

Template to prepare preprints and manuscripts using markdown and github actions

[Michael D. Catchen](#)^{1,2}

¹ McGill University ² Québec Centre for Biodiversity Sciences

Correspondance to:

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

This work is released by its authors under a CC-BY 4.0 license



Last revision: *October 13, 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 scripts. The markdown is handled with pandoc.

1 Introduction

2 This document is for brainstorming the Colorado bumblebee metaweb prediction project.

3 1 Data

4 The data consists of two parts, the field data collected at three stations across Colorado, and the GBIF
5 metaweb data aggregated and cleaned by Julian R. and a student.

6 1.1 Field data

7 This data consists of nearly-daily records of visits between bee and flower species for the months
8 August-October. Data is collected from three field sites: 1) Rocky Mountain Biological Lab (RMBL) 2)
9 Niwot Ridge 3) Pikes Peak

10 each with tens of plots each across an elevational gradient, collected for 7/4/3 summers total.

11 *To be determined:* specific number of plots per site and years per site

12 1.2 GBIF metaweb

13 This data of a network of bee/flower species which have a record of both occurring on GBIF for locations
14 at some spatial grain

15 *To be determined:* how does the scale of spatial aggregation affect properties of metaweb, how much
16 taxonomic overlap is there with field data?—We have a lot of field data on interactions, but we can't
17 validate predictions on taxa that aren't present in field data. As a result validation can only be done on
18 interactions for which both species appear in field data.

19 Methods

20 [Figure 1 about here.]

21 Where X_{ij} is and L_{abc} is .

22 as in Gravel et al 2017

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

23 Then decompose probability of co-occurrence as

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

24 **A predictive model to make spatially explicit network prediction**

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

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

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

32 The distribution-predictor model, $f_s(s_i, \vec{x}, t)$ is trained on GBIF data to predict the occurrence of species
33 with features s_i at a location in space x , and time t . Many options here. Here the species level features
34 could be Climatic variables derived from remote sensing products. Co-occurrence to make a JSDM
35 Potentially weighted by phenology information from field data. Time (only for the phenology model, see
36 3.2 and 3.3)

37 **Combining distribution-predictor and interaction-predictor models**

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

42 **Results**

43 After comparing different combinations of features/model structures and finding the ‘best’ performing
44 model on validation data.

45 **Figure one: spatial species pool and network prediction**

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

48 **Figure two: Phenology**

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

51 **Figure three: Climate**

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

54 We can get a CMIP6 forecast of temperature and precipitation, and then predict how many observed
55 interactions in the field data will no longer have their composing species’ distributions overlap.

56 Decompose temporal component of overlap from spatial component.

57 **Discussion**

58 **Acknowledgements**

59 **References**

