

Building a better metaweb: predicting spatiotemporally explicit plant-pollinator networks

Michael D. Catchen^{1,2}, Paul CaraDonna^{3,4}, Jane E. Ogilvie³, Francis Banville³, Dominique Caron^{1,2}, Philippe Desjardins-Proulx^{5,2}, Norma R. Forero-Muñoz^{5,2}, Dominique Gravel^{6,2}, Laura Pollock^{1,2}, Tanya Strydom^{5,2}, Timothée Poisot^{5,2}, Julian Resasco⁷, Andrew Gonzalez^{1,2}

¹ McGill University; ² Québec Centre for Biodiversity Sciences; ³ Rocky Mountain Biological Laboratory; ⁴ Chicago Botanic Garden; ⁵ Université de Montréal; ⁶ Université de Sherbrooke; ⁷ University of Colorado Boulder

Correspondance to:

Michael D. Catchen — michael.catchen@mail.mcgill.ca

Purpose: This template provides a series of scripts to render a markdown document into an interactive website and a series of PDFs.

Motivation: It makes collaborating on text with GitHub easier, and means that we never need to think about the output.

Internals: GitHub actions and a series of python scritpts. The markdown is handled with pandoc.

Keywords:
species interactions
ecological forecasting
pollinators
bumblebees
network ecology

0.1. Abstract Using a data set of [DESCRIBE EACH DATASET IN A NICE WAY], we predict a spatiotemporally explicit metaweb of interactions between bumblebees (*Bombus*) and wildflowers (within *find clade*). We integrate this data with crowdsourced occurrence data and climate data to [best paint the picture of the Colorado bumblebee-plant metaweb]. Using temporal climate data, we forecast how the spatiotemporal overlap of interacting species will change under proposed climate scenarios. We use this to estimate what interactions between bees and plants need the most attention to prevent the spatiotemporal decoupling of an interactions from threatening ecosystem functioning or the persistence of a species.

1

Introduction

Ecosystems emerge from sets of interactions between pairs of species. these interactions produce the dynamics and persistence of ecosystems, and the abundance and diversity of the species within them. Plant-pollinator interactions specifically drive the function and persistence of “architecture of biodiversity” (Bascompte & Jordano 2007). However, we are far from a robust understanding of plant-pollinator networks. This is because sampling interactions is costly. Interactions vary in space and time (Poisot *et al.* 2015)—particularly relevant in this system (CaraDonna *et al.* 2014). This is why there is interest in using models to predict interactions from sparse data (Strydom2021?). In this paper, we combine several datasets, each spanning several years, to produce spatially and temporally explicit predictions of the bumblebee (genus *Bombus*) and wildflower pollination network across the state of Colorado.

We do this in two parts: (1) metaweb prediction and (2) conditioning our metaweb prediction on co-occurrence probability. First, we build a model to predict the metaweb—the network of *all* interactions, aggregated across all times and spatial locations—of *Bombus* and wildflower species across Colorado. (Why do this? The metaweb is more predictable than local interactions.) We do this using network

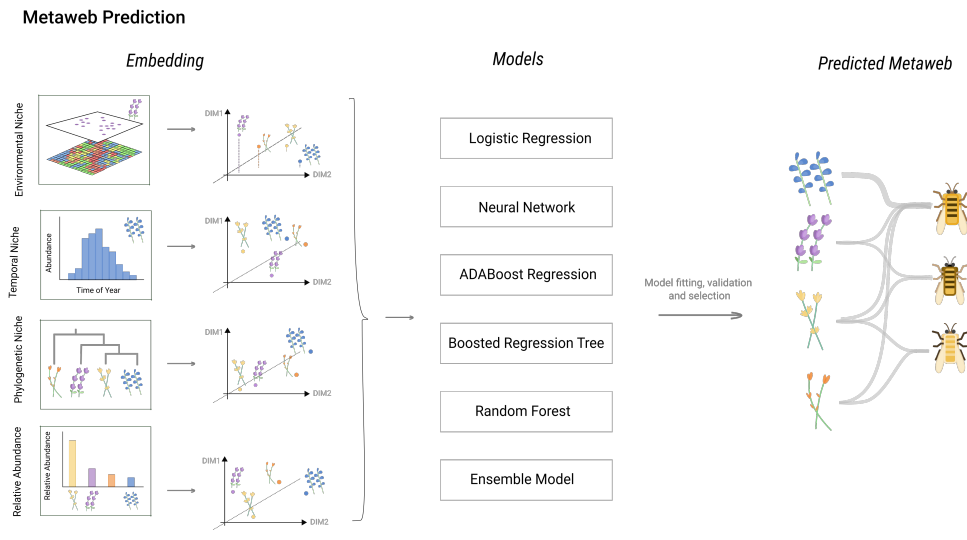
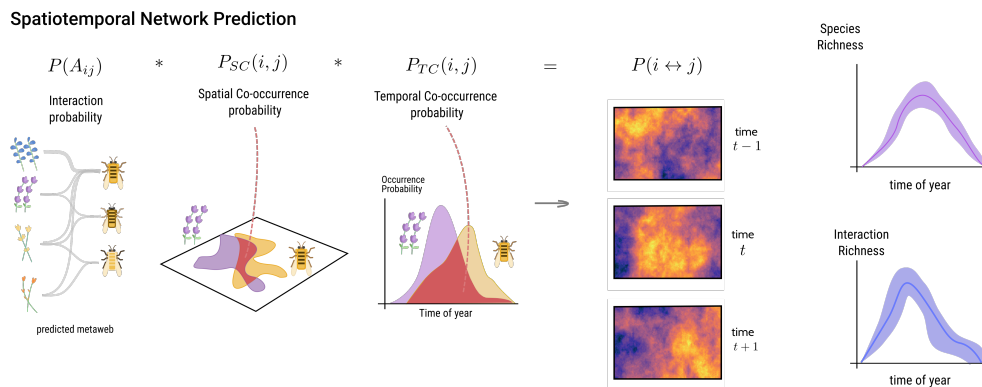


Figure 1 todo



embedding (**cite?**). Network embedding takes each node in the network (either a bumblebee or a wild-flower) and represents it in a latent n dimensional space. Combination of running models on Temporal niche (T), Phylogenetic niche (P), Environmental niche (E), and relative abundance in community (RA).

Second, we then use this metaweb to predict the structure of networks at specific locations and times of year (Gravel *et al.* 2019). Finally we suggest a map of sampling priority, which suggests the locations to sample that will best improve our understanding of the Colorado *Bombus* pollination metaweb.

Why is this good for science, what does this contribute to our understanding of plant-pollinator ints, networks, *Bombus*, predictive models, etc., and how can these results be useful.

2

Data and methodology

We use three separate field datasets to estimate the Colorado *Bombus* metaweb.

3

Predicting the metaweb

3.1. Feature Embedding

3.1.1 Environmental niche features We take the 19 BioClim layers from CHELSA (cite; 1km resolution) and a map of elevation and PCA them. A resulting 4 layers cover 99.5% of the variance. We use species occurrence data from GBIF, and consider each occurrence record as a point in environmental space. Then we fit a multivariate normal distribution to these points in environmental space.

3.1.2 Temporal niche features We take the mean and variance of the distribution of number of observations per week of year in the interaction field data.

3.1.3 Phylogenetic features

3.1.3.1 Phylogeny Construction We construct phylogenies for both *Bombus* and wildflower species using barcode markers, mitochondrial COI and chloroplast rbcL, respectively. These sequences were obtained from NCBI GenBank for all species. For species for which no sequence was available (only a handful of plants), their was substituted with a barcode from a member of the same genus. Justify why this is fine here.

These sequences were aligned using ClustalOmega v???, and then a posterior distribution of phylogenies and consensus tree was obtained via MrBayes v??, using XX substitution model with gamma-distributed rates. Run until convergence, which here we define as the standard-deviation of splits falling below 0.1.

3.1.3.2 Creating an embedding from phylogenies We simulate traits. Relationship between number of traits, num output PCA dimensions, and number of used dimensions in the model matter. Describe how that works.

3.1.4 Relative Abundance This embedding is the simplest.

3.2. Metaweb Model Fitting and Validation We fit a bunch of models using MLJ.jl. Some of them are bagged, some are not bagged.

We train each model on each possible combination of features.

AUC-ROC and AUC-PR values below in fig. 2

What does this tell us? The ensemble model is regularly the best for ROC, but not for PR. This illustrates an inherent trade-off between models being as “right” as possible versus a model being useful for discovering false-negatives.

Figure 3: Embedded networks pre and post prediction a la that virus paper

4

Predicting networks in space and time

Now that we have a metaweb, we can extend this to predict interactions at particular places and times by decomposing the probability of interaction at particular place and time into probability of interaction multiplied by probability of co-occurrence via properties of conditional probability (Gravel *et al.* 2019).

4.1. Spatial prediction Boosted regression tree using EvoTrees.jl, using data from GBIF.

How do we define? $P(i \leftrightarrow j)$ Is it $P(A_{ij})P(O_{ij})$ or $P(A_{ij})P(O_i)P(O_j)$

4.2. Temporal prediction **Figure 5: Maps over time figure and Prob(Connectance) vs. Month figure**

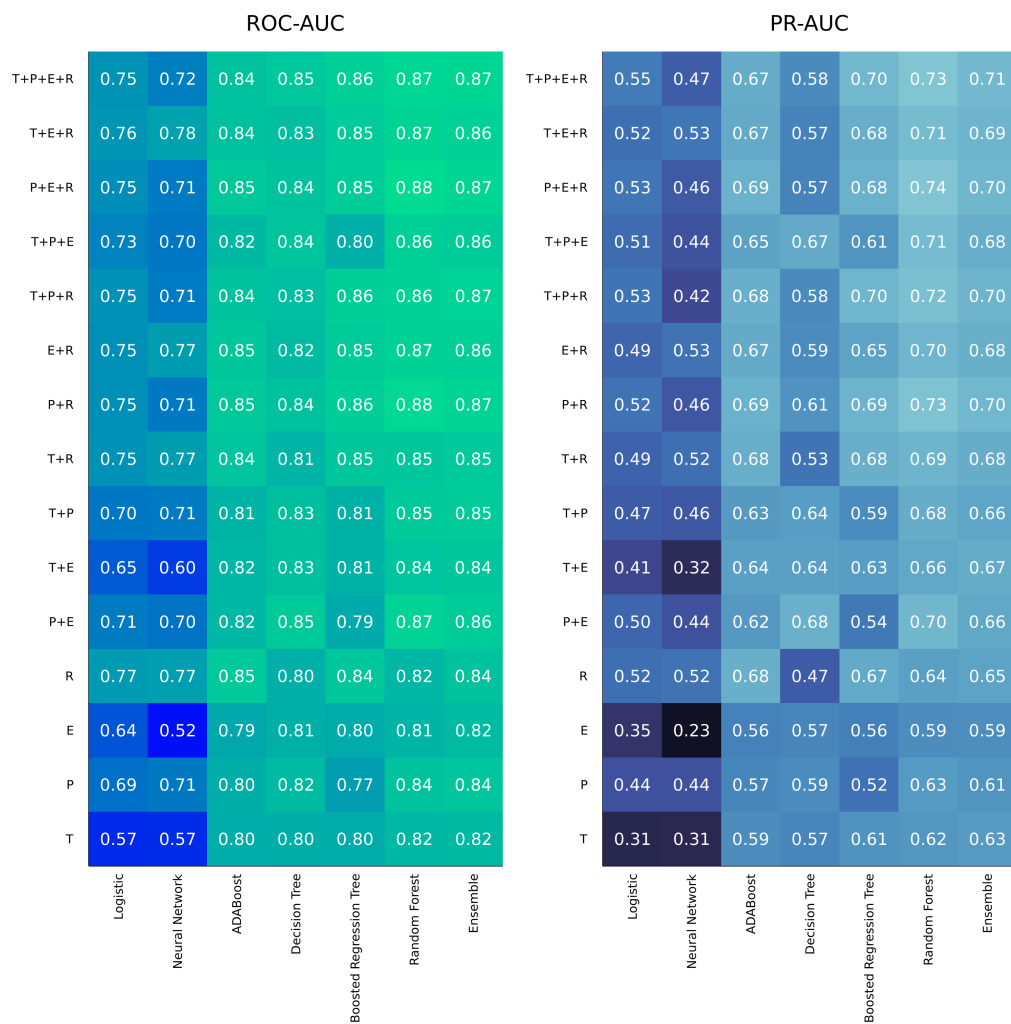
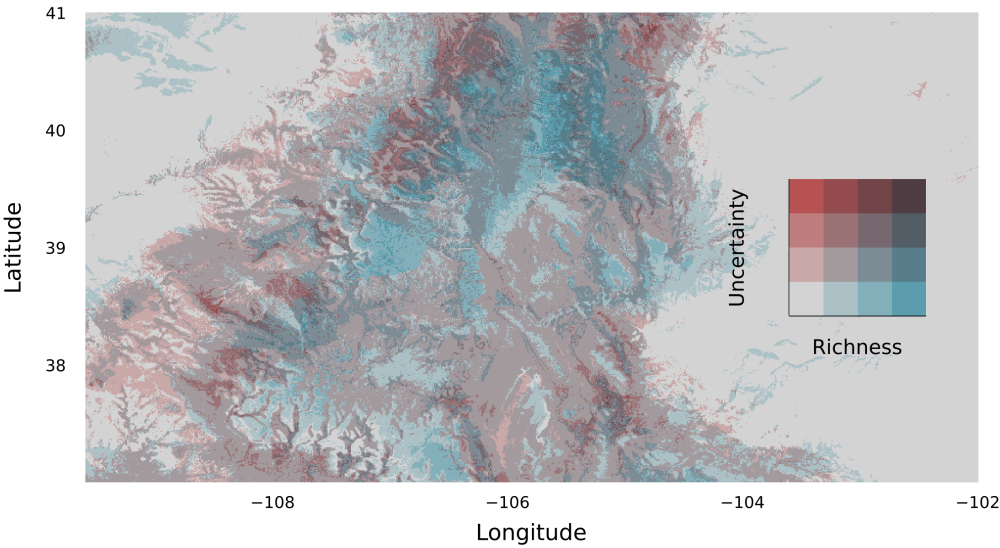


Figure 2 todo

Bombus



Flowers

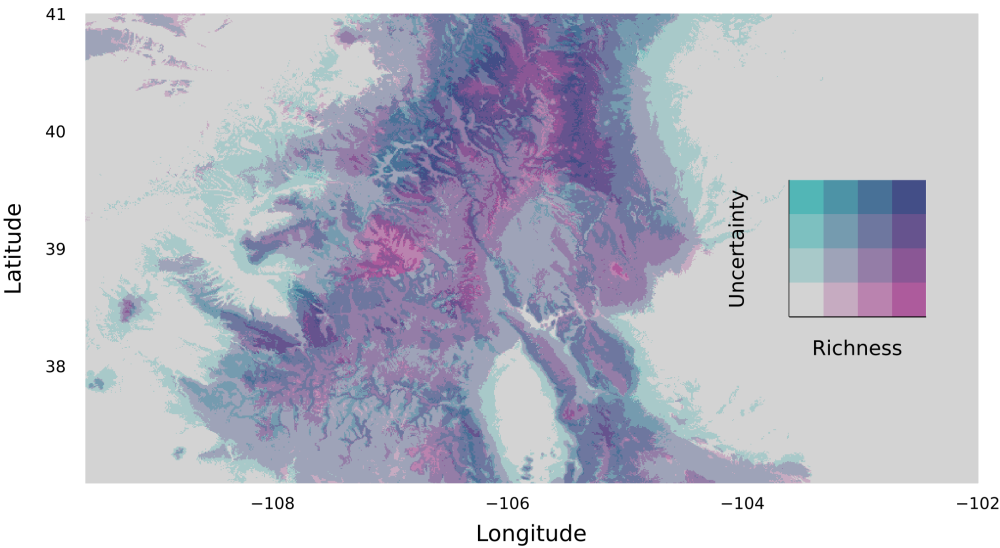


Figure 3 todo

Prioritizing spatial sampling of pollinator interactions

How do we improve our understanding of this pollination network, or determine if it is changing over time?

Figure 4: Uncertainty and sampling priority map

Discussion

We predict things alright. Emphasizes how heterogeneous data sources can improve interaction prediction (rocpr fig). Advances network embedding as a framework for prediction of species interaction networks.

Embedding is a good tool for x. How do we select the dimensionality of our predictors? [Paper on ecological dimensionality]. But this doesn't mean 7-dimensions is adequate to describe our predictors in models to predict species interactions.

Bascompte, J. & Jordano, P. (2007). Plant-Animal Mutualistic Networks: The Architecture of Biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, 38, 567–593.

CaraDonna, P.J., Iler, A.M. & Inouye, D.W. (2014). Shifts in flowering phenology reshape a subalpine plant community. *Proceedings of the National Academy of Sciences*, 111, 4916–4921.

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

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