41**, 21–38.

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.

Elith, J., Graham, C.H., Anderson, R.P. *et al.* (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.

Eviner, V.T. & Chapin, F.S., III (2003) Functional matrix: a conceptual framework for predicting multiple plant effects on ecosystem processes. *Annual Review of Ecology, Evolution, and Systematics*, **34**, 455–485.

Faegri, K. & van der Pijl, L. (1979) *The principles of pollination biology*. Pergamon Press, Oxford.

Fink, D., Hochachka, W.M., Zuckerberg, B., Winkler, D.W., Shaby, B., Munson, M.A., Hooker, G., Riedewald, M., Sheldon, D. & Kelling, S. (2010) Spatiotemporal exploratory models for broad-scale survey data. *Ecological Applications*, **20**, 2131–2147.

Finley, A.O. (2011) Comparing spatially-varying coefficients models for analysis of ecological data with non-stationary and anisotropic residual dependence. *Methods in Ecology and Evolution*, **2**, 143–154.

Finley, A.O., Banerjee, S. & McRoberts, R.E. (2009) Hierarchical spatial models for predicting tree species assemblages across large domains. *The Annals of Applied Statistics*, **3**, 1052–1079.

Fotheringham, A.S., Brunsdon, C. & Charlton, M. (2002) *Geographically weighted regression: the analysis of spatially varying relationships*. John Wiley & Sons, Chichester.

Gates, D.J. & Westcott, M. (1978) Zone of influence models for competition in plantations. *Advances in Applied Probability*, **10**, 499–537.

Gilman, S.E., Urban, M.C., Tewksbury, J., Gilchrist, G.W. & Holt, R.D. (2010) A framework for community interactions under climate change. *Trends in Ecology and Evolution*, **25**, 325–331.

Grenfell, B.T., Bjørnstad, O.N. & Kappey, J. (2001) Travelling waves and spatial hierarchies in measles epidemics. *Nature*, **414**, 716–723.

Guimerá, R., Stouffer, D.B., Sales-Pardo, M., Leicht, E.A., Newman, M.E.J. & Amaral, L.A.N. (2010) Origin of compartmentalization in food webs. *Ecology*, **91**, 2941–2951.

Guisan, A. & Thuiller, W. (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, **8**, 993–1009.

Gustafson, P. (2004) *Measurement error and misclassification in statistics and epidemiology*. Chapman & Hall/CRC, Boca Raton, FL.

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.

Hickler, T., Smith, B., Sykes, M.T., Davis, M.B., Sugita, S. & Walker, K. (2004) Using a generalized vegetation model to simulate vegetation dynamics in northeastern USA. *Ecology*, **85**, 519–530.

Hickler, T., Vohland, K., Feehan, J., Miller, P., Smith, B., Costa, L., Giesecke, T., Fronzek, S., Carter, T., Cramer, W., Kühn, I. & Sykes, M. (2012) Projecting the future distribution of European potential natural vegetation zones with a generalized, tree species-based dynamic vegetation model. *Global Ecology and Biogeography*, **21**, 50–63.

Holt, R.D. (2009) Bringing the Hutchinsonian niche into the 21st century: ecological and evolutionary perspectives. *Proceedings of the National Academy of Sciences USA*, **106**, 19,659–19,665.

Hothorn, T., Müller, J., Schröder, B., Kneib, T. & Brandl, R. (2011) Decomposing environmental, spatial, and spatio-temporal components of species distributions. *Ecological Monographs*, **81**, 329–347.

Hutchinson, G.E. (1957) Population studies: animal ecology and demography. Concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, **22**, 415–427.

Hutchinson, G.E. (1978) *An introduction to population biology*. Yale University Press, New Haven, CT.

Ings, T.C., Montoya, J.M., Bascompte, J., Blüthgen, N., Brown, L., Dormann, C.F., Edwards, F., Figueroa, D., Jacob, U., Jones, J.I., Lauridsen, R.B., Ledger, M.E., Lewis, H.M., Olesen, J.M., van Veen, F.J.F., Warren, P.H. & Woodward, G. (2009) Ecological networks – beyond food webs. *Journal of Animal Ecology*, **78**, 253–269.

Keeling, M.J., Woolhouse, M.E.J., Shaw, D.J., Matthews, L., Chase-Topping, M., Haydon, D.T., Cornell, S.J., Kappey, J., Wilesmith, J. & Grenfell, B.T. (2001) Dynamics of the 2001

UK foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. *Science*, **294,** 813–817.

Kissling, W.D., Rahbek, C. & Böhning-Gaese, K. (2007) Food plant diversity as broad-scale determinant of avian frugivore richness. *Proceedings of the Royal Society B: Biological Sciences*, **274,** 799–808.

Kissling, W.D., Field, R., Korntheuer, H., Heyder, U. & Böhning-Gaese, K. (2010) Woody plants and the prediction of climate-change impacts on bird diversity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **365,** 2035–2045.

Kissling, W.D., Sekercioglu, C.H. & Jetz, W. (2011) Bird dietary guild richness across latitudes, environments and biogeographic regions. *Global Ecology and Biogeography*, doi: 10.1111/j.1466-8238.2011.00679.x.

Kuparinen, A., Katul, G., Nathan, R. & Schurr, F.M. (2009) Increases in air temperature can promote wind-driven dispersal and spread of plants. *Proceedings of the Royal Society B: Biological Sciences*, **276,** 3081–3087.

Küster, E.C., Bierman, S.M., Klotz, S. & Kühn, I. (2011) Modelling the impact of climate and land use change on the geographical distribution of leaf anatomy in a temperate flora. *Ecography*, **34,** 507–518.

Latimer, A.M., Banerjee, S., Sang, H., Mosher, E.S. & Silander, J.A., Jr (2009) Hierarchical models facilitate spatial analysis of large data sets: a case study on invasive plant species in the northeastern United States. *Ecology Letters*, **12,** 144–154.

Lavorel, S. & Garnier, E. (2002) Predicting changes in community composition and ecosystem functioning from plant traits: revisiting the Holy Grail. *Functional Ecology*, **16,** 545–556.

Leathwick, J.R. & Austin, M.P. (2001) Competitive interactions between tree species in New Zealand's old-growth indigenous forests. *Ecology*, **82,** 2560–2573.

Lischke, H., Zimmermann, N.E., Bolliger, J., Rickebusch, S. & Löffler, T.J. (2006) TreeMig: a forest-landscape model for simulating spatio-temporal patterns from stand to landscape scale. *Ecological Modelling*, **199,** 409–420.

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

MacArthur, R.H. (1972) *Geographical ecology*. Princeton University Press, Princeton, NJ.

MacArthur, R.H. & Levins, R. (1967) Limiting similarity, convergence, and divergence of coexisting species. *The American Naturalist*, **101,** 377–385.

Manly, B.F.G. (2004) *Multivariate statistical methods: a primer*, 3rd edn. Chapman & Hall/CRC, Boca Raton, FL.

May, R.M. (1973) *Stability and complexity in model ecosystems*. Princeton University Press, Princeton, NJ.

McGill, B.J., Enquist, B.J., Weiher, E. & Westoby, M. (2006) Rebuilding community ecology from functional traits. *Trends in Ecology and Evolution*, **21,** 178–185.

Meier, E.S., Kienast, F., Pearman, P.B., Svenning, J.-C., Thuiller, W., Araújo, M.B., Guisan, A. & Zimmermann, N.E. (2010) Biotic and abiotic variables show little redundancy in explaining tree species distributions. *Ecography*, **33,** 1038–1048.

Meier, E.S., Edwards, T.C., Jr, Kienast, F., Dobbertin, M. & Zimmermann, N.E. (2011a) 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,** 371–382.

Meier, E.S., Lischke, H., Schmatz, D.R. & Zimmermann, N.E. (2011b) Climate, competition and connectivity affect future migration and ranges of European trees. *Global Ecology and Biogeography*, doi: 10.1111/j.1466-8238.2011.00669.x.

Montoya, J.M. & Solé, R.V. (2003) Topological properties of food webs: from real data to community assembly models. *Oikos*, **102,** 614–622.

Montoya, J.M., Pimm, S.L. & Solé, R.V. (2006) Ecological networks and their fragility. *Nature*, **442,** 259–264.

Montoya, J.M., Woodward, G., Emmerson, M.E. & Solé, R.V. (2009) Press perturbations and indirect effects in real food webs. *Ecology*, **90,** 2426–2433.

Mutshinda, C.M., O'Hara, R.B. & Woiwod, I.P. (2009) What drives community dynamics? *Proceedings of the Royal Society B: Biological Sciences*, **276,** 2923–2929.

Mutshinda, C.M., O'Hara, R.B. & Woiwod, I.P. (2011) A multispecies perspective on ecological impacts of climatic forcing. *Journal of Animal Ecology*, **80,** 101–107.

Nathan, R., Schurr, F.M., Spiegel, O., Steinitz, O., Trakhtenbrot, A. & Tsoar, A. (2008) Mechanisms of long-distance seed dispersal. *Trends in Ecology and Evolution*, **23,** 638–647.

Nicholson, A.J. & Bailey, V.A. (1935) The balance of animal populations.—Part I. *Proceedings of the Zoological Society of London*, **105,** 551–598.

O'Hara, R.B. & Sillanpää, M.J. (2009) A review of Bayesian variable selection methods: what, how and which. *Bayesian Analysis*, **4,** 85–118.

Olesen, J.M., Bascompte, J., Dupont, Y.L. & Jordano, P. (2007) The modularity of pollination networks. *Proceedings of the National Academy of Sciences of the USA*, **104,** 19,891–19,896.

Olesen, J.M., Bascompte, J., Elberling, H. & Jordano, P. (2008) Temporal dynamics in a pollination network. *Ecology*, **89,** 1573–1582.

Olesen, J.M., Bascompte, J., Dupont, Y.L., Elberling, H., Rasmussen, C. & Jordano, P. (2011) Missing and forbidden links in mutualistic networks. *Proceedings of the Royal Society B: Biological Sciences*, **278,** 725–732.

Ollerton, J., Alarcón, R., Waser, N.M., Price, M.V., Watts, S., Cranmer, L., Hingston, A., Peter, C.I. & Rotenberry, J. (2009) A global test of the pollination syndrome hypothesis. *Annals of Botany*, **103,** 1471–1480.

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.

Pagel, J. & Schurr, F.M. (2011) Forecasting species ranges by statistical estimation of ecological niches and spatial population dynamics. *Global Ecology and Biogeography*, doi: 10.1111/j.1466-8238.2011.00663.x.

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

envelope models useful? *Global Ecology and Biogeography*, **12**, 361–371.

Petanidou, T., Kallimanis, A.S., Tzanopoulos, J., Sgardelis, S.P. & Pantis, J.D. (2008) Long-term observation of a pollination network: fluctuation in species and interactions, relative invariance of network structure and implications for estimates of specialization. *Ecology Letters*, **11**, 564–575.

Peters, V.E., Mordecai, R., Carroll, C.R., Cooper, R.J. & Greenberg, R. (2010) Bird community response to fruit energy. *Journal of Animal Ecology*, **79**, 824–835.

Prentice, I.C., Cramer, W., Harrison, S.P., Leemans, R., Monserud, R.A. & Solomon, A.M. (1992) A global biome model based on plant physiology and dominance, soil properties and climate. *Journal of Biogeography*, **19**, 117–134.

Prentice, I.C., Bondeau, A., Cramer, W., Harrison, S.P., Hickler, T., Lucht, W., Sitch, S., Smith, B. & Sykes, M.T. (2007) Dynamic global vegetation modelling: quantifying terrestrial ecosystem responses to large-scale environmental change. *Terrestrial ecosystems in a changing world* (ed. by J.G. Canadell, D. Pataki and L.F. Pitelka), pp. 175–192. Springer, Berlin.

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.

Reineking, B. & Schröder, B. (2006) Constrain to perform: regularization of habitat models. *Ecological Modelling*, **193**, 675–690.

Rouget, M., Richardson, D.M., Lavorel, S., Vayreda, J., Gracia, C. & Milton, S.J. (2001) Determinants of distribution of six *Pinus* species in Catalonia, Spain. *Journal of Vegetation Science*, **12**, 491–502.

Rüger, N., Williams-Linera, G., Kissling, W.D. & Huth, A. (2008) Long-term impacts of fuelwood extraction on a tropical montane cloud forest. *Ecosystems*, **11**, 868–881.

Sauer, J.R., Hines, J.E. & Fallon, J. (2007) *The North American breeding bird survey, results and analysis 1966–2007, version 10.13.2007*. USGS Patuxent Wildlife Research Center, Laurel, MD.

Schemske, D.W., Mittelbach, G.G., Cornell, H.V., Sobel, J.M. & Roy, K. (2009) Is there a latitudinal gradient in the importance of biotic interactions? *Annual Review of Ecology, Evolution, and Systematics*, **40**, 245–269.

Schiffers, K., Tielbörger, K., Tietjen, B. & Jeltsch, F. (2011) Root plasticity buffers competition among plants: theory meets experimental data. *Ecology*, **92**, 610–620.

Schurr, F.M., Midgley, G.F., Rebelo, A.G., Reeves, G., Poschlod, P. & Higgins, S.I. (2007) Colonization and persistence ability explain the extent to which plant species fill their potential range. *Global Ecology and Biogeography*, **16**, 449–459.

Schurr, F.M., Spiegel, O., Steinitz, O., Trakhtenbrot, A., Tsoar, A. & Nathan, R. (2009) Long-distance seed dispersal. *Annual Plant Reviews*, **38**, 204–237.

Schweiger, O., Settele, J., Kudrna, O., Klotz, S. & Kühn, I. (2008) Climate change can cause spatial mismatch of trophically interacting species. *Ecology*, **89**, 3472–3479.

Schweiger, O., Heikkinen, R.K., Harpke, A., Hickler, T., Klotz, S., Kudrna, O., Kühn, I., Pöyry, J. & 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.

Scotti, M., Podani, J. & Jordán, F. (2007) Weighting, scale dependence and indirect effects in ecological networks: a comparative study. *Ecological Complexity*, **4**, 148–159.

Sebastián-González, E., Sánchez-Zapata, J.A., Botella, F. & Ovaskainen, O. (2010) Testing the heterospecific attraction hypothesis with time-series data on species co-occurrence. *Proceedings of the Royal Society B: Biological Sciences*, **277**, 2983–2990.

Simberloff, D. & Dayan, T. (1991) The guild concept and the structure of ecological communities. *Annual Review of Ecology and Systematics*, **22**, 115–143.

Smith, T.M., Shugart, H.H. & Woodward, F.I. (1997) *Plant functional types: their relevance to ecosystem properties and global change*. Cambridge University Press, Cambridge.

Solé, R.V. & Bascompte, J. (2006) *Self-organization in complex ecosystems*. Princeton University Press, Princeton, NJ.

Stang, M., Klinkhamer, P.G.L., Waser, N.M., Stang, I. & van der Meijden, E. (2009) Size-specific interaction patterns and size matching in a plant–pollinator interaction web. *Annals of Botany*, **103**, 1459–1469.

Suding, K.N., Lavorel, S., Chapin, F.S., III, Cornelissen, J.H.C., Díaz, S., Garnier, E., Goldberg, D., Hooper, D.U., Jackson, S.T. & Navas, M.-L. (2008) Scaling environmental change through the community-level: a trait-based response-and-effect framework for plants. *Global Change Biology*, **14**, 1125–1140.

Tibshirani, R. (1996) Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, **58**, 267–288.

Tylianakis, J.M., Didham, R.K., Bascompte, J. & Wardle, D.A. (2008) Global change and species interactions in terrestrial ecosystems. *Ecology Letters*, **11**, 1351–1363.

Vázquez, D.P., Morris, W.F. & Jordano, P. (2005) Interaction frequency as a surrogate for the total effect of animal mutualists on plants. *Ecology Letters*, **8**, 1088–1094.

Volterra, V. (1926) Variations and fluctuations of the number of individuals in animal species living together. *Animal ecology* (ed. by R.N. Chapman), pp. 409–448. Wiley, New York.

Wardle, D.A., Barker, G.M., Bonner, K.I. & Nicholson, K.S. (1998) Can comparative approaches based on plant ecophysiological traits predict the nature of biotic interactions and individual plant species effects in ecosystems? *Journal of Ecology*, **86**, 405–420.

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

Williams, R.J., Anandanadesan, A. & Purves, D. (2010) The probabilistic niche model reveals the niche structure and role of body size in a complex food web. *PLoS ONE*, **5**, e12092.

Woiwod, I.P. & Harrington, R. (1994) Flying in the face of change: the Rothamsted insect survey. *Long-term experiments*

in agricultural and ecological sciences (ed. by R.A. Leigh and A.E. Johnston), pp. 321–342. CAB International, Wallingford, UK.

Woodward, G., Ebenman, B., Emmerson, M., Montoya, J.M., Olesen, J.M., Valido, A. & Warren, P.H. (2005) Body size in ecological networks. *Trends in Ecology and Evolution*, **20**, 402–409.

Wootton, J.T. & Emmerson, M. (2005) Measurement of interaction strength in nature. *Annual Review of Ecology, Evolution, and Systematics*, **36**, 419–444.

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.