

Spatially explicit predictions of food web structure from regional level data

Gabriel Dansereau^{1,2} Ceres Barros³ Timothée Poisot^{1,2}

¹ Université de Montréal ² Québec Centre for Biodiversity Sciences ³ University of British Columbia

Correspondance to:

Gabriel Dansereau — gabriel.dansereau@umontreal.ca

Abstract: Knowledge about how ecological networks vary across global scales is currently limited given the complexity of acquiring repeated data for species interactions. Yet, recent developments of metawebs—lists of potential interactions between species in a given species pool—highlight efficient ways to document interactions first on broader scales. Downscaling metawebs towards local network predictions is therefore a promising approach to use current data to investigate the variation of networks across space. However, issues remain in how to represent the spatial variability and uncertainty of species interactions, especially for large scale food webs. Here, we present a probabilistic framework to downscale a metaweb based on the Canadian mammal metaweb and species occurrences from GBIF. We investigate how this approach can be used to represent the variability of networks and communities between ecoregions in Canada. Our results show that community metrics (species richness) and network metrics (number of links) differ in how they vary within and between ecoregions. This allows us to identify variability hotspots unique for different biodiversity aspects. Given recent developments on similar probabilistic metawebs, our approach highlights how there are now many opportunities in various systems for local predictions of networks across broad spatial scales.

This work is released by its authors under a CC-BY 4.0 license



Last revision: *September 14, 2023*

1 Introduction

2 Because species interactions vary in time and space, and because species show high turnover over larger spatial
3 extents, adequately capturing the diversity of ecological networks is a challenging task (Jordano 2016). Most
4 studies on food webs have previously focused on local networks limited in size and extent, and are rarely
5 replicated in space or time (Mestre *et al.* 2022). Interactions can show important variations in space (Poisot *et*
6 *al.* 2015; Zarnetske *et al.* 2017), yet available network data also show important geographical bias, limiting our
7 ability to answer questions in many biomes and over broad spatial extents (Poisot *et al.* 2021), by focusing
8 sampling efforts in a few areas or biomes. Moreover, global monitoring of biotic interactions is insufficient to
9 properly describe and understand how ecosystems are reacting to global change (Windsor *et al.* 2023).

10 Approaches to predict species interactions (e.g., Morales-Castilla *et al.* 2015; Desjardins-Proulx *et al.* 2017) are
11 increasingly used as an alternative to determine potential interactions; they can handle limited data to
12 circumvent data scarcity (Strydom *et al.* 2021), but are still rarely used to make explicitly spatial predictions.
13 As a result, there have been repeated calls for globally distributed interaction and network data coupled to
14 repeated sampling in time and space (Mestre *et al.* 2022; Windsor *et al.* 2023), which will help understand the
15 macroecological variations of food webs (Baiser *et al.* 2019).

16 Despite these limitations, food web ecologists often can infer a reasonable approximation of the network
17 existing within a region. This representation, called a metaweb, contains all possible interactions between
18 species in a given regional species pool (Dunne 2006), and provides a solid foundation to develop approaches to
19 estimate the structure of networks at finer spatial scales.

20 When assembled by integrating different data sources (potentially with additional predictive steps), the metaweb
21 allows to overcome sampling limitations to upscale network data to a global scale. For example, Albouy *et al.*
22 (2019) coupled data on fish distribution with a statistical model of trophic interaction to provide estimates of the
23 potential food web structure at the global scale. Recent studies have focused on assembling metawebs for
24 various taxa through literature surveys and expert elicitation (European tetrapods, Maiorano *et al.* 2020) or
25 using predictive tools (marine fishes, Albouy *et al.* 2019; Canadian mammals, Strydom *et al.* 2022a). At a finer
26 spatial scale, the locally observed food webs [“realization” of the metaweb; Poisot *et al.* (2012)] reflect local
27 environmental conditions, but nevertheless retain the signal of the metaweb to which they belong (Saravia *et al.*
28 2022). Given this, Strydom *et al.* (2022b) called the prediction of the metaweb structure the core goal of
29 predictive network ecology, as if there is a strong link between the metaweb and its local realizations, strong

30 predictions of the metaweb have the potential to bring us closer to producing accurate local (downscaled)
31 predictions. Therefore, establishing or predicting the metaweb should be the first target for systems lacking
32 information about local realizations. This is not the same as using interactions to improve predictions of species
33 distributions, as recent studies have done (Moens *et al.* 2022; Poggiato *et al.* 2022; Lucas *et al.* 2023), although
34 these are incredibly relevant and answer long-standing calls to include interactions within such models (Wisz *et*
35 *al.* 2013). Instead, predicting networks in space is a different task, and it serves another goal: focusing first on
36 the distribution of network structures and its drivers rather than on the distribution of species.

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

54 A key challenge remains in how to downscale a regional metaweb towards local network predictions reflecting
55 the spatial variability of interactions. Even when the metaweb is known, local networks may vary substantially
56 and differ both from one another and from the metaweb (McLeod *et al.* 2021), emphasizing the need for
57 methods to generate local, downscaled network predictions. A potential limitation to previous downscaling
58 approaches is that they assume interactions are constant across space, which ignores well-documented
59 interaction variability, and masks the effect of environmental conditions on interaction realization (Braga *et al.*

2019). In contrast, recent studies argued that seeing interactions as probabilistic events (rather than binary ones) allows us to account for their variability in space (Poisot *et al.* 2016) and that this should also be reflected in metawebs (Strydom *et al.* 2022b). Gravel *et al.* (2019) introduced a probabilistic framework describing how the metaweb can generate local realizations and showed how it could be used for interaction distribution modelling. This approach to downscaling is relevant when combined with in situ observations of interactions and local networks to train interaction models (in this case, with willow-galler-parasitoid networks). However, such data is rarely available across broad spatial extents (Hortal *et al.* 2015; Poisot *et al.* 2021; Windsor *et al.* 2023). Spatially replicated interaction data required for such a model are especially challenging to document with large food web systems such as European tetrapod and Canadian mammal metawebs (Maiorano *et al.* 2020; Strydom *et al.* 2022a), where hundreds of species result in tens of thousands of species pair that may potentially interact. We currently lack a downscaling framework that is both probabilistic and can be trained without replicated in situ interaction data. Additionally, a probabilistic view can allow propagating uncertainty, which can play a key role in evaluating the quality of the predictions. Assessing model uncertainty would enable us to determine to which degree we should trust our predictions and to identify what to do to improve the current knowledge. Here, we present a method to downscale a metaweb in space, and illustrate it by spatially reconstructing local instances of a probabilistic metaweb of Canadian mammals. We do so using a probabilistic approach to both species distributions and interactions in a system without spatially replicated interaction data. We then explore how the spatial structure of the downscaled metaweb varies in space and how the uncertainty of interactions can be made spatially explicit. We further show that the downscaled metaweb can highlight important biodiversity areas and bring novel ecological insights compared to traditional community measures like species richness.

Methods

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

[Figure 1 about here.]

88 **Data**

89 **Metaweb**

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

101 **Species occurrences**

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

114 **Environmental data**

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

130 **Analyses**

131 **Species distribution models**

132 Our selection criteria for choosing an SDM algorithm was to have a method that generated probabilistic results
133 (similar to Gravel *et al.* 2019), including both a probability of occurrence for a species in a specific site and the
134 uncertainty associated with the prediction. These were crucial to obtaining a probabilistic version of the
135 metaweb as they were used to create spatial variations in the localized interaction probabilities (see next
136 section). One suitable method for this is Gradient Boosted Trees with a Gaussian maximum likelihood from the
137 `EvoTrees.jl` *Julia* package (<https://github.com/EvoTrees/EvoTrees.jl>). This method returns a prediction for
138 every site with an average value and a standard deviation, which we used as a measure of uncertainty to build a
139 Normal distribution for the probability of occurrence of a given species at all sites (represented as probability
140 distributions on Fig. 1). We trained models across the extent chosen for occurrences (longitudes 175°W to
141 45°W and latitudes 10°N to 90°N), then predicted species distributions only for Canada. We used the 2021

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

144 **Building site-level instances of the metaweb**

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

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

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

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

170 sampled several random network realizations to represent the potential local realization process (Poisot *et al.*
171 2015). This resulted in a distribution of localized networks, which we averaged over the number of simulations
172 to obtain a single probabilistic network for the site.

173 **Downscaled metaweb**

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

186 **Analyses of results by ecoregions**

187 Since both species composition and network summary values display a high spatial variation and complex
188 patterns, we simplified the representation of their distribution by grouping sites by ecoregion, as species and
189 interaction composition have been shown to differ between ecoregions across large spatial scales (Martins *et al.*
190 2022). To do so, we rasterized the Canadian subset of the global map of ecoregions from Dinerstein *et al.*
191 (2017; also used by Martins *et al.* 2022), which resulted in 44 different ecoregions. For every measure we
192 report (e.g. species richness, number of links), we calculated the median site value for each ecoregion. We also
193 measured within-ecoregion variation as the 89% interquantile range of the site values in each ecoregion
194 (threshold chosen to avoid confusion with conventional significance tests; McElreath 2020).

195 **Analyses of ecological uniqueness**

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

217 We used *Julia* v1.9.0 (Bezanson *et al.* 2017) to implement all our analyses. We used packages `GBIF.jl`
218 (Dansereau & Poisot 2021) to reconcile species names using the GBIF Backbone Taxonomy,
219 `SpeciesDistributionToolkit.jl` (<https://github.com/PoisotLab/SpeciesDistributionToolkit.jl>) to handle
220 raster layers, species occurrences and generate pseudoabsences, `EvoTrees.jl`
221 (<https://github.com/EvoVest/EvoTrees.jl>) to perform the Gradient Boosted Trees, `EcologicalNetworks.jl`
222 (Poisot *et al.* 2019) to analyze network and metaweb structure, and `Makie.jl` (Danisch & Krumbiegel 2021) to
223 produce figures. Our data sources (CHELSA, EarthEnv, Ecoregions) were all unprojected, and we did not use a
224 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 (using the median value as a way to protect against long tails in the distributions). 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,

252 Northern Canadian Shield Taiga, Southern Hudson Bay Taiga) ranked higher for the number of links than for
253 species richness. On the other hand, within-ecoregion variability showed different bivariate relationships and a
254 less constant latitudinal gradient (Fig. 3B). This indicates that richness and links do not co-vary completely
255 (i.e. their variability is not closely connected) although they may show similar distributions for median values.

256 [Figure 3 about here.]

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

272 [Figure 4 about here.]

273 Discussion

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

278 the metaweb of Canadian mammals (Strydom *et al.* 2022a) and a probabilistic equivalent to how the European
 279 tetrapod metaweb (Maiorano *et al.* 2020) was used to predict localized networks in Europe (Braga *et al.* 2019;
 280 O'Connor *et al.* 2020; Galiana *et al.* 2021; Gaüzère *et al.* 2022; Botella *et al.* 2023). Therefore, our approach
 281 could open similar possibilities of investigations in North America with food webs of Canadian mammals, for
 282 instance, on the structure of food webs over space (Braga *et al.* 2019) and on the effect of land-use
 283 intensification on food webs (Botella *et al.* 2023). Interesting research applications include assessing climate
 284 change impacts on network structure or investigating linkages between network structure and stability.

285 As our approach is probabilistic, it does not assume species interact whenever they co-occur and incorporates
 286 variability based on environmental conditions, which could lead to different results by introducing a different
 287 association between species richness and network properties. Galiana *et al.* (2021) found that species richness
 288 had a large explanatory power over network properties but mentioned it could potentially be due to interactions
 289 between species being fixed in space. Here, we found mismatches in the distribution of species richness and
 290 interactions, especially apparent in their within-ecoregion variability (Fig. 3), highlighting that interactions
 291 might vary differently than species distributions even over continental-scale gradients. Network density (links
 292 on Fig. 3A) were also lower in the north, contrarily to previous studies where connectance was higher in the
 293 north, although those were in Europe for all tetrapods (Braga *et al.* 2019; Galiana *et al.* 2021) and
 294 willow-galler-parasitoid networks (Gravel *et al.* 2019). Further research should investigate why these results
 295 might differ between the two continents and whether it is due to the methodology, data, or biogeographical
 296 processes.

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

308 within-ecoregion variation are lower for networks than for species (Fig. 4). Additionally, higher values for
309 network LCBD also indicate that most ecoregions can hold ecologically unique sites.

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

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

338 other interaction types to highlight interactions hotspots on global scales (e.g. mapping geographical hotspots of
339 predicted host-virus interactions between bats and betacoronaviruses, Becker *et al.* 2022; predicting the
340 distribution of hidden interactions in the mammalian virome, Poisot *et al.* 2023). Our method offers the
341 potential to bring these global predictions down to the local scale where they can be made more actionable, and
342 vastly increases the diversity of ecological networks that can be projected in space.

343 **Acknowledgements**

344 We acknowledge that this study was conducted on land within the traditional unceded territory of the Saint
345 Lawrence Iroquoian, Anishinabewaki, Mohawk, Huron-Wendat, and Omàmiwininiwak nations. GD is funded
346 by the NSERC Postgraduate Scholarship – Doctoral (grant ES D – 558643), the FRQNT doctoral scholarship
347 (grant no. 301750), and the NSERC CREATE BIOS² program. TP is funded by the Wellcome Trust
348 (223764/Z/21/Z), NSERC through the Discovery Grant and Discovery Accelerator Supplements programs, and
349 the Courtois Foundation. This research was enabled in part by support provided by Calcul Québec
350 (calculquebec.ca) and the Digital Research Alliance of Canada (alliancecan.ca) through the Narval general
351 purpose cluster.

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.

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

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

382 Frelat, R., Kortsch, S., Kröncke, I., Neumann, H., Nordström, M.C., Olivier, P.E.N., *et al.* (2022). [Food web](#)
383 [structure and community composition: A comparison across space and time in the North Sea](#). *Ecography*,
384 2022.

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

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

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

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

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

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

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

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

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

402 Hortal, J., de Bello, F., Diniz-Filho, J.A.F., Lewinsohn, T.M., Lobo, J.M. & Ladle, R.J. (2015). [Seven Shortfalls](#)
403 [that Beset Large-Scale Knowledge of Biodiversity](#). *Annual Review of Ecology, Evolution, and Systematics*,
404 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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

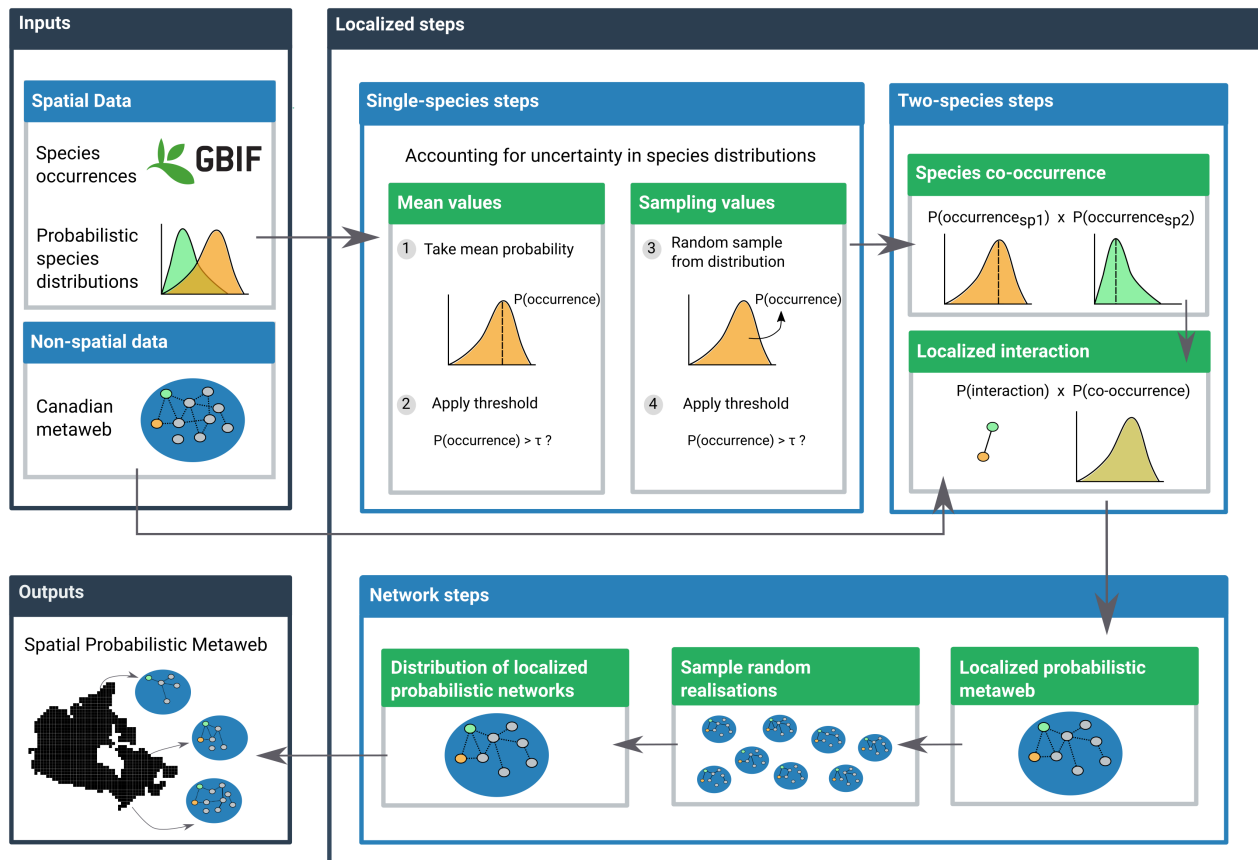


Figure 1: Conceptual figure of the workflow to obtain the spatial probabilistic metaweb (Chapter 1). The workflow has three components: the inputs, the localized steps, and the final spatial output. The inputs are composed of the spatial data (data with information in every cell) and the non-spatial 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.

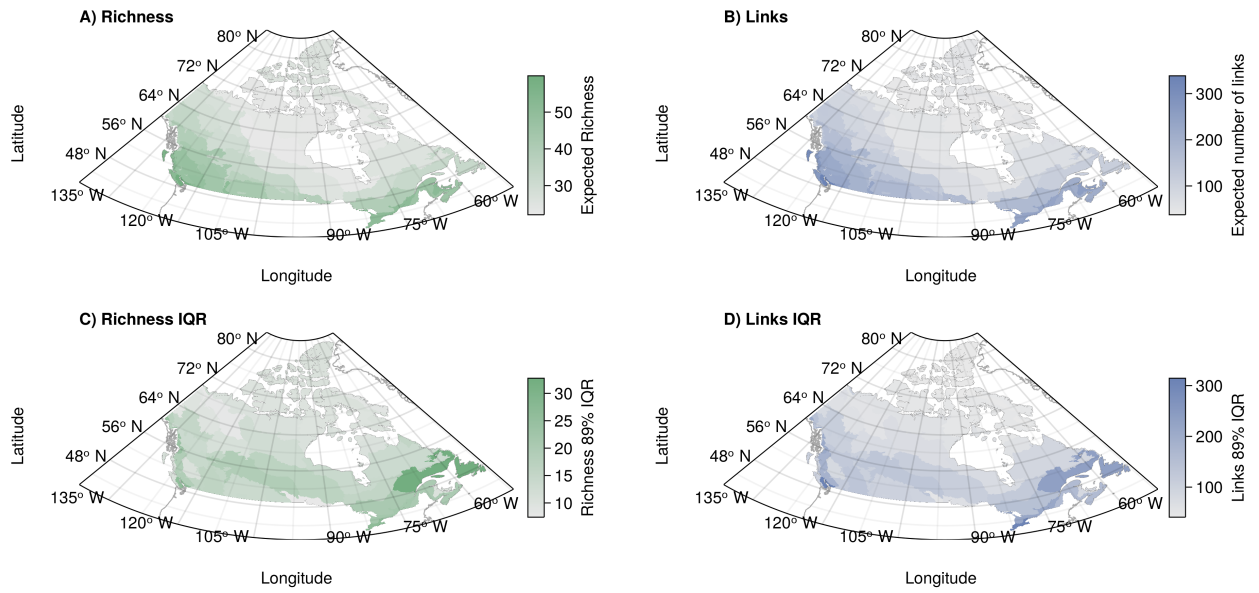


Figure 2: (A-B) Example of a community measure (A, expected species richness) and a network one (B, expected number of links). Both measures are assembled from the predicted probabilistic communities and networks, respectively. Values are first measured separately for all sites; then, the median value 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).

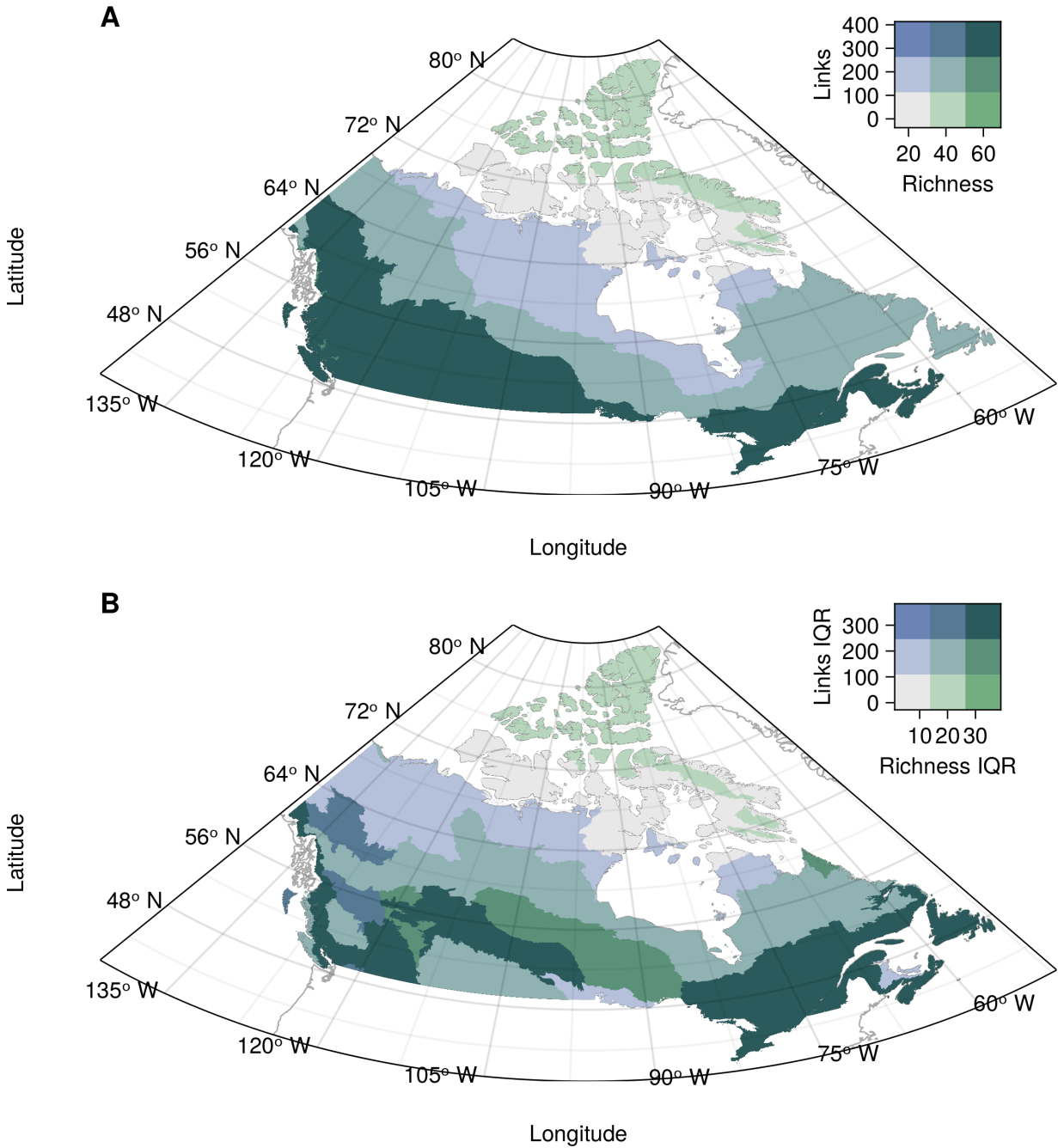


Figure 3: Bivariate relationship between community and network measures for the median ecoregion value (A) and the within-ecoregion 89% interquantile range (B). Values are grouped into three quantiles separately for each variable. The colour combinations represent the nine possible combinations of quantiles. 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).

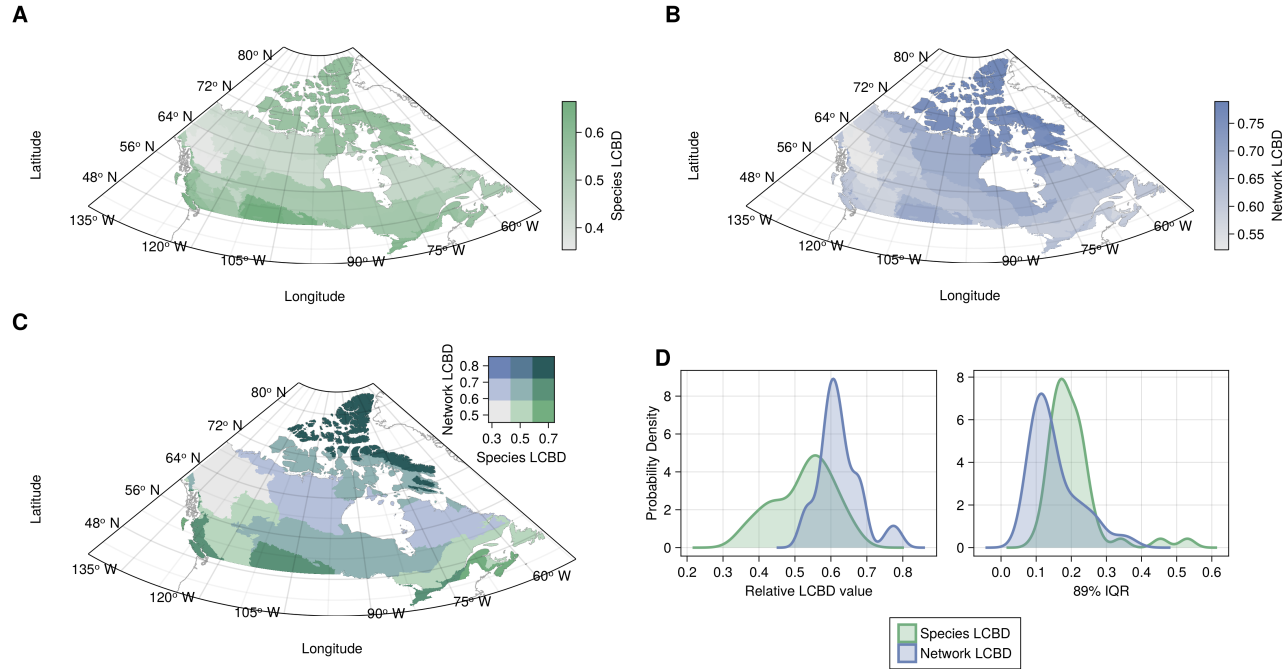


Figure 4: (A-B) Representation of the ecoregion uniqueness values based on species composition (a) and network composition (b). LCBD values were first computed across all sites and scaled relative to the maximum value observed. The ecoregion LCBD value is the median value for the sites in the ecoregion. (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.