

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

[Michael D. Catchen](#)^{1,2} [Paul CaraDonna](#)^{3,4} [Jane E. Ogilvie](#)³ [Francis Banville](#)³ [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 Resasco](#)⁷ [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 scripts. The markdown is handled with pandoc.

1 Abstract

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

9 Introduction

10 Ecosystems emerge from sets of interactions between pairs of species. these interactions produce the
11 dynamics and persistence of ecosystems, and the abundance and diversity of the species within them.
12 Plant-pollinator interactions specifically drive the function and persistence of “architecture of biodiversity”
13 (Bascompte & Jordano 2007). However, we are far from a robust understanding of plant-pollinator
14 networks. This is because sampling interactions is costly. Interactions vary in space and time (Poisot *et al.*
15 2015)—particularly relevant in this system (CaraDonna *et al.* 2014). This is why there is interest in using
16 models to predict interactions from sparse data (**Strydom2021?**). In this paper, we combine several
17 datasets, each spanning several years, to produce spatially and temporally explicit predictions of the
18 bumblebee (genus *Bombus*) and wildflower pollination network across the state of Colorado.

19 We do this in two parts: (1) metaweb prediction and (2) conditioning our metaweb prediction on
20 co-occurrence probability. First, we build a model to predict the metaweb—the network of *all*
21 interactions, aggregated across all times and spatial locations—of *Bombus* and wildflower species across
22 Colorado. (Why do this? The metaweb is more predictable than local interactions.) We do this using
23 network embedding (**cite?**). Network embedding takes each node in the network (either a bumblebee or a
24 wildflower) and represents it in a latent n dimensional space. Combination of running models on
25 Temporal niche (T), Phylogenetic niche (P), Environmental niche (E), and relative abundance in
26 community (RA).

27 Second, we then use this metaweb to predict the structure of networks at specific locations and times of

28 year (Gravel *et al.* 2019). Finally we suggest a map of sampling priority, which suggests the locations to
29 sample that will best improve our understanding of the Colorado *Bombus* pollination metaweb.
30 Why is this good for science, what does this contribute to our understanding of plant-pollinator ints,
31 networks, *Bombus*, predictive models, etc., and how can these results be useful.

32 **Data and methodology**

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

34 [Figure 1 about here.]

35 **Predicting the metaweb**

36 **Feature Embedding**

37 **Environmental niche features**

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

42 **Temporal niche features**

43 We take the mean and variance of the distribution of number of observations per week of year in the
44 interaction field data.

45 **Phylogenetic features**

46 **Phylogeny Construction** We construct phylogenies for both *Bombus* and wildflower species using
47 barcode markers, mitochondrial COI and chloroplast *rbcL*, respectively. These sequences were obtained
48 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.

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.

Relative Abundance

This embedding is the simplest.

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

[Figure 2 about here.]

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

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

71 multiplied by probability of co-occurrence via properties of conditional probability (Gravel *et al.* 2019).

72 **Spatial prediction**

73 Boosted regression tree using EvoTrees.jl, using data from GBIF.

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

75 **Temporal prediction**

76 [Figure 3 about here.]

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

78 **Prioritizing spatial sampling of pollinator interactions**

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

81 **Figure 4: Uncertainty and sampling priority map**

82 **Discussion**

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

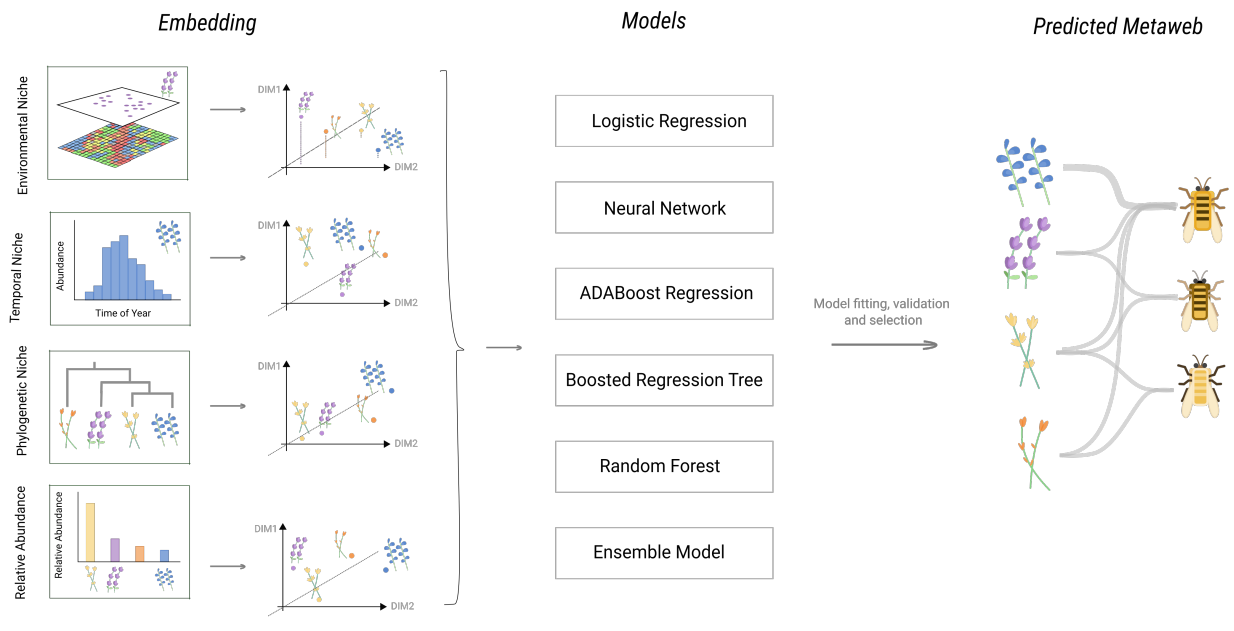
85 Embedding is a good tool for x. How do we select the dimensionality of our predictors? [Paper on
86 ecological dimensionality]. But this doesn't mean 7-dimensions is adequate to describe our predictors in
87 models to predict species interactions.

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

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

- 92 Gravel, D., Baiser, B., Dunne, J.A., Kopelke, J.-P., Martinez, N.D., Nyman, T., *et al.* (2019). Bringing Elton
93 and Grinnell together: A quantitative framework to represent the biogeography of ecological
94 interaction networks. *Ecography*, 42, 401–415.
- 95 Poisot, T., Stouffer, D.B. & Gravel, D. (2015). Beyond species: Why ecological interaction networks vary
96 through space and time. *Oikos*, 124, 243–251.

Metaweb Prediction



Spatiotemporal Network Prediction

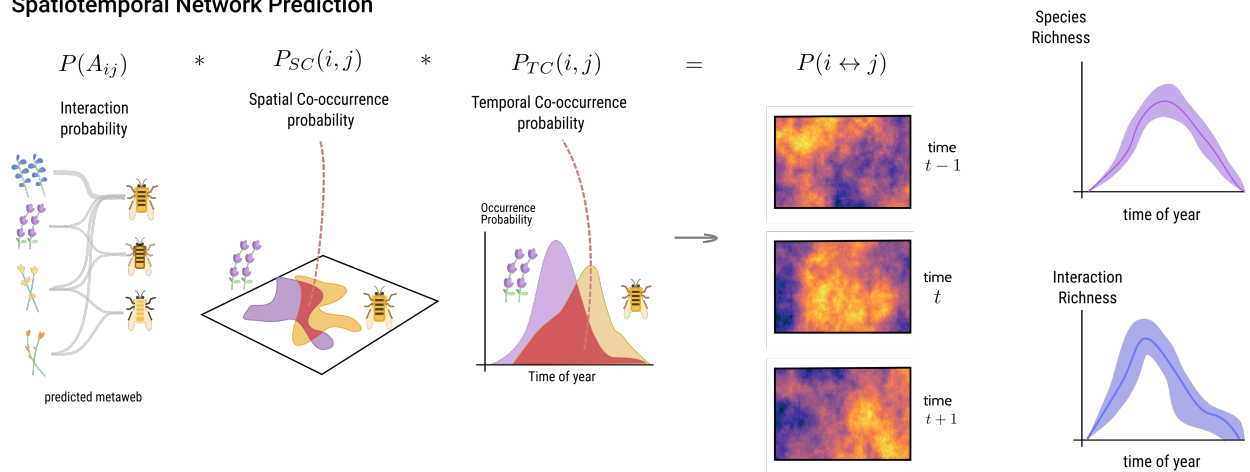


Figure 1: todo

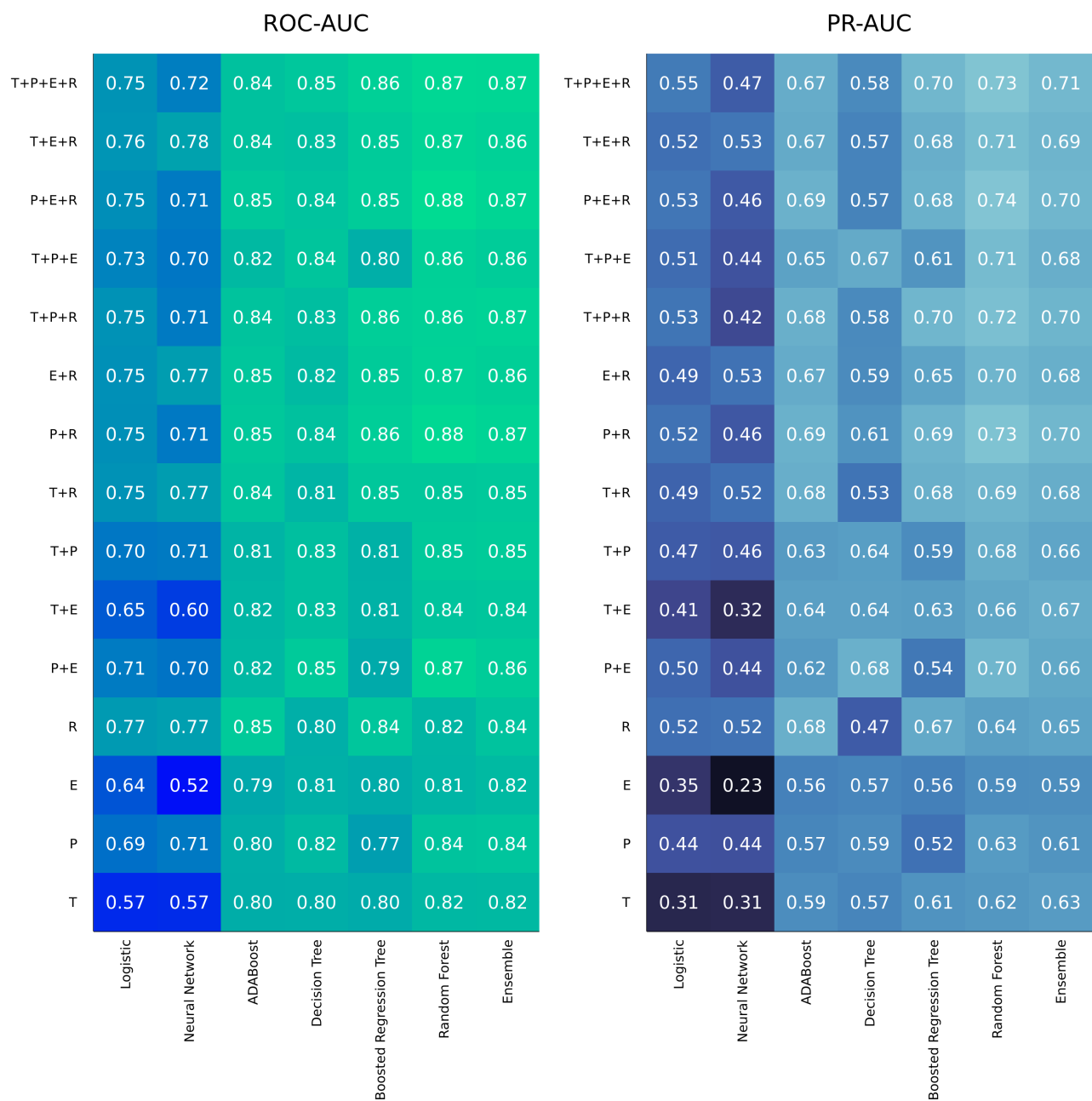
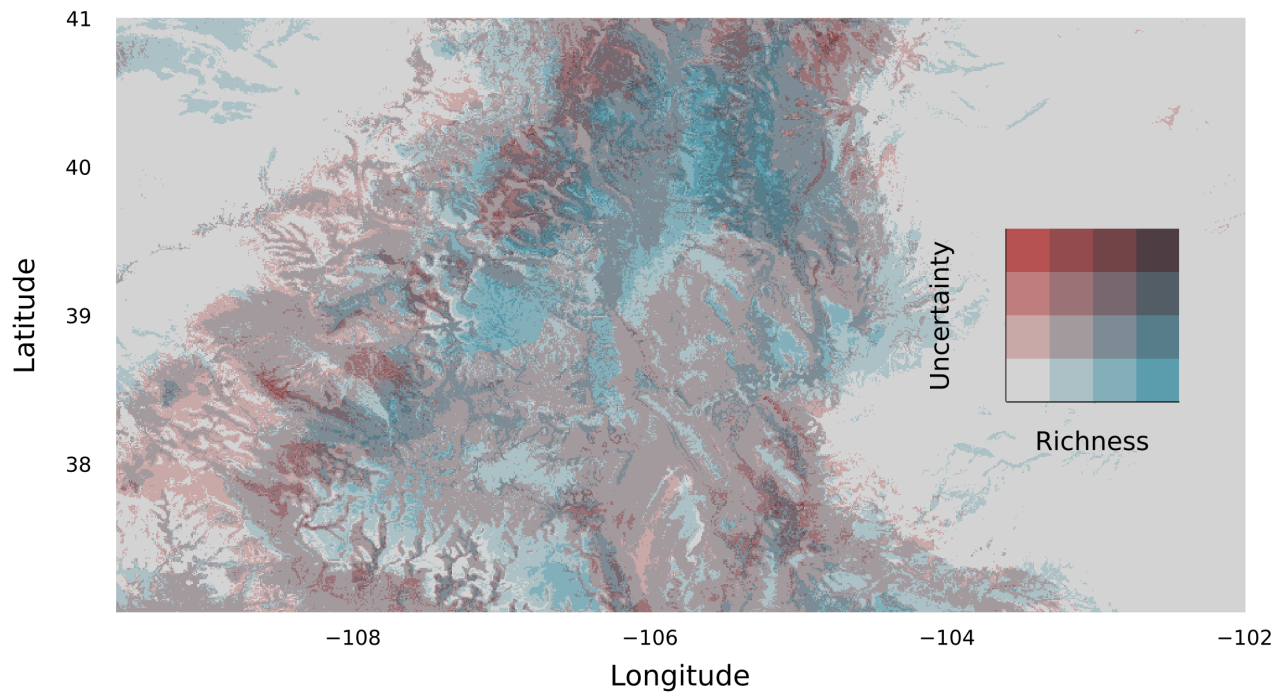


Figure 2: todo

Bombus



Flowers

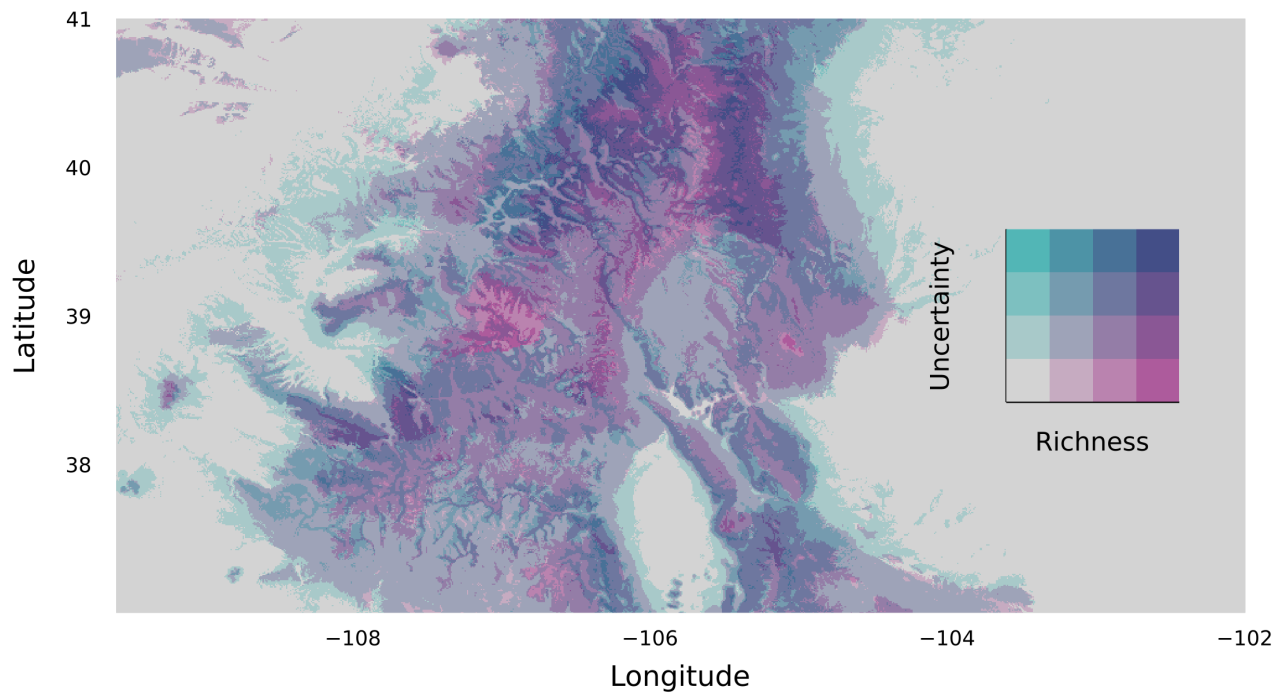


Figure 3: todo