

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

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

1 Introduction

2 Sampling ecological networks in space and time is a challenging task (Jordano 2016). Most studies on food
3 webs have previously focused on local webs limited in size and extent, and are rarely replicated in space and
4 time (Mestre *et al.* 2022). Interactions can show important variations in space (Poisot *et al.* 2015), yet available
5 network data also show important geographical bias, limiting our ability to answer questions in many biomes
6 and over broad spatial extents (Poisot *et al.* 2021). Moreover, global network monitoring is insufficient to
7 properly describe and understand how ecosystems are reacting to global change (Windsor *et al.* 2023).
8 Predictive approaches are increasingly used to predict species interactions (e.g., Morales-Castilla *et al.* 2015;
9 Desjardins-Proulx *et al.* 2017) and can handle limited data to circumvent data scarcity (Strydom *et al.* 2021),
10 but are rarely used to make explicitly spatial predictions. As a result, there have been repeated calls for globally
11 distributed interaction and network data and repeated sampling in time and space (Mestre *et al.* 2022; Windsor
12 *et al.* 2023).

13 The metaweb is an increasingly used concept to address the issue of data scarcity, and it further holds potential
14 to analyse networks at large spatial extents. A metaweb contains all possible interactions between species found
15 in a given regional species pool (Dunne 2006). Recent studies have focused on assembling metawebs for
16 various taxa through literature surveys and expert elicitation (European tetrapods, Maiorano *et al.* 2020) or
17 using predictive tools (Canadian mammals, Strydom *et al.* 2022a). In comparison, empirical networks are local
18 realizations of a regional metaweb (Poisot *et al.* 2012) and inherit the metaweb structure with little influence
19 from habitat and dynamical constraints (Saravia *et al.* 2022). Given this, Strydom *et al.* (2022b) called the
20 prediction of the metaweb structure the core goal of predictive network ecology and the key to produce accurate
21 downscaled and local predictions. Establishing or predicting the metaweb should therefore be the first target for
22 systems where we lack information about local realizations. This is not the same as using interactions to
23 improve predictions of species distributions as recent studies have done (for example, Moens *et al.* 2022;
24 Poggiato *et al.* 2022; Lucas *et al.* 2023), although these are incredibly relevant and answer long-standing calls
25 to include interactions within such models (Wisz *et al.* 2013). Instead, predicting networks in space is a
26 different task and it serves a different goal, focusing first on the distribution of networks and its drivers rather
27 than on the distribution of species.

28 A key challenge remains in how to downscale a regional metaweb towards local network predictions. Gravel *et*
29 *al.* (2019) introduced a mathematical framework describing how the metaweb can generate local realizations

30 and showed how it could be used for interaction distribution modelling. This approach to downscaling is useful
31 when combined with in situ observations of interactions and local networks (in this case with
32 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 model
33 is especially challenging to document with large food web systems such as European tetrapod and Canadian
34 mammal metawebs (Maiorano *et al.* 2020; Strydom *et al.* 2022a). In contrast, approaches to downscaling for
35 European tetrapods combined the metaweb with species distribution maps to generate local assemblages (Braga
36 *et al.* 2019; O'Connor *et al.* 2020; Galiana *et al.* 2021; Gaüzère *et al.* 2022). A potential limitation to this
37 approach is that it assumes that interactions are constant across space, which ignores behaviour variability and
38 does not consider the effect of environmental conditions on interaction realization (Braga *et al.* 2019). This is
39 an advantage of the probabilistic framework put forward by Gravel *et al.* (2019) and absent from the European
40 metaweb studies, as treating interactions as probabilistic events allows to account for their variability in space
41 (see Poisot *et al.* 2016 for probabilistic interactions; also Strydom *et al.* 2022b regarding probabilistic
42 metawebs). We currently lack a downscaling framework that is both probabilistic and does not require in situ
43 data. Additionally, a probabilistic view can allow propagating uncertainty, which can play a key role in
44 evaluating the quality of the predictions. Moreover, assessing model uncertainty would enable us to assess to
45 which degree we should trust our predictions and to identify what to do to improve the current knowledge. For
46 instance, we could locate where our knowledge and models are the most uncertain and determine new sampling
47 sites or areas where repeated sampling is necessary.

49 Explicit spatial predictions such as downscaled metaweb predictions are essential as they will allow
50 comparisons with extant work for species communities. Previous downscaling attempts allowed studying
51 network structures in novel ways, for instance, assessing changes in food web structure across space (Braga *et al.* 2019), the scaling of network area relationships (Galiana *et al.* 2021), or how sampling effort affects
52 measured network structure (McLeod *et al.* 2021). Further comparisons are relevant as they may go in
53 unexpected directions and highlight new elements regarding network biogeography. For instance, Frelat *et al.*
54 (2022) found a strong spatial coupling between community composition and food web structure but a temporal
55 mismatch depending on the spatial scale. Poisot *et al.* (2017) found that interaction uniqueness captures more
56 composition variability than community uniqueness and that sites with exceptional compositions might not be
57 the same for networks and communities. Spatialized network data will allow these comparisons and allow
58 identifying important conservation targets for networks and whether they differ geographically from areas
59

60 currently prioritized for biodiversity conservation.

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

67 **Methods**

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

74 [Figure 1 about here.]

75 **Data**

76 **Metaweb**

77 The main source of interaction data was the metaweb for Canadian mammals from Strydom *et al.* (2022a),
78 which is a-spatial, i.e., it represents interactions between mammals that can occur anywhere in Canada. The
79 species list for the Canadian metaweb was extracted from the International Union for the Conservation of
80 Nature (IUCN) checklist (Strydom *et al.* 2022a). Briefly, the metaweb was developed using graph embedding
81 and phylogenetic transfer learning based on the metaweb of European mammals, which is itself based on a
82 comprehensive survey of interactions reported in the scientific literature (Maiorano *et al.* 2020). The Canadian
83 metaweb is probabilistic, which has the advantage of reflecting the likelihood of an interaction taking place
84 given the phylogenetic and trait match between two species. This allows incorporating interaction variability

85 between species (i.e., taking into account that two species may not always interact whenever or wherever they
86 occur); however, we highlight that other factors beyond trait and phylogenetic matching (e.g., population
87 densities) will also contribute to observed interaction frequencies.

88 **Species occurrences**

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

101 **Environmental data**

102 We used species distribution models (SDMs, Guisan & Thuiller 2005) to project Canadian mammal habitat
103 suitability across the country, which we treated as information on potential distribution. For each species, we
104 related occurrences and pseudo-absences with 19 bioclimatic variables from CHELSA (Karger *et al.* 2017) and
105 12 consensus land-cover variables from EarthEnv (Tuanmu & Jetz 2014). The CHELSA bioclimatic variables
106 (*bio1-bio19*) represent various measures of temperature and precipitation (e.g., annual averages, monthly
107 maximum or minimum, seasonality) and are available for land areas across the globe. We used the most recent
108 version, the CHELSA v2.1 dataset (Karger *et al.* 2021), and subsetted it to land surfaces only using the
109 CHELSA v1.2 (Karger *et al.* 2018), which does not cover open water. The EarthEnv land-cover variables
110 represent classes such as Evergreen broadleaf trees, Cultivated and managed vegetation, Urban/Built-up, and
111 Open Water. Values range between 0 and 100 and represent the consensus prevalence of each class in

percentage within a pixel (hereafter called sites). We coarsened both the CHELSA and EarthEnv data from their original 30 arc-second resolution to a 2.5 arc-minute one (around 4.5 km at the Equator) using GDAL (GDAL/OGR contributors 2021). This resolution compromised capturing both local variations and broad scale patterns, while limiting computation costs to a manageable level as memory requirements increase rapidly with spatial resolution.

Analyses

Species distribution models

Our selection criteria for choosing an SDM algorithm was to have a method that generated probabilistic results (similar to Gravel *et al.* 2019), including both a probability of occurrence for a species in a specific site and the uncertainty associated with the prediction. These were crucial to obtaining a probabilistic version of the metaweb as they were used to create spatial variations in the localized interaction probabilities (see next section). One suitable method for this is Gradient Boosted Trees with a Gaussian maximum likelihood from the `EvoTrees.jl` Julia package (<https://github.com/Evovest/EvoTrees.jl>). This method returns a prediction for every site with an average value and a standard deviation, which we used as a measure of uncertainty to build a Normal distribution for the probability of occurrence of a given species at all sites (represented as probability distributions on Fig. 1). We trained models across the extent chosen for occurrences (longitudes 175°W to 45°W and latitudes 10°N to 90°N), then predicted species distributions only for Canada. We used the 2021 Census Boundary Files from Statistics Canada (Statistics Canada 2022) to set the boundaries for our predictions, which gave us 970,698 sites in total.

Building site-level instances of the metaweb

The next part of the method was the localized steps which produce local metawebs for every site. This component was divided into single-species, two-species, and network-level steps (*Localized steps* box on Fig. 1). The single-species steps represented four possible ways to account for uncertainty in the species distributions and bring variation to the spatial metaweb. We explored four different options to select a value ($P(\text{occurrence})$; Fig. 1) from the occurrence distributions obtained in the previous steps: 1) taking the mean from the distribution as the probability of occurrence (option 1 on Fig. 1); 2) converting the mean value to a binary one using a specific threshold per species (option 2); 3) sampling a random value within the Normal distribution

(option 3); or 4) converting a random value into a binary result (option 4, using a separate draw from option 3 and the same threshold as in option 2). The threshold (τ on Fig. 1) used was the value that maximized Youden's J informedness statistic (Youden 1950), the same metric used by Strydom *et al.* (2022a) at an intermediate step while building the metaweb. The four sampling options were intended to explore how uncertainty and variation in the species distributions can affect the metaweb result. We expected thresholding to have a more pronounced effect on network structure as it should reduce the number of links by removing many of the rare interactions (Poisot *et al.* 2016). Meanwhile, we expected random sampling to create spatial heterogeneity compared to the mean probabilities, as including some extreme values should confound the potential effects of environmental gradients. We chose option 1 as the default to present results as it is intuitive and essentially represents the result of a probabilistic SDM (as in Gravel *et al.* 2019).

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

The network-level steps then created the probabilistic metaweb for the site. We assembled all the local interaction probabilities (from the two-species steps) into a probabilistic network (Poisot *et al.* 2016). We then sampled several random network realizations to represent the potential local realization process (Poisot *et al.* 2015). This resulted in a distribution of localized networks, which we averaged over the number of simulations to obtain a single probabilistic network for the site.

Downscaled metaweb

The final output of our method was the downscaled metaweb, which contains a localized probabilistic metaweb in every site across the study area (Outputs box on Fig. 1). A metaweb essentially serves to set an upper bound on the potential interactions (Strydom *et al.* 2022b); therefore, the downscaled metaweb is a refined upper boundary at the local scale taking into account co-occurrences. It is still potential in nature and differs from a local realization, from which it should have a different structure. Nonetheless, from the downscaled metaweb we can create maps of network properties (e.g. number of links, connectance) measured on the local probabilities, display their spatial distribution, and compute some traditional community-level measures such as species

richness. We chose to compute and display the expected number of links (measured on probabilistic networks following Poisot *et al.* 2016; also see Gravel *et al.* 2019 for a similar example) as its relationship with species richness has been highlighted in a spatial context in recent studies (Galiana *et al.* 2021; **Galiana2022EcoNet?**). We also computed the uncertainty associated with the community and network measurements (richness variance and link variance, respectively) and compared their spatial distribution (see Supplementary Material).

Analyses of results by ecoregions

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

195 assembled a site-by-interaction matrix with the localized interaction values from the spatial probabilistic
 196 metaweb. We applied the Hellinger transformation on both matrices and computed the LCBD values from the
 197 total variance in the matrices (Legendre & De Cáceres 2013). High LCBD values indicate a high contribution
 198 to the overall variance and a unique species or interaction composition compared to other sites. Since values
 199 themselves are very low given our high number of sites (as in Dansereau *et al.* 2022), what matters primarily is
 200 the magnitude of the difference between the sites. Given this, we divided values by the maximum value in each
 201 matrix (species or network) and suggest that these should be viewed as relative contributions compared to the
 202 highest observed contribution. As with other measures, we then summarized the local uniqueness values by
 203 ecoregion by taking the median LCBD value and measuring the 89% interquantile range within all ecoregions.
 204 We used *Julia* v1.9.0 (Bezanson *et al.* 2017) to implement all our analyses. We used packages `GBIF.jl`
 205 (Dansereau & Poisot 2021) to reconcile species names using the GBIF Backbone Taxonomy,
 206 `SpeciesDistributionToolkit.jl` (<https://github.com/PoisotLab/SpeciesDistributionToolkit.jl>) to handle
 207 raster layers, species occurrences and generate pseudoabsences, `EvoTrees.jl`
 208 (<https://github.com/EvoVest/EvoTrees.jl>) to perform the Gradient Boosted Trees, `EcologicalNetworks.jl`
 209 (Poisot *et al.* 2019) to analyse network and metaweb structure, and `Makie.jl` (Danisch & Krumbiegel 2021) to
 210 produce figures. Our data sources (CHELSA, EarthEnv, Ecoregions) were all unprojected and we did not use a
 211 projection in our analyses, but we displayed the results using a Lambert conformal conic projection more
 212 appropriate for Canada using `GeoMakie.jl` (<https://github.com/MakieOrg/GeoMakie.jl>). All the code used to
 213 implement our analyses is available on GitHub (<https://github.com/PoisotLab/SpatialProbabilisticMetaweb>) and
 214 includes instructions on how to run a smaller example at a coarser resolution. Note that running our analyses at
 215 full scale is resource and memory intensive and required the use of compute clusters provided by Calcul Québec
 216 and the Digital Research Alliance of Canada. Final scripts ran for 0.0109 CPU core-years and peaked at 460 GB
 217 of RAM.

218 Results

219 Our method allowed us to display the spatial distribution of ecoregion-level community measures (here
 220 expected species richness) and network measures (expected number of links; Fig. 2). We highlight that the
 221 community and network-level measures presented here are not actual predictions of the measure itself (e.g., we
 222 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 species richness. On the other hand, within-ecoregion variability showed different bivariate relationships and a less constant latitudinal gradient (Fig. 3B). This indicates that richness and links do not co-vary completely (i.e. their variability is not closely connected) although they may show similar distributions for median values.

[Figure 3 about here.]

Our results also indicate a mismatch between the uniqueness of communities and networks (Fig. 4). Uniqueness was higher mostly in the north and along the south border for communities, but only in the north for networks (Fig. 4A-B). Consequently, ecoregions with both unique community composition and unique network composition were mostly in the north (Fig. 4C). Meanwhile, some areas were unique for one element but not the other. For instance, the New England-Acadian forests ecoregion (south-east, near 70°W and 48°N) had a highly unique species composition but a more common network composition (Fig. 4C). Opposite areas with unique network compositions only were observed at higher between latitudes 52°N and 70°N (Eastern Canadian Shield Taiga, Northern Canadian Shield Taiga, Canadian Low Arctic Tundra). Also, network uniqueness values for

ecoregions spanned a narrower range between the 44 ecoregions than species LCBD values (Fig. 4D, left). Within-ecoregion variation was also lower for network values with generally lower 89% interquantile ranges among the site-level LCBD values (Fig. 4D, right). Moreover, mismatched sites (unique for only one element) formed two distinct groups when evaluating the relationship between species richness and the number of links (see Supplementary Material). The areas only unique for their species composition had both a high richness and number of links. On the other hand, the sites only unique for their networks had both lower richness and a lower number of links, although they were not the sites with the lowest values for both.

[Figure 4 about here.]

Discussion

Our approach presents a way to downscale a metaweb, produce localized predictions using probabilistic networks as inputs and outputs, and incorporate uncertainty, as called for by Strydom *et al.* (2022b). It gives us an idea of what local metawebs or networks could look like in space, given the species distributions and their variability, as well as the uncertainty around the interactions. We also provide the first spatial representation of the metaweb of Canadian mammals (Strydom *et al.* 2022a) and a probabilistic equivalent to how the European tetrapod metaweb (Maiorano *et al.* 2020) was used to predict localized networks in Europe (Braga *et al.* 2019; O'Connor *et al.* 2020; Galiana *et al.* 2021; Gaüzère *et al.* 2022; Botella *et al.* 2023). Therefore, our approach could open similar possibilities of investigations in North America with food webs of Canadian mammals, for instance on the structure of food webs over space (Braga *et al.* 2019) and on the effect of land-use intensification on food webs (Botella *et al.* 2023). Interesting research areas could include assessing climate change impacts on network structure or investigating linkages between network structure and stability. Moreover, since our approach is probabilistic, it does not assume species interact whenever they co-occur, and incorporates variability based on environmental conditions, which could lead to different results by introducing a different association between species richness and network properties. Galiana *et al.* (2021) found that species richness had a large explanatory power over network properties but mentioned it could potentially be due to interactions between species being fixed in space. Here, we found mismatches in the distribution of species richness and interactions, and especially regarding their within-ecoregion variability (Fig. 3), highlighting that interactions might vary differently than species distributions in space. Network measures (links on Fig. 3A) were also lower in the north, contrarily to previous studies where connectance was higher in the north, although

279 those were in Europe for all tetrapods (Braga *et al.* 2019; Galiana *et al.* 2021) and willow-galler-parasitoid
 280 networks (Gravel *et al.* 2019). Further research should investigate why these results might be different between
 281 the two continents and whether it is due to the methodology, data, or biogeographical processes.

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

295 When to use the method we presented here will depend on the availability of interaction data or existing
 296 metawebs and on the intent to incorporate interaction variability, as well as ecoregion-level variability. In
 297 systems where in situ interaction and network data are available, the approach put forward by Gravel *et al.*
 298 (2019) achieves a similar purpose as we attempted here, but is more rigorous and allows modelling the effect
 299 of the environment on the interactions. Without such data, establishing or predicting the metaweb should be the
 300 first step towards producing localized predictions (Strydom *et al.* 2022b). Well-documented binary metawebs
 301 such as the European tetrapod metaweb could be partly be combined with our approach if used with
 302 probabilistic SDMs and summarized by ecoregions (as they would only lack an initial probabilistic metaweb,
 303 but would still obtain a more probabilistic output). Our approach will essentially differ from previous attempts
 304 in how it perceives uncertainty and variability. For instance, rare interactions should not be over-represented
 305 (Poisot *et al.* 2016) and should have lesser effects over computed network measures. Summarizing results by
 306 ecoregion allows showing variation within and between ecologically-meaningful biogeographic boundaries
 307 (Martins *et al.* 2022), which as our results showed is not constant across space and can help identify contrasting
 308 diversity hotspots.

References

- 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.
- 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.
- Dinerstein, E., Olson, D., Joshi, A., Vynne, C., Burgess, N.D., Wikramanayake, E., *et al.* (2017). [An Ecoregion-Based Approach to Protecting Half the Terrestrial Realm.](#) *BioScience*, 67, 534–545.
- Dunne, J. (2006). The network structure of food webs. In: *Ecological Networks: Linking Structure to Dynamics in Food Webs*. pp. 27–86.
- Frelat, R., Kortsch, S., Kröncke, I., Neumann, H., Nordström, M.C., Olivier, P.E.N., *et al.* (2022). [Food web structure and community composition: A comparison across space and time in the North Sea.](#) *Ecography*, 2022.

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

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

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

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

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

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

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

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

351 Hortal, J., de Bello, F., Diniz-Filho, J.A.F., Lewinsohn, T.M., Lobo, J.M. & Ladle, R.J. (2015). [Seven Shortfalls](#)
352 [that Beset Large-Scale Knowledge of Biodiversity](#). *Annual Review of Ecology, Evolution, and Systematics*,
353 46, 523–549.

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

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

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

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

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

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.](#)

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.

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.

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

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

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

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

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

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

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

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

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

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

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

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

416 Wisz, M.S., Pottier, J., Kissling, W.D., Pellissier, L., Lenoir, J., Damgaard, C.F., *et al.* (2013). [The role of biotic](#)
417 [interactions in shaping distributions and realised assemblages of species: Implications for species](#)

418 [distribution modelling](#). *Biological Reviews*, 88, 15–30.

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

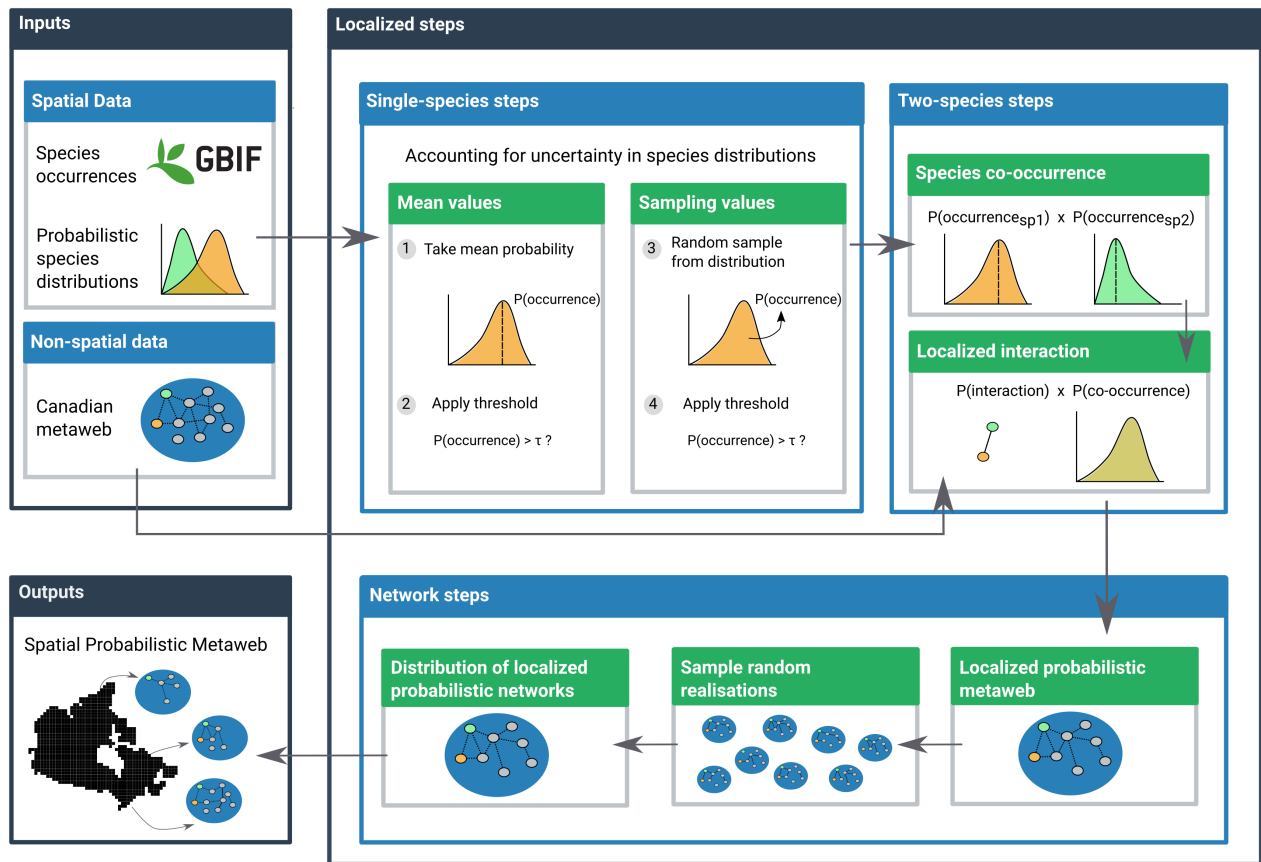


Figure 1: Conceptual figure of the workflow to obtain the spatial probabilistic metaweb (Chapter 1). The workflow has three components: the inputs, the localized steps, and the final spatial output. The inputs are composed of the spatial data (data with information in every cell) and the non-spatial 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 coming out of the network-level steps contains a spatialized probabilistic metaweb for every cell across the study extent.

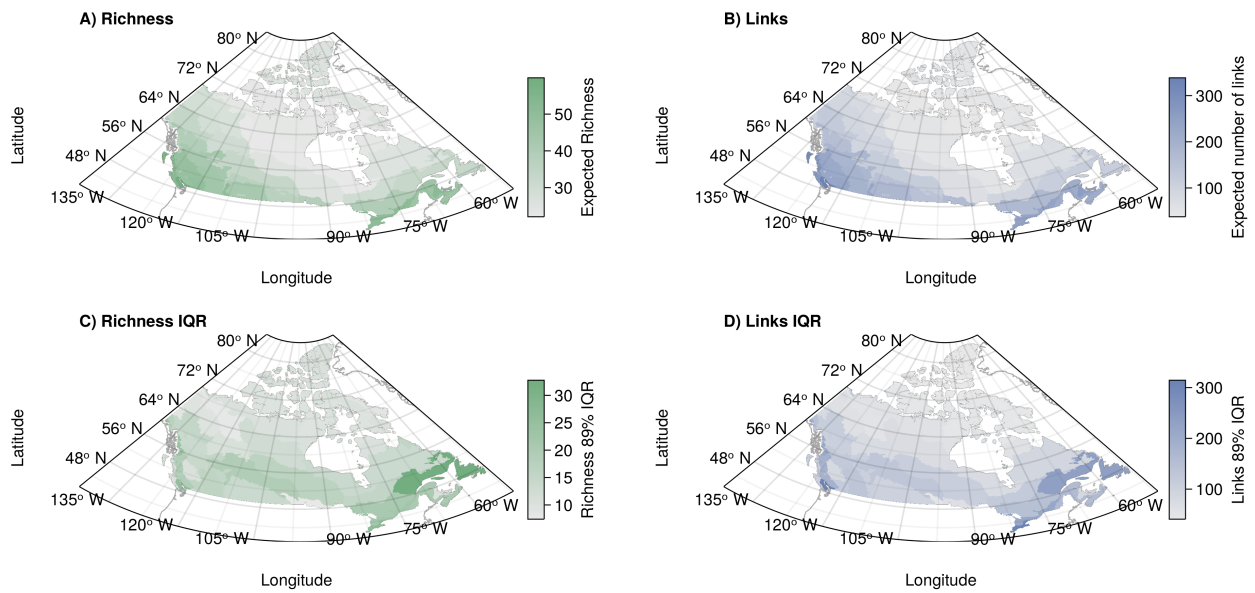


Figure 2: (A-B) Example of a community measure (A, expected species richness) and a network one (B, expected number of links). Both measures are assembled from the predicted probabilistic communities and networks, respectively. Values are first measured separately for all sites, then the median value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

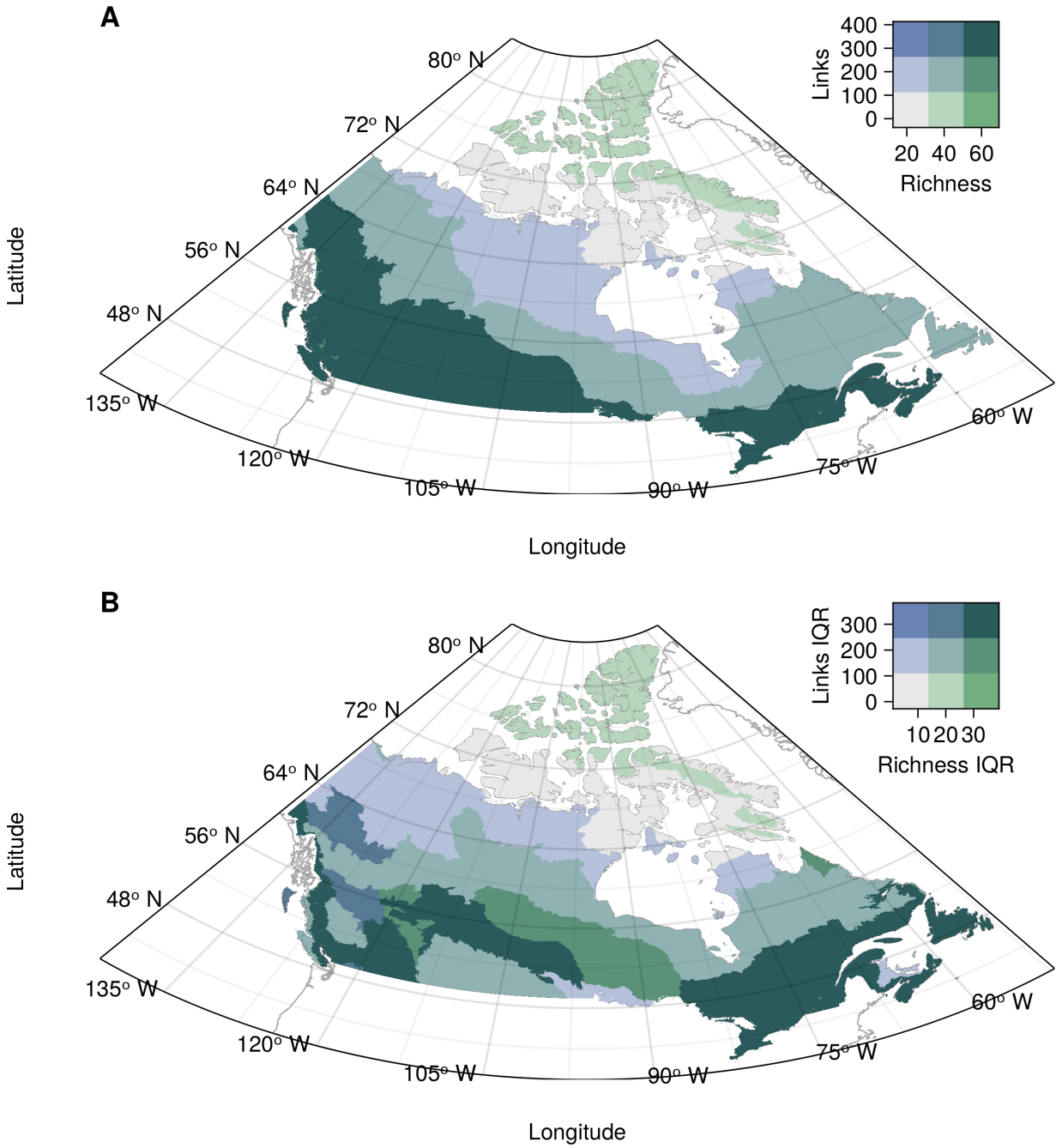


Figure 3: Bivariate relationship between community and network measures for the median ecoregion value (A) and the within-ecoregion 89% interquantile range (B). Values are grouped into three quantiles separately for each variable. The colour combinations represent the nine possible combinations of quantiles.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

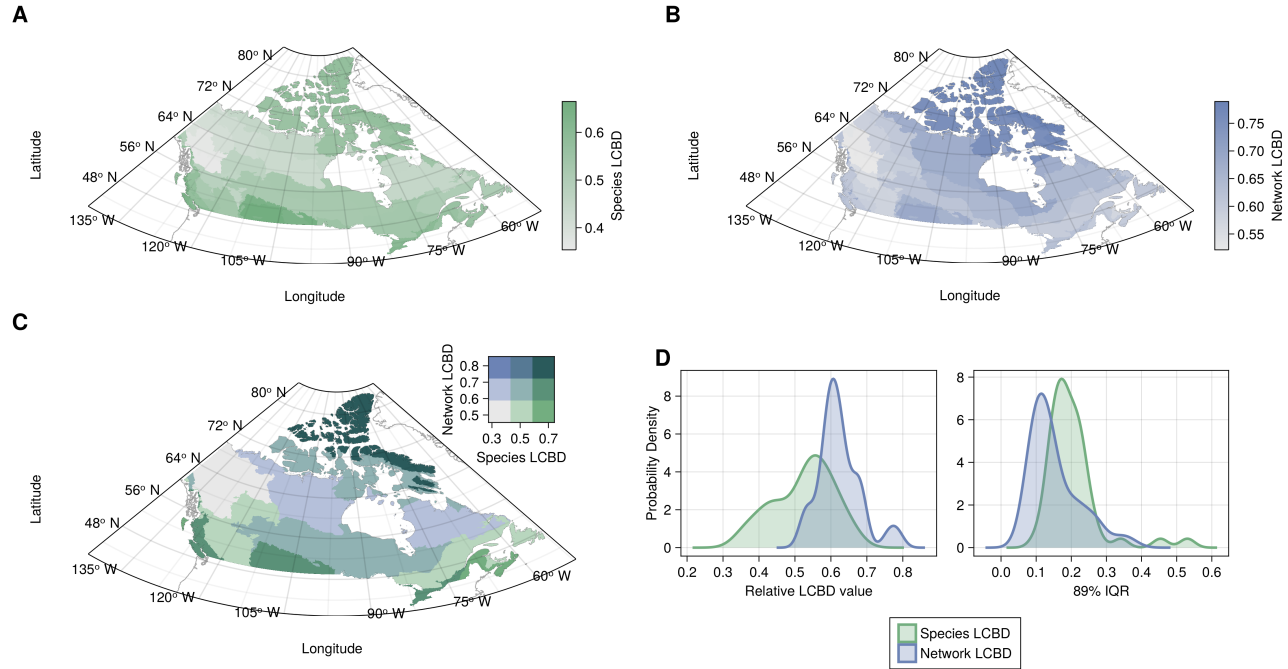


Figure 4: (A-B) Representation of the ecoregion uniqueness values based on species composition (a) and network composition (b). LCBD values were first computed across all sites and scaled relative to the maximum value observed. The ecoregion LCBD value is the median value for the sites in the ecoregion.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