

¹ **Title:** Bringing Elton and Grinnell together: a quantitative framework to represent
² the biogeography of ecological interactions
³ **Authors:** Dominique Gravel^{1,2,*}, Ben Baiser³, Jennifer A. Dunne⁴, Néo Martinez⁵,
⁴ Tommy Nyman⁶, Timothée Poisot^{2,7}, Tomas Roslin⁶, Spencer Wood⁸, Daniel B. Stouffer⁹,
⁵ Jason Tylianakis⁹

⁶
⁷ 1: Canada Research Chair in Integrative Ecology. Département de biologie, Uni-
⁸ versité de Sherbrooke, 2500 Boulevard l'Université, Sherbrooke (Québec). J1K 2R1

⁹
¹⁰ 2: Québec Centre for Biodiversity Sciences

¹¹
¹² 3: University of Canterbury at Christchurch, School of Biological Sciences,

¹³
¹⁴ 4:

¹⁵
¹⁶ 5:

¹⁷
¹⁸ 6:

¹⁹
²⁰ 7:

²¹
²² 8:

²³
²⁴ 9:

²⁵
²⁶ **Keywords:** networks, spatial ecology, co-occurrence, probability of interaction
²⁷

²⁸ **Words in the abstract:**

²⁹ **Words in the main text:**

³⁰ **Words in the legends:**

³¹ **Figures:**

³² **Tables:**

³³ **References:**

³⁴ **Abstract**

35 **Introduction**

36 Community ecology is defined in most textbooks as *the study of the interactions that*
37 *determine the distribution and abundance of organisms* (Krebs2001). Despite a general
38 concensus on this definition (Scheiner2007), it is surprising that most research on the
39 variation of community structure has focused mostly on the turnover of species com-
40 position (Anderson2011), neglecting variation in the way they interact with each other
41 (McGill2006). Not surprisingly given this fact, biogeographers are still struggling
42 to figure out if interactions do impact species distribution (Wisz2012; Kissling2012).
43 There has been a step ahead with recent methodological improvements accounting
44 for interactions in species distribution models (Pollock2014; Pelissier2013), but those
45 remain nonetheless a 'species- based' approach to communities where interactions are
46 fixed covariates affecting distribution. Here we plea for a more integrated approach
47 of species and interaction distribution.

48 The problem of community assembly is often formulated as *how do we sample a*
49 *regional pool of species to constitute a local community?* This question could be rewritten
50 to address the problem of network assembly, as *how do we sample a regional pool of*
51 *interactions to constitute a local interaction network?*. An illustration of this problem
52 for a food web is provided at Figure 1. The metaweb represents potential interactions
53 among all species that could be found in a given area. In this particular case, there are
54 274 nodes, and 1077 links among plants, herbivores and parasitoids from Northern
55 Europe. An instance of a local community is also illustrated, with 44 nodes and 160
56 interactions. Only 51.2% of all potential interactions are realized. Our objective here
57 is to provide a conceptual framework to explain the sampling of the regional pool
58 of interactions, along with a quantitative method to predict it. The problem could
59 be formalized by sequentially understanding why only a fraction of the species are
60 locally co-occurring, and after why these species are interacting or not.

61 There are multiple causes to spatial turnover in community composition. The first
62 and most studied driver is the effect of variation in the abiotic environment on species
63 performance. Combined with specific responses in demography, it generates varia-
64 tion among localities by locally selecting the fittest species (Leibold2004). Stochastic-
65 ity additionnaly plays a role, either because of the inherently unpredictable nature of
66 colonization and extinction events, or because of strong non-linear feedbacks generat-
67 ing alternative transients and equilibrium (Chase2007; Vellend2014). The analysis of
68 community turnover is usually performed with data represented in a table with rows
69 corresponding to sites (or measurements) and the columns to species. Beta- diversity
70 metrics quantify the variance of this community data (Legendre2005). Traditionnal
71 approaches rely on measures of the dissimilarity between communities, using indices
72 such as the Jaccard or the Bray Curtis measures of dissimilarity. Recent methods de-
73 compose the total variation of the community data into species and site contributions
74 to beta diversity (Legendre2013). Even though these methods compare whole list of
75 species between sites, or measurements, they remain fundamentally 'species-based'
76 since they are all based on the within column variation. None of them consider the
77 variation of associations (pairs or higher order motifs) explicitly.

78 The niche is by far the dominant concept to explain species distribution and com-
79 munity assembly, from the local to the global scale. Following Hutchinson, the niche
80 is viewed as the set of environmental conditions allowing a population to establish
81 and maintain a population (see also Holt2009). Community turnover arises following
82 successive species replacement along an environmental gradient, in agreement with
83 a Gleasonian view of communities (Gleason). The concept is straightforward to op-
84 erationalize with species distribution models, as it maps naturally on the available
85 data (both distribution and environmental data) and a vast array of statistical tools
86 representing it (e.g. Biomod, MaxEnt). It is however much harder to account for

87 ecological interactions in this approach (Peterson2011). These are often viewed as ex-
88 ternalities, constraining or expanding the range of environmental conditions required
89 for a species to maintain a population (Pulliam2000; Soberon2007).

90 The network approach proposes a convenient formalism to represent the structure
91 of local communities. Species are represented as nodes and interactions by links. The
92 data could also be represented by matrices, with each species in rows and columns,
93 and the entries representing the occurrence or the intensity of an interaction. Stud-
94 ies of network diversity are mostly concerned by the distribution of the interactions
95 within a location and not so much by the variation among locations (Dunne2005; Bas-
96 compte2007; Ings2007; Kefi2012). Network complexity is computed as the number
97 of interactions in the case of binary networks or interaction diversity in the case of
98 quantitative networks (Bersier2002). However, there is now evidence that ecological
99 interactions do also vary in space and time (Poisot2012; Trojelsgaard2015). The vari-
100 ability of community structure in this situation arises from the turnover of species
101 composition, along with the turnover of the interactions among pairs of species.
102 The occurrence and intensity of interactions could vary because of the environment,
103 species abundance and higher order interactions (Poisot2015a). The variation in com-
104 munity composition is often independent of the variation of ecological interactions,
105 suggesting these two components of network variability respond to different drivers
106 (Poisot2012).

107 Interestingly, the ecological network literature also has its own 'niche model' to
108 position a species in a community (Williams2000). The niche of a species in this
109 context represents the multidimensional space that could represent all of its interac-
110 tions. Each species is characterized by a niche position, an optimum and a range
111 over 3 to 5 different niche axes (Williams2000; Eklof2013). The niche model has been
112 successful at explaining the complexity of a variety of networks, from food webs to

¹¹³ plant-pollinator systems. The conceptual framework is however limited to local com-
¹¹⁴ munities and does not provide any explanation to the turnover of network structure
¹¹⁵ along environmental gradients.

¹¹⁶ Here we adopt the view that a community structure is best represented as an
¹¹⁷ ecological network of interactions and develop a theory to explain its turnover in
¹¹⁸ space and time. We propose a new description of the niche that integrates the effect
¹¹⁹ of the environment on species distribution and on ecological interactions. We first
¹²⁰ present the conceptual framework and then formalize it mathematically with a prob-
¹²¹ abilistic approach to the sampling of the regional pool of interactions. We apply the
¹²² framework to study the spatial variation of host-parasite interactions across Europe.
¹²³ We find that the variation of the environment causes both species and interaction
¹²⁴ turnover. The network structures changes systematically across the latitudinal gra-
¹²⁵ dient, with a peak of connectance at intermediate latitudes. At the pairwise level,
¹²⁶ the statistical approach could be conceived as an interaction distribution model. At
¹²⁷ the community level, the approach provides a likelihood based method to compare
¹²⁸ different hypotheses of network turnover.

¹²⁹ **The integrated niche**

¹³⁰ Correctly describing the niche is key to understand turnover in community structure.
¹³¹ Despite several attempts to refresh the conceptual basis of what ecological niches
¹³² are, ecologists have not moved far past the "n-dimensional hypervolume" formalism
¹³³ introduced by Hutchinson. Despite its intuitive interpretation and translation into
¹³⁴ species distribution models (Boulangeat2012; Blonder2014), the concept has been con-
¹³⁵ stantly criticized (Hardin1960; Peters1991; Chase2003; Silvertown2004; Soberon2007)
¹³⁶ and several attempts have been made to expand and reinforce it.

137 Part of the problem surrounding the definition of the niche has been clarified
138 with the distinction between Eltonian and Grinnellian definitions (Chase2003). The
139 Grinnellian dimension of the niche is the set of environmental conditions required
140 for a species to maintain a population in a location. The Grinnellian niche is the most
141 intuitive to apply and is the conceptual backbone of species distribution models. The
142 Eltonian niche on the other hand is the effect of a species on its environment. This
143 aspect of the niche is well known by community ecologists, but is trickier to turn into
144 predictive models. Nonetheless, the development of the niche model of food web
145 structure (Williams2000) and its parameterization (Williams2010; Gravel2013) made it
146 more operational.

147 These perspectives are rather orthogonal to each other and thus lead to consider-
148 able confusion in the literature (McIntyre). Chase2003 attempted to reconcile them
149 in their definition of the niche: *[The niche is] the joint description of the environmental*
150 *conditions that allow a species to satisfy its minimum requirements so that the birth rate of a*
151 *local population is equal or greater than its death rate along with the set of per capita effects*
152 *of that species on these environmental conditions.* Their representation merges zero-net
153 growth isolines figuring the Grinnellian niche (when do the population persists) with
154 impact vectors figuring the Eltonian niche (what is the per capita impact). While this
155 representation has been very influential in community ecology at the local scale (the
156 resource-ratio theory of coexistence - Tilman1982), it remains impracticable at the
157 large spatial scale because of the difficulties to measure it. The absence of any math-
158 ematical representation of the niche that could easily be fit to ecological data perhaps
159 explain why biogeographers are still struggling to develop species distribution mod-
160 els taking into account ecological interactions.

161 We propose to integrate the two perspectives of the niche with a visual repre-
162 sentation of both components. The underlying rationale is that, in addition to the

163 environmental constraints on demographic performance, any organism requires re-
164 sources to sustain its metabolic demand and reproduction. Abiotic environmental
165 axes are any non-consumable factors affecting the demographic performance of an
166 organism. Alternatively, the resource axes are traits of the resources allowing inter-
167 actions with the consumer. The niche should therefore be viewed as the set of abiotic
168 environmental factors (the Grinnelian component) along with the set of traits (the
169 Eltonian component) allowing a sustainable population to establish and maintain at
170 a location. Accordingly, each species could be characterized by an optimal position
171 in both the environmental (x-axis) and the trait (y-axis) plane. The integrated niche is
172 then the hypervolume where interactions could occur and sustain a population. This
173 approach radically change the representation of the niche, putting species distribution
174 and ecological interactions in the same formalism.

175 The limits of the niches could be independent of each other (as in the exam-
176 ple at Fig. 2), alternatively interact. For instance, the optimimal prey body size
177 for predatory fishes could reduce with increasing temperature (Lelong2015), which
178 would make diet boundaries functions of the environment. The other way around,
179 we could also consider that the growth rate of the predator could change with the
180 body size of the preys it feeds on, thereby altering the environmental boundaries.

₁₈₁ **A probabilistic representation of ecological interactions**

₁₈₂ **networks in space**

₁₈₃ We now formalize the integrated niche with a probabilistic approach to interactions
₁₈₄ and distribution. We seek to represent the probability an interaction between species i
₁₈₅ and j occurs at location y . We define L_{ijy} as a stochastic variable and are looking at the
₁₈₆ probability this event occurs, $P(L_{ijy})$. The occurrence of an interaction is dependent on
₁₈₇ the co-occurrence of species i and j . This argument might seem trivial at first, but the
₁₈₈ explicit consideration of this condition in the probabilistic representation of ecological
₁₈₉ interactions will prove fundamental to understand their variation. We thus define X_{iy}
₁₉₀ as a stochastic variable representing the occurrence of a species i at location y , and
₁₉₁ similarly X_{jy} the co-occurrence of species i and j . The quantity we seek to understand
₁₉₂ is the probability of a joint event:

$$P(X_{i,y}, X_{j,y}, L_{ij,y}) \quad (1)$$

₁₉₃ Or simply said, the probability of observing both species i and j , and an interaction
₁₉₄ from i to j . This probability could be decomposed in two parts using the product rule
₁₉₅ of probabilities:

$$P(X_{iy}, X_{jy}, L_{ijy}) = P(X_{iy}, X_{jy}|E_y)P(L_{ijy}|\mathbf{T}, X_{iy}, X_{jy}, E_y) \quad (2)$$

₁₉₆ The left term is the probability of observing the two species co- occurring at lo-
₁₉₇ cation y . It corresponds to the Grinnelian dimension of the niche. The right term
₁₉₈ is a conditional probability, representing the probability that an interaction occurs
₁₉₉ between species i and j , given their set of traits \mathbf{T} and they are co- occurring. It
₂₀₀ is referred as the metaweb and corresponds to the Eltonian dimension of the niche

described above. We will see below how this formalism could be directly fitted to empirical data. 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 about the temporal and spatial variation in network structure (see the explicit formulations at Table 1). The simplest model relates co-occurrence probability directly to the environment, $P(X_{ijy}|E_y)$. In this situation there is no underlying assumption about the ecological processes responsible for co-occurrence. It could arise because of the impact of ecological interactions on distribution (Pollock2014) or alternatively because of share environmental requirements. In the former case, species are not independent to each other and the conditional occurrence must be accounted for explicitly, $P(X_{ijy}|E_y) = P(X_i|E_y, X_j)P(X_j|E_y)$. In the later case, species are independent and only the marginal occurrence must be accounted for, $P(X_{ijy}|E_y) = P(X_i|E_y)P(X_j|E_y)$.

The co-occurrence probability itself could be dependent on ecological interactions. Direct pairwise interactions such as competition, facilitation and predation have long been studied for their impact on co-distribution (such as in the cases studied by Diamond1976, Connor1980, Gotelli2000 etc.). Second and higher order interactions (e.g. trophic cascade) could also impact co-occurrence. Co-occurrence in ecological networks is however a topic of its own, influenced by the degree distribution and species richness (Cazelles2015). Almost only first order and second order interactions do impact co-occurrence. The covariance of interacting species to an environmental gradient also influences co-occurrence (Cazelles2016). Because of the complexity of relating co-occurrence to the interaction network structure, we will focus here on the

226 variation of interactions and not on the distribution, and leave this issue for Discus-
227 sion and future research.

228 **Variants of the metaweb**

229 There are also variants of the metaweb. First, most documented metawebs have thus
230 far considered that ecological interactions are deterministic, not probabilistic (e.g.
231 Havens1992; Woods2015). Species are assumed to interact if they are found together
232 in a location, independently of their abundance and the environment. In other words,
233 $P(L_{ijy}) = 1$ if $X_{ijy} = 1$, and 0 otherwise. This approach might be a reasonable ap-
234 proximation when the sampling and inference scales are large enough so that prob-
235 abilities of observing at least one interaction converges to unity and that the only
236 variation of networks considered arises from species distribution.

237 Ecological interactions could also vary with the environment, such as $P(L_{ijy}|E_y)$.
238 Although it is not common to see a conditional representation of ecological interac-
239 tions, experimental studies of pairwise interactions revealing their sensitivity to the
240 environment are common (REF). For instance, it has been documented that the pre-
241 dation risks of shorebirds do vary at the continental scale, from the south to the north
242 (REF). The effect of the environment on interactions propagate up the community
243 and influence network structure (REF).

²⁴⁴ **Application: continental-scale variation of host-parasite**

²⁴⁵ **community structure**

²⁴⁶ In this section we provide an illustration of the framework with an empirical dataset
²⁴⁷ of host-parasitoid networks sampled throughout continental Europe along a south-
²⁴⁸ north gradient. The analysis targets networks composed of willows (genus *Salix*),
²⁴⁹ their galling insects, and the natural enemies of these gallers. The questions we
²⁵⁰ address with the framework and the dataset are: i) how much variation in the network
²⁵¹ structure is there across the gradient and ii) what is the primary driver of network
²⁵² turnover across the gradient?

²⁵³ **Data**

²⁵⁴ Communities of both willows and gallers are species-rich and widely distributed,
²⁵⁵ with pronounced variation in community composition across space. The genus *Salix*
²⁵⁶ includes over 400 species, most of which are shrubs or small trees (ref). The genus
²⁵⁷ is common in most habitats across the Northern Hemisphere (ref). Willows support
²⁵⁸ a highly diverse community of herbivorous insects, and one of the main herbivore
²⁵⁹ groups in this system are gall-inducing sawflies (Hymenoptera: Tenthredinidae: Ne-
²⁶⁰ matinae: Euurina (ref). Gall formation is induced by sawfly females during ovipo-
²⁶¹ sition, and gall formation includes marked manipulation of plant chemistry by the
²⁶² galler (ref). The enemy community of the gallers includes nearly 100 species belong-
²⁶³ ing to 17 insect families of four orders. These enemies encompass both inquilines and
²⁶⁴ parasitoids: inquiline larvae (Coleoptera, Lepidoptera, Diptera, and Hymenoptera)
²⁶⁵ feed primarily on gall tissue, but typically kill the galler larva in the process. Par-
²⁶⁶ asitoid larvae (representing many families in Hymenoptera) kill the galler larvae by
²⁶⁷ direct feeding (REF). In terms of associations between the trophic levels, phylogeny-

268 based comparative studies have demonstrated that galls represent "extended pheno-
269 types" of the gallers, meaning that gall form, location, and chemistry is determined
270 mainly by the galling insects and not by their host plants [ref]. Because galler par-
271 asitoids have to penetrate a protective wall of modified plant tissue in order to gain
272 access to their victims, gall morphology has been inferred to strongly affect the asso-
273 ciations between parasitoids and hosts (ref, ref). Thus, the set of parasitoids attacking
274 each host is presumptively constrained by the form, size, and thickness of its gall.

275 Local realizations of the willow–galler–enemy network were reconstructed from
276 samples collected between 1982 and 2010. Plant galls were collected by J.P. Kopelke
277 during this period at 374 sites across Central and Northern Europe. Sampling was
278 conducted in the summer months of June and/or July, during the latter stages of
279 larval development. Galler species were identified on the basis of willow host species
280 and gall morphology, as these are distinct for each sawfly species. At each site, galls
281 were randomly collected from several willow individuals in an area of about 0.1-
282 0.3 km². Most sites were visited only once, with a total of 641 site visits across the
283 374 sites. GPS coordinates were recorded for each location and the annual mean
284 temperature and annual precipitation were obtained from WorldClim. While other
285 covariates could have been considered, we figured that they are representative of
286 the most important axes of the European climate, and more easily interpretable than
287 reduced variables obtained by a PCA,

288 The methods used for rearing natural enemies from the galls have been previously
289 described by e.g. [ref, ref]. In brief, galls were opened to score the presence of galler
290 or parasitoid/inquiline larvae. Enemy larvae were classified to preliminary morphos-
291 pecies, and the identity of each morphospecies was determined by connecting them
292 to adults emerging after hibernation. The galls were reared by storing single galls in
293 small glass tubes (Kopelke 1985a, 1994a, 1999, 2003a, b). Hibernation of galls con-

294 taining parasitoids took place either within the glass tubes or between blotting paper
295 in flowerpots filled with clay granulate or a mixture of peat dust and sand. These
296 pots were stored over the winter in a roof garden and/or in a climatic chamber. In
297 most cases, the matching of larval morphospecies with adult individuals emerging
298 from the rearings allowed the identification of the natural enemies to the species
299 level. Nonetheless, in some cases, individuals could only be identified to one of the
300 (super)families Braconidae, Ichneumonidae, and Chalcidoidea. This was particularly
301 the case when only remains of faeces, vacant cocoons of parasitoids, and/or dead
302 host larvae were found, as was the case when parasitoids had already emerged from
303 the gall. As a result, the largest taxon in the data set, "Chalcidoidea indeterminate,"
304 represents a superfamily of very small parasitoids that are hard to distinguish.

305 In total, the sampling targeted 52 Salix taxa, yielding 146,622 galls for dissection
306 and rearing. These galls represented 96 galler species, and yielded 42,133 individually-
307 identified parasitoids and inquilines. Of these, 25,170 (60%) could be identified to the
308 species level. Overall, 126 parasitoid and inquiline taxa were distinguished in the ma-
309 terial. Data on host associations within subsets of this material have been previously
310 reported by (Kopelke ref ref b) and by Nyman et al. 2007, whereas the current study
311 represents the first analysis of the full data set from a spatial perspective.

312 Analysis

313 Computing the probability of observing an interaction involves fitting a set of bino-
314 mial models and collecting their estimated probabilities. We considered second-order
315 generalized linear models for the sake of the illustration, but more sophisticated fit-
316 ting algorithms (e.g. GAM or Random Forest) could be used as well, as long as the
317 algorithm can provide an estimated probability for each observation. The data con-
318 sists of a simple (albeit large) table with the observation of each species, X_i and X_j ,

319 their co-occurrence, X_{ij} , the observation of an interaction L_{ij} , and environmental co-
320 variates. There is one row per pair of species per site. We considered that an ab-
321 sence of a record of an interaction between co-occurring species at a site means a true
322 absence.

323 We compared three models for the co-occurrence probability. The first model fits
324 directly the co-occurrence probability conditional on the local environment, $P(X_{ijz}|E_z)$.
325 This model therefore does not make any hypothesis about the mechanisms driving
326 co-occurrence for any given environment and use the information directly available
327 in the data. The second models consider independent co-occurrence of species. In
328 this case, we fit independently $P(X_{iz}|E_z)$ and $P(X_{jz}|E_z)$ and take their product to
329 derive co-occurrence probability. This model should be viewed as a null hypothesis
330 with respect to the first model since their comparison reveals if there is significant
331 spatial association of the two species once considering the response to the environ-
332 ment (Cazelles2016). Finally the third model considers that co-occurrence probability
333 is independent of the environment and thus constant throughout the landscape. In
334 other words, $P(X_{ijz})$ is obtained by simply counting the number of observed co-
335 occurrences, divided by the total number of observations. The comparison between
336 the first and third model allows testing the hypothesis that co-occurrence is condi-
337 tional on the environment. Whenever the environment is included as a co-variate in
338 the glm, we considered a second-order polynomial response for both the temperature
339 and precipitations. There are consequently 5 parameters for the first model when
340 fitting a given pair of species, 10 parameters for the second model and only 1 for the
341 third model.

342 Following the same logic, we compared three models for the interaction prob-
343 ability. The first model fits the interaction probability conditional on the local en-
344 vironmental variables, $P(L_{ijz}|E_z, X_i, X_j)$. Consequently, the model was fitted on a

subset of the data, when the two species are found co- occurring. The second model fits the interaction probability independently of the local environmental variables, $P(L_{ijz}|X_i, X_j)$. It corresponds to the number of times the two species were observed interacting when co-occurring, divided by the number of times they co-occurred. The third model is an extreme case performed only to test the hypothesis that if two species are found to interact at least once, then they should interact whenever they co-occur, $P(L_{ijz}|X_i, X_j) = 1$. While not necessarily realistic, this model tests an hypothesis that is commonly done in the representation of local networks from knowledge of a deterministic metaweb (such as in Havens1992; Piechnik2008; Wood2015). There are consequently 5 parameters for the first model, a single parameter for the second model and no parameter to evaluate for the third model (the interaction probability is fixed by hypothesis).

The different models were fitted to each pair of species and the fitted probabilities were recorded. The joint probability $P(L_{ij}, X_i, X_j)$ was then computed from Eq. 2 and the likelihood of each observation was computed as $L(\theta|D) = P(L_{ij}, X_i, X_j)$ if an interaction was observed and $L(\theta|D) = 1 - P(L_{ij}, X_i, X_j)$ if no interaction was observed. The log- likelihood was summed over the entire dataset to compare the different models by AIC. Not surprisingly, it was impossible to compute this model for a very large number of pairs of species because they never co-occurred. These pairs were removed from the analysis because the co-occurrence probability is null and we have no information for the interaction probability. The reported likelihood across the entire dataset is summed over all pairs of species that were observed co- occurring at least one time. We considered separately the salix-galler and the galler-parasitoid interactions.

Finally, we used the full model (both the co-occurrence and the interaction are conditional on the environment) interpolate species distribution and the interaction

371 probability across the entire Europe. We reconstructed the expected network for each
372 location in a 1km X 1km grid. We then after computed the probabilistic connectance
373 following Poisot2015b.

374 All of the data are openly available in the database *mangal* (Poisot2015) and all R
375 scripts for querying and pre-processing the data, along with the analysis are provided
376 in supplementary material.

377 Results

378 We start with an example for a single pair of species that we selected because of a
379 sufficiently large number of times they were found co-occurring ($N_{ij} = 41$). Despite
380 the extent of the sampling, many pairs of species are found co-occurring only a few
381 times, making it hard to evaluate interaction probabilities with a reasonable confi-
382 dence interval. This particular example involves the interaction between *Phyllocolpa*
383 *plicalapporum* and *Pediobius saulius*, two fairly abundant species, observed respec-
384 tively $N_i = 53$ and $N_j = 129$ times across the 374 sites. These two species are found
385 interacting with marginal probability $P(L_{ij}) = 0.73$, which means they were found
386 interacting at 30 different locations. The model comparison (Table 1) reveals that the
387 interaction probability conditional on the co- occurrence better explain the distribu-
388 tion (Model 1 vs Model 2). The probabilistic representation of the metaweb yields
389 a much better fit to the data than the deterministic version. When the two species
390 co-occur, the occurrence of the interaction is insensitive to the environment (Model
391 2 vs Model 3). Alternatively, climatic variables significantly impact co- occurrence
392 (Model 3 vs Model 4). The neutral model performs worst than the non-random co-
393 occurrence model (Model 3 vs Model 6). The full model reveals that the greatest
394 interaction probability occurs at intermediate temperature and precipitations, simply
395 because this is where the two species are found co- occurring the most often (Fig.

396 3). The co-occurrence and the interaction probabilities could be represented in space,
397 where we find that the highest interaction probability occurs in central Europe (Fig.
398 4).

399 We did evaluated each model for all pairs of species in order to better understand
400 the large scale drivers of network turnover. Salix-gallers and gallers-parasitoids were
401 analyzed separately (Table 2). The results are comparable, albeit some very minor
402 details. We do find that across all pairs of species, the probabilistic representation of
403 interactions again does better than the deterministic (Model 1 vs Model 2). Interac-
404 tions do not happen systematically whenever the two species are found co-occurring,
405 meaning that the stochastic nature of interactions contribute to network variability
406 in addition to species turnover. There are 1077 recorded pairs of interactions, with
407 only 224 of them occurring less than 5 times. Out of these 224 interactions, only 77
408 are found systematically whenever the two species do co-occur. Even though inter-
409 actions are better represented probabilistically, the two environmental variables that
410 were considered are pretty poor predictors of their occurrence (Model 2 vs Model
411 3). Not surprisingly, the likelihood increases for both types of interactions when the
412 environment is considered. The extra number of parameters however exceed the gain
413 in likelihood, and therefore the best model excludes the effect of the environment.

414 According to the log-likelihood only, the co-occurrence is non-neutral for salix-
415 galler interactions, while it is neutral for the galler-parasitoid interaction. However,
416 the gain in log-likelihood for the neutral model of galler- parasitoid co-occurrence is
417 inferior to the extra number of parameters (twice as many since two species distribu-
418 tion models are fitted instead of just one), which has for consequence that the best
419 model according to AIC has non-random co-occurrence (Model 3 vs Model 6), for
420 both types of interactions.

421 The approach we present not only has implications for understanding the bio-

422 geography of pairwise interactions and interaction networks, but also the quality of
423 the evaluation of metawebs. We investigated the reliability of the estimated metaweb
424 across the entire dataset. As mentionned above, across the 32 412 pairs of species,
425 only 1077 pairs are interacting at least at a single location, for a connectance of 0.03.
426 However, only 8437 species are found co- occurring at least one time across all lo-
427 cations. There are consequently 23975 gaps of information in the metaweb (74.0%
428 - see Fig. 5). Given that we do not know if the non co-occurring species do in-
429 deed co-occur, it means that a more appropriate estimate of connectance would be
430 $C = 1077/8437 = 0.128$. This result reveals that the evaluation of the sampling
431 quality of ecological networks is a problem on its own that worths further attention.

432 Once we selected the best model (model 3, Table 2), we were able to reconstruct the
433 expected species richness across Europe, along with the most likely network for each
434 location, and therefore map the expected distribution of network properties (Fig. 6).
435 We simply considered connectance, as it could be easily computed from probabilistic
436 networks (Poisot2015b) and is also a good proxy for many other network properties
437 (Poisot2014). The diversity of *Salix* tends to increase toward boreal areas, and we
438 consequently find a peak in diversity in northern Europe. The distribution of the
439 expected number of interactions follows the distribution of species richness, but not
440 at the same rate. Consequently, connectance is peaking in central Europe and in
441 England.

442 **Discussion**

443 **Summary of the framework**

444

445 **Toward interaction distribution models**

446

447 **What are the drivers of network variation in space?**

448

449 **Forecasting network structure under global change**

450

451 **Investigating the realized niche and the impact of biotic interactions on distri-**
452 **bution**

453

454 **Guidance for empirical studies**

455

456 **Conclusion**

457 **Research agenda on probabilistic interactions**

458

459 • Need a new type of data

460 • Need to investigate the reliability of interactions

461 • From interaction probabilities to a distribution of network properties

462 • Deeper investigations of the environmental dependence of interactions

463 • Trait-based approach to interactions

464 • The effect of interaction on co-occurrence

⁴⁶⁵ Acknowledgements

⁴⁶⁶ This is a contribution to the working groups *Networks over ecological gradients* (Santa
⁴⁶⁷ Fe Institute) and *Continental-scale variation of ecological networks* (Canadian Institute for
⁴⁶⁸ Ecology and Evolution).

469 **Figure legends**

470 **Figure 1**

471 **Non-random sampling of the metaweb.** Network assembly could be viewed as a
472 sampling process of the regional pool of potential interactions. Species are sampled
473 first (indicated by colored nodes) and among the present species in the local network,
474 only some interactions are occurring (indicated by colored links). The challenge we
475 address with the quantitative framework proposed is to adequately characterize this
476 sampling process. The sampling of the metaweb is illustrated with a local interaction
477 network among Salix, gallers and parasitoids. Here, the metaweb was constructed
478 by aggregating observed interactions across 374 local networks. The color nodes
479 represent the species that were found in the most diverse of these 374 local networks.

480 **Figure 2**

481 **Visual representation of the integrated niche.** We represent visually the integration
482 between two views of the niche. In biogeography, the niche is considered the set of
483 environmental conditions where the intrinsic growth rate r is positive. The horizontal
484 axis represents an environmental gradient impacting the growth of the focal species
485 (in red). The location of the different species along this gradient represent their op-
486 timum, and the vertical dotted lines represent the limits of the [Grinnelian] niche of
487 the focal species. In food web ecology, the [Eltonian] niche represents the location of
488 a species in the food web, as determined by its preys and its predators. The vertical
489 axis represents a niche gradient, presumably a trait such as body size. The location of
490 each species along this gradient represent their niche position. The focal species will
491 feed on the different preys whose niche location falls within a given interval around
492 the optimum, represented by the horizontal dotted lines. The integrated niche [Grin-

nelian & Eltonian] corresponds to the square in the middle where an interaction is possible. According to our probabilistic framework, the central square represents the area where the joint probability of observing interactions and co-occurrence is positive.

Figure 3

Probabilistic representation of the interaction probability between a leaf galler (*Phyllocolpa plicalapporum*) and a parasitoid (*Pediobius saulius*) across a temperature and a precipitation gradient. The representation is based on predictions from model 3 (see Table 1). For the left panel: open circles represent the absence of both species or of an interaction, the closed circles represent co-occurrence and other symbols the occurrence of only one of the two species. For the other two panels the open circles represent co-occurrence but an absence of interaction and the closed circles represent the occurrence of an interaction.

Figure 4

Probabilistic representation of the interaction probability between a leaf galler (*Phyllocolpa plicalapporum*) and a parasitoid (*Pediobius saulius*) across Europe.

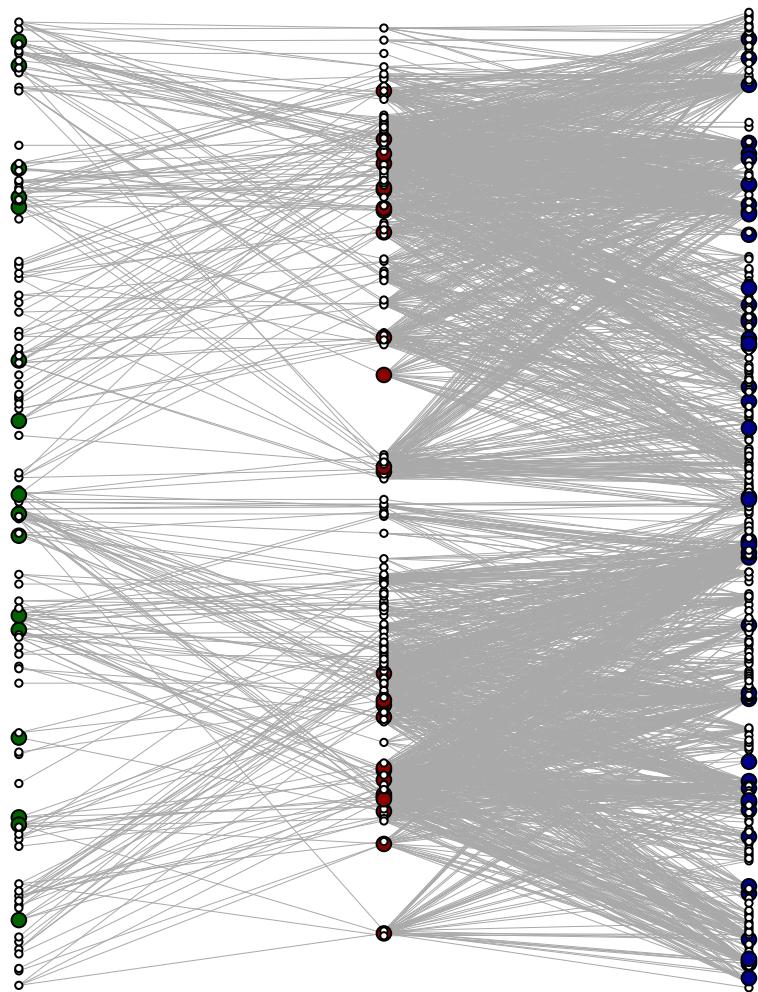
Figure 5

Representation of the A) metaweb and B) gaps of data in the metaweb. The salix and gallers were regrouped as 'victims' for the sake of the illustration, as the gallers and the parasitoids were regrouped as 'ennemis'.

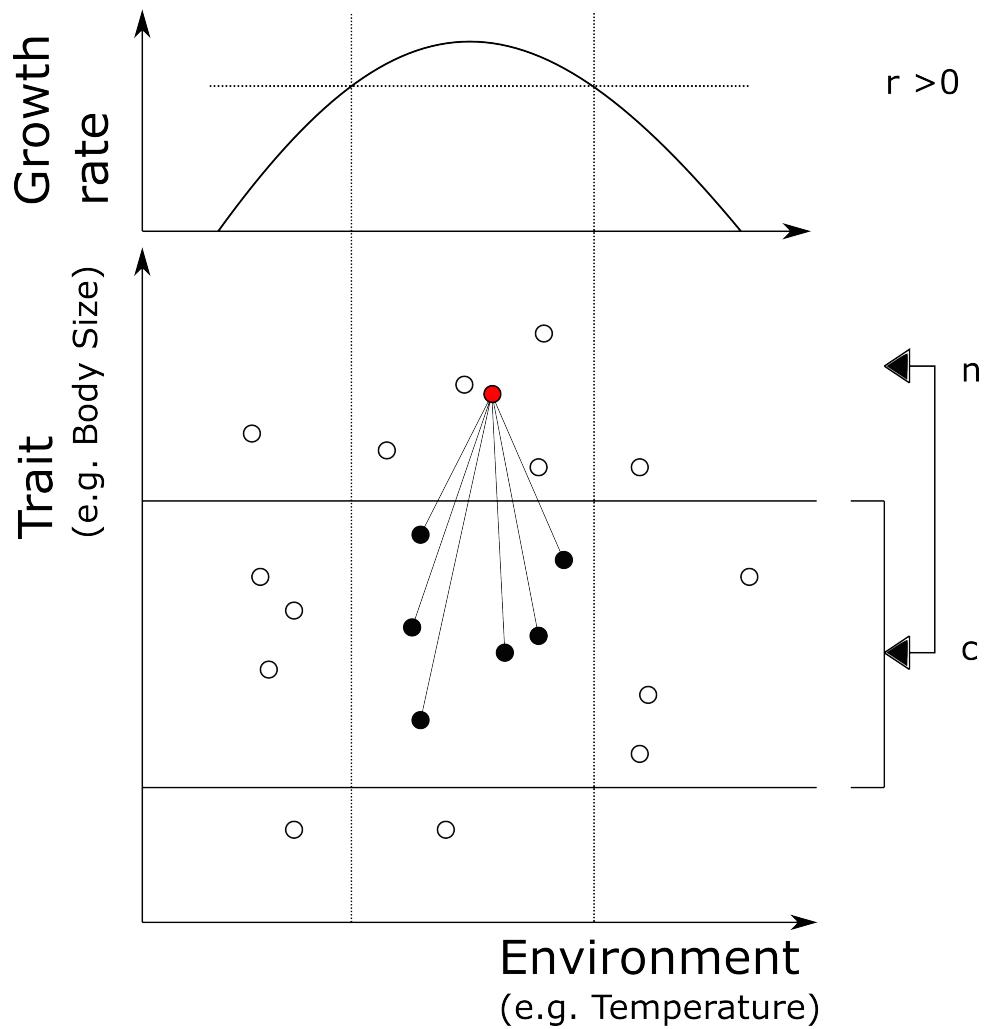
513 **Figure 6**

514 **Mapping the distribution of species richness, the number of links and connectance**
515 **across Europe.** The representation is based on predictions from model 3 (see Table
516 2).

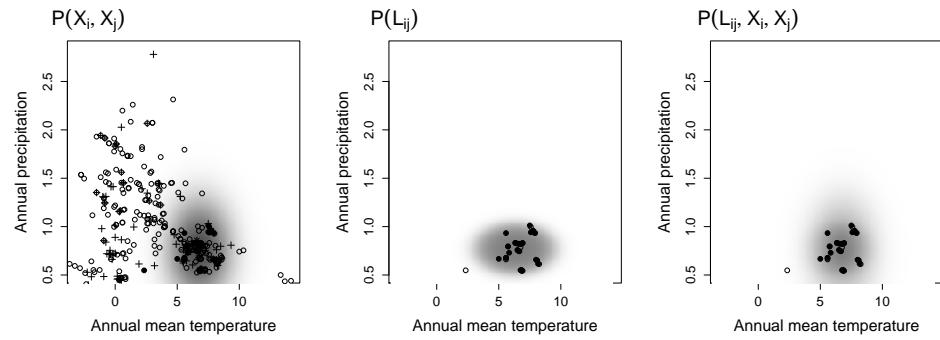
517 **Figure 1**



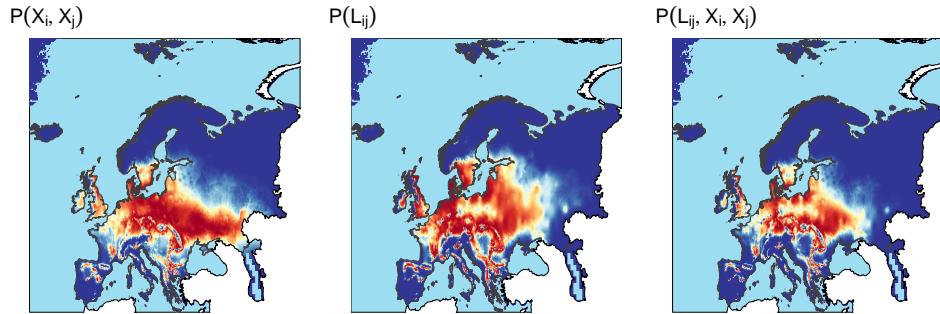
518 **Figure 2**



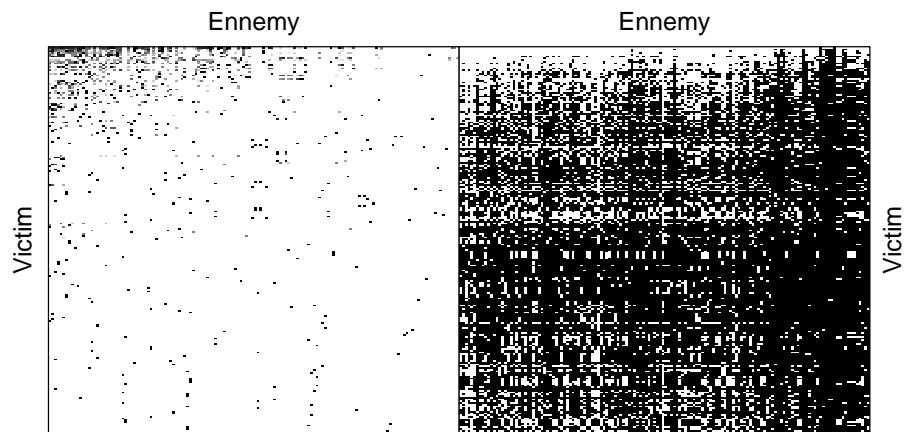
519 **Figure 3**



520 **Figure 4**



521 **Figure 5**



522 **Figure 6**

