Forecasting the spatio-temporal uncoupling of bumblebee-flower interaction networks

Michael D. Catchen [1,2] Francis Banville [3,4,2] Paul CaraDonna [5,6] Dominique Caron [1,2] Philippe Desjardins-Proulx [3,2] Norma R. Forero-Muñoz [3,2] Andrew Gonzalez [1,2] Dominique Gravel [4,2] Jane E. Ogilvie [5] Laura Pollock [1,2] Timothée Poisot [3,2] Tanya Strydom [3,2] Julian Resasco [7]

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

Correspondance to:

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

Last revision: *October 24, 2021*

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`.

title ideas: Forecasting the spatio-temporal uncoupling of bumblebee-flower interaction networks

Introduction

Species interactions and climate change.

Two dimensions: spatial and temporal.

1) Elevation gradients.

- range shifts in latitude context
- apply this to elevational gradients
- dispersal capacity and range shifts

2) Phenological uncoupling (**Olesen2011MisFor?**).

- Abundance is a function of time in the year

In this paper we.... combine data from a variety of sources: field data from several sites, crowd-sourced data (GBIF), and remotely-sensed data. to produce a *spatially and temporally explict* metaweb of bumblebee-flower interactions across Colorado. We then estimate the change in spatial and temporal overlap over time using the CMIP6 climate consensus forecast (**Karger2017CliHig?**).

Methods

[Figure 1 about here.]

Data

This project involves assembly and integration of data from a variety of (both structured and unstructured) sources. This data can be divided into four categories: field data, GBIF data, remote-sensing data, and phylogenetic data.

Field data

The field data consists of: (1) a seven year data-set from Rocky Mountain Biological Laboratory, consisting of season-long interaction and phenology data six plots along an elevation gradient. (2) a similar six year data set from Elk Meadows, CO, and (3) a year across a large elevation gradient at Pikes Peak.

Additional in-situ environmental sensors.

The partitioning of this data into training, test, and validation sets if described in the *Models* section.

GBIF data

The data from Global Biodiversity Information Facility (GBIF) itself comes in two forms: (1) spatial records of bumblebee and flower records (2) sparsely available records of the plants a bee was observed on (TODO details from Julian).

Remote-sensing data

The remote-sensing data consists of high-resolution LiDAR elevation data at 1 meter spatial resolution [(**NationalEcologicalObservatoryNetworkNEON2021EleLid?**);], and daily 1km resolution precipitation and temperature from CHELSA (**Karger2021GloDai?**).

Phylogenic data

The phylogenetic data consists of genomic barcodes available from NCBI GenBank.

A spatiotemporally explicit predictive metaweb model

What does it mean for it to be "spatiotemporally explicit?" Well the formal definition of a metaweb is total species pool and

We denote the predicted probability of two species, i and j, interacting a p_{ij}. The outcome is here is to build a model f, or rather a set of candidate models, that take i and j and inputs, and which potentially combine this with .features

$$p_{ij} = f(i, j)$$

Candidate models

True Neutral: $f(i, j) = \frac{1}{\sum_i \sum_j 1} = 1/(P \cdot F)$

Relative-abundance (interaction neutral): $f(i, j) = A_i A_j$ where A_x is the relative abundance of species x.

Relative-abundance + environment-embedding: $f(i, j) = g(i, j, E_i, E_j)$

Relative-abundance + phylogeny-embedding: $$

Relative-abundance + environment-embedding + phylogeny-embedding

In gravel et al 2017

$$P(X_{iy}, X_{jy}, L_{ijy}|E_y) = P(X_{iy}, X_{jy}P(L_{ijy}|X_{iy}, X_{jy}, E_y)$$

Then decompose probability of co-occurence as

$$P(X_{iy}, X_{jy}) = P(X_{iy})P(X_{jy})$$

A predictive model to make spatially explicit network prediction

The goal is two have two predictive models: interaction-predictor model and a distribution-predictor model (a la Strydom & Catchen et al. 2021, figure 2).

The interaction-predictor model, $f_i(s_i, s_j, \theta_i)$ predicts interaction based on species-level features (s_i, s_j), and is trained on the field-data.

These features could include Phylogeny (to be determined: how available are genomes or trees for these species) Environment/Climate Traits (to be determined: what trait data is available, how annoying is it to clean) Time (only for the phenology model, see 3.2 and 3.3)

The distribution-predictor model, $f_s(s_i, \vec{x}, t)$ is trained on GBIF data to predict the occurrence of species with features si at a location in space x, and time t. Many options here. Here the species level features could be Climatic variables derived from remote sensing products. Co-occurence to make a JSDM

Potentially weighted by phenology information from field data. Time (only for the phenology model, see 3.2 and 3.3)

Combining distribution-predictor and interaction-predictor models

Can split this into two based on how the distribution-predictor works. If f_s predicts co-occurrence, then draw the species pool first and predict interactions between the species in that pool. If f_s is a single-species SDM, get the occurrence probability for each species p_s and compute the probability of observing interaction as function of the product of occ. prob.

Results

After comparing different combinations of features/model structures and finding the 'best' performing model on validation data.

Figure one: spatial species pool and network prediction

Figure that is two panels: a map of total species richness and a map of network properties across Colorado. This model doesn't consider time, only other predictors.

Figure two: Phenology

Same as figure one but consists of maps but at different times of the year (e.g. March, June, August) and uses both an interaction-predictor and distribution-predictor that incorporate time into predictions

Figure three: Climate

Much as climate change has shifted temperature gradients to get warmer toward the poles, it has also moved temperature gradients up in elevation.

We can get a CMIP6 forecast of temperature and precipitation, and then predict how many observed interactions in the field data will no longer have their composing species' distributions overlap. Decompose temporal component of overlap from spatial component.

Discussion

Acknowledgements

References

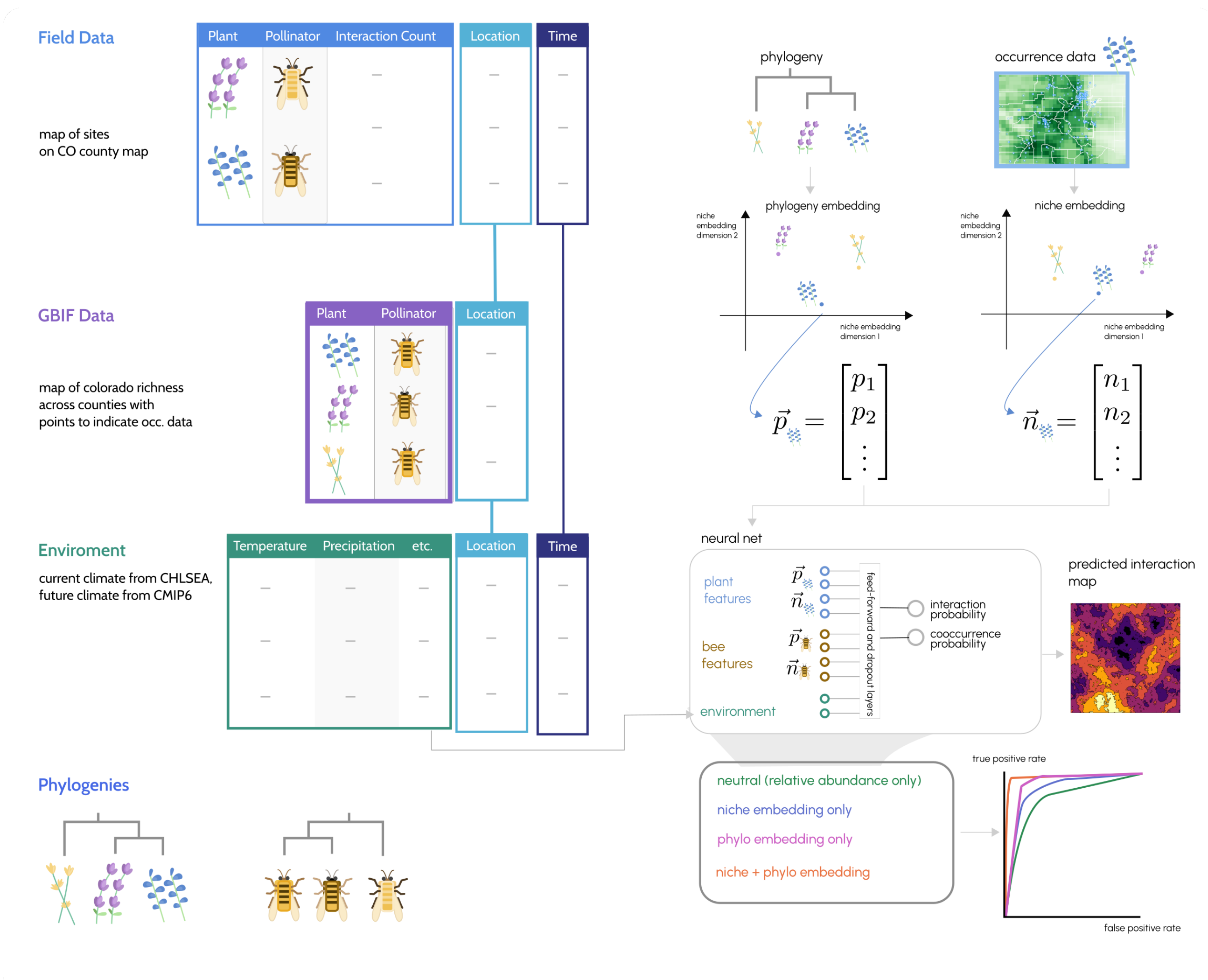

Figure 1: todo