

# Downscaling metawebs: propagation of uncertainties in species distribution and interaction probability

Gabriel Dansereau<sup>1,2</sup> Ceres Barros<sup>3</sup> Timothée Poisot<sup>1,2</sup>

<sup>1</sup> Université de Montréal <sup>2</sup> Québec Centre for Biodiversity Sciences <sup>3</sup> University of British Columbia

## Correspondance to:

Gabriel Dansereau — [gabriel.dansereau@umontreal.ca](mailto:gabriel.dansereau@umontreal.ca)

# 1 Introduction

2 Sampling ecological networks in space and time is a challenging task as species interactions display high  
3 turnover and low encounter rates, which require large sampling efforts to properly document (Jordano 2016).  
4 Most studies on food webs have previously focused on local webs limited in size and extent, and are rarely  
5 replicated in space and time (Mestre *et al.* 2022). Interactions can show important variations in space (Poisot *et al.* 2015; Zarnetske *et al.* 2017), yet available network data also show important geographical bias, limiting our  
6 ability to answer questions in many biomes and over broad spatial extents (Poisot *et al.* 2021). Moreover, global  
7 network monitoring is insufficient to properly describe and understand how ecosystems are reacting to global  
8 change (Windsor *et al.* 2023). Approaches to predict species interactions (e.g., Morales-Castilla *et al.* 2015;  
9 Desjardins-Proulx *et al.* 2017) are increasingly used as an alternative to determine potential interactions and  
10 can handle limited data to circumvent data scarcity (Strydom *et al.* 2021), but are rarely used to make explicitly  
11 spatial predictions. As a result, there have been repeated calls for globally distributed interaction and network  
12 data and repeated sampling in time and space (Mestre *et al.* 2022; Windsor *et al.* 2023), which will help  
13 understand the macroecological variations of food webs (Baiser *et al.* 2019). == (Tim: transition to metaweb  
14 concept) ==

16 The metaweb is an increasingly used concept to address the issue of data scarcity, and it further holds potential  
17 to analyze networks at large spatial extents. A metaweb contains all possible interactions between species in a  
18 given regional species pool (Dunne 2006). When assembled by integrating databases and computing tools, the  
19 metaweb allows to overcome sampling limitations to upscale network data to a global scale (Albouy *et al.*  
20 2019). Recent studies have focused on assembling metawebs for various taxa through literature surveys and  
21 expert elicitation (European tetrapods, Maiorano *et al.* 2020) or using predictive tools (marine fish food web,  
22 Albouy *et al.* 2019; Canadian mammals, Strydom *et al.* 2022a). However, the metaweb holds more information  
23 than the possible interactions and is also key to analyzing networks across space. Empirical networks, which are  
24 local realizations of a regional metaweb (Poisot *et al.* 2012), inherit the metaweb structure with little influence  
25 from habitat and dynamical constraints (Saravia *et al.* 2022). Given this, Strydom *et al.* (2022b) called the  
26 prediction of the metaweb structure the core goal of predictive network ecology and the key to producing  
27 accurate downscaled and local predictions. Therefore, establishing or predicting the metaweb should be the first  
28 target for systems lacking information about local realizations. This is not the same as using interactions to  
29 improve predictions of species distributions as recent studies have done (for example, Moens *et al.* 2022;

30 Poggiato *et al.* 2022; Lucas *et al.* 2023), although these are incredibly relevant and answer long-standing calls  
31 to include interactions within such models (Wisz *et al.* 2013). Instead, predicting networks in space is a  
32 different task, and it serves another goal, focusing first on the distribution of networks and its drivers rather than  
33 on the distribution of species.

34 Explicit spatial predictions such as downscaled metaweb predictions are essential as they will allow  
35 comparisons with extant work for species communities. Recent approaches to downscaling combined the  
36 metaweb with species distribution maps to generate local assemblages for European tetrapods (Braga *et al.*  
37 2019; O'Connor *et al.* 2020; Galiana *et al.* 2021; Gaüzère *et al.* 2022) and North Sea demersal fishes and  
38 benthic epifauna (Frelat *et al.* 2022). These downscaled assemblages allowed studying network structures in  
39 novel ways, for instance, assessing changes in food web structure across space (Braga *et al.* 2019) and  
40 describing the scaling of network area relationships (Galiana *et al.* 2021). Other examples have shown that the  
41 metaweb can be used to investigate large-scale variation in food web structure, indicating high geographical  
42 connections and heterogeneous robustness against species extinctions (marine fish food webs, Albouy *et al.*  
43 2019). Further comparisons are relevant as they may go in unexpected directions and highlight new elements  
44 regarding network biogeography. For instance, Frelat *et al.* (2022) found a strong spatial coupling between  
45 community composition and food web structure but a temporal mismatch depending on the spatial scale. Poisot  
46 *et al.* (2017) found that interaction uniqueness captures more composition variability than community  
47 uniqueness and that sites with exceptional compositions might differ for networks and communities. Spatialized  
48 network data will allow these comparisons, identifying important conservation targets for networks and whether  
49 they differ geographically from areas currently prioritized for biodiversity conservation.

50 A key challenge remains in how to downscale a regional metaweb towards local network predictions reflecting  
51 the spatial variability of interactions. Even when the metaweb is known, local networks may vary substantially  
52 and differ from the metaweb structure (McLeod *et al.* 2021), emphasizing the need for methods to generate  
53 local, downscaled network predictions. A potential limitation to previous downscaling approaches is that they  
54 assume interactions are constant across space, which ignores behaviour variability and does not consider the  
55 effect of environmental conditions on interaction realization (Braga *et al.* 2019). In contrast, recent studies  
56 argued that seeing interactions as probabilistic events (rather than binary ones) allows us to account for their  
57 variability in space (Poisot *et al.* 2016) and that this should also be reflected in metawebs (Strydom *et al.*  
58 2022b). Gravel *et al.* (2019) introduced a probabilistic framework describing how the metaweb can generate  
59 local realizations and showed how it could be used for interaction distribution modelling. This approach to

60 downscaling is relevant when combined with in situ observations of interactions and local networks to train  
61 interaction models (in this case, with willow-galler-parasitoid networks). However, such data is rarely available  
62 across broad spatial extents (Hortal *et al.* 2015; Poisot *et al.* 2021; Windsor *et al.* 2023). Spatially replicated  
63 interaction data required for such a model are especially challenging to document with large food web systems  
64 such as European tetrapod and Canadian mammal metawebs (Maiorano *et al.* 2020; Strydom *et al.* 2022a). We  
65 currently lack a downscaling framework that is both probabilistic and can be trained without replicated in situ  
66 interaction data. Additionally, a probabilistic view can allow propagating uncertainty, which can play a key role  
67 in evaluating the quality of the predictions. Assessing model uncertainty would enable us to determine to which  
68 degree we should trust our predictions and to identify what to do to improve the current knowledge.

69 Here, we present a method to downscale a metaweb in space by spatially reconstructing local instances of a  
70 probabilistic metaweb of Canadian mammals. We do so using a probabilistic approach to both species  
71 distributions and interactions in a system without spatially replicated interaction data. We then explore how the  
72 spatial structure of the downscaled metaweb varies in space and how the uncertainty of interactions can be  
73 made spatially explicit. We further show that the downscaled metaweb can highlight important biodiversity  
74 areas and bring novel ecological insights compared to traditional community measures like species richness.

## 75 **Methods**

76 Fig. 1 shows a conceptual overview of the methodological steps leading to the downscaled metaweb. The  
77 components were grouped as non-spatial and spatial inputs, localized site steps (divided into  
78 single-species-level, two-species-level, and network-level steps), and the final downscaled and spatialized  
79 metaweb. Throughout these steps, we highlight the importance of presenting the uncertainty of interactions and  
80 their distribution in space. We argue that this requires adopting a probabilistic view and incorporating variation  
81 between scales.

82 [Figure 1 about here.]

## 83 **Data**

### 84 **Metaweb**

85 The main source of interaction data was the metaweb for Canadian mammals from Strydom *et al.* (2022a),  
86 which is a-spatial, i.e., it represents interactions between mammals that can occur anywhere in Canada. The  
87 species list for the Canadian metaweb was extracted from the International Union for the Conservation of  
88 Nature (IUCN) checklist (Strydom *et al.* 2022a). Briefly, the metaweb was developed using graph embedding  
89 and phylogenetic transfer learning based on the metaweb of European mammals, which is itself based on a  
90 comprehensive survey of interactions reported in the scientific literature (Maiorano *et al.* 2020). The Canadian  
91 metaweb is probabilistic, which has the advantage of reflecting the likelihood of an interaction taking place  
92 given the phylogenetic and trait match between two species. This allows incorporating interaction variability  
93 between species (i.e., taking into account that two species may not always interact whenever or wherever they  
94 occur); however, we highlight that other factors beyond trait and phylogenetic matching (e.g., population  
95 densities) will also contribute to observed interaction frequencies.

### 96 **Species occurrences**

97 The downscaling of the metaweb involved combining it with species occurrence and environmental data. First,  
98 we extracted species occurrences from the Global Biodiversity Information Facility (GBIF; [www.gbif.org](http://www.gbif.org)) for  
99 the Canadian mammals after reconciling species names between the Canadian metaweb and GBIF using the  
100 GBIF Backbone Taxonomy (GBIF Secretariat 2021). This step removed potential duplicates by combining  
101 species listed in the Canadian metaweb which were considered as a single entity by GBIF. We collected  
102 occurrences for the updated species list (159 species) using the GBIF download API on October 21st 2022  
103 (GBIF.org 2022). We restricted our query to occurrences with coordinates between longitudes 175°W to 45°W  
104 and latitudes 10°N to 90°N. This was meant to collect training data covering a broader range than our prediction  
105 target (Canada only) and include observations in similar environments. Then, since GBIF observations  
106 represent presence-only data and most predictive models require absence data, we generated pseudo-absence  
107 data using the surface range envelope method, which selects random non-observed sites within the spatial range  
108 delimited by the presence data (Barbet-Massin *et al.* 2012).

## 109 **Environmental data**

110 We used species distribution models (SDMs, Guisan & Thuiller 2005) to project Canadian mammal habitat  
111 suitability across the country, which we treated as information on potential distribution. For each species, we  
112 related occurrences and pseudo-absences with 19 bioclimatic variables from CHELSA (Karger *et al.* 2017) and  
113 12 consensus land-cover variables from EarthEnv (Tuanmu & Jetz 2014). The CHELSA bioclimatic variables  
114 (*bio1-bio19*) represent various measures of temperature and precipitation (e.g., annual averages, monthly  
115 maximum or minimum, seasonality) and are available for land areas across the globe. We used the most recent  
116 version, the CHELSA v2.1 dataset (Karger *et al.* 2021), and subsetting it to land surfaces only using the  
117 CHELSA v1.2 (Karger *et al.* 2018), which does not cover open water. The EarthEnv land-cover variables  
118 represent classes such as Evergreen broadleaf trees, Cultivated and managed vegetation, Urban/Built-up, and  
119 Open Water. Values range between 0 and 100 and represent the consensus prevalence of each class in  
120 percentage within a pixel (hereafter called sites). We coarsened both the CHELSA and EarthEnv data from their  
121 original 30 arc-second resolution to a 2.5 arc-minute one (around 4.5 km at the Equator) using GDAL  
122 (GDAL/OGR contributors 2021). This resolution compromised capturing both local variations and broad-scale  
123 patterns while limiting computation costs to a manageable level as memory requirements rapidly increase with  
124 spatial resolution.

## 125 **Analyses**

### 126 **Species distribution models**

127 Our selection criteria for choosing an SDM algorithm was to have a method that generated probabilistic results  
128 (similar to Gravel *et al.* 2019), including both a probability of occurrence for a species in a specific site and the  
129 uncertainty associated with the prediction. These were crucial to obtaining a probabilistic version of the  
130 metaweb as they were used to create spatial variations in the localized interaction probabilities (see next  
131 section). One suitable method for this is Gradient Boosted Trees with a Gaussian maximum likelihood from the  
132 `EvoTrees.jl` *Julia* package (<https://github.com/EvoVest/EvoTrees.jl>). This method returns a prediction for  
133 every site with an average value and a standard deviation, which we used as a measure of uncertainty to build a  
134 Normal distribution for the probability of occurrence of a given species at all sites (represented as probability  
135 distributions on Fig. 1). We trained models across the extent chosen for occurrences (longitudes 175°W to  
136 45°W and latitudes 10°N to 90°N), then predicted species distributions only for Canada. We used the 2021

137 Census Boundary Files from Statistics Canada (Statistics Canada 2022) to set the boundaries for our  
138 predictions, which gave us 970,698 sites in total.

### 139 **Building site-level instances of the metaweb**

140 The next part of the method was the localized steps which produce local metawebs for every site. This  
141 component was divided into single-species, two-species, and network-level steps (*Localized steps* box on Fig. 1).

142 The single-species steps represented four possible ways to account for uncertainty in the species distributions  
143 and bring variation to the spatial metaweb. We explored four different options to select a value ( $P(\textit{occurrence})$ ;  
144 Fig. 1) from the occurrence distributions obtained in the previous steps: 1) taking the mean from the  
145 distribution as the probability of occurrence (option 1 on Fig. 1); 2) converting the mean value to a binary one  
146 using a specific threshold per species (option 2); 3) sampling a random value within the Normal distribution  
147 (option 3); or 4) converting a random value into a binary result (option 4, using a separate draw from option 3  
148 and the same threshold as in option 2). The threshold ( $\tau$  on Fig. 1) used was the value that maximized Youden's  
149  $J$  informedness statistic (Youden 1950), the same metric used by Strydom *et al.* (2022a) at an intermediate step  
150 while building the metaweb. The four sampling options were intended to explore how uncertainty and variation  
151 in the species distributions can affect the metaweb result. We expected thresholding to have a more pronounced  
152 effect on network structure as it should reduce the number of links by removing many of the rare interactions  
153 (Poisot *et al.* 2016). Meanwhile, we expected random sampling to create spatial heterogeneity compared to the  
154 mean probabilities, as including some extreme values should confound the potential effects of environmental  
155 gradients. We chose option 1 as the default to present results as it is intuitive and essentially represents the  
156 result of a probabilistic SDM (as in Gravel *et al.* 2019).

157 Next, the two-species steps were aimed at assigning a probability of observing an interaction between two  
158 species in a given site. For each species pair, we multiplied the product of the two species' occurrence  
159 probabilities ( $P(\textit{co-occurrence})$ ; Fig. 1) (obtained using one of the sampling options above) by their interaction  
160 probability in the Canadian metaweb. For cases where species in the Canadian metaweb were considered as the  
161 same species by the GBIF Backbone Taxonomy (the reconciliation step mentioned earlier), we used the highest  
162 interaction probabilities involving the duplicated species.

163 The network-level steps then created the probabilistic metaweb for the site. We assembled all the local  
164 interaction probabilities (from the two-species steps) into a probabilistic network (Poisot *et al.* 2016). We then

165 sampled several random network realizations to represent the potential local realization process (Poisot *et al.*  
166 2015). This resulted in a distribution of localized networks, which we averaged over the number of simulations  
167 to obtain a single probabilistic network for the site.

## 168 **Downscaled metaweb**

169 The final output of our method was the downscaled metaweb, which contains a localized probabilistic metaweb  
170 in every site across the study area (Outputs box on Fig. 1). A metaweb essentially serves to set an upper bound  
171 on the potential interactions (Strydom *et al.* 2022b); therefore, the downscaled metaweb is a refined upper  
172 boundary at the local scale taking into account co-occurrences. It is still potential in nature and differs from a  
173 local realization, from which it should have a different structure. Nonetheless, from the downscaled metaweb,  
174 we can create maps of network properties (e.g. number of links, connectance) measured on the local  
175 probabilities, display their spatial distribution, and compute some traditional community-level measures such as  
176 species richness. We chose to compute and display the expected number of links (measured on probabilistic  
177 networks following Poisot *et al.* 2016; also see Gravel *et al.* 2019 for a similar example) as its relationship with  
178 species richness has been highlighted in a spatial context in recent studies (Galiana *et al.* 2021, 2022). We also  
179 computed the uncertainty associated with the community and network measurements (richness variance and  
180 link variance, respectively) and compared their spatial distribution (see Supplementary Material).

## 181 **Analyses of results by ecoregions**

182 Since both species composition and network summary values display a high spatial variation and complex  
183 patterns, we simplified the representation of their distribution by grouping sites by ecoregion, as species and  
184 interaction composition have been shown to differ between ecoregions across large spatial scales (Martins *et al.*  
185 2022). To do so, we rasterized the Canadian subset of the global map of ecoregions from Dinerstein *et al.*  
186 (2017; also used by Martins *et al.* 2022), which resulted in 44 different ecoregions. For every measure we  
187 report (e.g. species richness, number of links), we calculated the median site value for each ecoregion. We also  
188 measured within-ecoregion variation as the 89% interquantile range of the site values in each ecoregion  
189 (threshold chosen to avoid confusion with conventional significance tests; McElreath 2020).



## Analyses of ecological uniqueness

We compared the compositional uniqueness of the networks and the communities to assess whether they indicated areas of exceptional composition. We measured uniqueness using the local contributions to beta diversity (LCBD, Legendre & De Cáceres 2013), which identify sites with exceptional composition by quantifying how much one site contributes to the total variance in the community composition. While many studies used LCBD values to evaluate uniqueness on local scales or few study sites (for example, da Silva & Hernández 2014; Heino & Grönroos 2017), recent studies used the measure on predicted species compositions over broad spatial extents and a large number of sites (Vasconcelos *et al.* 2018; Dansereau *et al.* 2022). LCBD values can also be used to measure uniqueness for networks by computing the values over the adjacency matrix, which has been shown to capture more unique sites and uniqueness variability than through species composition (Poisot *et al.* 2017). Here, we measured and compared the uniqueness of our localized community and network predictions. For species composition, we assembled a site-by-species community matrix with the probability of occurrence at every site from the species distribution models. For network composition, we assembled a site-by-interaction matrix with the localized interaction values from the spatial probabilistic metaweb. We applied the Hellinger transformation on both matrices and computed the LCBD values from the total variance in the matrices (Legendre & De Cáceres 2013). High LCBD values indicate a high contribution to the overall variance and a unique species or interaction composition compared to other sites. Since values themselves are very low given our high number of sites (as in Dansereau *et al.* 2022), what matters primarily is the magnitude of the difference between the sites. Given this, we divided values by the maximum value in each matrix (species or network) and suggest that these should be viewed as relative contributions compared to the highest observed contribution. As with other measures, we then summarized the local uniqueness values by ecoregion by taking the median LCBD value and measuring the 89% interquantile range within all ecoregions.

We used *Julia* v1.9.0 (Bezanson *et al.* 2017) to implement all our analyses. We used packages `GBIF.jl` (Dansereau & Poisot 2021) to reconcile species names using the GBIF Backbone Taxonomy, `SpeciesDistributionToolkit.jl` (<https://github.com/PoisotLab/SpeciesDistributionToolkit.jl>) to handle raster layers, species occurrences and generate pseudoabsences, `EvoTrees.jl` (<https://github.com/EvoVest/EvoTrees.jl>) to perform the Gradient Boosted Trees, `EcologicalNetworks.jl` (Poisot *et al.* 2019) to analyze network and metaweb structure, and `Makie.jl` (Danisch & Krumbiegel 2021) to produce figures. Our data sources (CHELSA, EarthEnv, Ecoregions) were all unprojected, and we did not use a projection in our analyses. However, we displayed the results using a Lambert conformal conic projection more

appropriate for Canada using `GeoMakie.jl` (<https://github.com/MakieOrg/GeoMakie.jl>). All the code used to implement our analyses is available on GitHub (<https://github.com/PoisotLab/SpatialProbabilisticMetaweb>) and includes instructions on how to run a smaller example at a coarser resolution. Note that running our analyses at full scale is resource and memory-intensive and required the use of compute clusters provided by Calcul Québec and the Digital Research Alliance of Canada. Final scripts required 900 CPU core-hours and peaked at 500 GB of RAM.

## Results

Our method allowed us to display the spatial distribution of ecoregion-level community measures (here, expected species richness) and network measures (expected number of links; Fig. 2). We highlight that the community and network-level measures presented here are not actual predictions of the measure itself (e.g., we do not present a prediction of actual species richness at each location). Instead, they are the reflection of these metrics from the localized predictions of the communities and networks obtained from the downscaling of the metaweb, then summarized for the ecoregions (median value). Expected ecoregion richness (Fig. 2A) and expected number of links (Fig. 2B) displayed similar distributions with a latitudinal gradient and higher values in the south. However, within-ecoregion variability was distributed differently, as some ecoregions along the coasts displayed higher interquantile ranges while ecoregions around the southern border displayed narrower ones (Fig. 2C-D). All results shown are based on the first sampling strategy (option 1) mentioned in the *Building site-level instances of the metaweb* section, where species occurrence probabilities were taken as the mean value of the distribution (results for other sampling strategies are discussed in Supplementary Material).

[Figure 2 about here.]

Direct comparison of the spatial distributions of species richness and expected number of links showed some areas with mismatches, both regarding the median estimates and regarding the within-ecoregion variability (Fig. 3). Median values for the ecoregions showed a similar bivariate distribution, with ecoregions in the south mostly displaying high species richness and a high number of links (Fig. 3A). The northernmost ecoregions (Canadian High Arctic Tundra and Davis Highlands Tundra) displayed higher richness (based on the quantile rank) compared to the number of links. Inversely, ecoregions further south (Canadian Low Arctic Tundra, Northern Canadian Shield Taiga, Southern Hudson Bay Taiga) ranked higher for the number of links than for

247 species richness. On the other hand, within-ecoregion variability showed different bivariate relationships and a  
248 less constant latitudinal gradient (Fig. 3B). This indicates that richness and links do not co-vary completely  
249 (i.e. their variability is not closely connected) although they may show similar distributions for median values.

250 [Figure 3 about here.]

251 Our results also indicate a mismatch between the uniqueness of communities and networks (Fig. 4). Uniqueness  
252 was higher mostly in the north and along the south border for communities, but only in the north for networks  
253 (Fig. 4A-B). Consequently, ecoregions with both unique community composition and unique network  
254 composition were mostly in the north (Fig. 4C). Meanwhile, some areas were unique for one element but not the  
255 other. For instance, the New England-Acadian forests ecoregion (south-east, near 70°W and 48°N) had a highly  
256 unique species composition but a more common network composition (Fig. 4C). Opposite areas with unique  
257 network compositions only were observed at higher between latitudes 52°N and 70°N (Eastern Canadian Shield  
258 Taiga, Northern Canadian Shield Taiga, Canadian Low Arctic Tundra). Also, network uniqueness values for  
259 ecoregions spanned a narrower range between the 44 ecoregions than species LCBD values (Fig. 4D, left).  
260 Within-ecoregion variation was also lower for network values with generally lower 89% interquantile ranges  
261 among the site-level LCBD values (Fig. 4D, right). Moreover, mismatched sites (unique for only one element)  
262 formed two distinct groups when evaluating the relationship between species richness and the number of links  
263 (see Supplementary Material). The areas only unique for their species composition had both a high richness and  
264 number of links. On the other hand, the sites only unique for their networks had both lower richness and a lower  
265 number of links, although they were not the sites with the lowest values for both.

266 [Figure 4 about here.]

## 267 Discussion

268 Our approach presents a way to downscale a metaweb, produce localized predictions using probabilistic  
269 networks as inputs and outputs, and incorporate uncertainty, as called for by Strydom *et al.* (2022b). It gives us  
270 an idea of what local metawebs or networks could look like in space, given the species distributions and their  
271 variability, as well as the uncertainty around the interactions. We also provide the first spatial representation of  
272 the metaweb of Canadian mammals (Strydom *et al.* 2022a) and a probabilistic equivalent to how the European

273 tetrapod metaweb (Maiorano *et al.* 2020) was used to predict localized networks in Europe (Braga *et al.* 2019;  
 274 O'Connor *et al.* 2020; Galiana *et al.* 2021; Gaüzère *et al.* 2022; Botella *et al.* 2023). Therefore, our approach  
 275 could open similar possibilities of investigations in North America with food webs of Canadian mammals, for  
 276 instance, on the structure of food webs over space (Braga *et al.* 2019) and on the effect of land-use  
 277 intensification on food webs (Botella *et al.* 2023). Interesting research areas include assessing climate change  
 278 impacts on network structure or investigating linkages between network structure and stability. Moreover, since  
 279 our approach is probabilistic, it does not assume species interact whenever they co-occur and incorporates  
 280 variability based on environmental conditions, which could lead to different results by introducing a different  
 281 association between species richness and network properties. Galiana *et al.* (2021) found that species richness  
 282 had a large explanatory power over network properties but mentioned it could potentially be due to interactions  
 283 between species being fixed in space. Here, we found mismatches in the distribution of species richness and  
 284 interactions, and especially regarding their within-ecoregion variability (Fig. 3), highlighting that interactions  
 285 might vary differently than species distributions in space. Network measures (links on Fig. 3A) were also lower  
 286 in the north, contrarily to previous studies where connectance was higher in the north, although those were in  
 287 Europe for all tetrapods (Braga *et al.* 2019; Galiana *et al.* 2021) and willow-galler-parasitoid networks (Gravel  
 288 *et al.* 2019). Further research should investigate why these results might differ between the two continents and  
 289 whether it is due to the methodology, data, or biogeographical processes.

290 Our LCBD and uniqueness results highlighted that areas with unique network composition might differ from  
 291 sites with unique species composition. In other words, the joint distribution of community and network  
 292 uniqueness highlights different diversity hotspots. Poisot *et al.* (2017) showed a similar result with host-parasite  
 293 communities of rodents and ecto-fleas. Our results further show how these differences could be distributed  
 294 across ecoregions and a broad spatial extent. Areas unique for only one element (species or network  
 295 composition) differed in their combination of species richness and number of links (supplementary material),  
 296 with species-unique sites displaying high values of both measures and network-unique sites displaying low  
 297 values. Moreover, LCBD scores essentially highlight variability hotspots and are a measure of the variance of  
 298 community or network structure. Here, they also serve as an inter-ecoregion variation measure, which can be  
 299 compared to the within-ecoregion variation highlighted by the interquantile ranges. The narrower range of  
 300 values for network LCBD values and the lower IQR values indicate that both the inter-ecoregion and  
 301 within-ecoregion variation are lower for networks than for species (Fig. 4). Additionally, higher values for  
 302 network LCBD also indicate that most ecoregions can hold ecologically unique sites.

303 When to use the method we presented here will depend on the availability of interaction data or existing  
304 metawebs and on the intent to incorporate interaction variability, as well as ecoregion-level variability. In  
305 systems where in situ interaction and network data are available, the approach put forward by Gravel *et al.*  
306 (2019) achieves a similar purpose as we attempted here, but is more rigorous and allows modelling the effect  
307 of the environment on the interactions. Without such data, establishing or predicting the metaweb should be the  
308 first step toward producing localized predictions (Strydom *et al.* 2022b). Well-documented binary metawebs  
309 such as the European tetrapod metaweb could be partly combined with our approach if used with probabilistic  
310 SDMs and summarized by ecoregions (as they would only lack an initial probabilistic metaweb, but would still  
311 obtain a more probabilistic output). Our approach will essentially differ from previous attempts in how it  
312 perceives uncertainty and variability. For instance, rare interactions should not be over-represented (Poisot *et al.*  
313 2016) and should have lesser effects over computed network measures. Furthermore, summarizing results by  
314 ecoregion allows for showing variation within and between ecologically meaningful biogeographic boundaries  
315 (Martins *et al.* 2022), which, as our results showed, is not constant across space and can help identify  
316 contrasting diversity hotspots.

317 The recent shift in focus towards building metawebs opens many opportunities for projections of networks in  
318 space through probabilistic downscaling, as we suggested here. Metawebs have been documented in many  
319 systems, allowing us to build new ones from predictions. How the European tetrapod metaweb (Maiorano *et al.*  
320 2020) was used to predict the Canadian mammal metaweb (Strydom *et al.* 2022a) is one such case, but recent  
321 examples also extend to other systems. Metawebs have been compiled for many marine food webs (e.g., Barents  
322 Sea, Kortsch *et al.* 2019; North Scotia Sea, López-López *et al.* 2022; Gulf of Riga, Kortsch *et al.* 2021) and  
323 used to predict the probability of novel interactions (Arctic food web of the Barents sea, Pecuchet *et al.* 2020).  
324 Olivier *et al.* (2019) built a temporally resolved metaweb of demersal fish and benthic epifauna but also  
325 suggested combining their approach with techniques estimating the probability of occurrence of trophic  
326 relationships to describe spatial and temporal variability more accurately. Lurgi *et al.* (2020) built a metaweb  
327 and probabilistic (occurrence-based) networks for rocky intertidal communities (and also showed that  
328 environmental factors do not affect the structure of binary and probabilistic networks in different ways). Albouy  
329 *et al.* (2019) predicted the global marine fish food web using a probabilistic model, showing the potential to  
330 describe networks across broad spatial scales. Similarly, predictive approaches are also increasingly used with  
331 other interaction types to highlight interactions hotspots on global scales (e.g. mapping geographical hotspots of  
332 predicted host-virus interactions between bats and betacoronaviruses, Becker *et al.* 2022; predicting the

333 distribution of hidden interactions in the mammalian virome, Poisot *et al.* 2023). Overall, these recent examples  
334 show that there is an opportunity for downscaling towards local network predictions through probabilistic  
335 metawebs and that many different systems can now be projected in space.

336 In this study, we presented a probabilistic framework to downscale a metaweb towards local networks based on  
337 the Canadian mammal metaweb and species occurrences from GBIF. Our approach showed that community and  
338 network structures do not always vary in the same way between and within ecoregions. Contrasting these spatial  
339 distributions highlighted variability hotspots unique for different aspects of their biodiversity. Our approach can  
340 be extended to many systems given recent developments in metaweb documentation and prediction, which will  
341 improve the description of ecological networks across broad spatial extents without additional sampling.

## References

- Albouy, C., Archambault, P., Appeltans, W., Araújo, M.B., Beauchesne, D., Cazelles, K., *et al.* (2019). [The marine fish food web is globally connected](#). *Nature Ecology & Evolution*, 3, 1153–1161.
- Baiser, B., Gravel, D., Cirtwill, A.R., Dunne, J.A., Fahimipour, A.K., Gilarranz, L.J., *et al.* (2019). [Ecogeographical rules and the macroecology of food webs](#). *Global Ecology and Biogeography*, geb.12925.
- Barbet-Massin, M., Jiguet, F., Albert, C.H. & Thuiller, W. (2012). [Selecting pseudo-absences for species distribution models: How, where and how many?](#) *Methods in Ecology and Evolution*, 3, 327–338.
- Becker, D.J., Albery, G.F., Sjodin, A.R., Poisot, T., Bergner, L.M., Chen, B., *et al.* (2022). [Optimising predictive models to prioritise viral discovery in zoonotic reservoirs](#). *The Lancet Microbe*, 3, e625–e637.
- Bezanson, J., Edelman, A., Karpinski, S. & Shah, V.B. (2017). [Julia: A fresh approach to numerical computing](#). *SIAM Review*, 59, 65–98.
- Botella, C., Gaüzère, P., O'Connor, L., Ohlmann, M., Renaud, J., Dou, Y., *et al.* (2023). [Land-use intensity influences European tetrapod food-webs](#) (Preprint). Authorea.
- Braga, J., Pollock, L.J., Barros, C., Galiana, N., Montoya, J.M., Gravel, D., *et al.* (2019). [Spatial analyses of multi-trophic terrestrial vertebrate assemblages in Europe](#). *Global Ecology and Biogeography*, 28, 1636–1648.
- da Silva, P.G. & Hernández, M.I.M. (2014). [Local and regional effects on community structure of dung beetles in a mainland-island scenario](#). *PLOS ONE*, 9, e111883.
- Danisch, S. & Krumbiegel, J. (2021). [Makie.jl: Flexible high-performance data visualization for Julia](#). *Journal of Open Source Software*, 6, 3349.
- Dansereau, G., Legendre, P. & Poisot, T. (2022). [Evaluating ecological uniqueness over broad spatial extents using species distribution modelling](#). *Oikos*, 2022, e09063.
- Dansereau, G. & Poisot, T. (2021). [SimpleSDMLayers.jl and GBIF.jl: A framework for species distribution modeling in Julia](#). *Journal of Open Source Software*, 6, 2872.
- Desjardins-Proulx, P., Laigle, I., Poisot, T. & Gravel, D. (2017). [Ecological interactions and the Netflix problem](#). *PeerJ*, 5, e3644.

368 Dinerstein, E., Olson, D., Joshi, A., Vynne, C., Burgess, N.D., Wikramanayake, E., *et al.* (2017). [An](#)  
369 [Ecoregion-Based Approach to Protecting Half the Terrestrial Realm](#). *BioScience*, 67, 534–545.

370 Dunne, J. (2006). The network structure of food webs. In: *Ecological Networks: Linking Structure to Dynamics*  
371 *in Food Webs*. pp. 27–86.

372 Frelat, R., Kortsch, S., Kröncke, I., Neumann, H., Nordström, M.C., Olivier, P.E.N., *et al.* (2022). [Food web](#)  
373 [structure and community composition: A comparison across space and time in the North Sea](#). *Ecography*,  
374 2022.

375 Galiana, N., Barros, C., Braga, J., Ficetola, G.F., Maiorano, L., Thuiller, W., *et al.* (2021). [The spatial scaling of](#)  
376 [food web structure across European biogeographical regions](#). *Ecography*, 44, 653–664.

377 Galiana, N., Lurgi, M., Bastazini, V.A.G., Bosch, J., Cagnolo, L., Cazelles, K., *et al.* (2022). [Ecological](#)  
378 [network complexity scales with area](#). *Nature Ecology & Evolution*, 1–8.

379 Gaüzère, P., O'Connor, L., Botella, C., Poggiato, G., Münkemüller, T., Pollock, L.J., *et al.* (2022). [The diversity](#)  
380 [of biotic interactions complements functional and phylogenetic facets of biodiversity](#). *Current Biology*.

381 GBIF Secretariat. (2021). [GBIF Backbone Taxonomy](#).

382 GBIF.org. (2022). [GBIF occurrence download](#).

383 GDAL/OGR contributors. (2021). *GDAL/OGR geospatial data abstraction software library*. Manual. Open  
384 Source Geospatial Foundation.

385 Gravel, D., Baiser, B., Dunne, J.A., Kopelke, J.-P., Martinez, N.D., Nyman, T., *et al.* (2019). [Bringing Elton](#)  
386 [and Grinnell together: A quantitative framework to represent the biogeography of ecological interaction](#)  
387 [networks](#). *Ecography*, 42, 401–415.

388 Guisan, A. & Thuiller, W. (2005). [Predicting species distribution: Offering more than simple habitat models](#).  
389 *Ecology Letters*, 8, 993–1009.

390 Heino, J. & Grönroos, M. (2017). [Exploring species and site contributions to beta diversity in stream insect](#)  
391 [assemblages](#). *Oecologia*, 183, 151–160.

392 Hortal, J., de Bello, F., Diniz-Filho, J.A.F., Lewinsohn, T.M., Lobo, J.M. & Ladle, R.J. (2015). [Seven Shortfalls](#)  
393 [that Beset Large-Scale Knowledge of Biodiversity](#). *Annual Review of Ecology, Evolution, and Systematics*,  
394 46, 523–549.



Jordano, P. (2016). [Sampling networks of ecological interactions](#). *Functional Ecology*, 30, 1883–1893.

Karger, D.N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., *et al.* (2017). [Climatologies at high resolution for the earth's land surface areas](#). *Scientific Data*, 4, 170122.

Karger, D.N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., *et al.* (2018). [Data from: Climatologies at high resolution for the earth's land surface areas](#).

Karger, D.N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., *et al.* (2021). [Climatologies at high resolution for the earth's land surface areas](#).

Kortsch, S., Frelat, R., Pecuchet, L., Olivier, P., Putnis, I., Bonsdorff, E., *et al.* (2021). [Disentangling temporal food web dynamics facilitates understanding of ecosystem functioning](#). *Journal of Animal Ecology*, 90, 1205–1216.

Kortsch, S., Primicerio, R., Aschan, M., Lind, S., Dolgov, A.V. & Planque, B. (2019). [Food-web structure varies along environmental gradients in a high-latitude marine ecosystem](#). *Ecography*, 42, 295–308.

Legendre, P. & De Cáceres, M. (2013). [Beta diversity as the variance of community data: Dissimilarity coefficients and partitioning](#). *Ecology Letters*, 16, 951–963.

López-López, L., Genner, M.J., Tarling, G.A., Saunders, R.A. & O’Gorman, E.J. (2022). [Ecological Networks in the Scotia Sea: Structural Changes Across Latitude and Depth](#). *Ecosystems*, 25, 457–470.

Lucas, P., Thuiller, W., Talluto, M., Polaina, E., Albrecht, J., Selva, N., *et al.* (2023). [Including biotic interactions in species distribution models improves the understanding of species niche: A case of study with the brown bear in Europe](#).

Lurgi, M., Galiana, N., Broitman, B.R., Kéfi, S., Wieters, E.A. & Navarrete, S.A. (2020). [Geographical variation of multiplex ecological networks in marine intertidal communities](#). *Ecology*, 101, e03165.

Maiorano, L., Montemaggiore, A., Ficetola, G.F., O’Connor, L. & Thuiller, W. (2020). [TETRA-EU 1.0: A species-level trophic metaweb of European tetrapods](#). *Global Ecology and Biogeography*, 29, 1452–1457.

Martins, L.P., Stouffer, D.B., Blendinger, P.G., Böhning-Gaese, K., Buitrón-Jurado, G., Correia, M., *et al.* (2022). [Global and regional ecological boundaries explain abrupt spatial discontinuities in avian frugivory interactions](#). *Nature Communications*, 13, 6943.

McElreath, R. (2020). [Statistical rethinking: A bayesian course with examples in R and Stan](#). Second. Chapman and Hall/CRC, New York.

McLeod, A., Leroux, S.J., Gravel, D., Chu, C., Cirtwill, A.R., Fortin, M.-J., *et al.* (2021). [Sampling and asymptotic network properties of spatial multi-trophic networks](#). *Oikos*, 130, 2250–2259.

Mestre, F., Gravel, D., García-Callejas, D., Pinto-Cruz, C., Matias, M.G. & Araújo, M.B. (2022). [Disentangling food-web environment relationships: A review with guidelines](#). *Basic and Applied Ecology*, 61, 102–115.

Moens, M., Biesmeijer, J., Huang, E., Vereecken, N. & Marshall, L. (2022). [The importance of biotic interactions in distribution models depends on the type of ecological relations, spatial scale and range](#).

Morales-Castilla, I., Matias, M.G., Gravel, D. & Araújo, M.B. (2015). [Inferring biotic interactions from proxies](#). *Trends in Ecology & Evolution*, 30, 347–356.

O'Connor, L.M.J., Pollock, L.J., Braga, J., Ficetola, G.F., Maiorano, L., Martinez-Almoyna, C., *et al.* (2020). [Unveiling the food webs of tetrapods across Europe through the prism of the Eltonian niche](#). *Journal of Biogeography*, 47, 181–192.

Olivier, P., Frelat, R., Bonsdorff, E., Kortsch, S., Kröncke, I., Möllmann, C., *et al.* (2019). [Exploring the temporal variability of a food web using long-term biomonitoring data](#). *Ecography*, 42, 2107–2121.

Pecuchet, L., Blanchet, M.-A., Frainer, A., Husson, B., Jørgensen, L.L., Kortsch, S., *et al.* (2020). [Novel feeding interactions amplify the impact of species redistribution on an Arctic food web](#). *Global Change Biology*, 26, 4894–4906.

Poggiato, G., Andréoletti, J., Shirley, L. & Thuiller, W. (2022). [Integrating food webs in species distribution models improves ecological niche estimation and predictions](#) (Preprint). Authorea.

Poisot, T., Bélisle, Z., Hoebeke, L., Stock, M. & Szefer, P. (2019). [EcologicalNetworks.jl: Analysing ecological networks of species interactions](#). *Ecography*, 42, 1850–1861.

Poisot, T., Bergeron, G., Cazelles, K., Dallas, T., Gravel, D., MacDonald, A., *et al.* (2021). [Global knowledge gaps in species interaction networks data](#). *Journal of Biogeography*, 48, 1552–1563.

Poisot, T., Canard, E., Mouillot, D., Mouquet, N. & Gravel, D. (2012). [The dissimilarity of species interaction networks](#). *Ecology Letters*, 15, 1353–1361.

Poisot, T., Cirtwill, A.R., Cazelles, K., Gravel, D., Fortin, M.-J. & Stouffer, D.B. (2016). [The structure of probabilistic networks](#). *Methods in Ecology and Evolution*, 7, 303–312.

Poisot, T., Guéveneux-Julien, C., Fortin, M.-J., Gravel, D. & Legendre, P. (2017). [Hosts, parasites and their interactions respond to different climatic variables](#). *Global Ecology and Biogeography*, 26, 942–951.

451 Poiset, T., Ouellet, M.-A., Mollentze, N., Farrell, M.J., Becker, D.J., Brierley, L., *et al.* (2023). [Network](#)  
452 [embedding unveils the hidden interactions in the mammalian virome](#). *Patterns*, 4, 100738.

453 Poiset, T., Stouffer, D.B. & Gravel, D. (2015). [Beyond species: Why ecological interaction networks vary](#)  
454 [through space and time](#). *Oikos*, 124, 243–251.

455 Saravia, L.A., Marina, T.I., Kristensen, N.P., De Troch, M. & Momo, F.R. (2022). [Ecological network](#)  
456 [assembly: How the regional metaweb influences local food webs](#). *Journal of Animal Ecology*, 91, 630–642.

457 Statistics Canada. (2022). *Boundary files, reference guide second edition, Census year 2021*. Second edition.  
458 Statistics Canada = Statistique Canada, Ottawa.

459 Strydom, T., Bouskila, S., Banville, F., Barros, C., Caron, D., Farrell, M.J., *et al.* (2022a). [Food web](#)  
460 [reconstruction through phylogenetic transfer of low-rank network representation](#). *Methods in Ecology and*  
461 *Evolution*, 13, 2838–2849.

462 Strydom, T., Bouskila, S., Banville, F., Barros, C., Caron, D., Farrell, M.J., *et al.* (2022b). [Predicting metawebs:](#)  
463 [Transfer of graph embeddings can help alleviate spatial data deficiencies](#).

464 Strydom, T., Catchen, M.D., Banville, F., Caron, D., Dansereau, G., Desjardins-Proulx, P., *et al.* (2021). [A](#)  
465 [roadmap towards predicting species interaction networks \(across space and time\)](#). *Philosophical*  
466 *Transactions of the Royal Society B: Biological Sciences*, 376, 20210063.

467 Tuanmu, M.-N. & Jetz, W. (2014). [A global 1-km consensus land-cover product for biodiversity and ecosystem](#)  
468 [modelling](#). *Global Ecology and Biogeography*, 23, 1031–1045.

469 Vasconcelos, T.S., Nascimento, B.T.M. do & Prado, V.H.M. (2018). [Expected impacts of climate change](#)  
470 [threaten the anuran diversity in the Brazilian hotspots](#). *Ecology and Evolution*, 8, 7894–7906.

471 Windsor, F.M., van den Hoogen, J., Crowther, T.W. & Evans, D.M. (2023). [Using ecological networks to](#)  
472 [answer questions in global biogeography and ecology](#). *Journal of Biogeography*, 50, 57–69.

473 Wisz, M.S., Pottier, J., Kissling, W.D., Pellissier, L., Lenoir, J., Damgaard, C.F., *et al.* (2013). [The role of biotic](#)  
474 [interactions in shaping distributions and realised assemblages of species: Implications for species](#)  
475 [distribution modelling](#). *Biological Reviews*, 88, 15–30.

476 Youden, W.J. (1950). [Index for rating diagnostic tests](#). *Cancer*, 3, 32–35.

477 Zarnetske, P.L., Baiser, B., Strecker, A., Record, S., Belmaker, J. & Tuanmu, M.-N. (2017). [The Interplay](#)  
478 [Between Landscape Structure and Biotic Interactions](#). *Current Landscape Ecology Reports*, 2, 12–29.

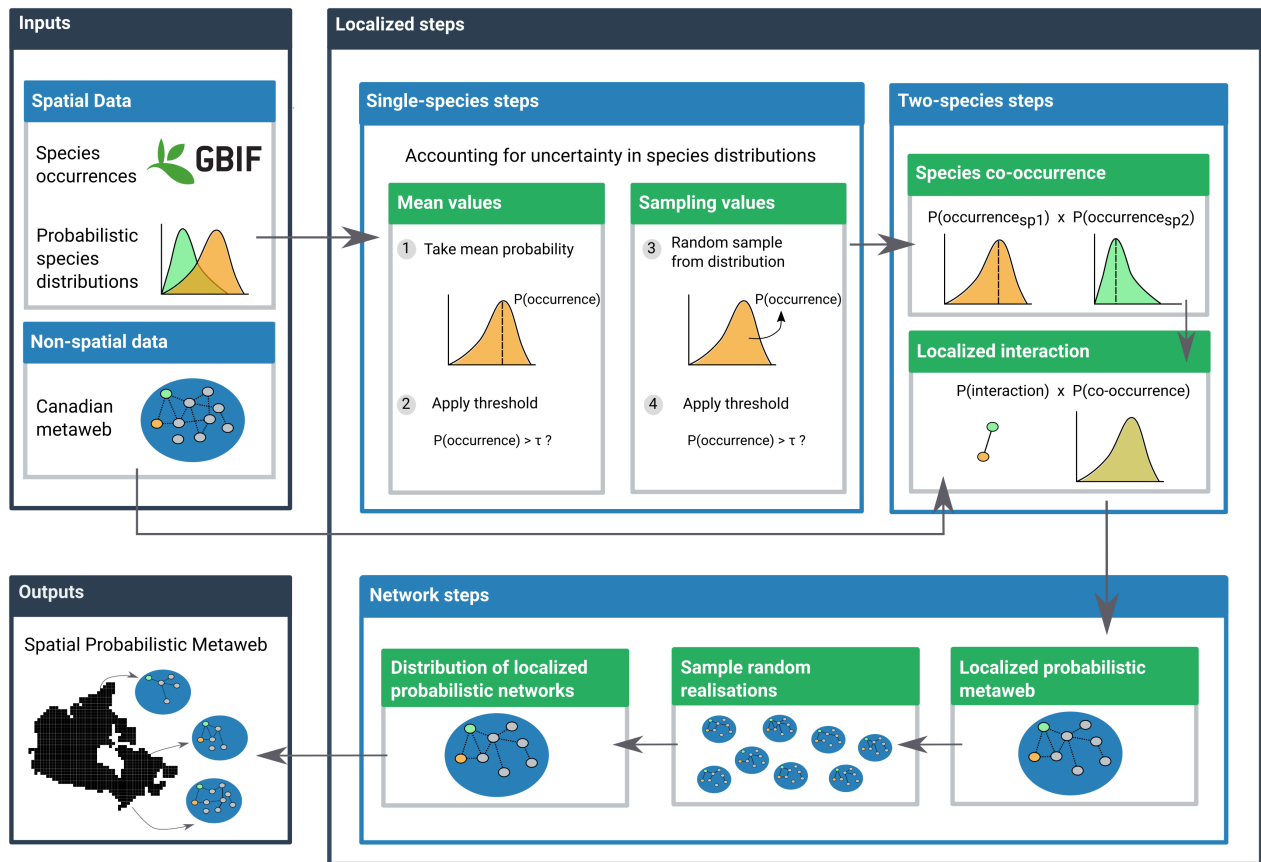


Figure 1: Conceptual figure of the workflow to obtain the spatial probabilistic metaweb (Chapter 1). The workflow has three components: the inputs, the localized steps, and the final spatial output. The inputs are composed of the spatial data (data with information in every cell) and the non-spatial data (constant for all of Canada). The localized steps use these data and are performed separately in every cell, first at a single-species level (using distribution data), then for every species pair (adding interaction data from the metaweb), and finally at the network level by combining the results of all species pairs. The final output of the network-level steps contains a downscaled probabilistic metaweb for every cell across the study extent.

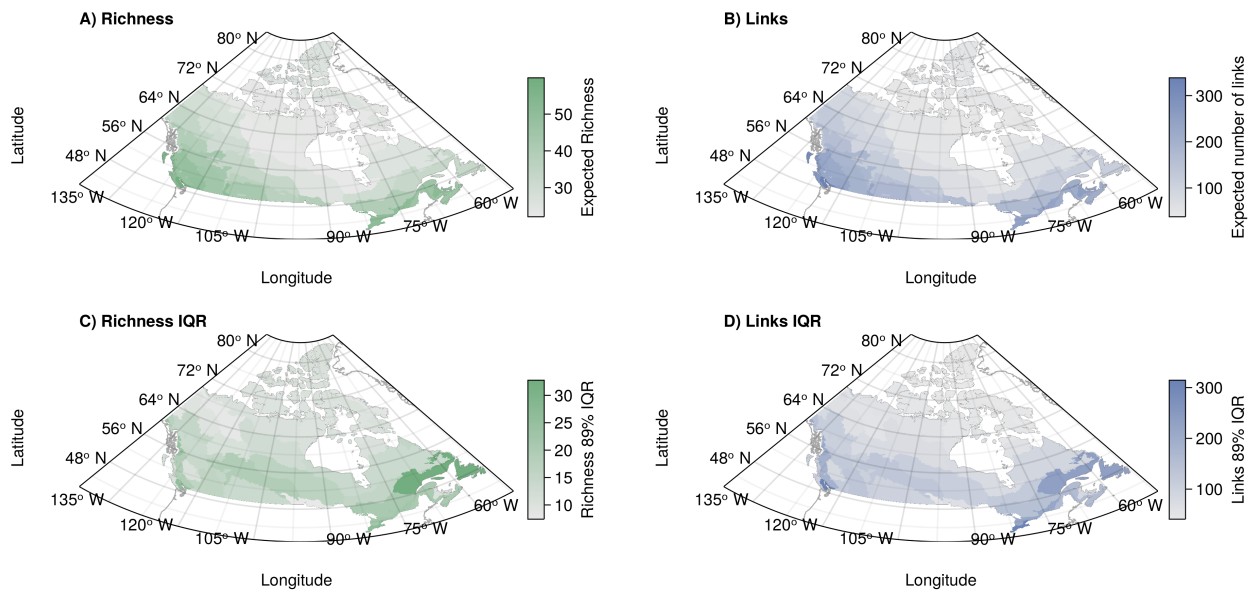


Figure 2: (A-B) Example of a community measure (A, expected species richness) and a network one (B, expected number of links). Both measures are assembled from the predicted probabilistic communities and networks, respectively. Values are first measured separately for all sites; then, the median value is taken to represent the ecoregion-level value. (C-B) Representation of the 89% interquantile range of values within the ecoregion for expected richness (C) and expected number of links (D).

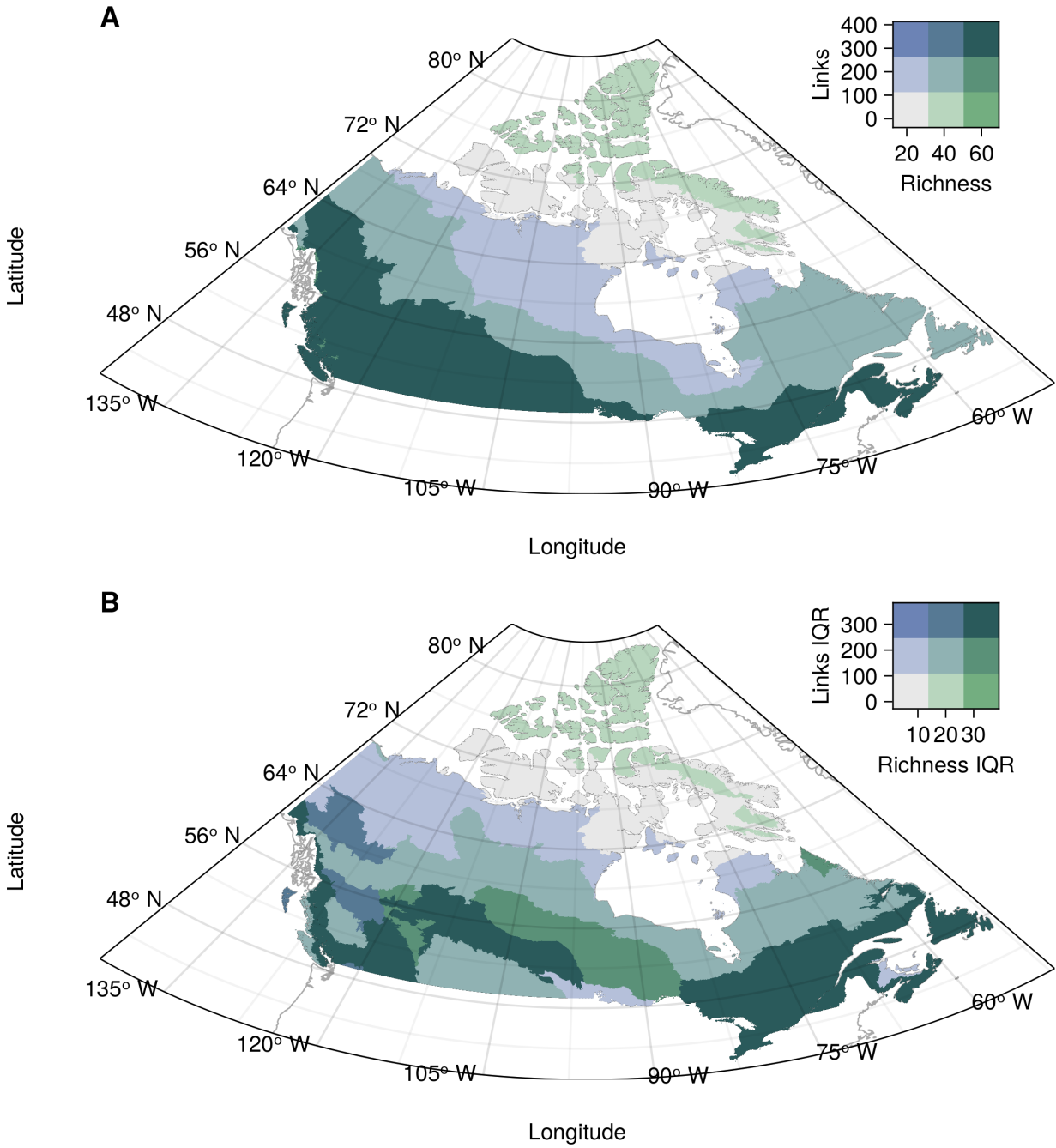


Figure 3: Bivariate relationship between community and network measures for the median ecoregion value (A) and the within-ecoregion 89% interquantile range (B). Values are grouped into three quantiles separately for each variable. The colour combinations represent the nine possible combinations of quantiles. Species richness (horizontal axis) goes left to right from low (light grey, bottom left) to high (green, bottom right). The number of links goes bottom-up from low (light grey, bottom left) to high (blue, top left).

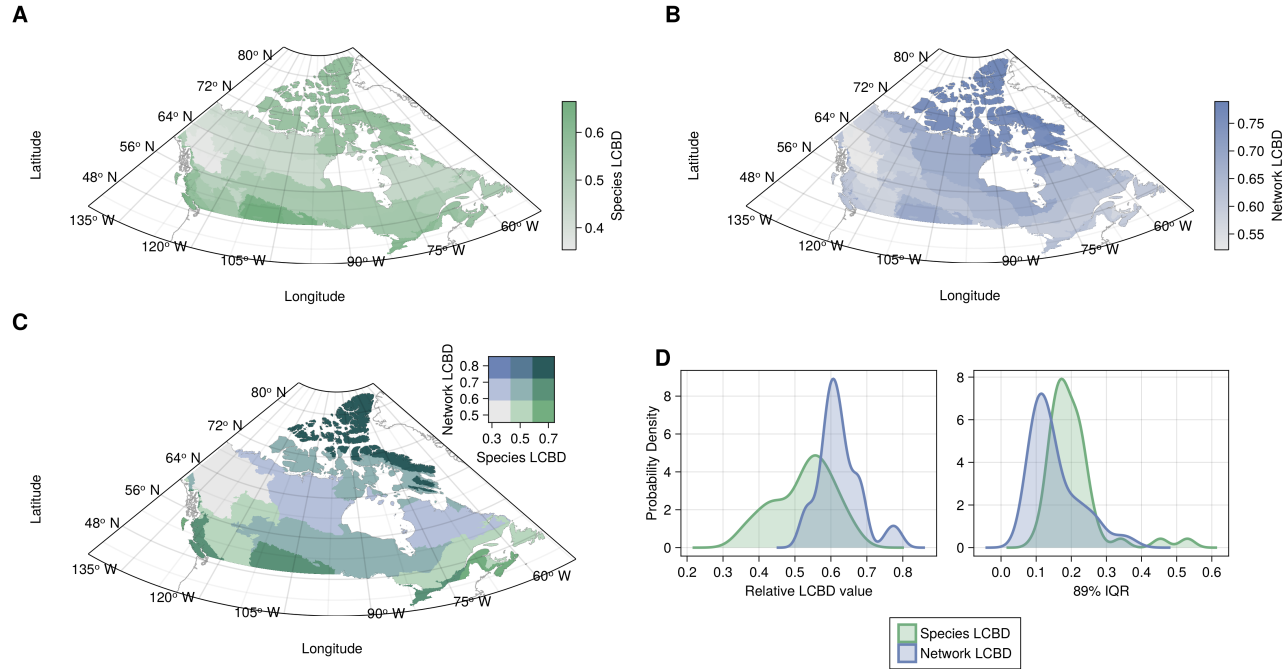


Figure 4: (A-B) Representation of the ecoregion uniqueness values based on species composition (a) and network composition (b). LCBD values were first computed across all sites and scaled relative to the maximum value observed. The ecoregion LCBD value is the median value for the sites in the ecoregion. (C) Bivariate representation of species and network composition LCBD. Values are grouped into three quantiles separately for each variable. The colour combinations represent the nine possible combinations of quantiles. The species uniqueness (horizontal axis) goes left to right from low uniqueness (light grey, bottom left) to high uniqueness (green, bottom right). The network composition uniqueness goes bottom-up from low uniqueness (light grey, bottom left) to high uniqueness (blue, top left). (D) Probability densities for the ecoregion LCBD values for species and network LCBD (left), highlighting the variability of the LCBD between ecoregions, and the 89% interquartile range of the values within each ecoregion (right), highlighting the variability within the ecoregions.