

ECOGRAPHY

Forum

Bringing Elton and Grinnell together: a quantitative framework to represent the biogeography of ecological interaction networks

Dominique Gravel, Benjamin Baiser, Jennifer A. Dunne, Jens-Peter Kopelke, Neo D. Martinez, Tommi Nyman, Timothée Poisot, Daniel B. Stouffer, Jason M. Tylianakis, Spencer A. Wood and Tomas Roslin

D. Gravel (<https://orcid.org/0000-0002-4498-7076>) (Dominique.Gravel@USherbrooke.ca) and *T. Poisot* (<http://orcid.org/0000-0002-0735-5184>), Québec Centre for Biodiversity Science, McGill Univ., Montréal, QC, Canada. DG also at: Canada Research Chair in Integrative Ecology, Dépt de Biologie, Univ. de Sherbrooke, Sherbrooke, QC, Canada. TP also at: Univ. de Montréal, Dépt des Sciences Biologiques, Montréal, QC, Canada. – *B. Baiser* (<https://orcid.org/0000-0002-3573-1183>), Univ. of Florida, Dept of Wildlife Ecology and Conservation, Gainesville, FL, USA. – *J. A. Dunne*, Santa Fe Inst., Santa Fe, NM, USA. – *J.-P. Kopelke*, Senckenberg Research Inst. and Natural History Museum, Frankfurt am Main, Germany. – *N. D. Martinez*, Ecology and Evolutionary Biology Dept, Univ. of Arizona, Tucson, AZ, USA. – *T. Nyman*, Norwegian Inst. of Bioeconomy Research, Dept of Ecosystems in the Barents Region, Svanvik, Norway. TN also at: Univ. of Eastern Finland, Dept of Environmental and Biological Sciences, Joensuu, Finland. – *D. B. Stouffer* and *J. M. Tylianakis*, Centre for Integrative Ecology, School of Biological Sciences, Univ. of Canterbury, Christchurch, New Zealand. JMT also at: Dept of Life Sciences, Imperial College London, Silwood Park Campus, Berkshire, UK. – *S. A. Wood*, Univ. of Washington, School of Environmental and Forest Sciences, Seattle, WA, USA. – *Tomas Roslin*, Dept of Ecology, Swedish Univ. of Agricultural Sciences, Uppsala, Sweden.

Ecography

41: 1–15, 2018

doi: 10.1111/ecog.04006

Subject Editor: Pablo A. Marquet

Editor-in-Chief: Miguel Araújo

Accepted 4 July 2018

Biogeography has traditionally focused on the spatial distribution and abundance of species. Both are driven by the way species interact with one another, but only recently community ecologists realized the need to document their spatial and temporal variation. Here, we call for an integrated approach, adopting the view that community structure is best represented as a network of ecological interactions, and show how it translates to biogeography questions. We propose that the ecological niche should encompass the effect of the environment on species distribution (the Grinnellian dimension of the niche) and on the ecological interactions among them (the Eltonian dimension). Starting from this concept, we develop a quantitative theory to explain turnover of interactions in space and time – i.e. a novel approach to interaction distribution modeling. We apply this framework to host–parasite interactions across Europe and find that two aspects of the environment (temperature and precipitation) exert a strong imprint on species co-occurrence, but not on species interactions. Even where species co-occur, interaction proves to be stochastic rather than deterministic, adding to variation in realized network structure. We also find that a large majority of host–parasite pairs are never found together, thus precluding any inferences regarding their probability to interact. This first attempt to explain variation of network structure at large spatial scales opens new perspectives at the interface of species distribution modeling and community ecology.

Keywords: networks, spatial ecology, co-occurrence



www.ecography.org

© 2018 The Authors. Ecography © 2018 Nordic Society Oikos

Introduction

Community ecology is ‘the study of the interactions that determine the distribution and abundance of organisms (Krebs 2009). Despite a general consensus on this definition (Scheiner and Willig 2007), research on variation in community structure (beta-diversity) has mostly focused on the spatial and temporal turnover of species composition (Anderson et al. 2011). Such research usually neglects variation in the way species interact with each other, despite accumulating empirical evidence that it is a major source of diversity (Poisot et al. 2015). Given this omission, it is perhaps not surprising that biogeographers are still struggling to establish whether interactions actually impacts the distribution of species at large spatial scales (Kissling et al. 2012). An interaction is conceived as the direct effect of the action of one species on the demography on another species; whether it scales up and impacts the (co-)distribution of species remains matter of debate at the moment (González-Salazar et al. 2013, Wisz et al. 2013, Cazelles 2016, Harris 2016, Godsoe et al. 2017). Treating interactions as fixed events nonetheless neglects a large part of the complexity of empirical communities, and will most likely deliver biased metrics (Poisot et al. 2016b). Recent attempts at accounting for interactions in species distribution models (Pellissier et al. 2013, Pollock et al. 2014, Ovaskainen et al. 2017) have brought some methodological advances, but are not sufficient for two reasons. First, these techniques are still based on a ‘species-based’ approach to communities, where interactions are merely treated as fixed covariates affecting distribution. Second, they fail to provide a conceptual step forward, both in their treatment of interactions and in the quality of the predictions they make.

Network approaches offer a convenient representation of communities because they simultaneously account for species composition and their interactions. Species are represented as nodes, so that networks already encompass all the information used by current approaches of species distribution modeling; in addition, interactions are represented by links, so that networks provide additional information on community structure. To date, studies of network diversity have mostly been concerned with the distribution of interactions within locations, and less so with variation among locations (Dunne 2006, Bascompte and Jordano 2007, Ings et al. 2009, Kéfi et al. 2012). There is, however, ample evidence that interaction networks vary in space and time (Laliberté and Tylianakis 2010, Poisot et al. 2012, 2016c, Schleuning et al. 2012, Albouy et al. 2014, Trøjelsgaard et al. 2015), even though there is no common framework with which to generalize these results. Metacommunity theory provides explanations for variation in the distribution of the nodes (Gravel et al. 2011, Pillai et al. 2011, Cazelles et al. 2015), but there are no such explanation to the variation of node and link occurrences. Consequently, we need theory to formalize these observations, as it is the only way towards fulfilling the goal of community ecology: providing cogent

predictions about, and understanding of, the structure of ecological communities.

Given the historically different approaches to modelling the distributions of species vs. interactions, there is a need to bring the two together. Here, we offer an integrated approach to do so, adopting the view that community structure is best represented as a network of ecological interactions. Based on this idea, we propose a description of the ecological niche that integrates the effect of the environment on species distribution and on the ecological interactions among them. Building on this concept, we develop a quantitative theory to explain turnover of interactions in space and time. We first present the concept and then formalize it mathematically, using a probabilistic model to represent the sampling of the regional pool of interactions. At the level of species pairs, the statistical approach could be conceived as an interaction distribution model. At the community level, the approach provides a likelihood-based method to compare different hypotheses of network turnover. As an illustrative example, we apply this novel framework to a large dataset on host-parasite interactions across Europe and find that two aspects of the environment (temperature and precipitation) exert a strong imprint on species co-occurrence, but not on species interactions. The network structure changes systematically across the latitudinal gradient, with a peak of connectance at intermediate latitudes.

The two dimensions of community structure

The problem of community assembly is often formulated as ‘how are species sampled from a regional pool to constitute a local community’ (Götzenberger et al. 2012)? This question could be rewritten to address the problem of network assembly, as ‘how do samples from a regional pool of interactions constitute a local interaction network?’ An illustration of this problem for a food web is provided in Fig. 1. The regional pool of interactions, the ‘metaweb’, represents potential interactions among all species that could be found in a given area. In this particular case, there are 275 nodes, and 1173 links among the plants (52 nodes), herbivores (96 nodes), and parasitoids (127 nodes) from northern Europe. An instance of a local community is also illustrated, with 45 nodes and 93 interactions. Only 55.0% of all potential interactions (plant-herbivore or herbivore parasitoid combinations) are realized in this local network, revealing the stochastic nature of ecological interactions. Our objective here is to provide theory to explain the sampling of the regional pool of interactions, along with a quantitative method to predict it. The problem could be formalized sequentially by understanding first why only a fraction of the species co-occur locally and second why these species do or do not interact.

There are multiple causes of spatial turnover of species co-occurrence. The first and most-studied driver is the effect of variation in the abiotic environment on species performance. Combined with specific responses in demography,

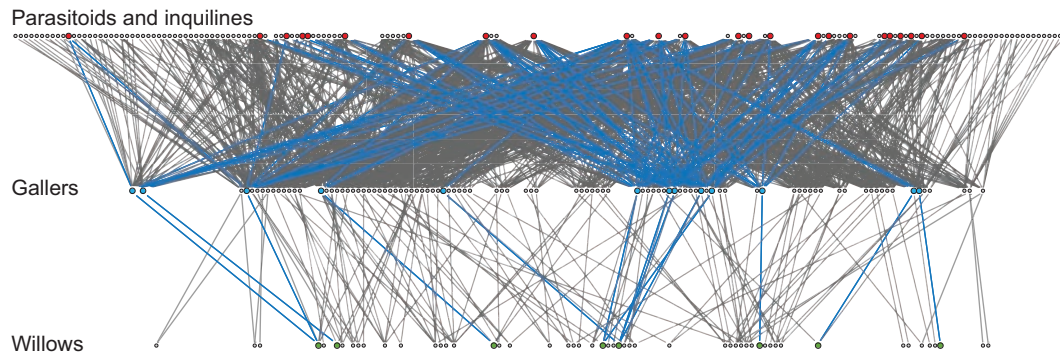


Figure 1. Non-random sampling of the metaweb. Network assembly can be viewed as a sampling process of the regional pool of potential interactions. Species (indicated by colored nodes) are sampled first, and among the species found in the local network, only some interactions (indicated by blue links) occur. We characterize these sampling processes with the quantitative framework proposed in this paper. As a concrete illustration of metaweb sampling, we here show a local interaction network among *Salix* (bottom/green), gallers (center/blue), and parasitoids (top/red). The metaweb was constructed by aggregating interactions observed across 370 local networks. Nodes were distributed to minimize crossing of interactions.

it generates variation among sites by selecting the locally fit-test species (Leibold et al. 2004). Stochasticity plays an additional role, either because colonization and extinction events (Hanski 1999) are inherently unpredictable or because strong non-linear feedbacks in community dynamics generate alternative transients and equilibria (Chase 2007, Vellend et al. 2014). Interaction themselves may impact co-distribution, as hypothesized long ago by Diamond (1975). Analyses of community turnover are usually performed with data represented in a table with rows corresponding to sites (or measurements) and columns to species. Metrics of beta diversity quantify the variance of this community data (Legendre et al. 2005). Traditional approaches rely on measures of dissimilarity among communities, such as the Jaccard or Bray–Curtis indices. More recent approaches decompose total variation of the community data into species and site contributions to beta diversity (Legendre and De Cáceres 2013), and further partition it into dissimilarity due to changes in species richness and dissimilarity due to actual species turnover (Baselga 2010). Even though all of these methods compare whole lists of species among sites or measurements, they remain fundamentally ‘species-based’. None of them explicitly considers variation of associations (i.e. of pairs or higher-order motifs, Stouffer et al. 2007). Some previous studies have considered how species distribution could be influenced by the joint effects of the abiotic and biotic environment (Stephens and Heau 2009, González-Salazar et al. 2013, Cazelles et al. 2015, Ovaskainen et al. 2017), here we inverse the problem and describe how the distribution of biotic interactions is influenced by species distribution and the environment.

We are now getting a better understanding of interaction turnover. As mentioned above, in the network approach to community structure, species and interactions are represented by nodes and links, respectively. Associations can also be represented by matrices in which entries represent the occurrence or intensity of interactions among species (rows and columns). Network complexity is then computed as the

number of interactions (in the case of binary networks) or interaction diversity (in the case of quantitative networks, Bersier et al. (2002)). Variability in community structure consequently arises from the turnover of species composition, along with turnover of interactions among pairs of species. The occurrence and intensity of interactions could vary because of the environment, species abundance, and higher-order ecological interactions (Poisot et al. 2015). Variation in community composition can be independent of variation of ecological interactions, suggesting that species and interaction distribution may well respond to different drivers (Poisot et al. 2012).

The ‘niche’ is by far the dominant concept invoked to explain species distributions and community assembly, from the local to the global scale. Following Hutchinson (1957), the niche is viewed as the set of environmental conditions allowing a population to establish and persist (see also Holt 2009). In other words, the niche is the location in that multidimensional space allowing a species to have a positive growth rate when rare (Godsoe et al. 2017). Community turnover arises as a result of successive replacement of species along an environmental gradient, in agreement with the Gleasonian view of communities (Gleason 1926). The concept is straightforward to put into practice with species distribution models, as it maps naturally on available distributional and environmental data. Consequently, a vast array of statistical tools have been developed to implement it (e.g. BIOMOD – Thuiller 2003, MaxEnt – Phillips et al. 2006). It is however much harder to account for ecological interactions within this approach (Townsend et al. 2011). As such, these interactions are often viewed as externalities constraining or expanding the range of environmental conditions required for a species to maintain a viable population (Pulliam 2000, Soberón 2007).

Interestingly, the food web literature also has its own ‘niche model’ to position a species in a community (Williams and Martinez 2000), generalized later to other types of interaction networks (Eklöf et al. 2013). The niche of a species in

this context represents the multidimensional space of all of its interactions. Each species is characterized by a niche position, an optimum and a range over three to five different axes (Williams and Martinez 2000, Eklöf et al. 2013). The niche model of food web structure and its variants have successfully explained the complexity of a variety of networks, from food webs to plant-pollinator systems (Allesina et al. 2008, Williams et al. 2010, Eklöf et al. 2013). This conceptual framework is, however, limited to local communities, and does not provide any explanation for the turnover of network structure along environmental gradients.

The integrated niche

Despite several attempts to update the concept of the ecological niche, ecologists have not moved far beyond the ‘*n*-dimensional hypervolume’ defined by Hutchinson (1957). Despite its intuitive interpretation and easy translation into species distribution models (Boulangeat et al. 2012, Blonder et al. 2014), the concept has been frequently criticized (Hardin 1960, Peters 1991, Silvertown 2004), and several attempts have been made to expand and improve it (Pulliam 2000, Chase and Leibold 2003, Soberón 2007, Holt 2009, McNerny and Etienne 2012a).

Part of the problem surrounding the niche concept has been clarified with the distinction between Eltonian and Grinnellian definitions (Chase and Leibold 2003). The Grinnellian dimension of the niche is the set of abiotic environmental conditions required for a species to maintain a population in a location. The Grinnellian niche is intuitive to apply, and constitutes the conceptual backbone of species distribution models. The Eltonian niche, on the other hand, is ‘the place of a species in its biotic environment, its relations to food and enemies’. While this aspect of the niche is well known by community ecologists, it is trickier to turn into predictive models. Nonetheless, the development of the niche model of food web structure (Williams and Martinez 2000) and its parameterization using functional traits (Gravel et al. 2013, Bartomeus et al. 2016) made it more operational.

These perspectives are rather orthogonal to each other, and this has resulted in considerable confusion in the literature (McNerny and Etienne 2012b). Chase and Leibold (2003) attempted to reconcile with the following definition: ‘[The niche is] the joint description of the environmental conditions that allow a species to satisfy its minimum requirements so that the birth rate of a local population is equal to or greater than its death rate along with the set of per capita effects of that species on these environmental conditions’. Their representation merges zero-net-growth isoclines delimiting the Grinnellian niche (‘when does the population persist?’) with impact vectors delimiting the Eltonian niche (‘what is the per-capita impact?’). While this representation has been very influential in local-scale community ecology (the resource-ratio theory of coexistence, Tilman 1982), it remains impractical at larger spatial scales because of the

difficulties in measuring it. The absence of any mathematical representation of the niche that can be easily fit to ecological data may explain why biogeographers are still struggling to develop species distribution models that also consider ecological interactions. Thus, a more integrative description of the niche will be key to understand spatial and temporal turnover in community structure.

We propose to integrate the two perspectives of the niche using a visual representation of both components (Fig. 2). The underlying rationale is that, in addition to the environmental constraints on demographic performance (Fig. 2, top panel), any organism requires resources to meet its metabolic demands and to sustain reproduction (Fig. 2, nodes in network of bottom panel). Abiotic environmental axes are any non-consumable factors affecting the demographic

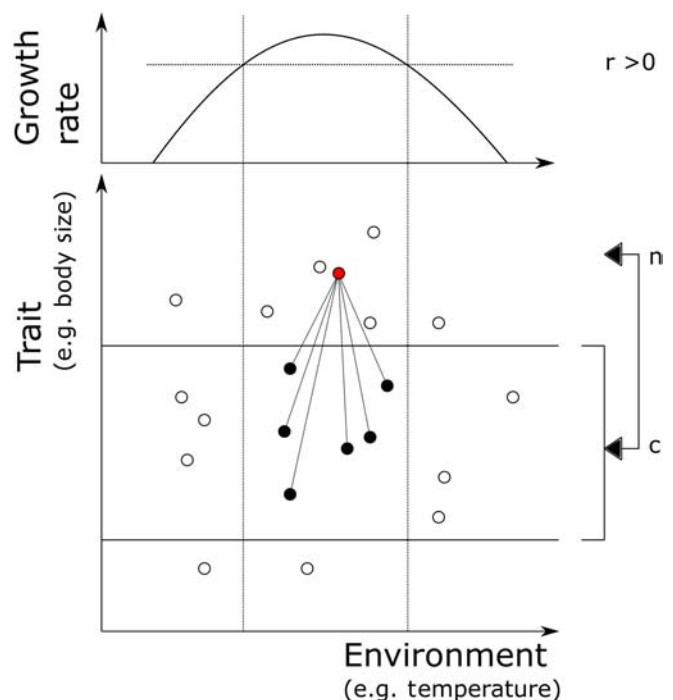


Figure 2. Visual representation of the integrated niche. In biogeography, the niche is considered the set of environmental conditions where the intrinsic growth rate r is positive (Holt 2009). The horizontal axis represents an environmental gradient impacting the growth of a focal species (red point). The location of each species (grey points) along this gradient represents their optimum, and the vertical dotted lines represent the limits of the Grinnellian niche of the focal species. In food web ecology, the Eltonian niche represents the location of a species in the food web, as determined by its niche position (n) and its niche optimum (c). The vertical axis represents a niche gradient, for example a trait such as body size. The location of each species along this gradient represents their niche position. The focal species will feed only on host/prey species occupying niche locations within a given interval around the optimum, represented by the horizontal lines. The integrated Grinnellian and Eltonian niche corresponds to the square in the middle where an interaction is possible owing to a match of traits and spatial distribution. The central square represents the area where the joint probability of observing co-occurrence and interactions is not null.

performance of an organism. Alternatively, the resource axes are traits of the resources that allow interactions with the consumers. By definition of what an interaction is, these axes also influence the growth rate. The niche can therefore be viewed as the set of abiotic environmental conditions (the Grinnellian component) along with the set of traits (the Eltonian component) that allow a population to establish and to grow at a location. This visual representation makes operational the theoretical interpretation of the niche and species ranges by Godsoe et al. (2017). Accordingly, each species can be characterized by an optimal position along both the environmental (x-axis) and the trait (y-axis) plane. The integrated niche is then the hypervolume where interactions can occur and where a population has a positive growth rate.

This approach changes the representation of the niche, putting species distributions and ecological interactions into the same formalism. Moreover, it allows the limits of the niche axes to be independent of each other (as in the example in Fig. 2), or to interact. Some of these axes may not be independent. For instance, the optimal prey size for predatory fishes could decline with increasing temperature (Gibert and DeLong 2014), which would make diet boundaries functions of the environment. Alternatively, we could also consider that the growth rate of the predator changes with the size of its prey items, thereby altering the environmental boundaries and changing the shape of the niche in the middle of this visual representation. It is also important that this conceptual representation depicts a reality that could be highly dimensional and sometimes difficult to represent statistically (Clark et al. 2007).

A probabilistic representation of interaction networks in space

We now formalize the integrated niche with a probabilistic approach to interactions and distributions. In particular, we seek to represent the probability that an interaction between species i and j occurs at location y . We define L_{ij} as a stochastic process taking a value of 1 when an interaction occurs and a value of 0 when it does not, and focus on the probability that this event occurs, $P(L_{ij}=1)$, over a given sampling interval and space. We note that the occurrence of an interaction is dependent on the co-occurrence of species i and j . This argument might seem trivial at first, but the explicit consideration of this condition in the probabilistic representation of ecological interactions will prove instrumental to understanding their variation. We note that this phenomenon we aim to describe differs fundamentally from the effect that interactions do have on co-occurrence (Cazelles et al. 2016). We similarly define X_{iy} as a stochastic process representing the occurrence of species i at location y , taking a value of 1 when the species occurs and a value of 0 when it is absent. The quantity we seek to understand is the probability of a joint event, conditional on the set of environmental conditions E_y :

$$P(X_{iy}, X_{jy}, L_{ij} | E_y)$$

Or simply said, the probability of observing both species i and j plus an interaction between i and j given the conditions E_y at location y . This probability could be decomposed into two parts using the product rule of probabilities:

$$P(X_{iy}, X_{jy}, L_{ij} | E_y) = P(X_{iy}, X_{jy} | E_y) P(L_{ij} | X_{iy}, X_{jy}, E_y)$$

The first term on the right-hand side of the equation is the probability of observing the two species co-occurring at location y . It corresponds to the Grinnellian dimension of the niche. The second term represents the probability that an interaction occurs between species i and j , given that they are co-occurring. This predicate can be refined using information on trait distribution and trait matching rules. For brevity, we have here avoided specific references to traits and considered that the interactions are provided in the data. But one could develop the model further to integrate such traits, inspired by the framework developed by Gravel et al. (2016). They proposed to represent the occurrence of an interaction between species as a joint probability event $P(L_{ij}, T_i, T_j | E)$, where T_i and T_j are trait vectors for species i and j respectively. In order to develop a trait-based representation of the spatial variation of interactions, one may thereafter consider how these traits are related to the environment and how they co-occur. Above, we referred to this entity as the ‘metaweb’ and note that it corresponds to the Eltonian dimension of the niche. Below, we will see how this formalism can be directly fit to empirical data, restricting ourselves to occurrences only and referring to the above citations for trait-matching and trait-environment relationships. But before turning to an application, we will discuss the interpretation of different variants of these two terms.

Variants of co-occurrence

There are several variants to the co-occurrence probability, representing different hypotheses concerning spatial variation in network structure (see the explicit formulations in Table 1). The simplest model relates the probability of co-occurrence directly to the environment, $P(X_{iy}, X_{jy} | E_y)$. In this situation, there are no underlying assumptions about the ecological processes responsible for co-occurrence. Spatial associations between species could arise because interactions constrain distribution, where in such case the co-occurrence would be conditional on L , or, alternatively, because of environmental requirements shared between i and j (Pollock et al. 2014, Cazelles et al. 2016). In the former case, species are not independent of each other and the conditional occurrence must be accounted for explicitly, $P(X_{iy}, X_{jy} | E_y) = P(X_{iy} | E_y, X_{jy}) P(X_{jy} | E_y)$. In the latter case, species are independent, and only the marginal occurrence must be accounted for, $P(X_{ij} | E_y) = P(X_{iy} | E_y) P(X_{jy} | E_y)$.

Table 1. Example of data used to perform the analysis. Traditional species distribution models represent the occurrence of single species as a function of the environment, with models such as $X_i \sim \text{Temp}$ performed one species after the other. Some also tried to represent both the abiotic and biotic niche, with models such as $X_i \sim \text{Temp} + X_j$ (González-Salazar et al. 2013), while the new joint species distribution models (Ovaskainen et al. 2017) consider the entire community at once ($\mathbf{X} \sim \text{Temp}$) and aims at representing co-distribution (X_{ij}). Here is the focus is on the occurrence of interactions, L_{ij} . The full model (see model 3, Tables 1–2) considers the result of two submodels, one representing the occurrence of interactions given co-occurrence, $L_{ij} \sim \text{Temp}$ (evaluated only on the subset of data where $X_{ij}=1$), and the other representing the co-occurrence, $X_{ij} \sim \text{Temp}$.

Location	Sp _i	Sp _j	X_i	X_j	X_{ij}	L_{ij}	Av. temp
Kühkopf	<i>Phyllocolpa prussica</i>	<i>Chrysocharis elongata</i>	1	1	1	1	8.25
Kühkopf	<i>Phyllocolpa leucapsis</i>	<i>Chrysocharis elongata</i>	0	1	0	0	8.25
Ötztal	<i>Phyllocolpa prussica</i>	<i>Chrysocharis elongata</i>	0	0	0	0	-1.50
Ötztal	<i>Phyllocolpa leucapsis</i>	<i>Chrysocharis elongata</i>	0	0	0	0	-1.50
Kilpisjärvi	<i>Phyllocolpa prussica</i>	<i>Chrysocharis elongata</i>	0	1	0	0	-1.92
Kilpisjärvi	<i>Phyllocolpa leucapsis</i>	<i>Chrysocharis elongata</i>	0	1	0	0	-1.92
Wiistensachsen	<i>Phyllocolpa prussica</i>	<i>Chrysocharis elongata</i>	1	1	1	0	5.58
Wiist ensachsen	<i>Phyllocolpa leucapsis</i>	<i>Chrysocharis elongata</i>	1	1	1	1	5.58
Steinamoen	<i>Phyllocolpa prussica</i>	<i>Chrysocharis elongata</i>	0	1	0	0	3.17
Steiriarioeri	<i>Phyllocolpa leucapsis</i>	<i>Chrysocharis elongata</i>	0	1	0	0	3.17

The co-occurrence probability itself could depend on ecological interactions. This should be viewed as the realized component of the niche (i.e. the distribution when accounting for species interactions). Direct pairwise interactions such as competition, facilitation, and predation have long been studied for their impact on co-distribution (Diamond 1975, Connor and Simberloff 1979, Gotelli 2000). Second and higher-order interactions (e.g. trophic cascades) could also affect co-occurrence (Harris 2016, Staniczenko et al. 2017). Co-occurrence of multiple species embedded in ecological networks is a topic of its own, however, and is influenced by both network topology and species richness (Cazelles et al. 2016). Not only direct interactions influence co-occurrence, but indirect interactions do as well (e.g. plant species sharing an herbivore, or herbivores sharing parasitoids, could repel each other in space, Holt and Lawton 1993). The impact of direct interactions and first-order indirect interactions on co-occurrence tends to vanish with increasing species richness in the community (Cazelles et al. 2016). Further, co-occurrence is also influenced by the covariance of interacting species to an environmental gradient (Cazelles et al. 2015).

The interpretation that interactions may impact co-occurrence, and that the realization of an interaction requires co-occurrence are not mutually exclusive. But since here the object is to describe the variation in the occurrence of interactions, which absolutely requires co-occurrence, we will give attention to this part of the problem. Relating co-occurrence to the structure of interaction networks is a problem on its own and we will therefore focus here on the variation of interactions and not on their distribution, and leave this specific issue for the Perspectives section and future research.

Variants of the metaweb

There are also variants of the metaweb. First, most documented metawebs have thus far considered ecological interactions to be deterministic, rather than probabilistic (Havens 1992, Wood et al. 2015). Species are assumed to interact

whenever they are found together in a location, independent of their local abundance and the environment. In other words, $P(L_{ij} | X_{ij} = 1, X_{ij} = 1) = 1$. This approach might be a reasonable approximation if the spatial or temporal scale of sampling and inference is so large that the probability of observing at least one interaction converges to unity. In this scenario, network variation arises solely from species distributions.

Second, ecological interactions could also vary with the environment, $P(L_{ij} | E_j)$. Although it is rare to see a conditional representation of pairwise ecological interactions, experimental studies have frequently revealed interactions to be sensitive to the environment. For instance, McKinnon et al. (2010) showed that predation risks of shorebirds vary at the continental scale, decreasing from the south to the north. It is also common to see increasing top-down control with temperature (Shurin et al. 2012, Gray et al. 2016). Effects of the environment on interactions (Gibert and DeLong 2014) also propagate up the community and influence network structure (Tylianakis et al. 2007, Petchey et al. 2010, Woodward et al. 2010).

Application: continental-scale variation of host-parasite community structure

We now turn to an illustration of our theory with the analysis of an empirical dataset of host-parasite networks sampled throughout the south-north environmental gradient in continental Europe (Kopelke et al. 2017). Our objective here is to illustrate potential applications of the approach outlined above, rather than to specifically describe the biogeography of this particular community. The focal system consists of local food webs of willows (genus *Salix*), their galling insects, and the natural enemies (parasitoids and inquilines) of gallers. We ask: 1) how much does network structure vary across the gradient, and 2) what is the primary driver of network turnover across the gradient?

Data

Communities of willows, gallers, and parasitoids are species-rich and widely distributed, with pronounced variation in community composition across space. The genus *Salix* includes over 400 species, most of which are shrubs or small trees (Argus 1997), and is common in moist habitats across the Northern Hemisphere (Skvortsov 1999). Willows support a highly diverse community of herbivorous insects, with one of the main herbivore groups being gall-inducing sawflies (Hymenoptera: Tenthredinidae: Nematinae: Euurina (Kopelke 1999). Gall formation is induced by sawfly females during oviposition, and includes marked manipulation of host-plant chemistry by the galler (Nyman and Julkunen-Tiitto 2000). The enemy community of the gallers includes nearly 100 species belonging to 17 insect families of four orders (Kopelke 2003). These encompass two main types: inquiline larvae (Coleoptera, Lepidoptera, Diptera, and Hymenoptera) feed primarily on gall tissue, but typically kill the galler larva in the process, while parasitoid larvae (representing many families in Hymenoptera) kill the galler larvae by direct feeding (Kopelke 2003).

Local realizations of the willow-galler-parasitoid network were reconstructed from community samples collected between 1982 and 2010. During this period, willow galls were collected at 370 sites across Central and Northern Europe. In total, 52, 96 and 127 *Salix*, galler and parasitoid and inquiline taxa were distinguished, respectively. The strength of this dataset is that observations were observed in situ, rather than inferred from expert knowledge or other sources of information, thereby allowing the analysis of their spatial variation. The drawback is that, because many species are rare, some of the pairs of species may have been observed co-occurring only a few times, or never. As a consequence, despite the extent of the sampling, there is significant uncertainty in the quantification of some of the links. The above described methodology explicitly aims at describing this uncertainty, and research should be done to develop methods to reduce it (see conclusion).

The current study represents the first analysis of the full data set from a spatial perspective. Full details about data collection and species identification are provided in (Kopelke et al. 2017). Annual mean temperature and precipitation were obtained from WorldClim using the R package *dismo* (Hijmans 2015) and GPS coordinates of the sampling locations. While other covariates could have also been considered to represent the high dimensionality of the niche, these two variables are likely representative of the most important axes of European climate, and are also more easily interpretable than reduced variables obtained, for example, by principal component analysis.

Analysis

Computing the probability of observing an interaction involves fitting a set of binomial models and collecting their estimated probabilities. For the sake of illustration, we

considered generalized linear models – although more flexible fitting algorithms (e.g. GAM or Random Forest) could equally well be used, as long as the algorithm can estimate the probability for each observation. The data consist of a simple (albeit large and sparse) table with the observation of each species, X_{ij} and X_{jy} , their co-occurrence, X_{ijy} , the observation of an interaction L_{ijy} , and environmental co-variables E_y (Table 1). Thus, there is one row per pair of species per site. We considered that an absence of a record of an interaction between co-occurring species at a site means a true absence (see below for a discussion on this issue).

We compared three models for the co-occurrence probability. The first one directly represents the co-occurrence probability conditional on the local environment, $P(X_{ij}, X_{jy} | E_y)$ (models are listed in Table 1, 2). Hence, this model makes no assumption about the mechanisms driving co-occurrence for any given environment, and instead uses the information directly available in the data. It thereby indirectly accounts for the effect of interactions on co-occurrence, if there are any. The second model considers independent occurrence of species. In this case, we independently fit $P(X_{ij} | E_y)$ and $P(X_{jy} | E_y)$, and we then take their product to derive the probability of co-occurrence. This model should be viewed as a null hypothesis with respect to the first model, since a comparison between them will reveal if there is significant spatial association of the two species beyond a joint response to the shared environment (Cazelles et al. 2016). Finally, the third model assumes that the probability of co-occurrence is independent of the environment and thus constant throughout the landscape. In other words, $P(X_{ij}, X_{jy})$, is obtained by simply counting the number of observed co-occurrences divided by the total number of observations. Thus, the comparison between the first and third model allows us to test the hypothesis that co-occurrence is conditional on the environment. Whenever the environment was included as a covariate in the GLM, we considered a second-order polynomial term for both temperature and precipitation in order to account for optima in environmental conditions. There are consequently five parameters for the first model when fitting a given pair of species, 10 parameters for the second, and only one for the third model.

Table 2. Summary of model comparison for the interaction between the leaf folder *Phyllocolpa prussica* and the parasitoid *Chrysocharis elongata*. Note that E is a short notation including both temperature and precipitation.

#	Metaweb model	Co-occurrence model		LL	npars	AIC
1	$P(L_{ijy})$	$P(X_{ij}, X_{jy} E_y)$		-71.1	6	154.2
2	$P(L_{ijy} X_{ij}, X_{jy})$	$P(X_{ij}, X_{jy} E_y)$		-65.7	6	143.4
3	$P(L_{ijy} X_{ij}, X_{jy}, E_y)$	$P(X_{ij}, X_{jy} E_y)$		-65.6	10	151.3
4	$P(L_{ijy} X_{ij}, X_{jy}, E_y)$	$P(X_{ij}, X_{jy})$		-84.5	6	183
5	$P(L_{ijy} X_{ij}, X_{jy}, E_y)$	$P(X_{ij})P(X_{jy})$		-80.7	7	173.4
6	$P(L_{ijy} X_{ij}, X_{jy}, E_y)$	$P(X_{ij} E_y)P(X_{jy} E_y)$		-68.8	15	167.6

Following the same logic, we compared three models of the interaction probability. The first model conditions the interaction probability on the local environmental variables, $P(L_{ij} | X_{ij}, X_{jy}, E_j)$. Consequently, the model was fit to the subset of the data where the two species co-occur. The second model fits the interaction probability independently of the local environmental variables, $P(L_{ij} | X_{ij}, X_{jy})$. It corresponds to the number of times the two species were observed to interact when co-occurring, divided by the number of times that they co-occurred. The third model is an extreme case evaluated only to test the hypothesis that if two species are found to interact at least once, then they should interact whenever they co-occur, $P(L_{ij} | X_{ij}, X_{jy}) = 1$. While not necessarily realistic, this model tests an assumption commonly invoked in the representation of local networks from the knowledge of a deterministic metaweb. It represents potential interactions, rather than realized ones. There are consequently five parameters for the first model, a single parameter for the second model and no parameter to evaluate for the third model (where the interaction probability is fixed by the hypothesis).

We fitted the different models to each pair of species and recorded the predicted probabilities. The joint probability $P(L_{ij}, X_{ij}, X_{jy})$ was then computed from eq. 2, and the likelihood of each observation D_{ij} given the model θ was computed as $L(D_{ij} | \theta) = P(L_{ij}, X_{ij}, X_{jy})$ if an interaction was observed, and as $L(D_{ij} | \theta) = 1 - P(L_{ij}, X_{ij}, X_{jy})$ if no interaction was observed. The log-likelihood was summed over the entire dataset to compare the different models by AIC. We therefore evaluated the likelihood of all local networks, given the model. Not surprisingly, there was a very large number of species pairs for which this model could not be computed, as they simply never co-occurred. For these pairs, we have no information of the interaction probability, and they were consequently removed from the analysis. The log-likelihood reported across the entire dataset was therefore summed over all pairs of species observed to co-occur at least once. Interactions between the first (*Salix*) and second (gallers) trophic layers and those between the second and third (parasitoids) were considered separately. Finally, we used the full model (in which both co-occurrence and the interaction are conditional on the environment) to interpolate species distributions and interaction probabilities across the entire European continent. We reconstructed the expected network for each location in a 1×1 km grid and computed the probabilistic connectance following Poisot et al. (2016b).

Data availability

All of the data are openly available in the database *mangal* (Poisot et al. 2016a) and all R scripts for running the analyses, are provided in the github repository <https://github.com/DominiqueGravel/ms_probaweb>.

Results

Despite the extensive sampling, many pairs of species were observed to co-occur only a few times. This made it difficult to evaluate interaction probabilities with any reasonable confidence. Thus, we start with an example of a single pair of species selected because of its high number of co-occurrences ($N_{ij} = 38$): the leaf folder *Phyllocolpa prussica* and the parasitoid *Chrysocharis elongata*. These two fairly abundant species were observed $N_i = 49$ and $N_j = 121$ times, respectively, across the 370 sites, and they were found to interact with a marginal probability $P(L_{ij}) = 0.55$, which means they interacted at 21 different locations. Here, a comparison of model fit (Table 2) reveals that conditioning the interaction probability on local environmental conditions adds no explanatory power beyond a model assuming the same probability of interaction anywhere in space (model 2 vs model 3). Moreover, when the two species co-occur, the occurrence of the interaction was insensitive to the environment (model 2 vs model 3). Alternatively, climatic variables significantly impacted co-occurrence (model 3 vs model 4). The independent model performed worse than the non-random co-occurrence model (model 3 vs model 6). The full model revealed that the greatest interaction probability occurred at intermediate temperature and precipitation, simply because this is where the two species most frequently co-occur (Fig. 3). The probabilities of co-occurrence and interaction can be represented in space, where we find that the highest interaction probability occurred in central Europe (Fig. 4).

We evaluated each model for all pairs of species in order to better understand the large-scale drivers of network turnover. The results were highly consistent among trophic levels (*Salix*-gallers and gallers- parasitoids; Table 3), despite large variability in the fit of the models to the different pairs of species. This variability was particularly important for the models of interactions given co-occurrence. Across all pairs of species, the conditional representation of interactions performed (model 2) better than the marginal one (model 1); that is, interactions did not occur systematically whenever the two species were found co-occurring. Hence, in addition to species turnover, the stochastic nature of interactions contributes to network variability. In total, we recorded 1173 pairs of interactions, only 290 of which occurred more than five times. Out of these 290 interactions, 143 were systematically detected whenever the two species co-occurred. In the instances when species co-occurred, the two environmental variables considered proved relatively poor predictors of their interactions (model 2 vs model 3). Not surprisingly, for both types of interactions (*Salix*-galler and galler-parasitoid), the log-likelihood increased when the environment was considered. However, the extra number of parameters exceeded the gain in log-likelihood and inflated AIC. Therefore, the most parsimonious model excluded the effect of the environment. On the basis of log-likelihood only, co-occurrence was non-independent for both *Salix*-galler and galler-parasitoid interactions. Thus, according to AIC, the best model was the one

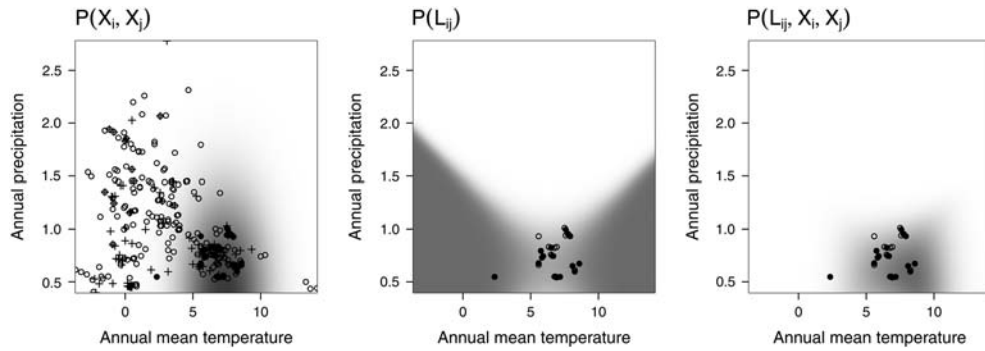


Figure 3. Probabilistic representation of the interaction probability between a leaf folder (*Phyllocolpa prussica*) and a parasitoid (*Chrysocharis elongata*) across gradients of annual average temperature and annual precipitation. The colour gradient represents the probability of observing the event, from 0 (white) to 1 (black). The representation is based on predictions from model 3 (see Table 1). In the left panel, open circles represent the absence of both species, whereas closed circles represent co-occurrence and plus signs the occurrence of only one of the two species. In the other two panels, open circles represent co-occurrence but an absence of interaction and closed circles the occurrence of an interaction.

of non-random co-occurrence (model 3 vs model 6) for both types of interactions.

Once we had selected the best model based on AIC (model 3, Table 2), we used it to reconstruct the expected species richness, along with the most likely network for each location. Using this approach, we mapped the expected distribution of network properties across Europe (Fig. 5). For simplicity, we chose to consider connectance as our descriptor of network configuration, as this metric can be easily computed from probabilistic networks (Poisot et al. 2016b) and is also a good proxy for many other network properties (Poisot and Gravel 2014). Overall, we found a peak in *Salix*, galls and parasitoid diversity in northern Europe. The expected number of interactions roughly followed the distribution of species richness, but accumulated at a rate different from species numbers. Connectance likewise peaked in northern Europe (Fig. 5).

Interpretation

We have proposed that the representation of community structure and its variation in space is best captured by the

formalism of ecological networks, as both the distribution of species and their interspecific interactions can then be accounted for. We consequently revised the niche concept in order to integrate its abiotic and biotic components. This integrated niche was represented visually with an ordination of species into an environmental space and a trait space. The fundamental niche of a species is represented as the set of environmental conditions and resources that allow a species to establish in a location, thereby integrating the Eltonian and the Grinnellian components of the niche. We then translated the concept mathematically by investigating the probability of the joint occurrences of species and their interaction, which should be interpreted as an interaction distribution model. We used this approach to characterize the turnover of the structure of ecological interactions in a species-rich tri-trophic network across Western Europe, finding that the primary driver of network variation is the turnover in species composition.

Applying the framework to our large data set on host-parasite interactions across Europe revealed key features in the interaction between *Salix* taxa, their herbivores, and the

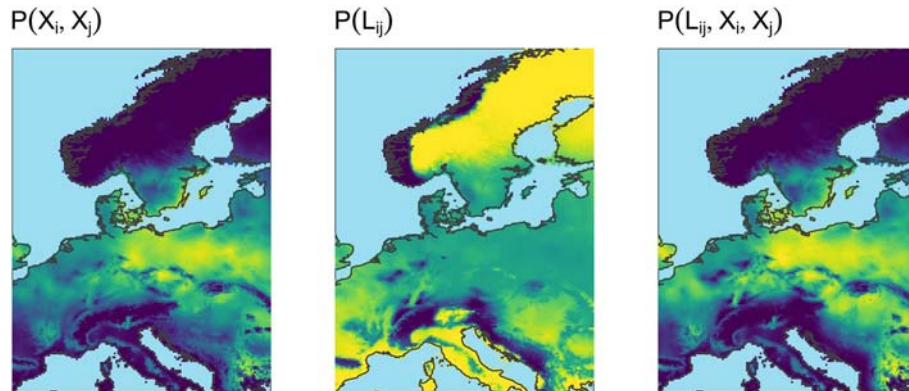


Figure 4. Probabilistic representation of the interaction probability between a leaf folder (*Phyllocolpa prussica*) and a parasitoid (*Chrysocharis elongata*) across Europe. Climate data are extracted from WorldClim over a 1 km \times 1 km grid and predictions made according to the model illustrated at Fig. 3. Colour gradient ranges from low values (dark blue) to high values (yellow).

Table 3. Summary of model comparison for the interaction across all pairs of *Salix*, gallers and parasitoids.

Interaction	#	Metaweb model	Co-occurrence model	LL	npars	AIC
<i>Salix</i> -Galler	1	$P(L_{ijv})$	$P(X_{ijv}, X_{jv} E_v)$	-6022.1	7548	27140.3
	2	$P(L_{ijv} X_{ijv}, X_{jv})$	$P(X_{ijv}, X_{jv} E_v)$	-5547.9	7548	26191.8
	3	$P(L_{ijv} X_{ijv}, X_{jv}, E_v)$	$P(X_{ijv}, X_{jv} E_v)$	-5364.0	12580	35888.0
	4	$P(L_{ijv} X_{ijv}, X_{jv}, E_v)$	$P(X_{ijv}, X_{jv})$	-5998.4	8806	30287.2
	5	$P(L_{ijv} X_{ijv}, X_{jv}, E_v)$	$P(X_{ijv})P(X_{jv})$	-6636.1	7548	27092.7
	6	$P(L_{ijv} X_{ijv}, X_{jv}, E_v)$	$P(X_{ijv} E_v)P(X_{jv} E_v)$	-6002.9	18870	49745.7
Galler-Parasitoid	1	$P(L_{ijv})$	$P(X_{ijv}, X_{jv} E_v)$	-22068.1	19206	82548.2
	2	$P(L_{ijv} X_{ijv}, X_{jv})$	$P(X_{ijv}, X_{jv} E_v)$	-19504.8	19206	77421.6
	3	$P(L_{ijv} X_{ijv}, X_{jv}, E_v)$	$P(X_{ijv}, X_{jv} E_v)$	-20217.0	32010	104454.1
	4	$P(L_{ijv} X_{ijv}, X_{jv}, E_v)$	$P(X_{ijv}, X_{jv})$	-19591.3	22407	77594.5
	5	$P(L_{ijv} X_{ijv}, X_{jv}, E_v)$	$P(X_{ijv})P(X_{jv})$	-22491.5	19206	89796.9
	6	$P(L_{ijv} X_{ijv}, X_{jv}, E_v)$	$P(X_{ijv} E_v)P(X_{jv} E_v)$	-18936.9	48015	133903.7

natural enemies of these herbivores. Consistent with a general increase in the diversity of *Salix* towards boreal areas (Cronk et al. 2015), overall species richness of the networks increased towards the north. The distribution of *Salix* species richness largely matched those of gallers and parasitoids. These observations within Europe are also matched

by the ones found at a global scale for *Salix* (Argus 1997, Cronk et al. 2015, Wu et al. 2015) and sawflies (Kouki et al. 1994, Kouki 1999). Species richness in a common group of parasitic wasps, the Ichneumonidae, was originally presumed to show a similar ‘reversed latitudinal gradient’, but this observation has been recently challenged by findings of rather

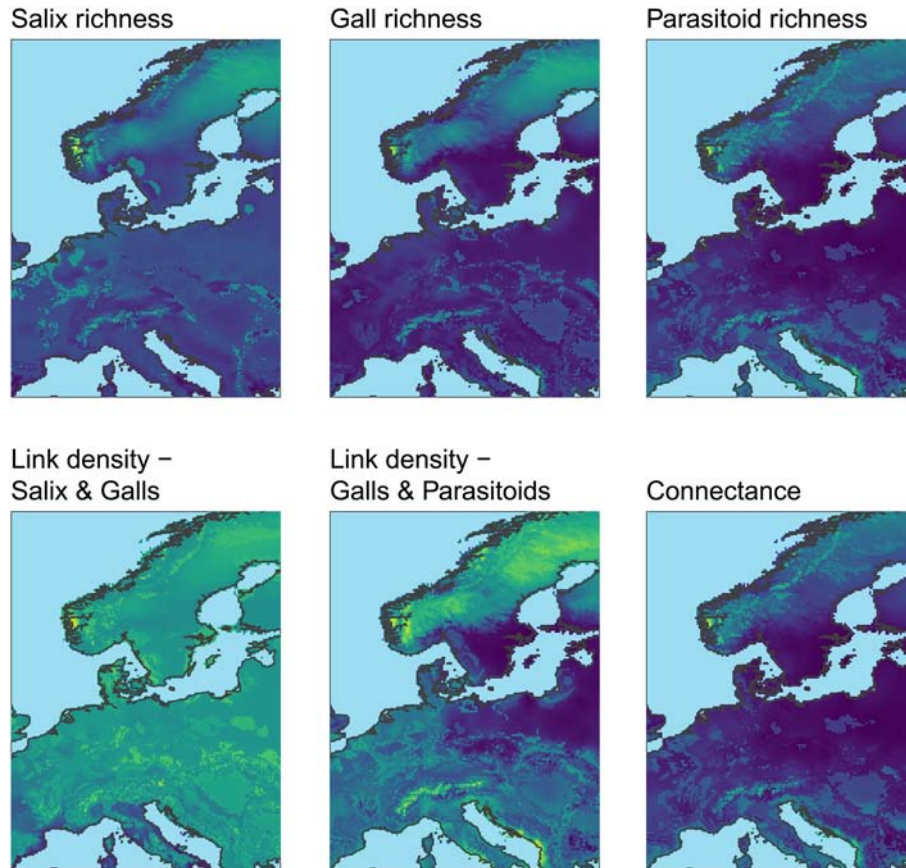


Figure 5. Mapping the distribution of species richness, the number of links and connectance across Europe. The representation is based on predictions from model 3 (see Table 2) using climate data extracted from WorldClim over a 1 km × 1 km grid. Species richness is obtained by summation of individual occurrence probabilities, and link density by summation of interaction probabilities. Color gradient ranges from low values (dark blue) to high values (yellow).

high ichneumonid diversity in the tropics (Veijalainen et al. 2013). Nevertheless, ichneumonid subfamilies specifically associated with sawflies (Ctenopelmatinae, Tryphoninae) are clearly less diverse in the south.

Exactly what processes are responsible for the distribution of species richness at different trophic levels is yet to be established (but see Roininen et al. 2005, Nyman et al. 2010, Leppänen et al. 2014), but as a net outcome of different latitudinal trends across trophic levels, the distribution of co-occurrence and therefore of potential interactions differed between the first and second layers of feeding links. The correlation between expected *Salix* and galler richness was 0.73, while it was 0.58 between gallers and their parasitoids. Therefore, the ratio of herbivores to *Salix* species is essentially constant across Europe, whereas each herbivore species is potentially attacked by a richer enemy community at higher latitudes. Consequently, overall connectance peaks in northern Europe (Fig. 5).

In terms of species interacting with each other, our analysis suggests that the environment leaves a detectable imprint on species co-occurrence, but only a slight mark on the occurrence of realized links among species in a specific place: the probability of finding a given combination of species at a higher and a lower trophic level at the same site was clearly affected by the environment, whereas the probability of observing an interaction between the two was, given co-occurrence, not significantly so. The interactions were highly uncertain, as only few pairs of species were systematically interacting when co-occurring but most not. This applies to the example species *Phyllocolpa prussica* and *Chrysocharis elongata* (Fig. 3, 4), but also to all species pairs more generally. For the example species pair, the full model revealed that the joint probability event (interaction and co-occurrence) peaks at intermediate temperature and precipitation, simply because this is where the two species co-occur most often. This does not imply that species will always interact when they meet – although this is a basic assumption in most documented metawebs to date (Havens 1992, Wood et al. 2015). Rather, an interaction is better represented as a stochastic process whose probability is also influenced by the probability with which species co-occur. What we cannot reliably know is how this stochasticity splits into different sources of uncertainty – i.e., the extent to which a species at the higher trophic level runs into a species at the lower level co-occurring at the site, the extent to which this interaction is detected by an observer collecting a finite sample, or simply the uncertainty arising from incomplete description of a highly dimensional niche. Future work will be required to document the relative importance of these sources of uncertainty in the occurrence of interactions.

Perspectives

Evidence that the structure of ecological networks does vary across habitats (Tylianakis et al. 2007), over environmental gradients (Lurgi et al. 2012) and in time (Simanonok and Burkle 2014) is accumulating rapidly. It is not clear, however,

to what extent the turnover of network structure is driven by a systematic change in species composition or of pairwise interactions (Poisot et al. 2012, 2015). Our model comparison of host-parasite interactions revealed that most of the turnover is driven by species-specific responses to the environment, impacting species richness, and that co-occurrence was mostly independent. Further, the occurrence of interactions among host and parasite is highly stochastic even when both are present, and not predictable by the variables considered by us. We know that interactions vary with the environment in other systems, for instance, herbivory (Shurin et al. 2012, Baskett et al. 2018) and predation (McKinnon et al. 2010, Legagneux et al. 2014) are often found to increase with temperature, resulting in spatial variation of trophic cascades (Gray et al. 2016). What remains unclear, however, is the extent to which such variation is driven by a turnover of species composition along gradients, or a turnover of the interactions. Here we found that interactions vary substantially but non-predictably along the annual temperature and the precipitation gradient. The lack of detectable signal may be due to our choice of covariates. Indeed, a previous study focusing on a subset of the system studied here identified habitat characteristics as the primary drivers of interactions (Nyman et al. 2015). New investigations with other systems will thus be required to challenge this result. Under all circumstances, documenting the relationship between the environment and the occurrence of interactions at continental scales is critical for understanding how large-scale variation of trophic regulation influences community dynamics and ecosystem functioning (Harfoot et al. 2014).

The framework we provide complements methods to compute network beta-diversity presented by Poisot et al. (2012) and Poisot et al. (2018). The total network turnover is partitioned into interaction turnover and species turnover – which in our approach would correspond to the three terms of eq. 2, respectively. Importantly, one could easily derive the different β of Poisot et al. (2012) using the fitted probabilities. For instance, for a pair of species i and j and sites x and y , the expectation for the fraction a for the interactions would be $P(L_{ijx} | X_{ix}, X_{jx})P(L_{ijy} | X_{iy}, X_{jy})$. The same could be computed for fractions b , and therefore the expected beta-diversity for an entire network could be recomputed this way. The novel feature of the approach presented here is then that the different contributions to network turnover can be attributed to different hypotheses (as performed here), in order to evaluate for instance the role of environmental heterogeneity or species co-distribution on network turnover.

We restricted our analysis to the effect of co-occurrence on ecological interactions, neglecting the inverse of the problem. We did not investigate in depth the drivers of co-occurrence and simply took it for granted from the data. Co-occurrence was indeed different from the expectation of independent species distributions. It thus begs the question of whether, once environmental effects on species-specific distribution have been accounted for, interactions come with significant effects on co-occurrence? We could rephrase this problem by asking whether the fundamental niche differs from the

realized niche, and how this applies to our framework. For example, we have considered above simply the co-occurrence probability, $P(X_{ij}, X_{jy} | E_y)$, which could be expanded as $P(X_{ij} | X_{jy}, E_y)P(X_{jy} | E_y)$. After some re-arrangement of eq. 2, the marginal occurrence probability, $P(X_{jy} | E_y)$, could be considered as a species distribution model taking into account the interaction between these species (in the same spirit as done by González-Salazar et al. 2013). This derivation would however critically depend on a strong *a priori* expectation of the conditional probability of observing a species given the distribution of the other species. This assumption seems reasonable for some situations, such as a parasitoid species that requires a host to develop. On the other hand, Cazelles (2016) found that the strength of this association is often rather weak if not inexistant (for instance, with the example pair presented in Table 1). The lack of an association could simply arise when the parasitoid is generalist enough that it is not constrained to track the distribution of any single/given host (Cazelles et al. 2015).

The approach we presented could easily be expanded to account for traits, in order to derive a complete and more mechanistic representation of the niche. In this context, both the Eltonian and Grinnellian dimensions of the niche could be represented. As mentioned earlier, Gravel et al. (2016) proposed to represent the occurrence of interactions based on some trait-matching rules (given by a function of the type $P(L_{ij}) | T_{ij}, T_{jy}, E_y$) and the co-distribution of traits (corresponding to $P(T_{ij}, T_{jy} | E_y)$). The key insight from the formalism proposed by Gravel et al. (2016) is that the co-distribution of traits will impact variation in network structure. As a result, it was found that network properties often co-vary with functional structure and functional diversity (Laigle et al. 2018). The next step in the development of this framework will be to develop the models for the co-distribution of traits, for instance following Ovaskainen et al. (2017). This may facilitate its accessibility and application, since the basic unit of analysis is no longer the species, but rather a combination of traits, thereby facilitating the statistical analysis by sharing information among species.

At present, there is only indirect support for the hypothesis that interacting species are conditionally distributed, but this possibility should be the topic of theoretical investigation with dynamical metacommunity models (Cazelles et al. 2015) and empirical hypothesis testing. The impact of ecological interactions on the distribution of co-occurrence has been the topic of many publications since Diamond (1975) seminal study on competition and ‘checkerboard’ distribution, but pairwise approaches have only recently received attention (Veech 2013). Whether two interacting species are more closely associated in space remains unclear, since most approaches based on null models consider community-level metrics (Gotelli 2000), such as the C-score, thereby making it hard to evaluate if specific interactions do indeed affect co-occurrence. The expansion of the framework we describe to account for the difference between the realized and the fundamental niche

will therefore require further investigation of the impact of interactions on co-occurrence.

Ecological networks are known to be extremely sparse, i.e. they have far more absences than presences of interactions. Absences of interactions, however, can come from different sources. The fact that unequal sampling at the local scale can affect our understanding of network structure is well documented (Martinez et al. 1999). In a spatial context, however, some interactions may be undocumented simply because the species involved have never been observed to co-occur. Although these cases are reported as a lack of interactions, in actuality we cannot make any reliable inference from them: since the species have never been observed together, it remains possible that they would interact if they did. A fundamentally different category of absences of interactions are then those reported after multiple observations of species co-occurrence. Thus, to gain confidence that the probability of an interaction is low, extensive sampling (that is, several records of co-occurrence) is needed. Generally, our confidence that the interaction is indeed impossible will increase with the number of observations of the species pair. Seeing that this is essentially a Bernoulli process (the probability that the species will interact given their presence), the breadth of the confidence interval is expected to saturate after a fixed number of observations, which can be set as a threshold above which a species pair has finally been observed “often enough”. For instance, if two species are observed together at $k=10$ locations but are never found interacting ($n=0$), this means that the maximum likelihood estimate for the probability of this binomial distribution is $p=0$. Using the Clopper-Pearson method to compute the 95% confidence interval, we find that the real value of the probability of the interaction occurring is somewhere between 0 and 0.31. This interval is rather large, indicating that a significant amount of sampling is required to have credible interaction networks. Fortunately, there are Bayesian methods which could be used to complement available information and reduce this uncertainty (Cirtwill et al. 2018).

Conclusion

Our representation of spatial variation of community structure offers a new approach for the study of the biogeography of ecological networks. We see the following key challenges and opportunities ahead in this exciting area of research:

1. New generation of network data. Investigating spatial variation of network structure will require high quality and highly replicated network data. We have investigated one the most comprehensive spatial network datasets we are aware of and nonetheless found immense gaps of knowledge in its resolution. Species richness accumulates much faster than observations of ecological interactions (Poisot et al. 2012). Each pair of species must be observed several times in order to obtain reliable estimates of their interaction probability.

2. Estimation of the reliability of interactions. We need quantitative tools to estimate the confidence intervals around inferred interaction probabilities, as well as estimators of the frequency of false absences. Bayesian methods are promising to that end because we could use information on the target species (e.g. if they are known as specialists or generalists) to provide prior estimates of the interaction probability.

3. From interaction probabilities to a distribution of network properties. Metrics are available to analyse the structure of probabilistic networks (Poisot et al. 2016b). These metrics are useful as first approximation, but they assume independence among interactions. This might not be the case in nature because of the role of co-occurrence and shared environmental requirements. We also need to better understand the distribution of network properties arising from probabilistic interactions.

4. Investigation of the environmental-dependence of ecological interactions. There is evidence that interactions can vary in space, but this problem has not been investigated in a systematic fashion. The paucity of currently available data precludes an extensive analysis of this question at present.

5. Effects of ecological interactions on co-occurrence. We have intentionally omitted the feedback of ecological interactions on co-occurrence in this framework. As abundance can impact the occurrence of interactions, and, conversely since interactions impact abundance (Canard et al. 2014), we could reasonably expect that interactions will also influence co-occurrence. Theory in this regard does exist for simple three-species modules (Cazelles et al. 2015), but its extension to entire co-occurrence networks will prove critical in the future, especially given the interest in using co-occurrence to infer ecological interactions (Morales-Castilla et al. 2015, Morueta-Holme et al. 2016).

Funding – This is a contribution to the working groups ‘continental-scale variation of ecological networks’ (supported by the Canadian Inst. for Ecology and Evolution) and the ‘next generation data, models, and theory working group’ (supported by the Santa Fe Inst., the Betsy and Jesse Fink Foundation, the ASU-SFI Center for Biosocial Complex Systems and NSF Grant PHY-1240192). DG also acknowledges financial support from NSERC-Discovery grant program and Canada Research Chair program.

Author contributions – DG, BB, JAD, NDM, TP, SAW, DBS and JMT contributed to the concepts; JPK, TN and TR contributed the data; DG developed the probabilistic approach and performed the analysis; DG, TP and TR wrote the first draft and all of the authors edited the manuscript.

References

Albouy, C. et al. 2014. From projected species distribution to food-web structure under climate change. – *Global Change Biol.* 20: 730–741.
 Allesina, S. et al. 2008. A general model for food web structure. – *Science* 320: 658–661.
 Anderson, M. J. et al. 2011. Navigating the multiple meanings of beta-diversity: a roadmap for the practicing ecologist. – *Ecol. Lett.* 14: 19–28.

Argus, G. 1997. Infrageneric classification of *Salix* (Salicaceae) in the New World. – *Syst. Bot. Monogr.* 52: 1–121.
 Bartomeus, I. et al. 2016. A common framework for identifying linkage rules across different types of interactions. – *Funct. Ecol.* 30: 1894–1903.
 Bascompte, J. and Jordano, P. 2007. Plant-animal mutualistic networks: the architecture of biodiversity. – *Annu. Rev. Ecol. Syst.* 38: 567–593.
 Baselga, A. 2010. Partitioning the turnover and nestedness components of beta diversity. – *Global Ecol. Biogeogr.* 19: 134–143.
 Baskett, C. A. et al. 2018. Latitudinal patterns of herbivore pressure in a temperate herb support the biotic interactions hypothesis. – *Ecol. Lett.* 21: 578–587.
 Bersier, L. F. et al. 2002. Quantitative descriptors of food-web matrices. – *Ecology* 83: 2394–2407.
 Blonder, B. et al. 2014. The n-dimensional hyper-volume. – *Global Ecol. Biogeogr.* 23: 595–609.
 Boulangeat, I. et al. 2012. Accounting for dispersal and biotic interactions to disentangle the drivers of species distributions and their abundances. – *Ecol. Lett.* 15: 584–593.
 Canard, E. et al. 2014. Empirical evaluation of neutral interactions in host-parasite networks. – *Am. Nat.* 183: 468–479.
 Cazelles, K. 2016. Influence des interactions biotiques sur la répartition géographique des espèces. – PhD thesis, Univ. du Québec à Rimouski, Rimouski.
 Cazelles, K. et al. 2016. A theory for species co-occurrence in interaction networks. – *Theor. Ecol.* 9: 39–48.
 Cazelles, K. et al. 2015. On the integration of biotic interaction and environmental constraints at the biogeographical scale. – *Ecography* 39: 921–931.
 Chase, J. and Leibold, M. 2003. *Ecological niches*. – Chicago Univ. Press.
 Chase, J. M. 2007. Drought mediates the importance of stochastic community assembly. – *Proc. Natl Acad. Sci. USA* 104: 17430–17434.
 Cirtwill, A. et al. 2018. A quantitative framework for investigating the reliability of network construction. – *bioRxiv*, doi:https://doi.org/10.1101/332536.
 Clark, J. S. et al. 2007. Resolving the biodiversity paradox. – *Ecol. Lett.* 10: 647–659.
 Connor, E. and Simberloff, D. 1979. The assembly of species communities: chance or competition? – *Ecology* 60: 1132–1140.
 Cronk, Q. et al. 2015. *Salix* transect of Europe: latitudinal patterns in willow diversity from Greece to arctic Norway. – *Biodivers. Data J.* 3: e6258.
 Diamond, J. 1975. Assembly of species communities. – In: Cody, M. and Diamond, J. (eds), *Ecology and evolution of communities*. Harvard Univ. Press, Cambridge, pp. 342–444.
 Dunne, J. A. 2006. The network structure of food webs. – In: Pascual, M. and Dunne, J. A. (eds), *Ecological networks: linking structure and dynamics*. Oxford Univ. Press, pp. 27–86.
 Eklöf, A. et al. 2013. The dimensionality of ecological networks. – *Ecol. Lett.* 16: 577–583.
 Gibert, J. P. and DeLong, J. P. 2014. Temperature alters food web body-size structure. – *Biol. Lett.* 10: 20140473.
 Gleason, H. 1926. The individualistic concept of the plant association. – *Bull. Torrey Bot. Club* 53: 7–26.
 Godsoe, W. et al. 2017. Integrating biogeography with contemporary niche theory. – *Trends Ecol. Evol.* 32: 488–499.
 González-Salazar, C. et al. 2013. Comparing the relative contributions of biotic and abiotic factors as mediators of species’ distributions. – *Ecol. Model.* 248: 57–70.

- Gotelli, N. J. 2000. Null model analysis of species co-occurrence patterns. – *Ecology* 81: 2606–2621.
- Götzenberger, L. et al. 2012. Ecological assembly rules in plant communities—approaches, patterns and prospects. – *Biol. Rev. Cambridge Phil. Soc.* 87: 111–127.
- Gravel, D. et al. 2016. The meaning of functional trait composition of food webs for ecosystem functioning. – *Phil. Trans. R. Soc. B* 371: 20150268.
- Gravel, D. et al. 2011. Trophic theory of island biogeography. – *Ecol. Lett.* 14: 1010–1016.
- Gravel, D. et al. 2013. Inferring food web structure from predator-prey body size relationships. – *Methods Ecol. Evol.* 4: 1083–1090.
- Gray, S. M. et al. 2016. Temperature and trophic structure are driving microbial productivity along a biogeographical gradient. – *Ecography* 39: 981–996.
- Hanski, I. 1999. *Metapopulation ecology*. – Oxford Univ. Press.
- Hardin, G. 1960. The competitive exclusion principle. – *Science* 131: 1292–1297.
- Harfoot, M. B. J. et al. 2014. Emergent global patterns of ecosystem structure and function from a mechanistic general ecosystem model. – *PLoS Biol.* 12: e1001841.
- Harris, D. J. 2016. Inferring species interactions from co-occurrence data with markov networks. – *Ecology* 97: 3308–3314.
- Havens, K. 1992. Scale and structure in natural food webs. – *Science* 257: 1107–1109.
- Hijmans, R. J. 2015. raster: Geographic data analysis and modeling R package ver. 2.5-2. – <<https://cran.r-project.org/web/packages/raster/index.html>>.
- Holt, R. D. 2009. Bringing the Hutchinsonian niche into the 21st century: ecological and evolutionary perspectives. – *Proc. Natl Acad. Sci. USA* 106: 19659–19665.
- Holt, R. D. and Lawton, J. H. 1993. Apparent competition and enemy-free space in insect host-parasitoid communities. – *Am. Nat.* 142: 623–645.
- Hutchinson, G. 1957. Concluding remarks. – *Cold Spring Harbour Symposium* 22: 415–427.
- Ings, T. C. et al. 2009. Ecological networks—beyond food webs. – *J. Anim. Ecol.* 78: 253–269.
- Kéfi, S. et al. 2012. More than a meal. Integrating non-feeding interactions into food webs. – *Ecol. Lett.* 15: 291–300.
- Kissling, W. D. et al. 2012. Towards novel approaches to modelling biotic interactions in multispecies assemblages at large spatial extents. – *J. Biogeogr.* 39: 2163–2178.
- Kopelke, J. et al. 2017. Food-web structure of willow-galling sawflies and their natural enemies across Europe. – *Ecology* 98: 1730–1730.
- Kopelke, J. P. 1999. Gallenerzeugende Blattwespen Europas – Taxonomische Grundlagen, Biologie und Ökologie (Tenthredinidae: Nematinae: Euura, Phyllocolpa, Pontania). – *Courier Forschungsinst. Senckenberg* 212: 1–183.
- Kopelke, J. P. 2003. Natural enemies of gall-forming sawflies on willows (*Salix* spp.). – *Entomol. Gen.* 26: 277–312.
- Kouki, J. 1999. Latitudinal gradients in species richness in northern areas: some exceptional patterns. – *Ecol. Bull.* 47: 30–37.
- Kouki, J. et al. 1994. Reversed latitudinal gradient in species richness of sawflies (Hymenoptera, Symphyta). – *Ann. Zool. Fenn.* 31: 83–88.
- Krebs, C. 2009. *Ecology: the experimental analysis of distribution and abundance*. 6th ed. – Pearson, San Francisco.
- Laigle, I. et al. 2018. Species traits as drivers of food web structure. – *Oikos* 127: 316–326.
- Laliberté, E. and Tylianakis, J. M. 2010. Deforestation homogenizes tropical parasitoid–host networks. – *Ecology* 91: 1740–1747.
- Legagneux, P. et al. 2014. Arctic ecosystem structure and functioning shaped by climate and herbivore body size. – *Nat. Clim. Change* 4: 379–383.
- Legendre, P. et al. 2005. Analyzing beta diversity: partitioning the spatial variation of composition data. – *Ecol. Monogr.* 75: 435–450.
- Legendre, P. and De Cáceres, M. 2013. Beta diversity as the variance of community data: dissimilarity coefficients and partitioning. – *Ecol. Lett.* 16: 951–963.
- Leibold, M. et al. 2004. The metacommunity concept: a framework for multi-scale community ecology. – *Ecol. Lett.* 7: 601–613.
- Leppänen, S. et al. 2014. A comparative analysis of genetic differentiation across six shared willow host species in leaf- and bud-galling sawflies. – *PLoS One* 9: e116286.
- Lurgi, M. et al. 2012. Novel communities from climate change. – *Phil. Trans. R. Soc. B* 367: 2913–2922.
- Martinez, N. D. et al. 1999. Effects of sampling effort on characterization of food-web structure. – *Ecology* 80: 1044–1055.
- McInerny, G. J. and Etienne, R. S. 2012a. Ditch the niche – is the niche a useful concept in ecology or species distribution modelling? – *J. Biogeogr.* 39: 2096–2102.
- McInerny, G. J. and Etienne, R. S. 2012b. Stitch the niche – a practical philosophy and visual schematic for the niche concept. – *J. Biogeogr.* 39: 2103–2111.
- McKinnon, L. et al. 2010. Lower predation risk for migratory birds at high latitudes. – *Science* 327: 326–327.
- Morales-Castilla, I. et al. 2015. Inferring biotic interactions from proxies. – *Trends Ecol. Evol.* 30: 347–356.
- Morueta-Holme, N. et al. 2016. A network approach for inferring species associations from co-occurrence data. – *Ecography* 39: 1139–1150.
- Nyman, T. and Julkunen-Tiitto, R. 2000. Manipulation of the phenolic chemistry of willows by gall-inducing sawflies. – *Proc. Natl Acad. Sci. USA* 97: 13184–13187.
- Nyman, T. et al. 2015. Determinants of parasitoid communities of willow-galling sawflies: habitat overrides physiology, host plant and space. – *Mol. Ecol.* 24: 5059–5074.
- Nyman, T. et al. 2010. How common is ecological speciation in plant-feeding insects? A ‘Higher’ Nematinae perspective. – *BMC Evol. Biol.* 10: e266.
- Ovaskainen, O. et al. 2017. How to make more out of community data? A conceptual framework and its implementation as models and software. – *Ecol. Lett.* 20: 561–576.
- Pellissier, L. et al. 2013. Combining food web and species distribution models for improved community projections. – *Ecol. Evol.* 3: 4572–4583.
- Petchey, O. L. et al. 2010. Predicting the effects of temperature on food web connectance. – *Phil. Trans. R. Soc. B* 365: 2081–2091.
- Peters, R. 1991. *A critique for ecology*. – Cambridge Univ. Press.
- Phillips, S. J. et al. 2006. Maximum entropy modeling of species geographic distributions. – *Ecol. Model.* 190: 231–259.
- Pillai, P. et al. 2011. Metacommunity theory explains the emergence of food web complexity. – *Proc. Natl Acad. Sci. USA* 108: 19293–19298.
- Poisot, T. and Gravel, D. 2014. When is an ecological network complex? Connectance drives degree distribution and emerging network properties. – *PeerJ* 2: e251.
- Poisot, T. et al. 2012. The dissimilarity of species interaction networks. – *Ecol. Lett.* 15: 1353–1361.

- Poisot, T. et al. 2015. Beyond species: why ecological interactions vary through space and time. – *Oikos* 124: 243–251.
- Poisot, T. et al. 2016a. mangal – making ecological network analysis simple. – *Ecography* 39: 384–390.
- Poisot, T. et al. 2016b. The structure of probabilistic networks. – *Methods Ecol. Evol.* 7: 303–312.
- Poisot, T. et al. 2016c. Synthetic datasets and community tools for the rapid testing of ecological hypotheses. – *Ecography* 39: 402–408.
- Poisot, T. et al. 2018. Hosts, parasites and their interactions respond to different climatic variables. – *Global Ecol. Biogeogr.* 26: 942–951.
- Pollock, L. J. et al. 2014. Understanding co-occurrence by modelling species simultaneously with a joint species distribution model (jsdm). – *Methods Ecol. Evol.* 5: 397–406.
- Pulliam, H. R. 2000. On the relationship between niche and distribution. – *Ecol. Lett.* 3: 349–361.
- Roininen, H. et al. 2005. Biology, ecology, and evolution of gall-inducing sawflies (Hymenoptera: Tenthredinidae and Xyelidae). – In: Raman, A. et al. (eds), *Biology, ecology and evolution of gall-inducing arthropods*. Science Publishers, pp. 467–494.
- Scheiner, S. M. and Willig, M. R. 2007. A general theory of ecology. – *Theor. Ecol.* 1: 21–28.
- Schleuning, M. et al. 2012. Specialization of mutualistic interaction networks decreases toward tropical latitudes. – *Curr. Biol.* 22: 1925–1931.
- Shurin, J. B. et al. 2012. Warming shifts top-down and bottom-up control of pond food web structure and function. – *Phil. Trans. R. Soc. B* 367: 3008–3017.
- Silvertown, J. 2004. Plant coexistence and the niche. – *Trends Ecol. Evol.* 19: 605–611.
- Simanonok, M. P. and Burkle, L. A. 2014. Partitioning interaction turnover among alpine pollination networks: spatial, temporal, and environmental patterns. – *Ecosphere* 5: 1–17.
- Skvortsov, A. K. 1999. Willows of Russia and adjacent countries. Taxonomical and geographical revision. – *Univ. Joensuu Fac. Math. Natl Sci. Rep. Ser.* 39: 1–307.
- Soberón, J. 2007. Grinnellian and eltonian niches and geographic distributions of species. – *Ecol. Lett.* 10: 1115–1123.
- Staniczenko, P. P. et al. 2017. Linking macroecology and community ecology: refining predictions of species distributions using biotic interaction networks. – *Ecol. Lett.* 20: 693–707.
- Stephens, C. and Heau, J. 2009. Using biotic interaction networks for prediction in biodiversity and emerging diseases. – *PloS One* 4: e5725.
- Stouffer, D. B. et al. 2007. Evidence for the existence of a robust pattern of prey selection in food webs. – *Proc. R. Soc. B* 274: 1931–1940.
- Thuiller, W. 2003. Biomod – optimizing predictions of species distributions and projecting potential future shifts under global change. – *Global Change Biol.* 9: 1353–1362.
- Tilman, D. 1982. Resource competition and community structure. – Princeton Univ. Press.
- Townsend, P. A. et al. 2011. Ecological niches and geographic distributions. – Princeton Univ. Press.
- Trøjsgaard, K. et al. 2015. Geographical variation in mutualistic networks: similarity, turnover and partner fidelity. – *Proc. R. Soc. B* 282: 20142925.
- Tylianakis, J. M. et al. 2007. Habitat modification alters the structure of tropical host–parasitoid food webs. – *Nature* 445: 202–205.
- Veech, J. A. 2013. A probabilistic model for analysing species co-occurrence. – *Global Ecol. Biogeogr.* 22: 252–260.
- Veijalainen, A. et al. 2013. Subfamily composition of Ichneumonidae (Hymenoptera) from western Amazonia: insights into diversity of tropical parasitoid wasps. – *Insect Conserv. Divers.* 6: 28–37.
- Vellend, M. et al. 2014. Assessing the relative importance of neutral stochasticity in ecological communities. – *Oikos* 123: 1420–1430.
- Williams, R. and Martinez, N. 2000. Simple rules yield complex food webs. – *Nature* 404: 180–183.
- Williams, R. J. et al. 2010. The probabilistic niche model reveals the niche structure and role of body size in a complex food web. – *PloS One* 5: e12092.
- Wisz, M. S. et al. 2013. The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. – *Biol. Rev.* 88: 15–30.
- Wood, S. A. et al. 2015. Effects of spatial scale of sampling on food web structure. – *Ecol. Evol.* 5: 3769–3782.
- Woodward, G. et al. 2010. Climate change and freshwater ecosystems: impacts across multiple levels of organization. – *Phil. Trans. R. Soc. B* 365: 2093–2106.
- Wu, J. et al. 2015. Phylogeny of *Salix* subgenus *Salix* s.l. (Salicaceae): delimitation, biogeography, and reticulate evolution. – *BMC Evol. Biol.* 15: e31.