

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

‡ Equal contributions

Correspondance to:

Gabriel Dansereau — gabriel.dansereau@umontreal.ca

1 Introduction

2 Sampling species interactions and ecological networks in repeated locations in space and time is a challenging
3 task (Jordano 2016). Most studies on food webs have previously focused on local webs limited in size and
4 extent and are rarely replicated in space and time (Mestre *et al.* 2022). Available network data also show
5 important geographical bias, limiting our ability to answer questions in many biomes and over broad spatial
6 extents (Poisot *et al.* 2021). Yet, we know that interactions can show important variations in space (Poisot *et al.*
7 2015) and that they are critical to evaluating how ecosystems will react to global change. Even predictive
8 approaches (Morales-Castilla *et al.* 2015; Desjardins-Proulx *et al.* 2017), seen as a way to circumvent data
9 scarcity (Strydom *et al.* 2021), often do not have an explicit spatial component. As a result, there have been
10 repeated calls for globally distributed interaction and network data and repeated samplings in time and space
11 (Poisot *et al.* 2021; Mestre *et al.* 2022).

12 A promising concept gaining momentum for the analysis of networks in space is that of the metaweb. A
13 metaweb contains all the possible interactions between the species found in a given regional species pool
14 (Dunne 2006). Studies are now directly focused on assembling the metaweb for various taxa through extensive
15 literature surveys or predictions (Maiorano *et al.* 2020; Strydom *et al.* 2022a). These new data sources have
16 allowed studying network structures in novel ways, for instance, through the structure of food webs over space
17 (Braga *et al.* 2019), the scaling of network area relationships (Galiana *et al.* 2021), or the effect of sampling
18 effort to capture network structure (McLeod *et al.* 2021). A key element that emerged from these studies is that
19 the structure of empirical food webs is inherited from the metaweb with little influence from habitat and
20 dynamical constraints (Saravia *et al.* 2022). This makes the metaweb the core goal of predictive network
21 ecology (Strydom *et al.* 2022b), i.e. the first approximation of networks in space we should aim for. Yet, rather
22 than directly predicting the metaweb and localized networks, some studies have instead focused on using
23 interactions to improve predictions of species distributions (Moens *et al.* 2022; for example, Poggiato *et al.*
24 2022; Lucas *et al.* 2023), answering long-standing calls to include interactions within such models (Wisz *et al.*
25 2013). However, refining predictions of species distributions through interactions differs from predicting
26 networks in space; it serves a different goal and does not fully consider networks as spatial elements.

27 Two network aspects are essential to develop a proper spatial perspective. First, networks are local realizations
28 of a regional metaweb (Poisot *et al.* 2012, 2015), which reinforces why the metaweb should be predicted first.
29 Second, to represent them correctly, interactions must be seen as probabilistic events (Poisot *et al.* 2016) to

30 account for their variability in space. In contrast, many studies assume interactions are binary events, whether at
31 the interaction, network or metaweb level (Gaüzère *et al.* 2022; Mendoza & Araujo 2022). However, a
32 probabilistic view can additionally allow propagating uncertainty, which can play a key role in evaluating the
33 quality of the predictions. Moreover, assessing model uncertainty would enable us to assess to which degree we
34 should trust our predictions and to identify what to do to improve the current knowledge. For instance, we could
35 locate where our knowledge and models are the most uncertain and determine new sampling locations
36 accordingly.

37 Repeated data collections of networks and explicit spatial predictions are essential as they will allow
38 comparisons with extant work for species communities. These comparisons are relevant as they may go in
39 unexpected directions and highlight new elements regarding network biogeography. For instance, Frelat *et al.*
40 (2022) found a strong spatial coupling between community composition and food web structure but a temporal
41 mismatch depending on the scale. Poisot *et al.* (2017) found that interaction uniqueness captures more
42 variability than community uniqueness and that the sites with exceptional compositions might not be the same
43 for networks and communities. Therefore, spatialized network data will allow these comparisons and
44 potentially identify important conservation targets for networks, which might differ from the areas most relevant
45 for maintaining communities.

46 Here, we present a method to downscale a metaweb in space by developing an explicit spatial probabilistic
47 metaweb for Canadian mammals. We present how the spatial structure of the downscaled metaweb varies in
48 space and how the uncertainty of interactions can be made spatially explicit. We further show that the
49 downscaled metaweb can highlight important biodiversity areas and bring novel ecological insights compared
50 to community measures.

51 **Methods**

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

[Figure 1 about here.]

Inputs

The inputs were divided into two main categories: the spatial and non-spatial ones (*Inputs* box on Fig. 1).

Non-spatial inputs

The main building block for the interaction data was the metaweb for Canadian mammals from Strydom *et al.* (2022a), a non-spatial input (represented as nodes and links on Fig. 1). A metaweb contains all the possible interactions between the species found in a given regional species pool (Dunne 2006). The species list for the Canadian metaweb was extracted from the International Union for the Conservation of Nature (IUCN) checklist (Strydom *et al.* 2022a). Briefly, the metaweb was developed using graph embedding and phylogenetic transfer learning based on the metaweb of European mammals, which is itself based on a comprehensive survey of interactions reported in the scientific literature (Maiorano *et al.* 2020). The Canadian metaweb is probabilistic, which has the advantage of taking into account that species do not necessarily interact whenever they co-occur (Blanchet *et al.* 2020). However, the Canadian metaweb is not explicitly spatial: it only gives information on interactions in Canada as a whole and does not represent networks at specific locations. Local networks, on the other hand, are realizations from the metaweb resulting from sorting the species and the interactions (Poisot *et al.* 2015). A spatial and localized metaweb is not equivalent to the local networks, as it will have a different structure and a higher connectance (Strydom *et al.* 2022b). Therefore, producing a spatial metaweb requires additional steps to account for species composition and interaction variability in space.

Spatial inputs

The spatial data used to develop the spatial component of the metaweb were species occurrences and environmental data. First, we extracted species occurrences from the Global Biodiversity Information Facility (GBIF; www.gbif.org) for the Canadian mammals after reconciling species names between the Canadian metaweb and GBIF using the GBIF Backbone Taxonomy (GBIF Secretariat 2021). Doing so, we removed potential duplicates where species listed in the Canadian metaweb are considered as a single entity by GBIF. We collected occurrences for our species list (159 species) using the GBIF download API on October 21st 2022 (GBIF.org 2022). We restricted our query to occurrences with coordinates between longitudes 175°W to 45°W

84 and latitudes 10°N to 90°N. This was meant to collect training data covering a broader range than our prediction
 85 target (Canada only) and include observations in similar environments. Then, since GBIF observations
 86 represent presence-only data and most predictive models require absence data, we generated pseudo-absence
 87 data using the surface range envelope method available in `SimpleSDMLayers.jl` (Dansereau & Poisot 2021).
 88 This method generates pseudo-absences by selecting random non-observed locations within the spatial range
 89 delimited by the presence data (Barbet-Massin *et al.* 2012).

90 We used environmental data and species distribution models (SDMs, Guisan & Thuiller 2005) to predict the
 91 distribution of Canadian mammals across the whole country. The environmental data we used were the 19
 92 bioclimatic variables from CHELSA (Karger *et al.* 2017) and the 12 consensus land cover variables from
 93 EarthEnv (Tuanmu & Jetz 2014). The CHELSA bioclimatic variables (*bio1-bio19*) represent various measures
 94 of temperature and precipitation (e.g., annual averages, monthly maximum or minimum, seasonality) and are
 95 available for land areas across the globe. Therefore, they can be used to capture the climatic tolerance of species
 96 and model habitat suitability in new locations. We used the most recent version, the CHELSA v2.1 dataset
 97 (Karger *et al.* 2021). However, this version also includes bioclimatic data for open water, while we decided here
 98 to focus only on land surfaces. We used the previous version, CHELSA v1.2 (Karger *et al.* 2018), which shares
 99 a similar grid but does not cover open water, as a mask to clip the v2.1 data to land surfaces only. The EarthEnv
 100 land cover variables represent classes such as Evergreen broadleaf trees, Cultivated and managed vegetation,
 101 Urban/Built-up, and Open Water. Values range between 0 and 100 and represent the consensus prevalence of
 102 each class in percentage within a pixel. We coarsened both the CHELSA and EarthEnv data from their original
 103 30 arc-second resolution to a 2.5 arc-minute one (around 4.5 km at the Equator) using GDAL/OGR
 104 contributors (2021). This represented a compromise to catch both local variations and broad scale patterns
 105 while limiting computation costs to a manageable level, as memory requirements on localized interactions rise
 106 very quickly.

107 Our selection criteria for choosing an SDM algorithm was to have a method that generated probabilistic results,
 108 including both a probability of occurrence for a species in a specific location and the uncertainty associated
 109 with the prediction. These were crucial to obtaining a probabilistic version of the metaweb as they were used to
 110 create spatial variations in the localized interaction probabilities (see next section). One promising method for
 111 this is Gradient Boosted Trees with a Gaussian maximum likelihood from the `EvoTrees.jl` *Julia* package
 112 (<https://github.com/Evovest/EvoTrees.jl>). This method returns a prediction for every pixel with an average value
 113 and a standard deviation, which we used as a measure of uncertainty to build a Normal distribution for the

114 probability of occurrence of a given species at all pixels (represented as probability distributions on Fig. 1). We
115 trained models across the extent chosen for occurrences (longitudes 175°W to 45°W and latitudes 10°N to
116 90°N), then predicted species distributions only for Canada. We used the 2021 Census Boundary Files from
117 Statistics Canada (Statistics Canada 2022) to set the boundaries for our predictions, which gave us 970,698 sites
118 in total.

119 **Localized steps**

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

122 The single-species steps represented four possible ways to account for uncertainty in the species distributions
123 and bring variation to the spatial metaweb. We explored four different options to select a value from the
124 occurrence distributions obtained in the previous steps (Inputs section): 1) taking the mean from the distribution
125 as the probability of occurrence (option 1 on Fig. 1); 2) converting the mean value to a binary one using a
126 specific threshold per species (option 2); 3) sampling a random value within the Normal distribution (option 3);
127 4) converting the random value into a binary result (option 4). The threshold (τ on Fig. 1) used was the value
128 that maximized Youden's J informedness statistic (Youden 1950), the same metric used by Strydom *et al.*
129 (2022a) at an intermediate step while building the metaweb. The four sampling options were intended to
130 explore how uncertainty and variation in the species distributions can affect the metaweb result and reproduce
131 some of the filterings that create the local network realizations (Poisot *et al.* 2015). We expected thresholding to
132 have a more pronounced effect on network structure as it should reduce the number of links by removing many
133 of the rare interactions (Poisot *et al.* 2016). Meanwhile, we expected random sampling to create spatial
134 heterogeneity compared to the mean probabilities, as including some extreme values should disrupt the
135 potential effects of environmental gradients.

136 Next, the two-species steps aimed to give the probability of observing a given interaction in a location. For all
137 species pairs, we multiplied the two species' occurrence probability obtained using the sampling options
138 described in the previous paragraph, then multiplied the co-occurrence probability by the interaction probability
139 from the Canadian metaweb. For cases where species in the Canadian metaweb were considered as the same
140 species by the GBIF Backbone Taxonomy (the reconciliation step mentioned earlier), we used the highest
141 interaction probabilities involving the duplicated species.

142 The network-level steps then created the probabilistic metaweb for the location. We assembled all the local
143 interaction probabilities (from the two-species steps) into a probabilistic network (Poisot *et al.* 2016). We then
144 sampled several random network realizations to represent the potential local realization process (Poisot *et al.*
145 2015). Finally, this resulted in a distribution of localized networks, which we averaged over the number of
146 simulations to obtain a probabilistic network.

147 **Outputs**

148 The final output of our method was the spatial probabilistic metaweb, which contains a localized probabilistic
149 metaweb in every cell across the student extent (Outputs box on Fig. 1). This gives us an idea of the possible
150 networks in all locations as the metaweb essentially serves to set an upper bound on the potential interactions
151 (Strydom *et al.* 2022b), but with the added benefit of accounting for co-occurrence probabilities in this case.
152 From there, we can create maps of network properties (e.g. number of links, connectance) measured on the
153 local realizations, display their spatial distribution, and compute some community-level measures such as
154 species richness. We can also calculate the uncertainty associated with the network and community
155 measurements and contrast their spatial distribution (see Supplementary Material).

156 **Ecoregions**

157 Since both species composition and network summary values display a high spatial variation and complex
158 patterns, we simplified the representation of their distribution by grouping sites by ecoregion, as species and
159 interaction composition have been shown to differ between ecoregions across large spatial scales (Martins *et al.*
160 2022). To do so, we used the global map of ecoregions from (Dinerstein *et al.* 2017; also used by Martins *et al.*
161 2022), rasterized it, and clipped it to Canada, which selected 44 different ecoregions. For every measure we
162 report (e.g. species richness, number of links), we first calculated the measure for every site separately, then we
163 extracted the median value for each ecoregion. We also measured the within-ecoregion variation by measuring
164 the 89% interquantile range of the values in each ecoregion (threshold chosen to avoid confusion with
165 conventional significance tests, inspired by McElreath 2020).

166 Ecological uniqueness

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

188 Software

189 We used *Julia* v1.9.0 (Bezanson *et al.* 2017) to implement all our analyses. We used packages `GBIF.jl`
190 (Dansereau & Poisot 2021) to reconcile species names using the GBIF Backbone Taxonomy,
191 `SpeciesDistributionToolkit.jl` to handle raster layers and species occurrences, `EcologicalNetworks.jl`
192 (Poisot *et al.* 2019) to analyse network and metaweb structure, and `Makie.jl` (Danisch & Krumbiegel 2021) to
193 produce figures. Our data sources (CHELSA, EarthEnv, Ecoregions) were all unprojected and we did not use a

194 projection in our analyses, but we displayed the results using a Lambert conformal conic projection more
195 appropriate for Canada using `GeoMakie.jl`. All the code used to implement our analyses is available on GitHub
196 (<https://github.com/PoisotLab/SpatialProbabilisticMetaweb>) and includes instructions on how to run a smaller
197 example at a coarse resolution. Note that running our analyses at full scale is resource and memory intensive and
198 required the use of compute clusters provided by Calcul Québec and the Digital Research Alliance of Canada.

199 Results

200 Our method allowed us to display the spatial distribution of ecoregion-level measures (Fig. 2), either for
201 community measures (e.g. expected species richness) or for network measures (e.g. expected number of links).
202 Importantly, both community and network-level measures presented are not predictions of the measure itself but
203 were instead computed over localized predictions of the communities and networks, then summarized for the
204 ecoregions. Expected ecoregion richness (Fig. 2A), which is the median of the expected species richness of the
205 sites within the ecoregion, and expected number of links (Fig. 2B) displayed similar distributions with a
206 latitudinal gradient and higher values in the south. However, within-ecoregion variability was distributed
207 differently, as some ecoregions along the coasts displayed higher interquantile ranges while ecoregions around
208 the southern border displayed narrower ones (Fig. 2C-D). All results shown are based on the first sampling
209 strategy (option 1) mentioned in the Localized steps section, where species occurrence probabilities were taken
210 as the mean value of the distribution (results for other sampling strategies are discussed in Supplementary
211 Material).

212 [Figure 2 about here.]

213 Direct comparison of the spatial distributions of species richness and expected number of links showed some
214 areas with mismatches, both regarding the median estimates and regarding the within-ecoregion variability
215 (Fig. 3). Median values for the ecoregions showed a similar bivariate distribution with ecoregions in the south
216 mostly displaying high species richness and a high number of links (Fig. 3A). The northernmost ecoregions
217 (Canadian High Arctic Tundra and Davis Highlands Tundra) displayed higher richness (based on the quantile
218 rank) compared to the number of links. Inversely, ecoregions further south (Canadian Low Arctic Tundra,
219 Northern Canadian Shield Taiga, Southern Hudson Bay Taiga) ranked higher for the number of links than for
220 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 link do not vary hand in hand (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 higher north between latitudes 52°N and 70°N (Eastern Canadian Shield Taiga, Northern Canadian Shield Taiga, Canadian Low Arctic Tundra). When comparing the values, 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

The spatial probabilistic metaweb we produced here is a first attempt at downscaling a metaweb and producing localized predictions, 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. It is the first representation in space of the metaweb of Canadian mammals (Strydom *et al.* 2022a), which is not spatialized. Conceptually, this is similar 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.*

247 2020; Galiana *et al.* 2021; Gaüzère *et al.* 2022; Botella *et al.* 2023). Therefore, our approach could open similar
248 possibilities of investigations in North America with food webs of Canadian mammals, for instance on the
249 structure of food webs over space (Braga *et al.* 2019), on the scaling of network area relationships (Galiana *et*
250 *al.* 2021), and on the effect of land-use intensification on food webs (Botella *et al.* 2023). Moreover, our
251 approach is probabilistic, does not assume species interact whenever they co-occur, and incorporates variability
252 based on environmental conditions, which could lead to different results by introducing a different association
253 between species richness and network properties. Galiana *et al.* (2021) found that species richness had a large
254 explanatory power over network properties but mentioned it could potentially be due to interactions between
255 species being fixed in space. Here, we found mismatches in the distribution of species richness and interactions,
256 and especially regarding their within-ecoregion variability (Fig. 3), highlighting that interactions might vary
257 differently than species distributions in space. Network measures (links on Fig. 3A) were also lower in the north,
258 contrarily to previous studies (e.g. connectance higher in the north, Braga *et al.* 2019; Galiana *et al.* 2021).

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

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.
- Blanchet, F.G., Cazelles, K. & Gravel, D. (2020). [Co-occurrence is not evidence of ecological interactions.](#) *Ecology Letters*, 23, 1050–1063.
- 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). Preprints.
- 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.

298 Frelat, R., Kortsch, S., Kröncke, I., Neumann, H., Nordström, M.C., Olivier, P.E.N., *et al.* (2022). [Food web](#)
 299 [structure and community composition: A comparison across space and time in the North Sea.](#) *Ecography*,
 300 2022.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

327 Martins, L.P., Stouffer, D.B., Blendinger, P.G., Böhning-Gaese, K., Buitrón-Jurado, G., Correia, M., *et al.*
328 (2022). [Global and regional ecological boundaries explain abrupt spatial discontinuities in avian frugivory](#)
329 [interactions](#). *Nature Communications*, 13, 6943.

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

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

334 Mendoza, M. & Araujo, M.B. (2022). [Biogeography of bird and mammal trophic structures](#). *Ecography*, 2022,
335 e06289.

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

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

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

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

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

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

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

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

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

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

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

359 Saravia, L.A., Marina, T.I., Kristensen, N.P., De Troch, M. & Momo, F.R. (2022). [Ecological network](#)
360 [assembly: How the regional metaweb influences local food webs](#). *Journal of Animal Ecology*, n/a.

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

363 Strydom, T., Bouskila, S., Banville, F., Barros, C., Caron, D., Farrell, M.J., *et al.* (2022a). [Food web](#)
364 [reconstruction through phylogenetic transfer of low-rank network representation](#). *Methods in Ecology and*
365 *Evolution*, n/a.

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

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

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

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

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

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

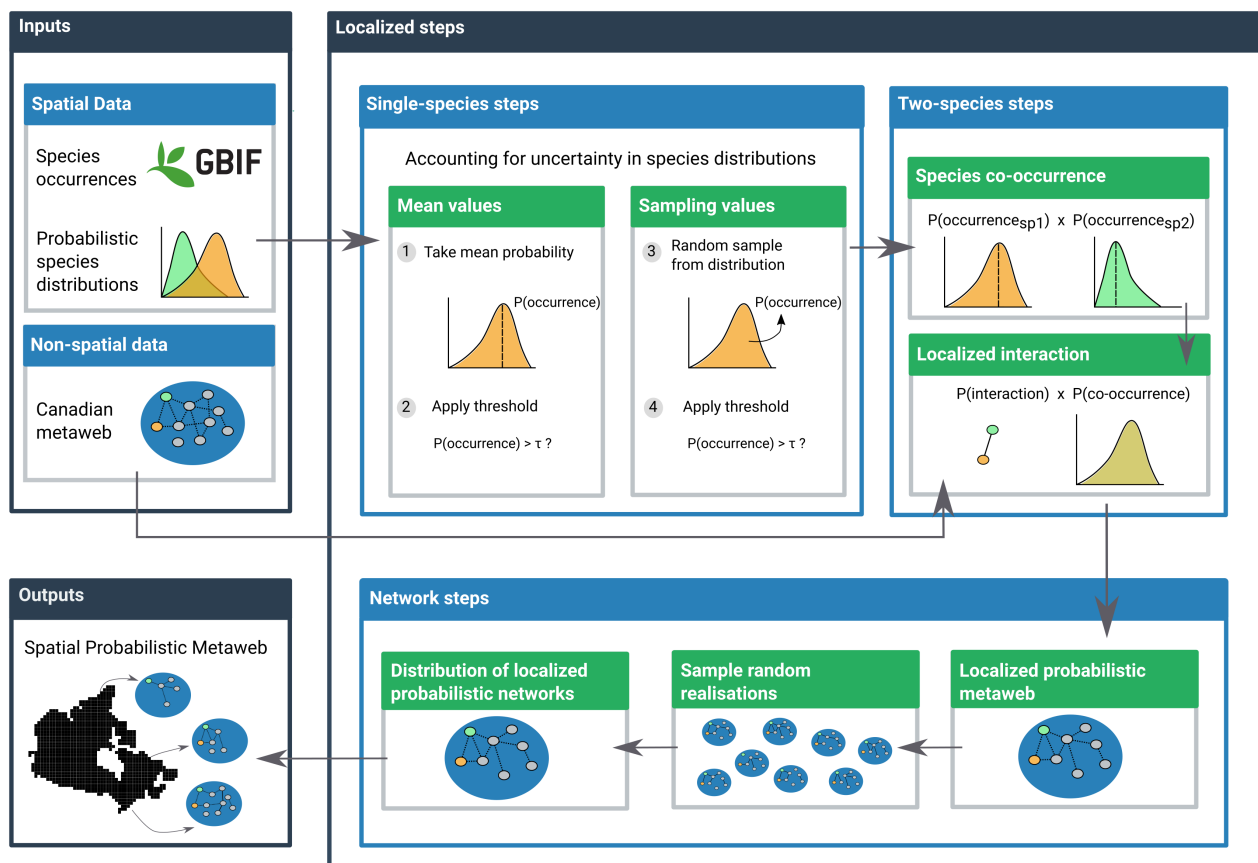


Figure 1: Conceptual figure of the workflow to obtain the spatial probabilistic metaweb (Chapter 1). The workflow has three components: the inputs, the localized steps, and the final spatial output. The inputs are composed of the spatial data (data with information in every cell) and the non-spatial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

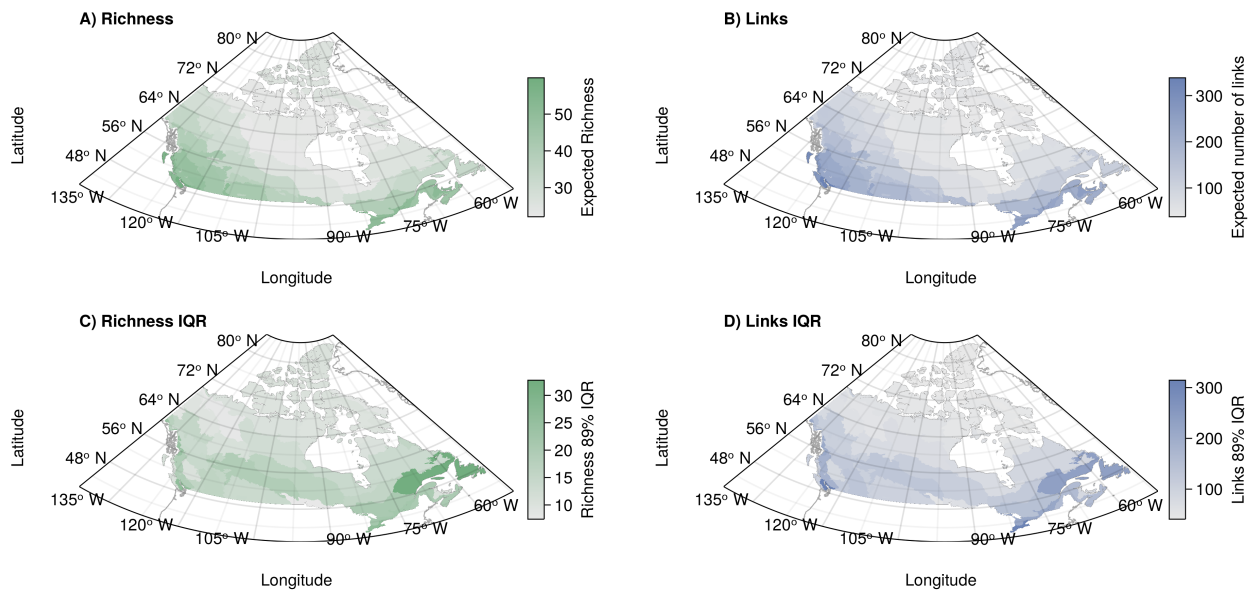


Figure 2: (A-B) Example of a community measure (A, expected species richness) and a network one (B, expected number of links). Both measures are assembled from the predicted probabilistic communities and networks, respectively. Values are first measured separately for all sites, then the median value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

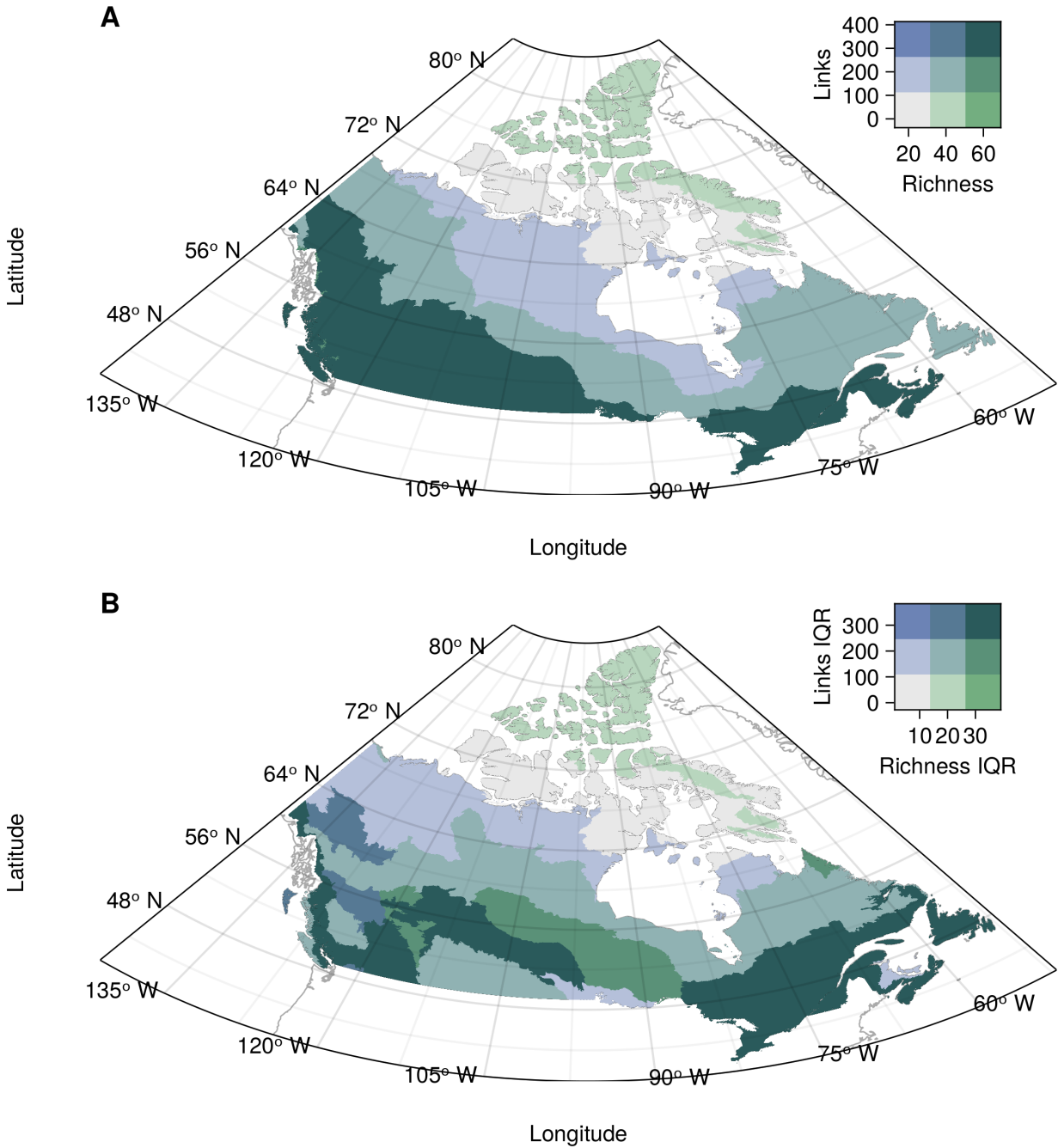


Figure 3: Bivariate relationship between community and network measures for the median ecoregion value (A) and the within-ecoregion 89% interquantile range (B). Values are grouped into three quantiles separately for each variable. The colour combinations represent the nine possible combinations of quantiles.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

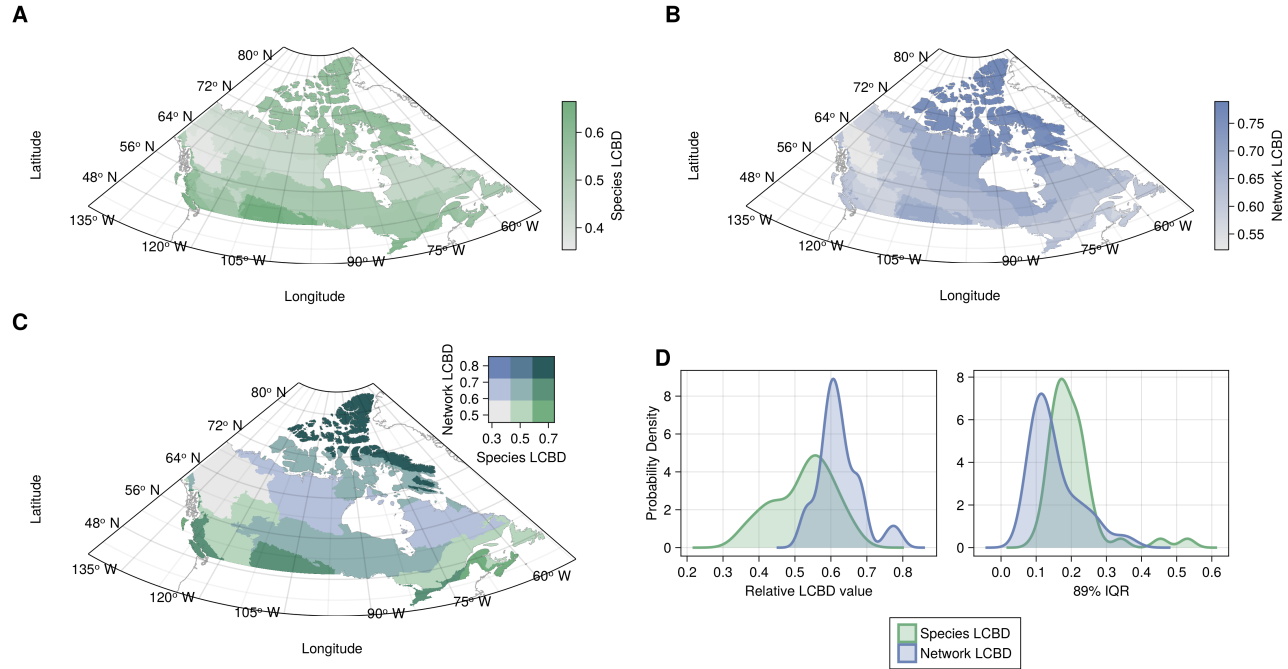


Figure 4: (A-B) Representation of the ecoregion uniqueness values based on species composition (a) and network composition (b). LCBD values were first computed across all sites and scaled relative to the maximum value observed. The ecoregion LCBD value is the median value for the sites in the ecoregion.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