

¹ **Title:** Bringing Elton and Grinnell together: a quantitative framework to represent
² the biogeography of ecological interaction networks

³

⁴ Abstract

⁵ Biogeography has traditionally focused on the spatial distribution and abundance of
⁶ species. Both are driven by the way species interact with one another, but also by the way
⁷ these interactions vary across time and space. Here, we call for an integrated approach,
⁸ adopting the view that community structure is best represented as a network of ecological
⁹ interactions, and show how it translates to biogeography questions. We propose that the
¹⁰ ecological niche should encompass the effect of the environment on species distribution
¹¹ (the Grinnellian dimension of the niche) and on the ecological interactions among them
¹² (the Eltonian dimension). Starting from this concept, we develop a quantitative theory to
¹³ explain turnover of interactions in space and time – *i.e.* a novel approach to interaction
¹⁴ distribution modeling. We apply this framework to host–parasite interactions across Eu-
¹⁵ rope and find that two aspects of the environment (temperature and precipitation) exert
¹⁶ a strong imprint on species co-occurrence, but not on species interactions. Even where
¹⁷ species co-occur, interaction proves to be stochastic rather than deterministic, adding to
¹⁸ variation in realized network structure. We also find that a large majority of host-parasite
¹⁹ pairs are never found together, thus precluding any inferences regarding their probability
²⁰ to interact. This first attempt to explain variation of network structure at large spa-
²¹ tial scales opens new perspectives at the interface of species distribution modeling and
²² community ecology.

²³ **Keywords:** networks, spatial ecology, co-occurrence, probability of interaction

²⁴

25 Introduction

26 Community ecology is *the study of the interactions that determine the distribution and*
27 *abundance of organisms* (?). Despite a general consensus on this definition (?), research
28 on variation in community structure (beta-diversity) has mostly focused on the spatial
29 and temporal turnover of species composition (?). Such research usually neglects variation
30 in the way species interact with each other, despite accumulating empirical evidence that
31 is a major source of diversity (?). Given this omission, it is perhaps not surprising that
32 biogeographers are still struggling to establish whether interactions actually impacts the
33 distribution of species at large spatial scales (?). An interaction is conceived as the direct
34 effect of the action of one species on the demography on another species; whether it
35 scales up and impact the (co-)distribution of species remains matter of debate at the
36 moment (?????). Treating interactions as fixed events nonetheless neglects a large part
37 of the complexity of empirical communities, and will most likely deliver biased metrics
38 (?). Recent attempts at accounting for interactions in species distribution models (???)
39 have brought some methodological advances, but are not sufficient for two reasons. First,
40 these techniques are still based on a ‘species-based’ approach to communities, where
41 interactions are merely treated as fixed covariates affecting distribution. Second, they fail
42 to provide a conceptual step forward, both in their treatment of interactions and in the
43 quality of the predictions they make.

44 Network approaches offer a convenient representation of communities because they
45 simultaneously account for species composition and their interactions. Species are repre-
46 sented as nodes, so that networks already encompass all the information used by current
47 approaches of species distribution modeling; in addition, interactions are represented by
48 links, so that networks provide additional information on community structure. To date,
49 studies of network diversity have mostly been concerned with the distribution of inter-
50 actions within locations, and less so with variation among locations (????). There is,

51 however, ample evidence that interaction networks vary in space and time (??????),
52 even though there is no common framework with which to generalize these results. Meta-
53 community theory provides explanations for variation in the distribution of the nodes
54 (???), but there are no such explanation to the variation of node and link occurrences.
55 Consequently, we need theory to formalize these observations, as it is the only way to-
56 wards fulfilling the goal of community ecology: providing cogent predictions about, and
57 understanding of, the structure of ecological communities.

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

75 The two dimensions of community structure

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

91 There are multiple causes of spatial turnover of species co-occurrence. The first and
92 most-studied driver is the effect of variation in the abiotic environment on species perfor-
93 mance. Combined with specific responses in demography, it generates variation among
94 sites by selecting the locally fittest species (?). Stochasticity plays an additional role,
95 either because colonization and extinction events (?) are inherently unpredictable or be-
96 cause strong non-linear feedbacks in community dynamics generate alternative transients
97 and equilibria (??). Interaction themselves may impact co-distribution, as hypothesized
98 long ago by ?. Analyses of community turnover are usually performed with data rep-
99 resented in a table with rows corresponding to sites (or measurements) and columns to
100 species. Metrics of beta diversity quantify the variance of this community data (?). Tra-

ditional 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 (?), and further partition it into dissimilarity due to changes in species richness and dissimilarity due to actual species turnover (?). Even though all of these methods compare whole lists of species among sites or measurements, they remain fundamentally "species-based", since they report variation within columns. None of them explicitly considers variation of associations (i.e., of pairs or higher-order motifs – ?). The only exceptions are the Joint Species Distribution Models (??), which further account for the covariance among species after representing their response to 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, ?). 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 (?). Variation in community composition can be independent of variation of ecological interactions, suggesting that species and interaction distribution may well respond to different drivers (?).

The "niche" is by far the dominant concept invoked to explain species distributions and community assembly, from the local to the global scale. Following ?, the niche is viewed as the set of environmental conditions allowing a population to establish and persist (see also ?). In other words, the niche is the location in that multidimensional

space allowing a species to have a positive growth rate when rare (?). Community turnover arises as a result of successive replacement of species along an environmental gradient, in agreement with the Gleasonian view of communities (?). 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 ?, MaxEnt ?). It is however much harder to account for ecological interactions within this approach (?). 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 (??).

Interestingly, the food web literature also has its own "niche model" to position a species in a community (?), generalized later to other types of interaction networks (?). 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 niche axes (??). 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 (???). This conceptual framework is, however, limited to local communities, and does not provide any explanation for the turnover of network structure along environmental gradients.

145 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 ?. Despite its intuitive interpretation and easy translation into species distribution models (??), the concept has been frequently criticized (???), and several attempts have been made to expand and improve it (?????).

151 Part of the problem surrounding the niche concept has been clarified with the distinc-
152 tion between Eltonian and Grinnellian definitions (?). The Grinnellian dimension of the
153 niche is the set of abiotic environmental conditions required for a species to maintain a
154 population in a location. The Grinnellian niche is intuitive to apply, and constitutes the
155 conceptual backbone of species distribution models. The Eltonian niche, on the other
156 hand, is 'the place of a species in its biotic environment, its relations to food and ene-
157 mies'. While this aspect of the niche is well known by community ecologists, it is trickier
158 to turn into predictive models. Nonetheless, the development of the niche model of food
159 web structure (?) and its parameterization using functional traits (??) made it more
160 operational.

161 These perspectives are rather orthogonal to each other, and this has resulted in con-
162 siderable confusion in the literature (?). ? attempted to reconcile with the following
163 definition: "*[The niche is] the joint description of the environmental conditions that allow*
164 *a species to satisfy its minimum requirements so that the birth rate of a local population*
165 *is equal to or greater than its death rate along with the set of per capita effects of that*
166 *species on these environmental conditions*". Their representation merges zero-net-growth
167 isoclines delimiting the Grinnellian niche ("when does the population persist?") with im-
168 pact vectors delimiting the Eltonian niche ("what is the per-capita impact?"). While this
169 representation has been very influential in local-scale community ecology (the resource-
170 ratio theory of coexistence, ?), it remains impractical at larger spatial scales because of
171 the difficulties in measuring it. The absence of any mathematical representation of the
172 niche that can be easily fit to ecological data may explain why biogeographers are still
173 struggling to develop species distribution models that also consider ecological interac-
174 tions. Thus, a more integrative description of the niche will be key to understand spatial
175 and temporal turnover in community structure.

176 We propose to integrate the two perspectives of the niche using a visual representa-

tion 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 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 make operational the theoretical interpretation of the niche and species ranges by ?. 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 (?), 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 (?).

201 A probabilistic representation of interaction networks in
202 space

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

$$P(X_{iy}, X_{jy}, L_{ijy}|E_y) \quad (1)$$

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

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

220 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 (?). Above, we referred to this entity as the "metaweb" and it corresponds to the Eltonian dimension of the niche. Below, we will see how this formalism can be directly fit 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 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 (??). 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, $P(X_{ijy}|E_y) = P(X_{iy}|E_y)P(X_{jy}|E_y)$.

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

(e.g. ???). Second- and higher-order interactions (e.g. trophic cascades) could also affect co-occurrence (??). 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 (?). 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 (?)). The impact of direct interactions and first-order indirect interactions on co-occurrence tends to vanish with increasing species richness in the community (?). Further, co-occurrence is also influenced by the covariance of interacting species to an environmental gradient (?).

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 vice-versa). 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 (e.g. ??). Species are assumed to interact whenever they are found together in a location, independent of their local abundance and the local environment. In other words, $P(L_{ijy}|X_{ijy} = 1) = 1$ and $P(L_{ijy}|X_{ijy} = 0) = 0$. 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.

271 Second, ecological interactions could also vary with the environment, so that $P(L_{ijy}|E_y)$.
272 Although it is rare to see a conditional representation of pairwise ecological interactions,
273 experimental studies have frequently revealed interactions to be sensitive to the environ-
274 ment. For instance, ? showed that predation risks of shorebirds vary at the continental
275 scale, decreasing from the south to the north. It is also common to see increasing top-
276 down control with temperature (e.g. ??). Effects of the environment on interactions (?)
277 also propagate up the community and influence network structure (???).

278 Application: continental-scale variation of host-parasite 279 community structure

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

288 Data

289 Communities of willows, gallers, and parasitoids are species-rich and widely distributed,
290 with pronounced variation in community composition across space. The genus *Salix* in-
291 cludes over 400 species, most of which are shrubs or small trees (?), and is common
292 in moist habitats across the Northern Hemisphere (?). Willows support a highly di-
293 verse community of herbivorous insects, with one of the main herbivore groups being

294 gall-inducing sawflies (Hymenoptera: Tenthredinidae: Nematinae: Eurina (?). Gall for-
295 mation is induced by sawfly females during oviposition, and includes marked manipulation
296 of host-plant chemistry by the galler (?). The enemy community of the gallers includes
297 nearly 100 species belonging to 17 insect families of four orders (?). These encompass two
298 main types: inquiline larvae (Coleoptera, Lepidoptera, Diptera, and Hymenoptera) feed
299 primarily on gall tissue, but typically kill the galler larva in the process, while parasitoid
300 larvae (representing many families in Hymenoptera) kill the galler larvae by direct feeding
301 (?).

302 Local realizations of the willow-galler-parasitoid network were reconstructed from
303 community samples collected between 1982 and 2010. During this period, willow galls
304 were collected at 370 sites across Central and Northern Europe. In total, 52, 96 and
305 127 *Salix*, galler and parasitoid and inquiline taxa were distinguished, respectively. The
306 strength of this dataset is that observations were observed in situ, rather than inferred
307 from expert knowledge or other sources of information, thereby allowing the analysis of
308 their spatial variation. The drawback is that, because many species are rare, some of the
309 pairs of species may have been observed co-occurring only a few times, or never. As a
310 consequence, despite the extent of the sampling, there is significant uncertainty in the
311 quantification of some of the links. The above described methodology explicitly aims at
312 describing this uncertainty, and research should be done to develop methods to reduce it
313 (see Conclusion).

314 The current study represents the first analysis of the full data set from a spatial per-
315 spective. Full details about data collection and species identification are provided in (?).
316 Annual mean temperature and precipitation were obtained from WorldClim using the R
317 package dismo (?) and GPS coordinates of the sampling locations. While other covariates
318 could have also been considered to represent the high dimensionality of the niche, these
319 two variables are likely representative of the most important axes of European climate,

320 and are also more easily interpretable than reduced variables obtained, for example, by
321 principal component analysis.

322 Analysis

323 Computing the probability of observing an interaction involves fitting a set of binomial
324 models and collecting their estimated probabilities. For the sake of illustration, we con-
325 sidered generalized linear models – although more flexible fitting algorithms (e.g. GAM
326 or Random Forest) could equally well be used, as long as the algorithm can estimate the
327 probability for each observation. The data consist of a simple (albeit large and sparse)
328 table with the observation of each species, X_{iy} and X_{jy} , their co-occurrence, X_{ijy} , the
329 observation of an interaction L_{ijy} , and environmental co- variates E_y . Thus, there is
330 one row per pair of species per site. We considered that an absence of a record of an
331 interaction between co-occurring species at a site means a true absence (see below for a
332 discussion on this issue).

333 We compared three models for the co-occurrence probability. The first one directly
334 represents the co-occurrence probability conditional on the local environment, $P(X_{iy}, X_{jy}|E_y)$
335 (models are listed in Tables 1 and 2). Hence, this model makes no assumption about the
336 mechanisms driving co- occurrence for any given environment, and instead uses the in-
337 formation directly available in the data. It thereby indirectly accounts for the effect of
338 interactions on co-occurrence, if there are any. The second model considers independent
339 occurrence of species. In this case, we independently fit $P(X_{iy}|E_y)$ and $P(X_{jy}|E_y)$, and
340 we then take their product to derive the probability of co-occurrence. This model should
341 be viewed as a null hypothesis with respect to the first model, since a comparison be-
342 tween the respective models will reveal if there is significant spatial association of the
343 two species beyond a joint response to the shared environment (?). Finally, the third
344 model assumes that the probability of co-occurrence is independent of the environment

and thus constant throughout the landscape. In other words, $P(X_{iy}, X_{jy})$ 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.

Following the same logic, we compared three models of the interaction probability. The first model conditions the interaction probability on the local environmental variables, $P(L_{ijy}|X_{iy}, X_{jy}, E_y)$. 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, $P(L_{ijy}|X_{iy}, X_{jy})$. 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, $P(L_{ijy}|X_{iy}, X_{jy}) = 1$. 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 $P(L_{ijy}, X_{iy}, X_{jy})$ was then computed from Eq. 2, and the likelihood of each observation D_{ijy} was computed as $\mathcal{L}(\theta_{ijy}|D_{ijy}) = P(L_{ij}, X_{iy}, X_{jy})$ if

371 an interaction was observed, and as $\mathcal{L}(\theta_{ijy}|D_{ijy}) = 1 - P(L_{ijy}, X_{iy}, X_{jy})$ if no interaction
372 was observed. The log-likelihood was summed over the entire dataset to compare the
373 different models by AIC. We therefore evaluate the likelihood of all local networks, given
374 the model. Not surprisingly, there was a very large number of species pairs for which
375 this model could not be computed, as they simply never co-occurred. For these pairs, we
376 have no information of the interaction probability, and they were consequently removed
377 from the analysis. The log-likelihood reported across the entire dataset was therefore
378 summed over all pairs of species observed to co-occur at least once. Interactions between
379 the first (*Salix*) and second (gallers) trophic layers and those between the second and
380 third (parasitoids) were considered separately. Finally, we used the full model (in which
381 both co-occurrence and the interaction are conditional on the environment) to interpolate
382 species distributions and interaction probabilities across the entire European continent.
383 We reconstructed the expected network for each location in a 1 X 1 km grid and computed
384 the probabilistic connectance following ?.

385 All of the data are openly available in the database *mangal* (?) and all R scripts
386 for running the analyses, are provided in the github repository [https://github.com/
387 DominiqueGravel/ms_probaweb](https://github.com/DominiqueGravel/ms_probaweb).

388 Results

389 Despite the extensive sampling, many pairs of species were observed to co-occur only a
390 few times. This made it difficult to evaluate interaction probabilities with any reasonable
391 confidence. Thus, we start with an example of a single pair of species selected because
392 of its high number of co-occurrences ($N_{ij} = 38$): the leaf folder *Phyllocolpa prussica* and
393 the parasitoid *Chrysocharis elongata*. These two fairly abundant species were observed
394 $N_i = 49$ and $N_j = 121$ times, respectively, across the 370 sites, and they were found to
395 interact with a marginal probability $P(L_{ij}) = 0.55$, which means they interacted at 21

396 different locations. Here, a comparison of model fit (Table 1) reveals that conditioning
397 the interaction probability on local environmental conditions adds no explanatory power
398 beyond a model assuming the same probability of interaction anywhere in space (Model 1
399 vs Model 2). Moreover, when the two species co-occur, the occurrence of the interaction
400 was insensitive to the environment (Model 2 vs Model 3). Alternatively, climatic variables
401 significantly impacted co-occurrence (Model 3 vs Model 4). The neutral model performed
402 worse than the non-random co-occurrence model (Model 3 vs Model 6). The full model
403 revealed that the greatest interaction probability occurred at intermediate temperature
404 and precipitation, simply because this is where the two species most frequently co-occur
405 (Fig. 3). The probabilities of co-occurrence and interaction can be represented in space,
406 where we find that the highest interaction probability occurred in Central Europe (Fig.
407 4).

408 We evaluated each model for all pairs of species in order to better understand the
409 large-scale drivers of network turnover. The results were highly consistent among trophic
410 levels (*Salix*-gallers and gallers-parasitoids; Table 2), despite large variability in the fit of
411 the models to the different pairs of species. This variability was particularly important for
412 the models of interactions given co-occurrence. Across all pairs of species, the conditional
413 representation of interactions performed better than the marginal one (Model 1 vs Model
414 2); that is, interactions did not occur systematically whenever the two species were found
415 co-occurring. Hence, in addition to species turnover, the stochastic nature of interactions
416 contributes to network variability. In total, we recorded 1,173 pairs of interactions, only
417 290 of which occurred more than five times. Out of these 290 interactions, 143 were
418 systematically detected whenever the two species co-occurred. In the instances when
419 species co-occurred, the two environmental variables considered proved relatively poor
420 predictors of their interactions (Model 2 vs Model 3). Not surprisingly, for both types
421 of interactions (*Salix*-galler and galler-parasitoid), the log-likelihood increased when the

422 environment was considered. However, the extra number of parameters exceeded the gain
423 in log-likelihood and inflated AIC. Therefore, the most parsimonious model excluded the
424 effect of the environment. On the basis of log-likelihood only, co-occurrence was non-
425 neutral for both *Salix*-galler and galler-parasitoid interactions. Thus, according to AIC,
426 the best model was the one of non-random co-occurrence (Model 3 vs Model 6) for both
427 types of interactions.

428 Once we had selected the best model based on AIC (Model 3, Table 2), we used
429 it to reconstruct the expected species richness, along with the most likely network for
430 each location. Using this approach, we mapped the expected distribution of network
431 properties across Europe (Fig. 5). For simplicity, we chose to consider connectance
432 as our descriptor of network configuration, as this metric can be easily computed from
433 probabilistic networks (?) and is also a good proxy for many other network properties (?).
434 Overall, we found a peak in *Salix*, gallers and parasitoid diversity in Northern Europe.
435 The expected number of interactions roughly followed the distribution of species richness,
436 but accumulated at a rate different from species numbers. Connectance likewise peaked
437 in Northern Europe (Fig. 5).

438 Interpretation

439 We have proposed that the representation of community structure and its variation in
440 space is best captured by the formalism of ecological networks, as both the distribution
441 of species and their interspecific interactions can then be accounted for. We consequently
442 revised the niche concept in order to integrate its abiotic and biotic components. This
443 integrated niche was represented visually with an ordination of species into an environ-
444 mental space and a trait space. The fundamental niche of a species is represented as
445 the set of environmental conditions and resources that allow a species to establish in a

446 location, thereby integrating the Eltonian and the Grinnellian components of the niche.
447 We then translated the concept mathematically by investigating the probability of the
448 joint occurrences of species and their interaction, which should be interpreted as an in-
449 teraction distribution model. We used this approach to characterize the turnover of the
450 structure of ecological interactions in a species-rich tri-trophic network across Western
451 Europe, finding that the primary driver of network variation is the turnover in species
452 composition.

453 Applying the framework to our large data set on host-parasite interactions across Eu-
454 rope revealed key features in the interaction between *Salix* taxa, their herbivores, and the
455 natural enemies of these herbivores. Consistent with a general increase in the diversity of
456 *Salix* towards boreal areas (?), overall species richness of the networks increased towards
457 the north. The distribution of *Salix* species richness largely matched those of gallers and
458 parasitoids. These observations within Europe are also matched by the ones found at a
459 global scale for *Salix* (???) and sawflies (??). Species richness in a common group of
460 parasitic wasps, the Ichneumonidae, was originally presumed to show a similar "reversed
461 latitudinal gradient", but this observation has been recently challenged by findings of
462 rather high ichneumonid diversity in the tropics (?). Nevertheless, ichneumonid subfam-
463 ilies specifically associated with sawflies (Ctenopelmatinae, Tryphoninae) are clearly less
464 diverse in the south.

465 Exactly what processes are responsible for the distribution of species richness at dif-
466 ferent trophic levels is yet to be established (but see e.g. ???), but as a net outcome
467 of different latitudinal trends across trophic levels, the distribution of co-occurrence and
468 therefore of potential interactions differed between the first and second layers of feed-
469 ing links. The correlation between expected *Salix* and galler richness was 0.73, while
470 it was 0.58 between gallers and their parasitoids. Therefore, the ratio of herbivores to
471 *Salix* species is essentially constant across Europe, whereas each herbivore species is po-

472 potentially attacked by a richer enemy community higher latitudes. Consequently, overall
473 connectance peaks in Northern Europe (Fig. 5).

474 In terms of species interacting with each other, our analysis suggests that the en-
475 vironment leaves a detectable imprint on species co-occurrence, but only a slight mark
476 on the occurrence of realized links among species in a specific place: the probability of
477 finding a given combination of species at a higher and a lower trophic level at the same
478 site was clearly affected by the environment, whereas the probability of observing an in-
479 teraction between the two was, given co-occurrence, not significantly so. The interactions
480 were highly uncertain, as only few pairs of species were systematically interacting when
481 co-occurring but most not. This applies to the example species *Phyllocolpa prussica* and
482 *Chrysocharis elongata* (Figs. 3 and 4), but also to all species pairs more generally. For
483 the example species pair, the full model revealed that the joint probability event (inter-
484 action and co-occurrence) peaks at intermediate temperature and precipitation, simply
485 because this is where the two species co-occur most often. This does not imply that
486 species will always interact when they meet – although this is a basic assumption in most
487 documented metawebs to date (e.g. ??). Rather, an interaction is better represented as
488 a stochastic process whose probability is also influenced by the probability with which
489 species co-occur. What we cannot reliably know is how this stochasticity splits into dif-
490 ferent sources of uncertainty – i.e., the extent to which a species at the higher trophic
491 level runs into a species at the lower level co-occurring at the site, the extent to which
492 this interaction is detected by an observer collecting a finite sample, or simply the un-
493 certainty arising from incomplete description of a highly dimensional niche. Future work
494 will be required to document the relative importance of these sources of uncertainty in
495 the occurrence of interactions.

⁴⁹⁶ Perspectives

⁴⁹⁷ Evidence that the structure of ecological networks does vary across habitats (e.g. ?), over
⁴⁹⁸ environmental gradients (?) and in time (?) is accumulating rapidly. It is not clear,
⁴⁹⁹ however, to what extent the turnover of network structure is driven by a systematic
⁵⁰⁰ change in species composition or of pairwise interactions (??). Our model comparison of
⁵⁰¹ host-parasite interactions revealed that most of the turnover is driven by species-specific
⁵⁰² responses to the environment, impacting species richness, and that co-occurrence was
⁵⁰³ mostly neutral. Further, the occurrence of interactions among host and parasite is highly
⁵⁰⁴ stochastic even when both are present, and not predictable by the variables considered by
⁵⁰⁵ us. We know that interactions vary with the environment in other systems, for instance,
⁵⁰⁶ herbivory (??) and predation (??) are often found to increase with temperature, resulting
⁵⁰⁷ in spatial variation of trophic cascades (?). What remains unclear, however, is the extent
⁵⁰⁸ to which such variation is driven by a turnover of species composition along gradients,
⁵⁰⁹ or a turnover of the interactions. Here we found that interactions vary substantially but
⁵¹⁰ non-predictably along the annual temperature and the precipitation gradient. The lack
⁵¹¹ of detectable signal may be due to our choice of covariates. Indeed, a previous study
⁵¹² focusing on a subset of the system studied here identified habitat characteristics as the
⁵¹³ primary drivers of interactions (?). New investigations with other systems will thus be
⁵¹⁴ required to challenge this result. Under all circumstances, documenting the relationship
⁵¹⁵ between the environment and the occurrence of interactions at continental scales is critical
⁵¹⁶ for understanding how large- scale variation of trophic regulation influences community
⁵¹⁷ dynamics and ecosystem functioning (?).

⁵¹⁸ We restricted our analysis to the effect of co-occurrence on ecological interactions,
⁵¹⁹ neglecting the inverse of the problem. We did not investigate in depth the drivers of
⁵²⁰ co-occurrence and simply took it for granted from the data. Co-occurrence was indeed
⁵²¹ 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, $P(X_{iy}, X_{jy}|E_y)$, which could be expanded as $P(X_{iy}|X_{jy}, E_y)P(X_{jy}|E_y)$. 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. 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, ? found that the strength of this association is often rather weak if not neutral (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 (?).

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 (?) and empirical hypothesis testing. The impact of ecological interactions on the distribution of co-occurrence has been the topic of many publications since ? seminal study on competition and "checkerboard" distribution, but pairwise approaches have only recently received attention (?). Whether two interacting species are more closely associated in space remains unclear, since most approaches based on null models consider community-level metrics (e.g. ?), 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

548 of the impact of interactions on co-occurrence.

549 Ecological networks are known to be extremely sparse, *i.e.* they have far more ab-
550 sences than presences of interactions. Absences of interactions, however, can come from
551 different sources. The fact that unequal sampling at the local scale can affect our under-
552 standing of network structure is well documented (?). In a spatial context, however, some
553 interactions may be undocumented simply because the species involved have never been
554 observed to co-occur. Although these cases are reported as a lack of interactions, in actu-
555 ality we cannot make any reliable inference from them: since the species have never been
556 observed together, it remains possible that they would interact if they did. A fundamen-
557 tally different category of absences of interactions are then those reported after multiple
558 observations of species co-occurrence. Thus, to gain confidence that the probability of an
559 interaction is low, extensive sampling (that is, several records of co-occurrence) is needed.
560 Generally, our confidence that the interaction is indeed impossible will increase with the
561 number of observations of the species pair. Seeing that this is essentially a Bernoulli
562 process (the probability that the species will interact given their presence), the breadth
563 of the confidence interval is expected to saturate after a fixed number of observations,
564 which can be set as a threshold above which a species pair has finally been observed
565 "often enough". This will allow us to deal with both confirmed absences of interactions
566 and mere absence of evidence.

567 Conclusion

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

571 **1. New generation of network data.** Investigating spatial variation of network

572 structure will require high quality and highly replicated network data. We have in-
573 vestigated one the most comprehensive spatial network datasets we are aware of and
574 nonetheless found immense gaps of knowledge in its resolution. Species richness accumu-
575 lates much faster than observations of ecological interactions (?). Each pair of species
576 must be observed several times in order to obtain reliable estimates of their interaction
577 probability.

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

583 **3. From interaction probabilities to a distribution of network properties.**
584 Metrics are available to analyse the structure of probabilistic networks (?). These met-
585 rics are useful as first approximation, but they assume independence among interactions.
586 This might not be the case in nature because of the role of co-occurrence and shared en-
587 vironmental requirements. We also need to better understand the distribution of network
588 properties arising from probabilistic interactions.

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

593 **5. Effects of ecological interactions on co-occurrence.** We have intentionally
594 omitted the feedback of ecological interactions on co-occurrence in this framework. As
595 abundance can impact the occurrence of interactions, and, conversely since interactions
596 impact abundance (?), we could reasonably expect that interactions will also influence
597 co-occurrence. Theory in this regard does exist for simple three-species modules (?), but

598 its extension to entire co-occurrence networks will prove critical in the future, especially
599 given the interest in using co- occurrence to infer ecological interactions (??).

600 **Acknowledgements**

601 This is a contribution to the working groups *Continental-scale variation of ecological net-*
602 *works* supported by the Canadian Institute for Ecology and Evolution), and the *Next*
603 *Generation Data, Models, and Theory Working Group*, supported by the Santa Fe Insti-
604 tute, the Betsy and Jesse Fink Foundation, the ASU-SFI Center for Biosocial Complex
605 Systems, and NSF Grant PHY-1240192. DG also acknowledges financial support from
606 NSERC- Discovery grant program and Canada Research Chair program.

Table 1: Summary of model comparison for the interaction between the leaf folder *Phyllocolpa prussica*) and the parasitoid *Chrysocharis elongata*

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

Table 2: 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	$P(L_{ijy})$	$P(X_{iy}, X_{jy} E_y)$	-6022.1	7548	27140.3
	2	$P(L_{ijy} X_{iy}, X_{jy})$	$P(X_{iy}, X_{jy} E_y)$	-5547.9	7548	26191.8
	3	$P(L_{ijy} X_{iy}, X_{jy}, E_y)$	$P(X_{iy}, X_{jy} E_y)$	-5364.0	12580	35888.0
	4	$P(L_{ijy} X_{iy}, X_{jy}, E_y)$	$P(X_{iy}, X_{jy})$	-5998.4	8806	30287.2
	5	$P(L_{ijy} X_{iy}, X_{jy}, E_y)$	$P(X_{iy})P(X_{jy})$	-6636.1	7548	27092.7
	6	$P(L_{ijy} X_{iy}, X_{jy}, E_y)$	$P(X_{iy} E_y)P(X_{jy} E_y)$	-6002.9	18870	49745.7
Galler-Parasitoid	1	$P(L_{ijy})$	$P(X_{iy}, X_{jy} E_y)$	-22068.1	19206	82548.2
	2	$P(L_{ijy} X_{iy}, X_{jy})$	$P(X_{iy}, X_{jy} E_y)$	-19504.8	19206	77421.6
	3	$P(L_{ijy} X_{iy}, X_{jy}, E_y)$	$P(X_{iy}, X_{jy} E_y)$	-20217.0	32010	104454.1
	4	$P(L_{ijy} X_{iy}, X_{jy}, E_y)$	$P(X_{iy}, X_{jy})$	-19591.3	22407	77594.5
	5	$P(L_{ijy} X_{iy}, X_{jy}, E_y)$	$P(X_{iy})P(X_{jy})$	-22491.5	19206	89796.9
	6	$P(L_{ijy} X_{iy}, X_{jy}, E_y)$	$P(X_{iy} E_y)P(X_{jy} E_y)$	-18936.9	48015	133903.7

607 **Figure legends**

608 **Figure 1**

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

617 **Figure 2**

618 **Visual representation of the integrated niche.** In biogeography, the niche is con-
619 sidered the set of environmental conditions where the intrinsic growth rate r is positive
620 (?). The horizontal axis represents an environmental gradient impacting the growth of
621 a focal species (red point). The location of each species (grey points) along this gradi-
622 ent represents their optimum, and the vertical dotted lines represent the limits of the
623 Grinnellian niche of the focal species. In food web ecology, the Eltonian niche represents
624 the location of a species in the food web, as determined by its niche position (n) and
625 its niche optimum (c). The vertical axis represents a niche gradient, for example a trait
626 such as body size. The location of each species along this gradient represents their niche
627 position. The focal species will feed only on host/prey species occupying niche locations
628 within a given interval around the optimum, represented by the horizontal lines. The
629 integrated Grinnellian and Eltonian niche corresponds to the square in the middle where
630 an interaction is possible owing to a match of traits and spatial distribution. The central

631 square represents the area where the joint probability of observing co-occurrence and
632 interactions is positive.

633 **Figure 3**

634 **Probabilistic representation of the interaction probability between a leaf folder**
635 (*Phyllocolpa prussica*) **and a parasitoid (*Chrysocharis elongata*) across gradi-**
636 **ents of annual average temperature and annual precipitation.** The colour gra-
637 dient represents the probability of observing the event, from 0 (white) to 1 (black). The
638 representation is based on predictions from Model 3 (see Table 1). In the left panel,
639 open circles represent the absence of both species, whereas closed circles represent co-
640 occurrence and plus signs the occurrence of only one of the two species. In the other
641 two panels, open circles represent co-occurrence but an absence of interaction and closed
642 circles the occurrence of an interaction.

643 **Figure 4**

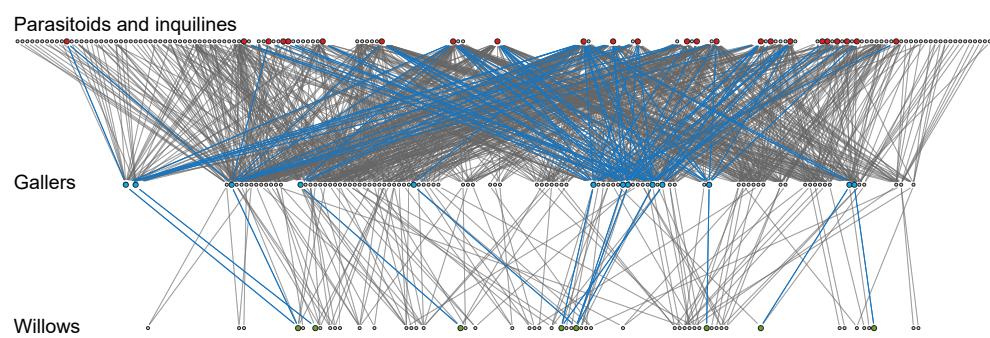
644 **Probabilistic representation of the interaction probability between a leaf folder**
645 (*Phyllocolpa prussica*) **and a parasitoid (*Chrysocharis elongata*) across Eu-**
646 **rope.** Climate data are extracted from WorldClim over a 1km x 1km grid and predic-
647 tions made according to the model illustrated at Fig. 3. Colour gradient ranges from low
648 values (dark blue) to high values (yellow).

649 **Figure 5**

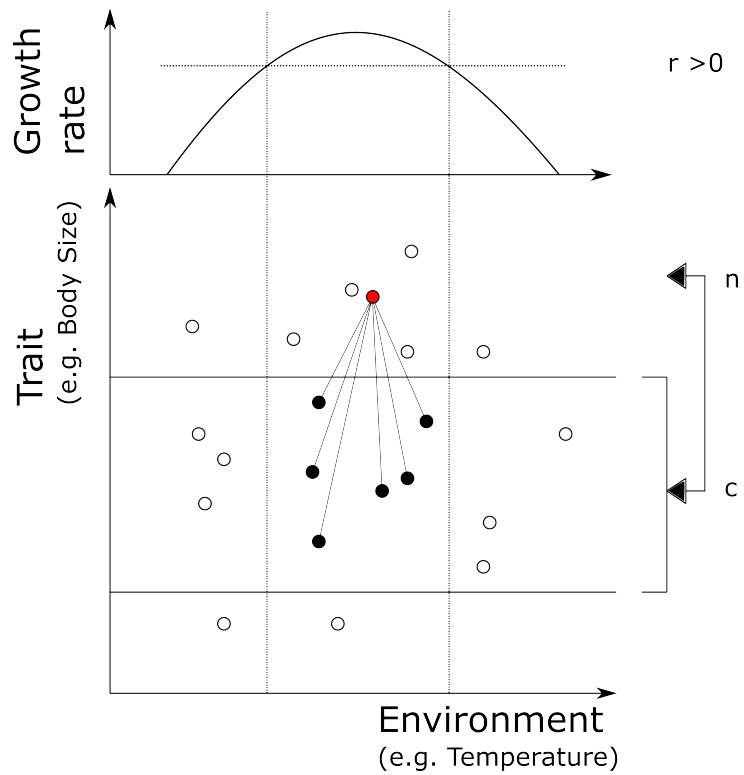
650 **Mapping the distribution of species richness, the number of links and con-**
651 **nnectance across Europe.** The representation is based on predictions from Model 3 (see
652 Table 2) using climate data extracted from WorldClim over a 1km x 1km grid. Species
653 richness is obtained by summation of individual occurrence probabilities, and link density

654 by summation of interaction probabilities. Colowr gradient ranges from low values (dark
655 blue) to high values (yellow).

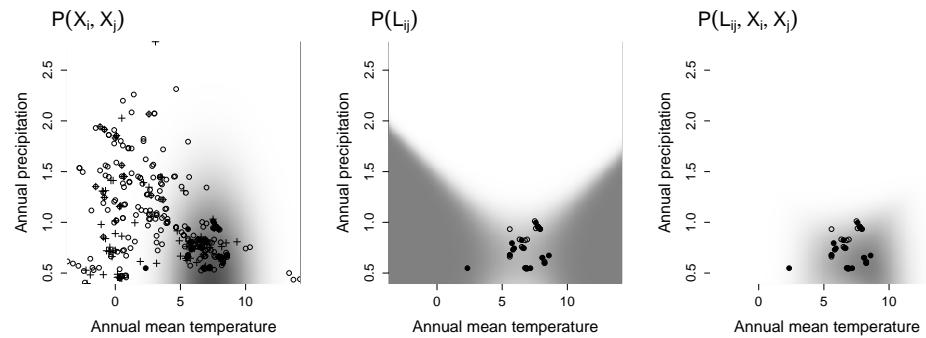
656 **Figure 1**



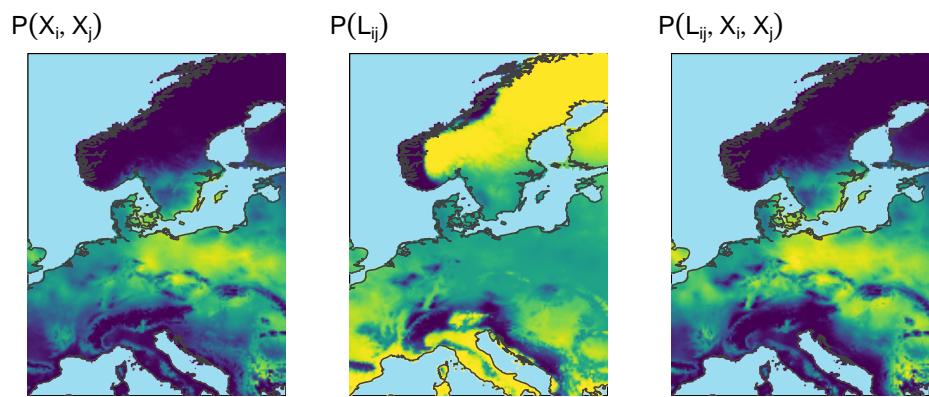
657 Figure 2



658 **Figure 3**



659 **Figure 4**



660 **Figure 5**

