Introduction

Because species interactions vary in time and space, and because species show high turnover over larger spatial extents, adequately capturing the diversity of ecological networks is a challenging task [@Jordano2016SamNet]. Most studies on food webs have previously focused on local networks limited in size and extent, and are rarely replicated in space or time [@Mestre2022DisFoo]. Interactions can show important variations in space [@Poisot2015SpeWhy; @Zarnetske2017IntLan], yet available network data also show important geographical bias by focusing sampling efforts in a few areas or biomes, limiting our ability to answer questions in many biomes and over broad spatial extents [@Poisot2021GloKno]. Moreover, global monitoring of biotic interactions is insufficient to properly describe and understand how ecosystems are reacting to global change [@Windsor2023UsiEco]. Approaches to predict species interactions [e.g., @Morales-Castilla2015InfBio; @Desjardins-Proulx2017EcoInt] are increasingly used as an alternative to determine potential interactions; they can handle limited data to circumvent data scarcity [@Strydom2021RoaPre], but are still rarely used to make explicitly spatial predictions. As a result, there have been repeated calls for globally distributed interaction and network data coupled to repeated sampling in time and space [@Mestre2022DisFoo; @Windsor2023UsiEco], which will help understand the macroecological variations of food webs [@Baiser2019EcoRul].

Despite these limitations, food web ecologists often can infer a reasonable approximation of the network existing within a region. This representation, called a metaweb, contains all possible interactions between species in a given regional species pool [@Dunne2006NetStr], and provides a solid foundation to develop approaches to estimate the structure of networks at finer spatial scales.

When assembled by integrating different data sources (and potentially with additional predictive steps), the metaweb allows to overcome sampling limitations and to raise network data to a global scale. For example, @Albouy2019MarFis coupled data on fish distributions with a statistical model of trophic interactions to provide estimates of the potential food web structure at the global scale. Recent studies have focused on assembling metawebs for various taxa through literature surveys and expert elicitation [European terrestrial tetrapods, @Maiorano2020TetSpe] or using predictive tools [marine fishes, @Albouy2019MarFis; Canadian mammals, @Strydom2022FooWeb]. At a finer spatial scale, the local food webs [i.e. the local "realization" of the metaweb when combined with species distributions, @Poisot2012DisSpe] reflect local environmental conditions but still retain the signal of the metaweb to which they belong [@Saravia2022EcoNet]. Given this, @~~Strydom2022PreMet~~ Strydom2023GraEmb defended that predicting the metaweb's structure should be the core goal of predictive network ecology~~, as if~~. Assuming there is a strong link between the metaweb and its local realizations, more ~~accurante~~ accurate predictions of the metaweb will have the potential to bring us closer to producing accurate local (downscaled) predictions [@Strydom2023GraEmb]. Therefore, establishing or predicting the metaweb should be the first target in systems lacking information about local realizations. This is not the same as using interactions to improve predictions of species distributions, as recent studies have done [@Moens2022ImpBio; @Poggiato2022IntFoo; @Lucas2023IncBio], although these are incredibly relevant and answer long-standing calls to include interactions within such models [@Wisz2013RolBio]. Instead, predicting networks in space is a different task, and it serves a different goal: focusing first on the distribution of network structures and its drivers rather than on the distribution of species.

Explicit spatial predictions (such as downscaled metaweb predictions) are essential as they will allow comparisons with extant work for species-rich communities. Recent approaches to metaweb downscaling combined a regional metaweb with species distribution maps to generate local assemblages for European tetrapods [@Braga2019SpaAna; @OConnor2020UnvFoo; @Galiana2021SpaSca; @Gauzere2022DivBio], Barents Sea marine taxa [@Kortsch2019FooStr], and North Sea demersal fishes and benthic epifauna [@Frelat2022FooWeb]. These downscaled assemblages allowed studying network structures in novel ways, for instance, assessing changes in food web structure across space [@Braga2019SpaAna], describing the scaling of network area relationships [@Galiana2021SpaSca]. Other examples have shown that the metaweb can be used to investigate large-scale variation in food web structure, indicating high geographical connections and heterogeneous robustness against species extinctions [@Albouy2019MarFis], which are only apparent when the local and global networks are both available. Further comparisons between network structure and other community properties are relevant as they may highlight new and surprising elements regarding network biogeography. For instance, @Frelat2022FooWeb found a strong spatial coupling between community composition and food web structure, but a temporal mismatch depending on the spatial scale. @Poisot2017HosPar found that

interaction uniqueness captures more composition variability than community uniqueness, and that sites with exceptional compositions might differ for networks and communities, because species distributions and species interactions had different bioclimatic drivers. Spatialized network data will allow these comparisons, identifying important conservation targets for networks and whether they differ geographically from areas currently prioritized for biodiversity conservation.

A key challenge remains in how to downscale a regional metaweb towards local network predictions that reflect the spatial variability of interactions. Even when the metaweb is known, local networks may vary substantially and differ both amongst themselves and from the metaweb [@McLeod2021SamAsy], emphasizing the need for methods to generate local, downscaled network predictions. A potential limitation to previous downscaling approaches is that they assume interactions are constant across space, which ignores well-documented interaction variability, and masks the effect of environmental conditions on interaction realization [@Braga2019SpaAna]. In contrast, recent studies argued that seeing interactions as probabilistic (rather than binary) events allows us to account for their variability in space [@Poisot2016StrPro] and that this should also be reflected at the metaweb level [@~~Strydom2022PreMet~~Strydom2023GraEmb]. @Gravel2019BriElt introduced a probabilistic framework describing how the metaweb can generate local realizations and showed how it could be used for modelling interaction distributions. 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 [@Hortal2015SevSho; @Poisot2021GloKno; @Windsor2023UsiEco]. Spatially replicated interaction data required for such models are especially challenging to document with large food web systems such as European tetrapod and Canadian mammal metawebs [@Maiorano2020TetSpe; @Strydom2022FooWeb], where hundreds of species result in tens of thousands of species pairs 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 workflow 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 insight compared to traditional community measures like species richness.

Methods

@Fig:conceptual shows a conceptual overview of the methodological workflow leading to the downscaled metaweb. Its components were grouped as non-spatial and spatial data, 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.

Data

Metaweb

The main source of interaction data was the metaweb for Canadian mammals from @Strydom2022FooWeb, which is a-spatial, i.e., it represents interactions between mammals that can occur anywhere in Canada. It contains 163 species and 3280 links with a probability of interaction. The species list for the Canadian metaweb was extracted from the International Union for the Conservation of Nature (IUCN) checklist [@Strydom2022FooWeb]. Briefly, the metaweb was developed using graph embedding and phylogenetic transfer
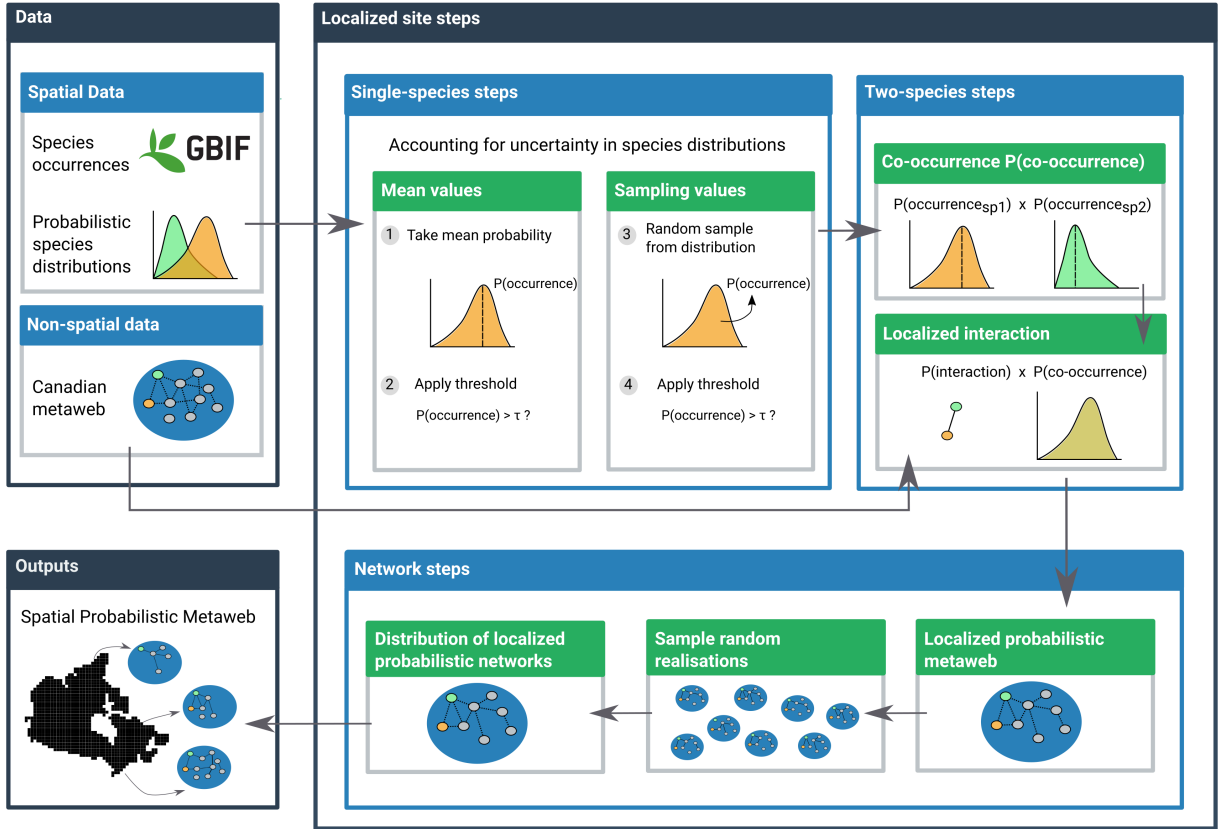
Figure 1: Conceptual figure of the proposed workflow used to downscale the probabilistic metaweb in space. The workflow has three components: the data, the localized steps, and the final spatial output. The data are composed of spatial data (with information in every cell) and non-spatial data (constant 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.

learning based on the metaweb of European terrestrial mammals, which is itself based on a comprehensive survey of interactions reported in the scientific literature [@Maiorano2020TetSpe]. The Canadian metaweb showed a 91% success rate for known interactions between Canadian mammals recorded in global databases [@Strydom2022FooWeb]. This metaweb is probabilistic, which has the advantage of reflecting the likelihood of an interaction taking place given the phylogenetic and trait match between two species. This allows incorporating interaction variability between species (i.e., taking into account that two species may not always interact whenever or wherever they occur); however, we highlight that other factors beyond trait and phylogenetic matching (e.g., population densities) will also contribute to observed interaction frequencies.

Species occurrences

The downscaling of the metaweb involved combining it with species occurrence 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 [@GBIFSecretariat2021GbiBac]. This step removed potential duplicates by combining species listed in the Canadian metaweb which were considered as a single entity by GBIF. We collected occurrences for the updated species list (159 species) using the GBIF download API on October 21st 2022 [@GBIF.org2022GbiOcc]. We restricted our query to occurrences with coordinates between longitudes 175°W to 45°W and latitudes 10°N to 90°N. This was meant to collect training data covering a broader range than our prediction target (Canada only) and include observations in similar environments. Then, since GBIF observations represent presence-only data and most predictive models require absence data, we generated ~~pseudo-absence data using the surface range envelope method, which selects random non-observed sites within the spatial range delimited by the presence data~~ the same number of pseudo-absence data as occurrences for every species [@Barbet-Massin2012SelPse]. We weighted candidate sites by their distance to a known observation (separately for each species) using the `DistanceToEvent` method from the *Julia* package `SpeciesDistributionToolkit` https://github.com/PoisotLab/SpeciesDistributionToolkit.jl , making it more likely to select sites further away from an observation and the known species range. This is because our intent was to model the potential distribution of species, capturing wider responses to the environment, as the downscaled metaweb we aimed to produce is potential in nature (see *Downscaled metaweb* section below).

Environmental data

We used species distribution models [SDMs, @Guisan2005PreSpe] to project Canadian mammal habitat suitability across the country, which we treated as information on potential distribution. For each species, we related occurrences and pseudo-absences with 19 bioclimatic variables from CHELSA [@Karger2017CliHig] and 12 consensus land-cover variables from EarthEnv [@Tuanmu2014Glo1km]. The CHELSA bioclimatic variables (*bio1-bio19*) represent various measures of temperature and precipitation (e.g., annual averages, monthly maximum or minimum, seasonality) and are available for land areas across the globe. We used the most recent version, the CHELSA v2.1 dataset [@Karger2021CliHig], and subsetted it to land surfaces only using the CHELSA v1.2 [@Karger2018DatCli], which does not cover open water. The EarthEnv land-cover variables represent classes such as Evergreen broadleaf trees, Cultivated and managed vegetation, Urban/Built-up, and 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/OGRcontributors2021GdaOgr]. This resolution compromised capturing both local variations and broad-scale patterns while limiting computation costs to a manageable level as memory requirements rapidly increase 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 @Gravel2019BriElt], 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~~ considered as a measure of uncertainty. We used the average value and standard deviation to build a Normal distribution for the probability of occurrence of a given species at ~~all sites~~ each site (represented as probability distributions on @Fig:conceptual). 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 [@StatisticsCanada2022BouFil] to set the boundaries for our predictions, which gave us 970,698 sites in total. Performance evaluation for the single species SDMs are available on GitHub.

Building site-level instances of the metaweb

The next part of the workflow was to produce local metawebs for every site (*Localized steps* box on @Fig:conceptual). This component was divided into single-species, two-species, and network-level steps.

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(occurrence)*; [@Fig:conceptual]) from the occurrence distributions obtained in the previous steps: 1) taking the mean from the distribution as the probability of occurrence (option 1 in @Fig:conceptual); 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 in @Fig:conceptual) used was the value that maximized Youden's J informedness statistic [@Youden1950IndRat], the same metric used by @Strydom2022FooWeb 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 [@Poisot2016StrPro]. On the other hand, we expected random sampling to create higher spatial heterogeneity compared to the mean probabilities, as including some extreme values should confound the ~~potential effects of~~ main trends promoted by environmental gradients. We chose option 1 to present our results as it is intuitive and essentially represents the result of a probabilistic SDM [as in @Gravel2019BriElt], but results obtained with other sampling strategies are available in Supplementary Material~~, Fig. S1~~.

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(co-occurrence)*; @Fig:conceptual) (obtained using 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 [@Poisot2016StrPro]. We then sampled several random network realizations to represent the potential local realization process [@Poisot2015SpeWhy]. 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 workflow was the downscaled metaweb, which contains a localized probabilistic metaweb in every site across the study area (*Outputs* box in @Fig:conceptual). The metaweb sets an upper bound on the potential interactions [@~~Strydom2022PreMet~~Strydom2023GraEmb], therefore, the downscaled metaweb is a refined upper boundary at the local scale that takes 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 of species interactions and occurrences, 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 @Poisot2016StrPro; see @Gravel2019BriElt for a similar example] as its relationship with species richness has been highlighted in a spatial context in recent studies [@Galiana2021SpaSca; @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 [@Martins2022GloReg]. To do so, we rasterized the Canadian subset of the global map of ecoregions from [@Dinerstein2017EcoApp; also used by @Martins2022GloReg], 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, as a way to avoid bias due to long tails in the distributions. We also measured within-ecoregion variation as the 89% interquantile range of the site values in each ecoregion [threshold chosen to avoid confusion with conventional significance tests\; @McElreath2020StaRet].

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, @Legendre2013BetDiv], 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, @daSilva2014LocReg; @Heino2017ExpSpe], recent studies used the measure on predicted species compositions over broad spatial extents and a large number of sites [@Vasconcelos2018ExpImp; @Dansereau2022EvaEco]. 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 [@Poisot2017HosPar]. 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 (970,698 sites by 159 species) with the probability of occurrence of each species at every site from obtained in the species distribution models. For network composition, we assembled a site-by-interaction matrix with the localized interaction values from the spatial probabilistic metaweb probability of interaction at every site given by the downscaled metaweb (therefore 970,698 sites by 3,108 interactions with defined probabilities in the metaweb). We applied the Hellinger transformation on both matrices and computed the LCBD values from the total variance in the matrices [@Legendre2013BetDiv]. High LCBD values indicate a high contribution to the overall variance and a unique species or interaction composition compared to other sites. Since the values themselves are very low given our high number of sites [as in @Dansereau2022EvaEco], what matters primarily is the magnitude of the difference between the sites. Given this, we divided values by the maximum value in each matrix (species or network) and suggest that these should be viewed as relative contributions compared to the highest observed contribution. As with other measures, we then summarized the local uniqueness values by ecoregion by taking the median LCBD value and measuring the 89% interquantile range.

Analyses of network motifs

To further explore network structure in space, we investigated the distribution of network motifs across space. Motifs are defined sets of interaction between species [@Milo2002NetMot; @Stouffer2007EviExi], for instance two predators sharing one prey, which are repeated within larger and more complex food webs. Motifs are linked to community persistence [@Stouffer2010UndFoo] and community structure [@Cirtwill2015ConPre; @Simmons2019MotBip], are conserved across scales [@Baker2015SpeRol; @Baiser2016MotAss], and are part of a common backbone of interactions among all food webs [@Mora2018IdeCom]. We focused on four of

the most studied three-species motifs [@Stouffer2007EviExi; @Stouffer2010UndFoo; @Baiser2016MotAss] : S1 (tri-trophic food chains), S2 (omnivory), S4 (exploitative competition) and S5 (apparent competition). These motifs can be grouped into two pairs according to the ecological information they display: S1 and S2 highlight different trophic structures, while S4 and S5 identicate different competition types. Therefore, we compared the spatial distribution of the motifs in each pair to see which ones were dominant across all our sites. First, we computed the expected motif count for each of the four motifs for all sites using the localized probabilistic networks from the downscaled metaweb [following @Poisot2016StrPro]. Then, we compared the expected counts of the motifs within the two pairs. To do so, we used a measure of normalized difference similar to the Normalized Difference Vegetation Index (NDVI), where we compute the difference between the two motif counts over their sum. We called the index comparing the two trophic motifs (S1 and S2) the Normalized Difference Trophic Index (NDTI), and the one comparing the competition motifs (S4 and S5) the Normalized Difference Competition Index. We defined both indexes as:

$$NDTI = \frac{(S1 - S2)}{(S1 + S2)}$$

$$NDCI = \frac{(S4 - S5)}{(S4 + S5)}$$

Values for both indexes are bounded between -1 and 1. A value of 0 indicates that both motifs have the same expected counts. Positive values indicate that the first motif in each index (S1 and S4) is dominant and has a higher expected count, while negative values indicate that the second motif (S2 and S5) is dominant. As with previous measures, we then summarized both index values by ecoregion by taking the ecoregion median and measuring its within-ecoregion variation with the 89% interquantile range. Ecoregion values therefore indicate if one type of trophic structure (for NDTI) and one type of competition (for NDCI) is dominant in the ecoregion, while the interquantile range values measures whether the dominant type varies within the ecoregion.

We used *Julia* v1.9.0 [@Bezanson2017JulFre] to implement all our analyses. We used packages `GBIF.jl` [@Dansereau2021SimJl] to reconcile species names using the GBIF Backbone Taxonomy, `SpeciesDistributionToolkit.jl` (https://github.com/PoisotLab/SpeciesDistributionToolkit.jl) to handle raster layers, species occurrences and generate ~~pseudoabsences~~pseudo-absences (using the `DistanceToEvent` method), `EvoTrees.jl` (https://github.com/Evovest/EvoTrees.jl) to perform the Gradient Boosted Trees, `EcologicalNetworks.jl` [@Poisot2019EcoJl] to analyze network and metaweb structure, and `Makie.jl` [@Danisch2021MakJl] to produce figures. Our data sources (CHELSA, EarthEnv, Ecoregions) were all unprojected, and we did not use a projection in our analyses. However, we displayed the results using a Lambert conformal conic projection more appropriate for Canada using `GeoMakie.jl` (https://github.com/MakieOrg/GeoMakie.jl). All the code used to implement our analyses is archived on Zenodo ~~()~~[<https://doi.org/10.5281/zenodo.8350065>\; @Dansereau2023PoiSpa] 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 computer clusters provided by Calcul Québec and the Digital Research Alliance of Canada. Full-scale computations (excluding motifs) required 900 CPU core-hours and peaked at 500 GB of RAM. Computations for network motifs were even more intensive and required 12 CPU core-years.

Results

Our workflow allowed us to display the spatial distribution of ecoregion-level community measures (here, expected species richness) and network measures (expected number of links; [@Fig:ecoregion_measures]). We highlight that the ~~community and network-level~~ measures presented here are ~~not actual predictions~~ first computed over the predicted communities and networks obtained when downscaling the metaweb, then summarized across the ecoregions (taking the median within each ecoregion). They are not a direct prediction 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).~~ Expected ecoregion richness ([@Fig:ecoregion_measures]A) and expected number of links ([@Fig:ecoregion_measures]B) displayed similar distributions with a latitudinal gradient and higher values in the south. ~~However, within-ecoregion~~ Within-ecoregion variability was distributed ~~differently, as some ecoregions along the coast displayed higher interquantile ranges , while ecoregions around~~ slightly differently with a less constant latitudinal gradient, notably lower interquantile ranges near the southern border ~~displayed narrower ones (~~(for example, near Vancouver Island and the Rockies on the West Coast, and near the Ontario Peninsula, the Saint-Lawrence Valley, and Central New-Brunswick in the East; [@Fig:ecoregion_measures]C-D). Bivariate comparison of the distributions of species richness and expected number of links and of their respective within-ecoregion variability further shows some areas of mismatches, indicating that richness and links do not co-vary completely although they may show similar distributions for median values (see Supplementary Material, Fig. S1). All results shown are based on the first sampling strategy (option 1) mentioned in the *Building site-level instances of the metaweb* section, where we used the mean value of the species distributions as the species occurrence probabilities (results for other sampling strategies are shown in Supplementary Material, Fig. ~~S1~~S2). Site-level results (before summarizing by ecoregion) are also provided in Supplementary Material (Figs. ~~S2-S5).~~

~~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:ecoregion_bivariates). 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:ecoregion_bivariatesA). The northernmost ecoregions (Canadian High Artic 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 Artic 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:ecoregion_bivariatesB). This indicates that richness and links do not co-vary completely (i.e. their variability is not highly correlated)although they may show similar distributions for median values~~S3-S6).

Our results also indicate a mismatch between the uniqueness of communities and networks ([@Fig:ecoregion_lcbd]). Uniqueness was higher mostly in the north and along the south border for communities, but ~~only~~ mostly in the north for networks ([@Fig:ecoregion_lcbd]A-B). Consequently, ecoregions with both unique community composition and unique network composition were mostly in the north ([@Fig:ecoregion_lcbd]C). 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~~ ecoregions along the south border had a unique species composition but a more common network composition ([@Fig:ecoregion_lcbd]C). ~~Opposite areas with~~ Two ecoregions showed the opposite (unique network compositions only~~were observed at higher between latitudes 52°N and~~) at higher latitudes (Davis Highlands tundra, near 70°N~~(Eastern Canadian Shield Taiga, Northern Canadian Shield Taiga, Canadian Low Artic Tundra). Also~~, and Southern Hudson Bay taiga, near 54°N). Moreover, network uniqueness values for ecoregions spanned a narrower range between the 44 ecoregions than species LCBD values ([@Fig:ecoregion_lcbd]D, left). Within-ecoregion variation was also lower for network values with generally lower 89% interquantile ranges among the site-level LCBD values ([@Fig:ecoregion_lcbd]D, 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, Fig. S5). The areas only unique for their species composition had both a high richness and number of links~~

Comparing the distribution of dominant network motifs revealed additional areas of variation in network structure ([@Fig:motifs]). NDTI displayed a latitudinal gradient between the trophic motifs. Northern ecoregions showed positive NDTI values and high dominance of S1 (tri-trophic chains) expected counts compared to S2 (omnivory) but ecoregions along the south border showed a reduced dominance ([@Fig:motifs]A). Ecoregions near the Ontario Peninsula and Saint-Lawrence Valley showed values close to zero, indicating a balance between two motifs, while Central New Brunswick had slighty negative values, indicating a low dominance of S2. In comparison, NDCI values showed an evenly high dominance of S5 (apparent competition) over S4 (exploitative competition) across all ecoregions ([@Fig:motifs]B).
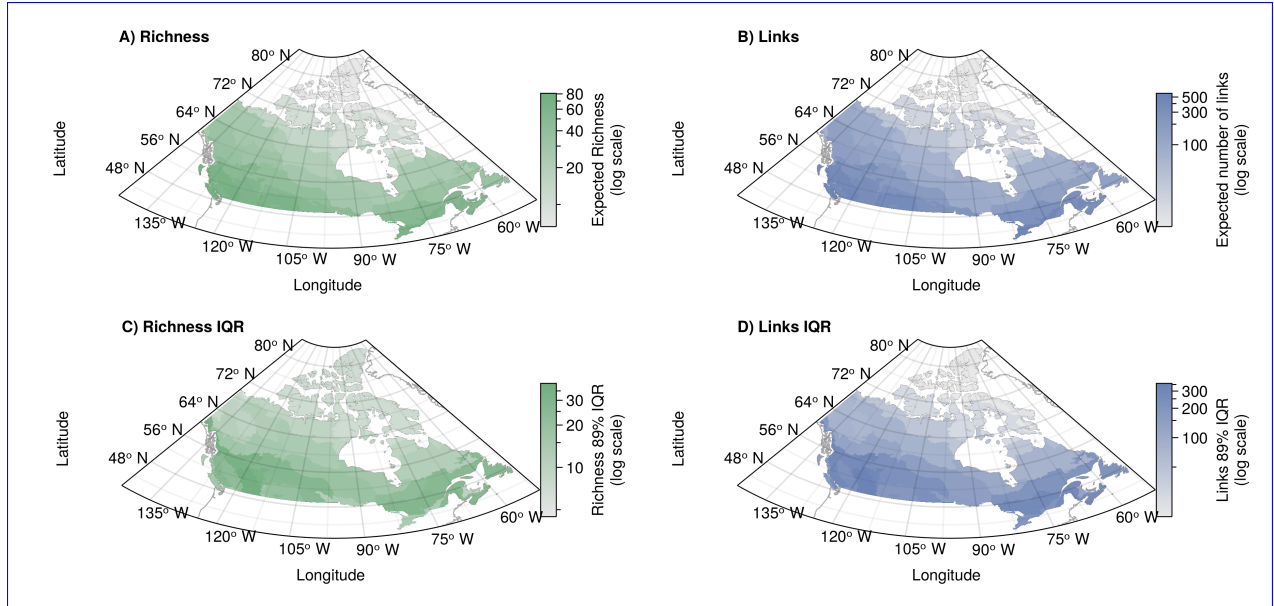
Figure 2: (A-B) Example of a community measure (A, expected species richness) and a network one (B, expected number of links). Both measures are assembled from the predicted probabilistic communities and networks, respectively. Values are first measured separately for all sites; then, the median value within each ecoregion was 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). All color bars follow a log scale with tick marks representing even intervals. Real values (non-log transformed) are shown beside major tick marks while minor ticks represent half increments.

Meanwhile, within-ecoregion variance displayed a different spatial distribution from the median values. NDTI interquantile ranges spanned a wide range of values and were higher both in the north and in the south (although not in the ecoregions with higher NDTI median values) ([@Fig:motifs]C). 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~~NDCI interquantile ranges showed lower within-ecoregion variance in most ecoregions except in the northernmost one (Canadian High Arctic tundra), which has a notably higher value ([@Fig:motifs]D). Although this higher variance does not reflect in the NDCI median values, it does appear when looking at the site-level, where this ecoregion is the only one with patches with high positive NDCI values (indicating a dominance of S4) surrounded by highly negative values (indicating a dominance of S5) as in other ecoregions (Fig. S6B).
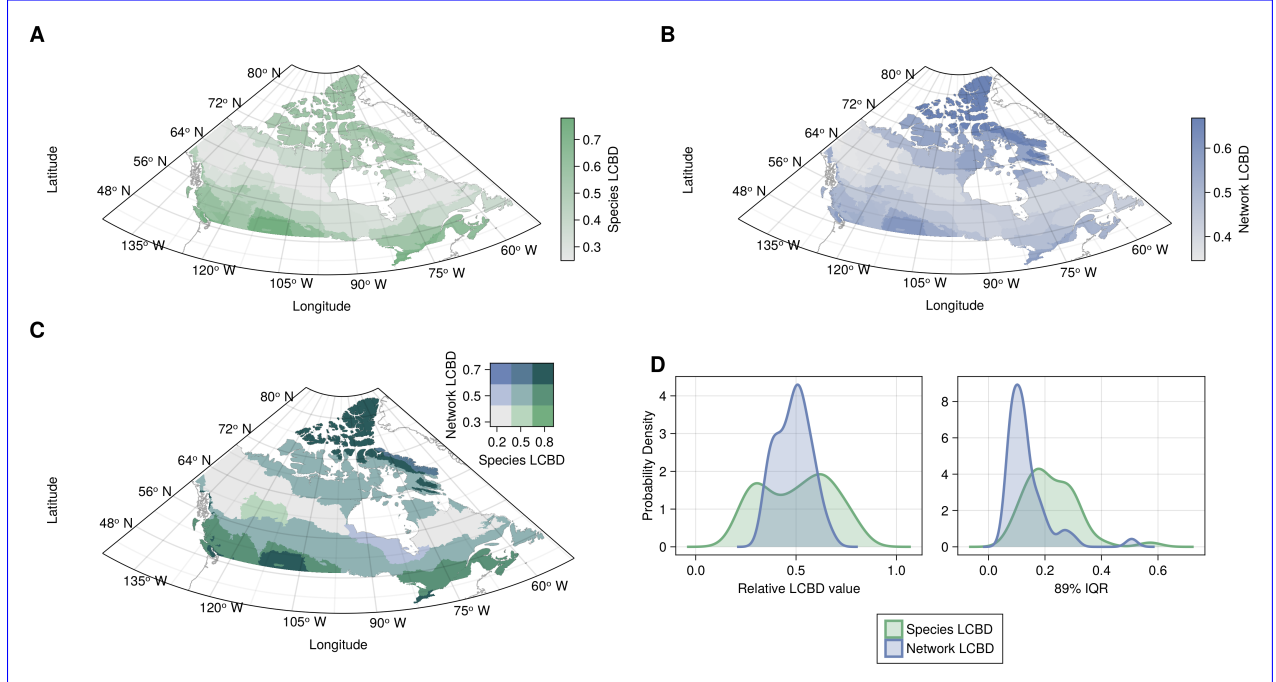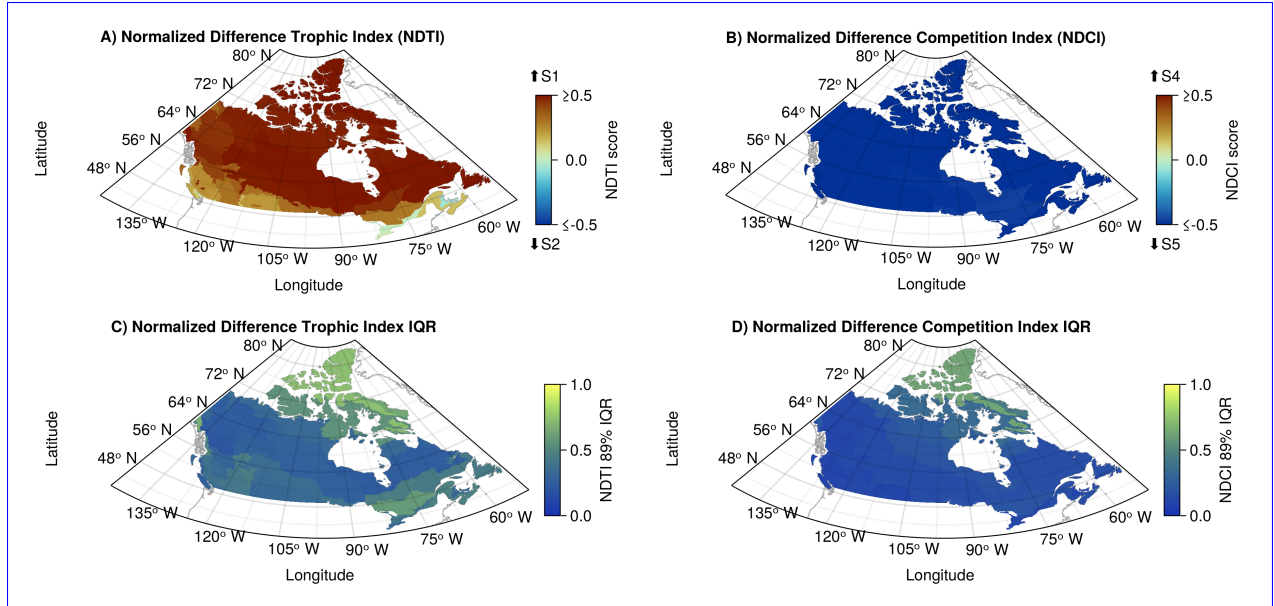
Figure 3: (A-B) Representation of the ecoregion uniqueness values based on species composition (A) and network composition (B). LCBD values were first computed across all sites and scaled relative to the maximum value observed. The ecoregion LCBD value is the median value for the sites in the ecoregion. (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 LCBD values between ecoregions, and the 89% interquartile range of the values within each ecoregion (right), highlighting the variability within the ecoregions.

A) Normalized Difference Trophic Index (NDTI)

B) Normalized Difference Competition Index (NDCI)

C) Normalized Difference Trophic Index IQR

D) Normalized Difference Competition Index IQR

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 @~~Strydom2022PreMet~~Strydom2023GraEmb. It gives us an idea of what local metawebs or networks could look like in space, given species distributions and their variability, as well as the uncertainty around species interactions. We also provide the first spatial representation of the metaweb of Canadian mammals [@Strydom2022FooWeb] and a probabilistic equivalent to how the European tetrapod metaweb [@Maiorano2020TetSpe] was used to predict localized networks in Europe [@Braga2019SpaAna; @OConnor2020UnvFoo; @Galiana2021SpaSca; @Gauzere2022DivBio; @Botella2023LanInt]. Therefore, our approach could open similar possibilities of investigations on the variation of structure in space [@Braga2019SpaAna] and on the effect of land-use intensification [@Botella2023LanInt] on North American food webs, particularly Canadian mammal food webs. Other interesting research applications include assessing climate change impacts on network structure [e.g., @Kortsch2015CliCha] or investigating linkages between network structure and stability [@Windsor2023UsiEco].

As our approach is probabilistic, it does not assume species interact whenever they co-occur and incorporates variability based on environmental conditions (via projected species distributions), which could lead to different results by introducing a different association between species richness and network properties. @Galiana2021SpaSca found that species richness had a large explanatory power over network properties, but mentioned this could potentially be due to interactions between species being constant across space. Here, we found ~~mismatches in the distribution of~~ that potential species richness and ~~interactions that were especially apparent in their~~ number of links displayed similar distributions following a latitudinal gradient, but that the within-ecoregion ~~variability (~~ variance was lower along the southern border than the measures themselves [(@Fig:ecoregion_~~bivariates~~measures]))~~, highlighting that interactions might vary differently than species distributions even over continental-scale gradients. Network~~. Why the relationship between the median ecoregion values and the ecoregion variance is not constant across space could be verified through more tests in future studies (for instance, testing the effect of higher urban density in the south). Examples of metaweb structure in space are rare and not ideal for comparison with our results. For instance, we found that network density (links on [@Fig:ecoregion_~~bivariates~~measures]~~A) were also~~ B) were lower in the north, ~~contrarily~~ which is contrary to what was observed ~~for all European terrestrial tetrapods~~ in Europe for the terrestrial tetrapod metaweb [@Braga2019SpaAna; @Galiana2021SpaSca] and for willow-galler-parasitoid networks [@Gravel2019BriElt], where connectance was higher in northern regions. However, those are systems with different number of species and environmental conditions (e.g. Europe and Canada could differ due to varying climatic conditions at the same latitudes). Further research should investigate why these results might differ between ~~the two continents and~~ continents and ecological systems and whether it is due to the methodology, data, or biogeographical processes.

Our LCBD and uniqueness results highlighted that areas with unique network composition differ from sites with unique species composition. In other words, the joint distribution of community and network uniqueness highlights different diversity hotspots. @Poisot2017HosPar showed a similar result with host-parasite communities of rodents and ectoparasitic fleas. Our results further show ~~how~~ that these differences could be distributed across ecoregions and ~~in~~ over a broad spatial extent ~~. Areas unique for only one element (species or network composition) differed in their combination of species richness and number of links (Supplementary Material, Fig. S5), with species-unique sites displaying high values of both measures, and network-unique sites displaying low values. Moreover,~~ for mammal food webs. LCBD scores essentially highlight variability hotspots and are a measure of the variance of community or network structure. Here, they also serve as an inter-ecoregion variation measure, which can be compared to the within-ecoregion variation highlighted by the interquantile ranges. The narrower range of values for network LCBD values and the lower IQR values indicate that both the inter-ecoregion and within-ecoregion variation are lower for networks than for species ([@Fig:ecoregion_lcbd]). ~~Additionally, higher values for network LCBD also indicate that most ecoregions can hold ecologically unique sites~~

Our analysis of the distribution of dominant network motifs revealed additional areas of variation in network structure. Trophic motifs (S1 and S2, measured through NDTI values) showed a latitudinal gradient different from the richness and links ones, with high dominance of tri-trophic chains (S1) in the north and higher omnivory counts (S2) only in a few ecoregions in the south. This results did not seem related to ecoregion variance, which once again showed a very different distribution from the median values. Meanwhile, competition motifs (S4 and S5, measured through NDCI values) showed an even dominance of apparent competition (S5) but high variance in a single ecoregion. Overall, our results show that dominant motifs within a mammal food web vary between ecoregions and vary differently across space.

When to use the workflow we presented here will depend on the availability of interaction data or existing metawebs, and on the intent to incorporate interaction variability, as well as ecoregion-level variability. In systems where *in situ* interaction and complete network data are available, the approach put forward by @Gravel2019BriElt achieves a similar purpose as we attempted here, but is more rigourous and allows modelling the effect of the environment on the interactions themselves. Without such data, establishing or predicting the metaweb (e.g. using transfer learning) should be the first step toward producing localized predictions [@~~Strydom2022PreMet~~Strydom2023GraEmb]. Our framework then downscales the metaweb towards the localized predictions, here using the probabilistic Canadian mammal one, but it can also use other metawebs generated through various means. Well-documented binary ~~metawebs~~ ones such as the European tetrapod metaweb could be partly combined with our approach if used with probabilistic SDMs and summarized by ecoregions (as they would only lack an initial probabilistic metaweb, but would still obtain a more probabilistic output). Our approach will essentially differ from previous attempts in how it perceives uncertainty and variability. For instance, rare interactions should not be over-represented [@Poisot2016StrPro] and should have lesser effects over computed network measures. Furthermore, summarizing results by ecoregion allows for showing variation within and between ecologically meaningful biogeographic boundaries [@Martins2022GloReg], which, as our results showed, is not constant across space and can help identify contrasting diversity hotspots.

Adding validation to our framework will be highly important for the predictions to be actionable. We see it as the next step since data availability currently remains an issue. Developing a way to generate actionable information when information is initially scarce, as we present here, is highly important in itself. Our manuscript highlights the application of a predictive pipeline in a case where interactions are almost fully predicted, but it can equally be applied when there are additional data about the interaction network. Moreover, the predictions we make are already actionable, as any future sampling of food web structure can be used to validate them, and fed into the model to iterate these results again. As @Strydom2023GraEmb point out, validation of metaweb predictions, empirical sampling, and method design should all proceed jointly, and making conceptual progress in one of these areas helps all the others.

The recent shift in focus towards building metawebs opens many opportunities for projections of networks in space through probabilistic downscaling, as we suggested here. Metawebs have been documented in many systems, allowing us to build new ones from predictions. How the European tetrapod metaweb [@Maiorano2020TetSpe] was used to predict the Canadian mammal metaweb [@Strydom2022FooWeb] is

one such case, but recent examples also extend to other systems. Metawebs have been compiled for many marine food webs [e.g., Barents Sea, @Kortsch2019FooStr; North Scotia Sea, @Lopez-Lopez2022EcoNet; Gulf of Riga, @Kortsch2021DisTem] and used to predict the probability of novel interactions [Artic food web of the Barents sea, @Pecuchet2020NovFee]. @Olivier2019ExpTem built a temporally resolved metaweb of demersal fish and benthic epifauna but also suggested combining their approach with techniques estimating the probability of occurrence of trophic relationships to describe spatial and temporal variability more accurately. @Lurgi2020GeoVar built a metaweb and probabilistic (occurrence-based) networks for rocky intertidal communities (and doing so they also showed that environmental factors do not affect the structure of binary and probabilistic networks in different ways). @Albouy2019MarFis predicted the global marine fish food web using a probabilistic model, showing the potential to describe networks across broad spatial scales. Similarly, predictive approaches are also increasingly used with other interaction types to highlight interactions hotspots on global scales [e.g. mapping geographical hotspots of predicted host-virus interactions between bats and betacoronaviruses, @Becker2022OptPre; predicting the distribution of hidden interactions in the mammalian virome, @Poisot2023NetEmb]. Our workflow offers the potential to bring these global predictions down to the local scale where they can be made more actionable, and vastly increases the diversity of ecological networks that can be projected in space.

Acknowledgements

References