

RSTB-2023-0166 Author response to reviewers

Here is the response from the authors.

Main changes:

- Change in pseudoabsence generation method
- Updated all results accordingly
- Added motifs analysis to further explore network structure in space
- **Moved the richness-links bivariate maps to supp.mat.**

Reviewer 1

Thank you for a great paper. I really like the design of your workflow, and I think it provides a sound and solid way of downscaling the metawebs into ecoregion or local webs, that could foster new analyses of different taxa and regions. Just to be clear, I believe the structure of your workflow seems quite sound to me, and the way it embeds probabilistic measures of potential distributions and potential interactions may help modelling the variability that local networks may have in nature in a realistic way. In fact, I do look forward to see comparisons of these potential networks with empirical networks (and where do realized networks lie within the possible networks), but that's a different step, that first requires developing the analytical framework that you are presenting here. I only have two major suggestions which, I believe, would sharpen and strengthen your workflow, and a specific comment.

We are grateful to Reviewer 1 for their kind words and understanding of our manuscript.

Why do you use random pseudoabsences within the species range to construct the Species Distribution Models if you want to represent the potential distribution? You do it throughout all the range of the data, but that includes both areas within the species range, and areas outside of it. I'm telling this because the decision is not meaningless in terms of whether you model current distributions (where including absences in places within the range to include areas that could host presences but are unoccupied by the species matters) or potential distributions (where capturing the wider responses to the environment is what matters). That is, capturing where the species is vs where the species could be. A consequence of this is that how you pick pseudoabsences is a key choice, and I believe you may have taken the wrong direction. There are examples in the literature about how to make this choice, but if you would create them from outside of the observed range of the species, or assigning lower probability to choosing non-sampled places within the known species range I believe your models will represent potential distribution much better. If you do so, this should be much clear in the text.

- We understand the reviewer's concerns and agree with them
- Our intent is to model the potential distribution
- Consequently we changed the pseudoabsence generation method for the DistanceToEvent one, which weight candidate sites by their distance to an observation, making it more likely

- to select sites further away from observed ones
- Add justification from Barbet-Massin et al.?

Your spatial comparison between richness and network size is so great that I wonder whether you do not explore other aspects of network structure, such as nestedness and modularity. These metrics may help understanding better the spatial mismatches between richness and number of links shown in figs 3 and 4, as if I'm getting it correctly, these may be caused by the selection of a handful modules in the northern ecoregions, compared to more connected networks towards the south-southwest.

- We added an analysis of motifs
- We chose motifs over nestedness and modularity as they are ecologically meaningful and have a clear interpretation
- **Brush over mismatch between richness and links?**

Line 162; rather than potential effects (which to me includes all potential variation associated to the probabilistic models) I would say “main/overall trends promoted by environmental gradients”, or something similar, emphasizing that the variability creates “noise” in local networks, but that such “noise” is realistic.

- We added this suggestion

Besides these three things, my only concern is that the results could perhaps allow to explore further the gradients in network structure (related also with my second major comment), which me being a biogeographer is perhaps unsurprising. I however would like to see you further explorations of this facet of your results; if not here, in other papers.

- See description of motifs

Reviewer 2

General comment

Dansereau et al. use a predicted tetrapod metaweb to predict the structure of mammal food webs for Canada at the local scale. I enjoyed reading the paper and applaud the authors for their efforts in predicting species interactions in space, which is a monumental task, and hard, maybe even impossible, to achieve empirically. All in all, the work by Dansereau et al. 2023 is very impressive and deals with a topic of high relevance and importance for food web ecology, and beyond. Especially, the enormous pipeline of computations behind this work is truly impressive and inspiring. The framework presented opens many new opportunities and holds great potential for improving predictive food web ecology. It especially nice to see that this approach allows for capturing variability in trophic interactions at the local food webs.

We are thankful to Reviewer 2 for their acknowledgement of the work we put into this manuscript.

That said, I do have some concerns and questions regarding the realism and validity of the predictions, based on species distribution data outside of the target region (Canada), and on a predicted metaweb based on species interactions from a different continent, which results in predictive algorithms trained on data outside the target region(s). Moreover, the study bases prediction upon prediction in a long pipeline of predictions, are there any risk involved with this in terms of how realistic the networks are? Could you add some reflections about this in the main text.

- Point out that the Canadian metaweb was validated
- SDMs are also validated
- We intend to represent potential networks and distributions
- We want to show the potential variation in space

In the following, I would like to ask some clarifying questions about the premise of this work and the choices made, which I also think should be made clearer in the main text. As far as I understand it, the accuracy of the metaweb is important for the accuracy of the local food web predictions, which in turn is important for the local webs to be actionable, one of the ultimate goals with this work.

Without comparing the predicted output to observed data, how sure can you be that the predicted food webs reflect actual food web realizations? Therefore, a validation of the results should be added to the paper, and this should also be discussed. There must be some field sampled information on mammals in Canada for some of the defined ecoregions which you could use for comparing the predicted species composition of the food webs to observed ones. For the predictions to be actionable, this would seem an important step.

- As Reviewer 1 pointed out we present an analytical framework first and foremost
- Summary results, such as motifs, will be easier to validate than entire networks, especially at the ecoregion level

- Validation would be relevant, but data availability is an issue
- Issues with available data
 - Newfoundland: not spatial, already used for Canadian metaweb
 - Arctic food webs: few species
 - Marine motifs: marine species

If I understand correctly, for the predictions behind this study to work, field sampled observations are necessary - that is the European tetrapod metaweb based on field observations were essential to make the predicted Canadian metaweb. I find this fundamental premise of the metaweb construction underreported in this paper. Despite all the help computers can provide in making highly resolved predictions of species interactions in space, they still need high quality input data based on field observations to work properly. This point is not trivial, especially not for the “actionability” of the predictions. This should be acknowledged in the paper.

- The European tetrapod metaweb is based on expert knowledge and literature review, not field observations, although this also requires tremendous work
- Yes, our framework does require a metaweb, but it could be applied to any one. We use the probabilistic value here, but it doesn't have to be. (as we discuss in the Discussion)
- Lack of input data is one the point of Tanya's manuscript: when we don't know, we can infer from somewhere else
 - But this is not the point of our paper
- Also from Tanya's perspective manuscript: a metaweb is sometimes “simpler” to obtain than localized network values in multiple locations

Generally, I find the description of the metaweb and the assumptions behind it too sparse. How big is the metaweb? How many species, links and trophic levels does it contain? Does it only contain mammals? What resource species (prey) does it contain? Is it just mammals preying on other mammals? Could you provide a network of the metaweb in the SI with some more information?

- We added more information on the metaweb relevant to this paper
- But we are not creating the metaweb here, we're reusing it from another manuscript

This leads me to another important question, why were only species richness and number of link density calculated as network properties. Strictly, speaking you would not have needed a food web approach to calculate these two properties. It would have been more interesting if you would have provided analyses of connectance, and in- and out-degree, in space? How was the choice of metrics made, could you please add a justification to the paper.

- We added a motifs analysis which makes use of the food web approach
- Connectance is in supp.mat. The spatial distribution exactly matches the number of links as, following Poisot et al. 2016's probabilistic definitions, it is the number of link divided by the number of species in the metaweb (159)

- Previous studies such as Gravel et al. 2019, Braga et al. 2019 use number of links/connectance as their main network property in space

Finally, while the paper is generally well-written, the clarity of some parts of the text could still be improved.

Detailed comments

We incorporated most of the suggestions from Reviewer 2 in the updated manuscript. All changes are listed in the additional Track Changes file. Here, we further provide responses to a few specific comments from the Reviewer.

Data:

L. 111. Why did you include species distributions and occurrences (long. and lat.) outside of Canada, and as far south as central America? Does it improve your predictions? Did you make a test run with species ranges restricted to Canada, or North America? What happens to your predictions/results if you are more conservative in your species range? I wonder whether including data from so far south can result in an over- or under- representation of certain species in some of your ecoregions, especially in the southern parts of Canada, could it? Could you provide a test with a more conservative species range more closely reflecting the biogeography of Canada, or at minimum justify your choice and its implications.

- Training SDMs on a wider range is a standard practice
- Species were selected for the Canadian Metaweb according to IUCN
- We want to include species even if only the southern part of their range is in Canada

L. 114. Can the random choice of pseudo-absences in space influence the outcome, or does this computational step not have any effect on the results at all?

- The method to select pseudo-absences does have an effect, as shown with our updated result
- We believe the DistanceToEvent method will be more appropriate to model the potential distribution of the species, as suggested by Reviewer 1

L. 213. What are the localized interaction values? What do you mean by “interaction value”, the interaction probability, mean/median no. of links at a given site? This is not clear.

L. 242. Are the expected median values per ecoregions based on a median binary summary web of the many versions of the localized metawebs, or did you calculate species richness and number of links per species for the many versions of the localized metawebs and then take the median? Could you make this clearer. How are the probabilities of the interactions reflected in these median values? ... because rare interactions are less often predicted and hence the median number of links per species would be lower?

- The median values represent the median across all sites in the ecoregion
- Species richness and number of links are first calculated on the localized metawebs
- Then we calculate the median value across all sites in the ecoregion and attribute this value for the ecoregion
- An ecoregion with a high median for the number of links could either have interactions with a very high probability or many interactions with a lower probability

Discussion:

L. 301. Are you comparing your predicted mammal food web from Canada to the observed tetrapod food web in Europe, the latter which includes more species, especially at lower trophic levels (I assume)? Is this meaningful at all?

- We recognize this is not the best comparison (as we mention in the text)
- However, we think it's relevant to mention here as it is the closest comparison of a metaweb structure in space

L. 302. It would have been nice if you could reflect upon and elaborate in which way the data and methodology could have influenced the results. The same for biogeography.

L. 303. There should be differences between Europe and Canada due to varying climatic conditions at the same latitudes. Due to the Gulf stream northern areas of Europe are much milder in e.g., temperature than regions in Canada, so it makes sense that the food webs differ in connection given that temperatures influence species trophic interactions and food web properties.

- We emphasized that we show examples of the potential distribution of the network metrics, not a detailed analysis of network structure in Canada.

Figures:

The figures are pretty, but it was impossible to read the color code (blue-green shading) on a grey color scale print out.

- We chose this specific color palette as there are very few colorblind-friendly ones

Fig. 3b. Could the higher variation in the predictions in the southernmost part of Canada be because the predictions were trained on data of species occurrences (distributions) outside the range of Canada?