

Predicting the spatial structure of St. Lawrence food webs

David Beauchesne, Kévin Cazelles, Guillaume Blanchet, Philippe Archambault, Dominique Gravel

15 April 2017

Network level evaluation of St. Lawrence networks

"The goal of this document is to combine the use of joint species distribution models (JSDMs) with a machine learning algorithm built to predict biotic interactions in order to predict the spatial structure of interactions networks, effectively bringing to "fruition" the integrated niche concept with predictive tools. At the moment, there are interactions that simply do not make any sense... Like Porifera eating *Gadus morhua*"

Overview

These are essentially the kind of final results we wish to obtain for the 2nd objective of my thesis. The results presented combine the use of joint species distribution models and a biotic interactions prediction machine learning algorithm in order to predict the spatial structure of ecological networks in the St. Lawrence. At the moment, though, there are still many things that need to be finished before we are remotely close to calling these final results and discussing them from an ecological standpoint rather than a purely methodological one:

- There are multiple additional datasets to incorporate to the HMSC model and how this will be achieved is still food for thoughts!
 - Most important of all, how will different datasets be incorporated in the same analysis.
 - My first naïve thoughts:
 1. Format data in a way that each dataset considered only included species that can be readily captured by fishing gears considered (see next point)
 2. Include fishing gear type and survey ID as random factors in the analyses.
 3. Consider individual observations for the analysis spatially rather than at the scale of the fishing activity itself. I mean that within a year, all fishing activity, whether they be from fishing or scientific trawl surveys, could be grouped as single observations if they fall within the same grid cell of e.g. 1km². We could also use something akin to a spatial moving window that create those observations in order to consider mobile gear types and spatial uncertainty.
- Complete data formatting in order to exclude taxa that are should not be considered as readily capture by each fishing activity considered. For example, coastal species shouldn't be included in the trawl survey, which does not operate at depth over 50 meters.
- Incorporate other or better environmental covariables, among them certain categorical variables that I haven't done yet (because I didn't know how to do it, which I've discussed with Guillaume. I should now be able to do it once I've figured it out appropriately)
- A new interactions catalogue that includes nutrients and detritus will need to be built to maximize results accuracy and to avoid weird predictions like "Porifera consumes *Gadus morhua*".
 - Short term, the large groups like Porifera could be split into their constituent species in order to avoid these mistakes.
 - The same could be true for all species groups and they could be reaggregated back at the group level after interactions predictions are performed

- At the moment, I believe that interactions prediction using *iEat* over predicts the amount of links, especially in areas where there are few species. At the moment, maps of taxa richness are the exact opposite of maps for link density and connectance. I believe that the fact that we are using a KNN algorithm forces the predictions to include a minimum number of interactions for taxa, meaning that areas with lower richness will have seemingly higher rates of link. I may be wrong though and results may actually represent a true ecological fact. I need to delve deeper into the results to find out, which I haven't had time to do.
- Use probabilities instead of binary values for both spatial occurrences and biotic interactions. Ultimately, this could allow us to obtain, for each grid cell, a matrix of length equal to the number of taxa indicating the probability of species interacting based on both interaction and co-occurrence probabilities.
 - At the moment, results from *iEat* are deterministic and obtaining probabilities would be highly valuable (recurring conversation, I haven't had time to truly get into it)
 - Results from JSDBMs are probabilistic, but I transformed them into binary occurrence probabilities using a 0.5 threshold value. This should be modified. In the meantime, we could derive taxa specific threshold values that minimize False Positive Rates and maximize True Positive Rates using ROC curves. I haven't looked into this yet.

Spatial description of St. Lawrence ecological networks

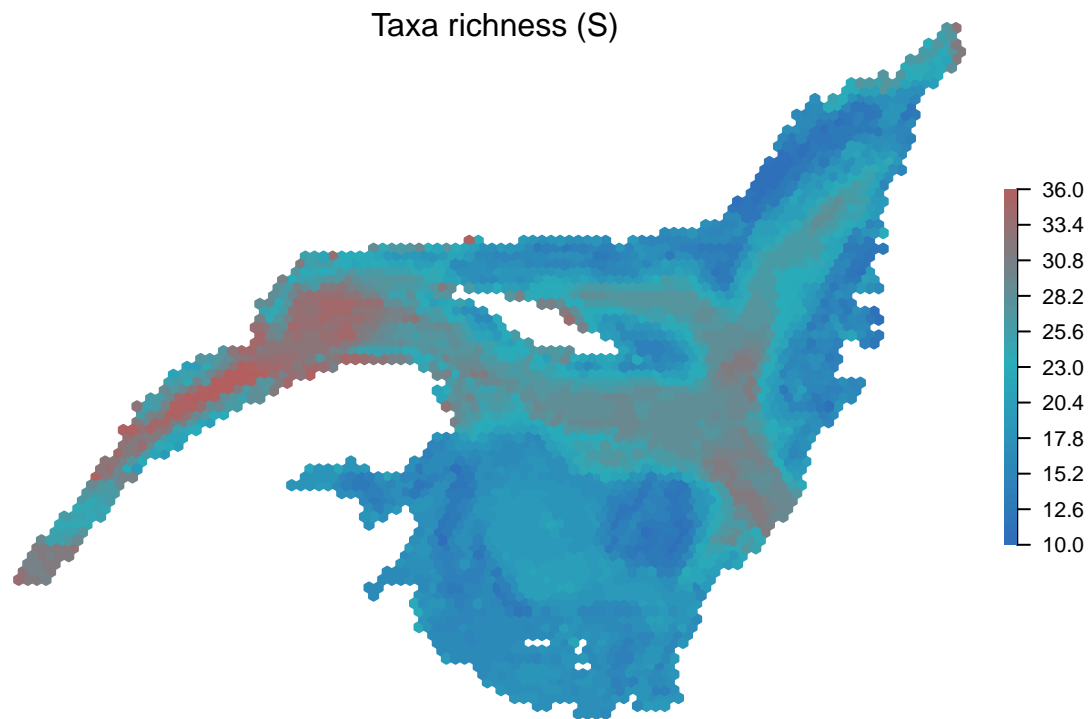


Figure 1. Taxa richness (S) evaluated using a metacommunity of 124 taxa.

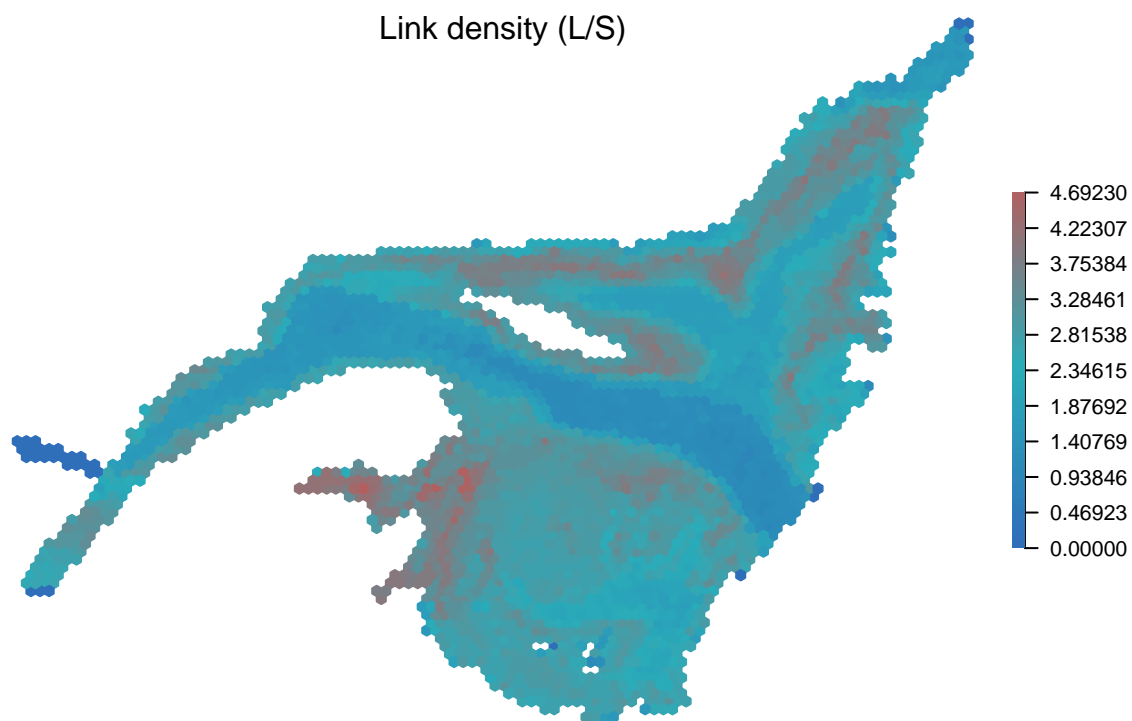


Figure 2. Predicted link density (L/S) evaluated using a metacommunity of 124 taxa.

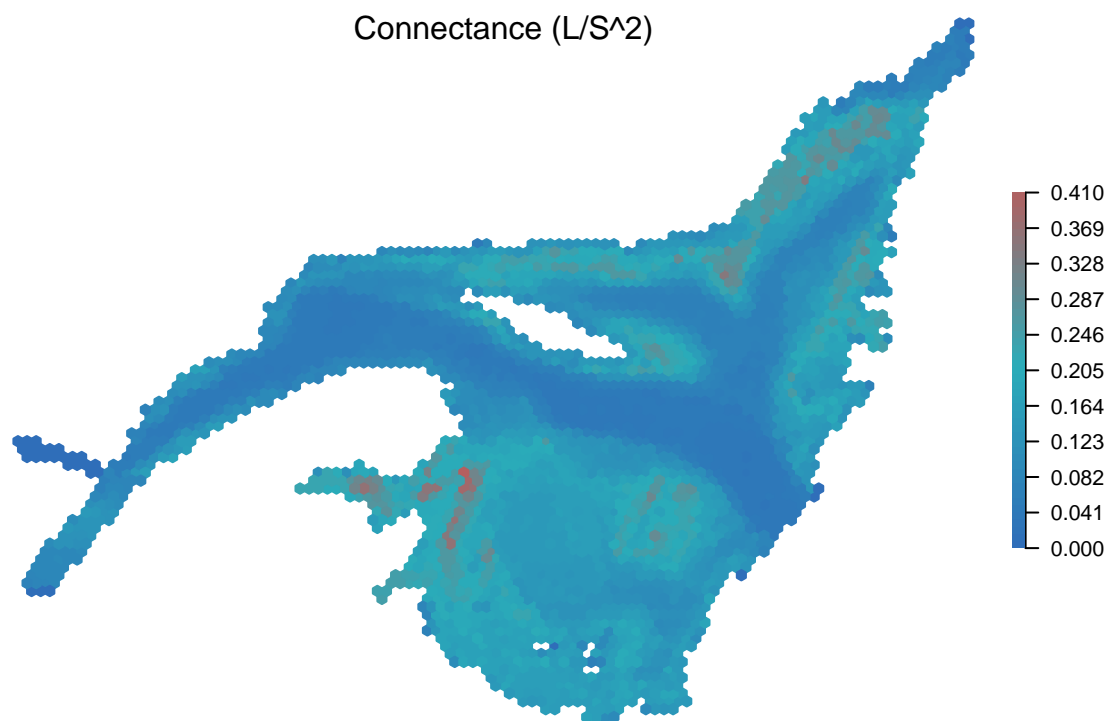


Figure 3. Predicted connectance (L/S^2) evaluated using a metacommunity of 124 taxa.