# Thesis proposal

Michael D. Catchen [1,2]

[1] McGill University [2] Québec Centre for Biodiversity Sciences

**Correspondance to:**

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



Last revision: *November 20, 2021*

The proposal for my thesis, *Simulation models for predictive ecology*

# Introduction

## P1

Within the last several hundred years, human activity has rapidly changed Earth's atmosphere, oceans, and surface. Greenhouse gas emissions have caused an increase the temperature of both Earth's terrestial surface and its oceans, and both agricultural and urban development has rapidly reshaped the cover of Earth's surface. These the bulk of this change has occurred within the last several hundred years, a geological instant, potentially inducing shocks to ecosystems that could threated their integrity (**Scheffer?**). As a result understanding and predicting how ecosystems will change in the future, *ecological forecasting*, and making making descisions based on these predictions mitigating the consequences of this change, on ecosystems has emerged as an imperative for ecology and environmental science [Dietze (2017);].

## P2

However, robust forecasting of ecological processes will change in the future is, to say the least, quite difficult (Beckage *et al.* 2011; Petchey *et al.* 2015). This difficultly is compounded by a few factors, the first being that sampling ecosystems is not easy. Ecological data is often biased, and noisey, spatially and temporally sparse. As a result *ecosystem monitoring* (Makiola *et al.* 2020) has emerged as an imperative. Developing a system for ecological observation, which is able to coordinate across locations. (**AndyUrbanBiomonitoring?** paper).

The second major challenge in forecasting ecosystems is that the underlying dynamics of most ecological processes are fundementally unknown (and unknowable) and instead must be inferred.

Much of the history of quantitatively modeling ecosystems have been done in the language of dynamical systems, describing how the value of an observable state of the system, represented by a vector of numbers $[x_1, x_2, \dots, x_n]^T = \vec{x}$ changes as over time. It turns out to be much more effective to, rather than attempt to directly model $\vec{x}(t)$ itself, to instead describe how $\vec{x}$ changes from one timestep to the next, yielding models in the form of differential equations in continuous-time settings–$\frac{dx}{dt} = f(x)$– or difference equations in discrete-time settings–$x_t = f(x_{t-1})$–where $f : \mathbb{R}^n \to \mathbb{R}^n$ is an arbitrary function describing how the system changes on a moment-to-moment basis (e.g. in the context of communities, $f$ could be Lotka-Voltera, Holling-Type-III or DeAngelis-Beddington functional response). The form of this functional response in real systems is effectively unknown, and some forms are inherently more "forecastable" than others (Chen *et al.* 2019).

**P3**

However, we run into many problems when aiming to apply this type of model to empirical data in ecology.

The initial success of ODE models can be traced back to the larger program of ontological reductionism, which became the de facto apporoach model physical sciences after its early success in physics, which, by the time ecology was becoming a quantitative science (sometime in the 20th century, depending on who you ask), became the foundation for early quantitative models in ecology.

But ecosystems are perhaps the quintessential example of system that cannot be understood simply by iterative reduction of its components. Emergent phenomena, mechanisms at different scales, etc.

Some have been explored in the ecological literature: (1) Some applications of dynamic models in ecology assume long-run equilibrium. (2) Stochasticity

(3) Ecological processes vary across more variables than the tools of analytic models are suited for. As the number of variables in an analytic model increases, so does the ability of the scientist to decern clear relationships between them, and so does overfitting potential. Curse of dimensionality— Until the 20th century, no theory of the gravitational dynamics of more than 2 bodies. Understanding the gravitational dynamics of more than two planets with any reliability proved difficult. Using the same models (diffeqs), how could we adequately predict ecosystems?

**P4**

The term *ecological forecasting* implicitly creates an analogy between predicting how ecosystems will change in the future by using the term "forecasting"—the most immediate analog being the success story of weather forecasting via numerical weather prediction (NWP).

Although it is become almost hack to complain about the dang weather forecast being wrong, over the least 50 years the (Bauer *et al.* 2015).

The success of NWP, and the Earth observations that support it should serve as a template for development of a system for monitoring Earth's biodiversity. Much like ecology, NWP is faced with high-dimensional systems that are governed by different mechanisms at different scales.

NWP has worked because it incorporates information about data and meteorological processes collected at difference scales into models that. Use of computational methods in NWP.

Much as one would not aim to forecast the weather in Quebec by applying Navier-Stokes, forecasting ecological systems must

Transition to simulation as the solution: shift toward approach of building models that *generate* data.

(resolving the semantic ambuity of what differentiates "mechanistic" vs "phenomological" models is out of scope for now).

More broadly a reflection reflect ecology lagging behind the statistical methods used in sciences that face similar challenges (many dimensions, many mechanisms at different scales, each with stochasticity). Chaotic dynamics emerge from simple analytic models, and . Whether ecosystems actually exhibit chaotic behavior is a different question.

**P5**

But forecasting isn't the only difficult problem here.

Transition to theme of optimization given unknown information. A forecast gives us a range of future values with uncertainty around them. Further a convenient property that a forecasting model's uncertainty goes up over time (if we assume the underlying process is Markov–this is a strong assumption but oft true of the models we fit to temporal data)

In face of uncertainty, decision making is an optimization problem. We have some goal state for the future, and some estimate of what the state of the world will be given a set of actions. Frame optimization problem mathematically an introduce concept of solution-space and constraint.

Indeed Marx's most well known quote that "philosophers have hitherto only interpreted the world in various ways; the point is to change it." and a necessary step toward establishing a just and sustainable world.

[Figure 1 about here.]

**P6 – final intro para**

Three major components here: 1) Ecosystem monitoring, 2) Forecasting using the products of that monitoring, and 3) Choosing the best possible mitigation strategy.

This flow is outlined in the left panel of fig. 1

# Chapter One: Forecasting the spatial uncoupling of a plant-pollinator network

Plants and pollinators form interaction networks, called the "architecture of biodiversity" (**Jordano2007?**).

The stability, function, and persistance of ecosystems relies on the structure of these interactions. Antropogenic change threatens to unravel these networks. Two aspects to this change: spatial and temporal. Spatially, range shifts along elevational gradient, and temporall, phenological shifts.

The issue is that we don't really know what interactions are like now. So not only do we need to predict with data that is spatially and temporally sparse and likely to contain many interaction "false-negatives" (**Strydom2021?**)

This chapter uses several years of data on bee-flower phenology and interactions, combined with spatial records of species occurrence via GBIF, to forecast how much overlap there will be between plants/pollinators in space/time.

In stages, (1) take data from multiple sites to predict a spatial metaweb of *Bombus*-flower interactions across Colorado. (2) Predict how these spatial distributions will change under CMIP6. and (3) quantify the lack of overlap between species for which there is a predicted

[Figure 2 about here.]

## Data

The data for this chapter is derived from multiple souces and can be split into three categories. (1) Field data from three different locations acvross Colorado. All field sites have multiple plots across an elevational gradient.

System description: lots of data on *Bombus* (bumblebees) and wildflowers. Three different sites, (7/7/3) years each, each covering an elevational gradient.

## Methods

Split the process into parts.

1) Building an interaction prediction model. 2) Make it spatial based on distributions. 3) Forecast distributions based on CMIP6.

## Preliminary Results

1) we got a tree

Transition to next chapter by discussing uncertainty in interaction prediction across space.

# Chapter Two: Optimizing spatial sampling of species interactions

There are false-negatives in interation data. Co-occurrence is not the same thing as interaction (**cite?**), but often is used as a proxy.

This chapter unravels the relationship between a given species relative abundance and the sampling effort needed to adequately understand this species distribution and interactions.

There is more than one way to observe a false-negative.

[Figure 3 about here.]

It begins with a conceptual framework for understanding the difference in false-negatives in occurrence, co-occurrence, and interactions (fig. 3). We use a null model of the relative-abundance distribution (Hubbell 2001) to simulate realized false-negatives as a function of varying sampling effort.

This also goes on to includes testing some assumptions of the model with empirical data fig. 4. which indicate our neutral model, if anything, underestimates the probability of false-negatives due to positive correlations in co-occurrence in two spatially replicated networks (Thompson & Townsend 2000; Hadfield *et al.* 2014)—further I'm planning to add the field data from chapter one into this anlysis once complete.

[Figure 4 about here.]

new addition: - simulate species distribution and efficacy of detection given a set of observation points where the dist from observation site decays. optimize set of repeated sampling locations L for a *known* distribution D. address SDM not being the territory

### Results

- nonrandom association figure sampling effort under neutral model

## Chapter Three: Optimizing corridor placement against ecological dynamics

Promoting landscape connectivity is important to mitigate the effects of land-use change on Earth's biodiversity. However, the practical realities of conservation mean that there is a limitation on how much we can modify landscapes in order to do this. So what is the best place to put a corridor given a constraint on how much surface-area you can change in a landscape? This is the question this chapter seeks to answer. Models for proposing corridor locations have been developed, but are limited in that are not developed around promoting some element of ecosystem function, but instead by trying to find the path of least resistance given a resistance surface (Peterman 2018).

This chapter proposes a general algorithm for optimizing corridor placement based on a measurement of ecosystem functioning derived from simulations run on a proposed landscape modification. We propose various landscape modifications which alter the cover of a landscape, represented as a raster (fig. 6, left). We then compute a new resistance surface based on the proposed landscape modification, and based on the values of resistance to dispersal between each location we simulate spatially-explicit metapopulation dynamics model (Hanski & Ovaskainen 2000; Ovaskainen *et al.* 2002) to estimate a distribution of time until extinction for each landscape modification (fig. 6, right).

### Methods

- land cover -> resistance -> extinction time simulated annealing to
- optimize landscape optimization

## CH4 a software note on the resulting packages.

(MetacommunityDynamics.jl: a virtual laboratory for community ecology): a collection of modules in the Julia language for different aspects of metacommunity ecology, including most of the code used for the preceding chapters.

[Figure 5 about here.]

## Conclusion

## Appendix

[Figure 6 about here.]

## References

Bauer, P., Thorpe, A. & Brunet, G. (2015). The quiet revolution of numerical weather prediction. *Nature*, 525, 47–56.

Beckage, B., Gross, L.J. & Kauffman, S. (2011). The limits to prediction in ecological systems. *Ecosphere*, 2, art125.

Chen, Y., Angulo, M.T. & Liu, Y.-Y. (2019). Revealing Complex Ecological Dynamics via Symbolic Regression. *BioEssays*, 41, 1900069.

Dietze, M.C. (2017). Prediction in ecology: A first-principles framework. *Ecological Applications*, 27, 2048–2060.

Hadfield, J.D., Krasnov, B.R., Poulin, R. & Nakagawa, S. (2014). A Tale of Two Phylogenies: Comparative Analyses of Ecological Interactions. *The American Naturalist*, 183, 174–187.

Hanski, I. & Ovaskainen, O. (2000). The metapopulation capacity of a fragmented landscape. *Nature*, 404, 755–758.

Hubbell, S.P. (2001). *The unified neutral theory of biodiversity and biogeography*. Monographs in population biology. Princeton University Press, Princeton.

Makiola, A., Compson, Z.G., Baird, D.J., Barnes, M.A., Boerlijst, S.P., Bouchez, A., *et al.* (2020). Key Questions for Next-Generation Biomonitoring. *Frontiers in Environmental Science*, 7.

Ovaskainen, O., Sato, K., Bascompte, J. & Hanski, I. (2002). Metapopulation Models for Extinction Threshold in Spatially Correlated Landscapes. *Journal of Theoretical Biology*, 215, 95–108.

Ovaskainen, O., Sato, K., Bascompte, J. & Hanski, I. (2002). Metapopulation Models for Extinction Threshold in Spatially Correlated Landscapes. *Journal of Theoretical Biology*, 215, 95–108.

Petchey, O.L., Pontarp, M., Massie, T.M., Kéfi, S., Ozgul, A., Weilenmann, M., *et al.* (2015). The ecological forecast horizon, and examples of its uses and determinants. *Ecology Letters*, 18, 597–611.

Peterman, W.E. (2018). ResistanceGA: An R package for the optimization of resistance surfaces using genetic algorithms. *Methods in Ecology and Evolution*, 9, 1638–1647.

Thompson, R.M. & Townsend, C.R. (2000). Is resolution the solution?: The effect of taxonomic resolution on the calculated properties of three stream food webs. *Freshwater Biology*, 44, 413–422.

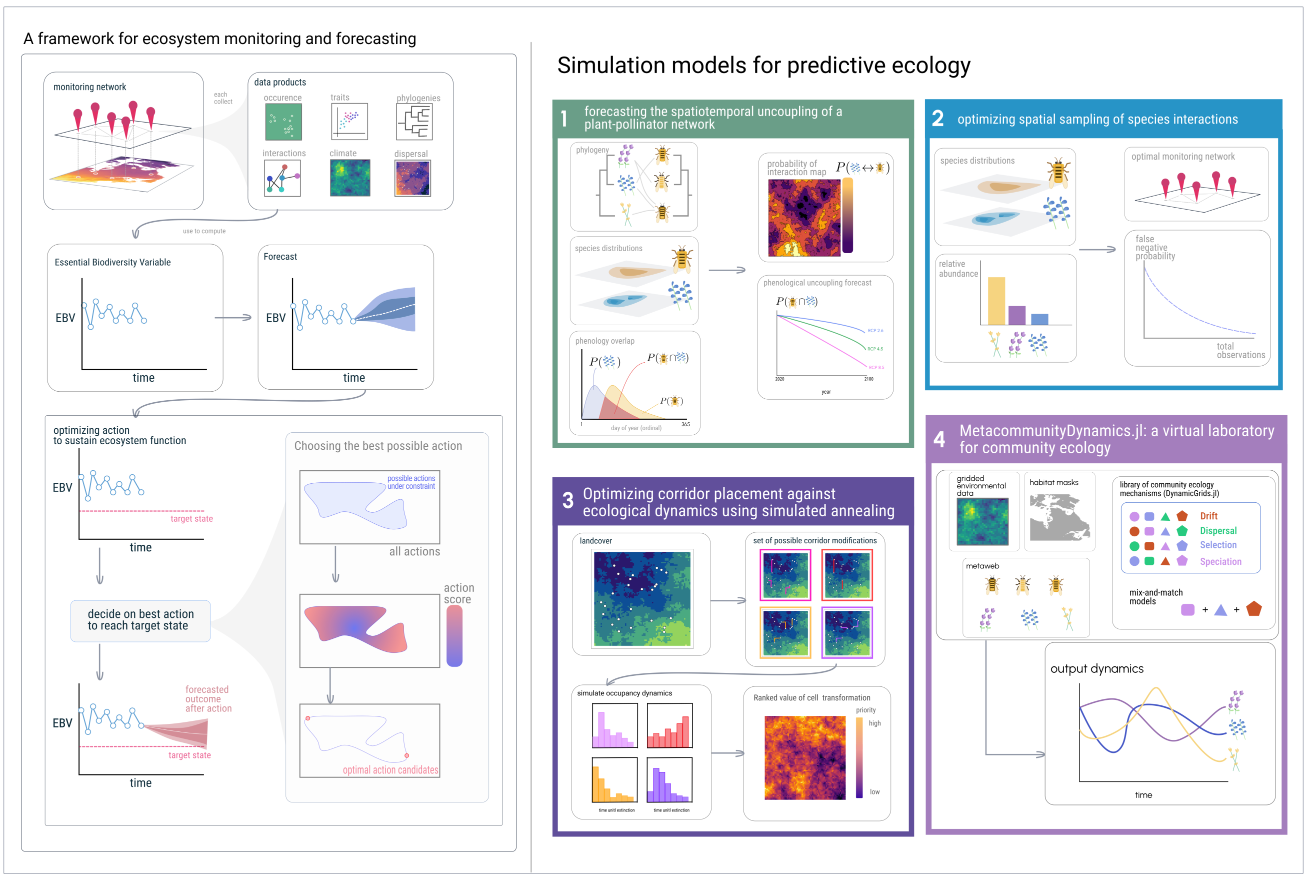


Figure 1: thesis concept

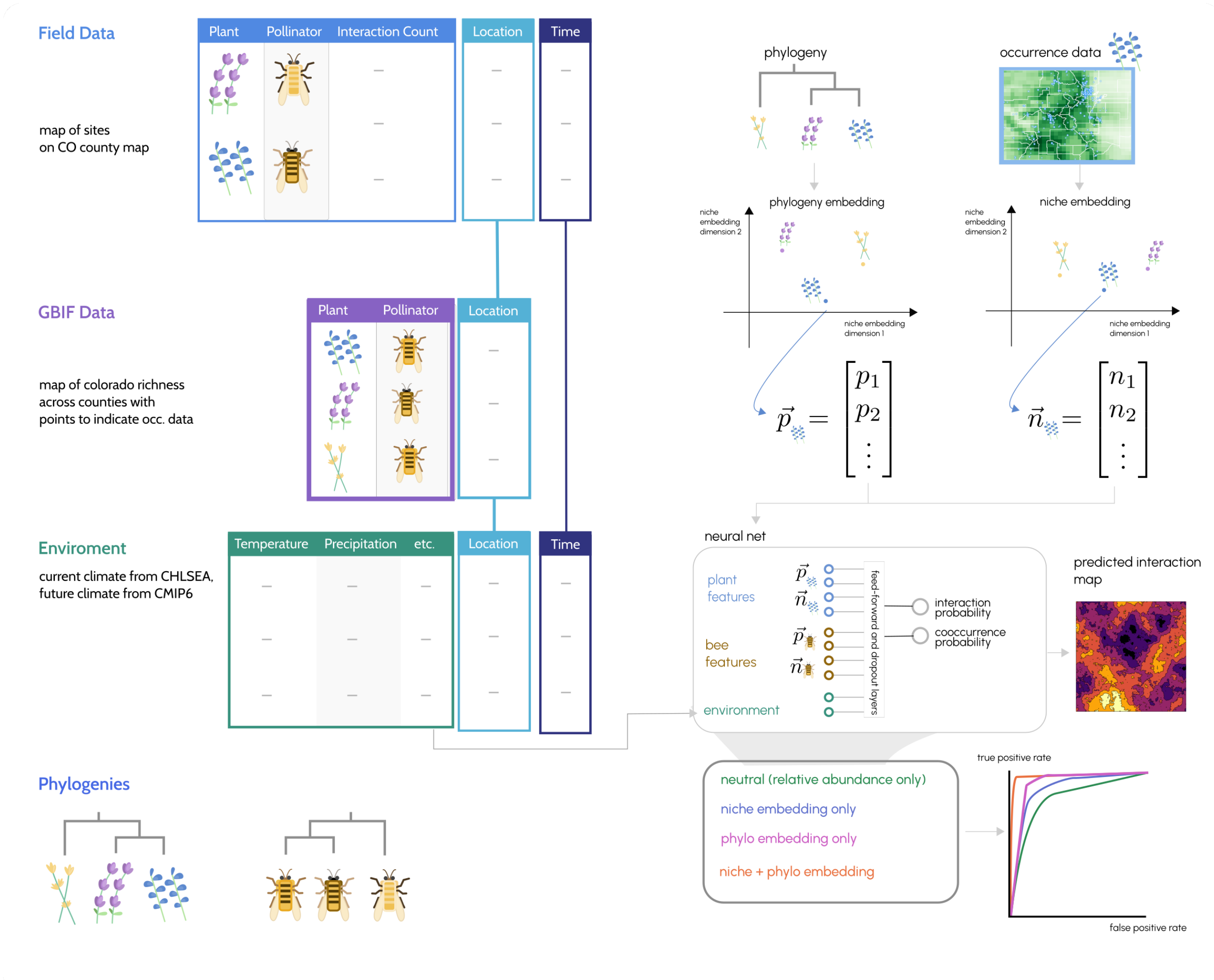


Figure 2: chapter one concept fig

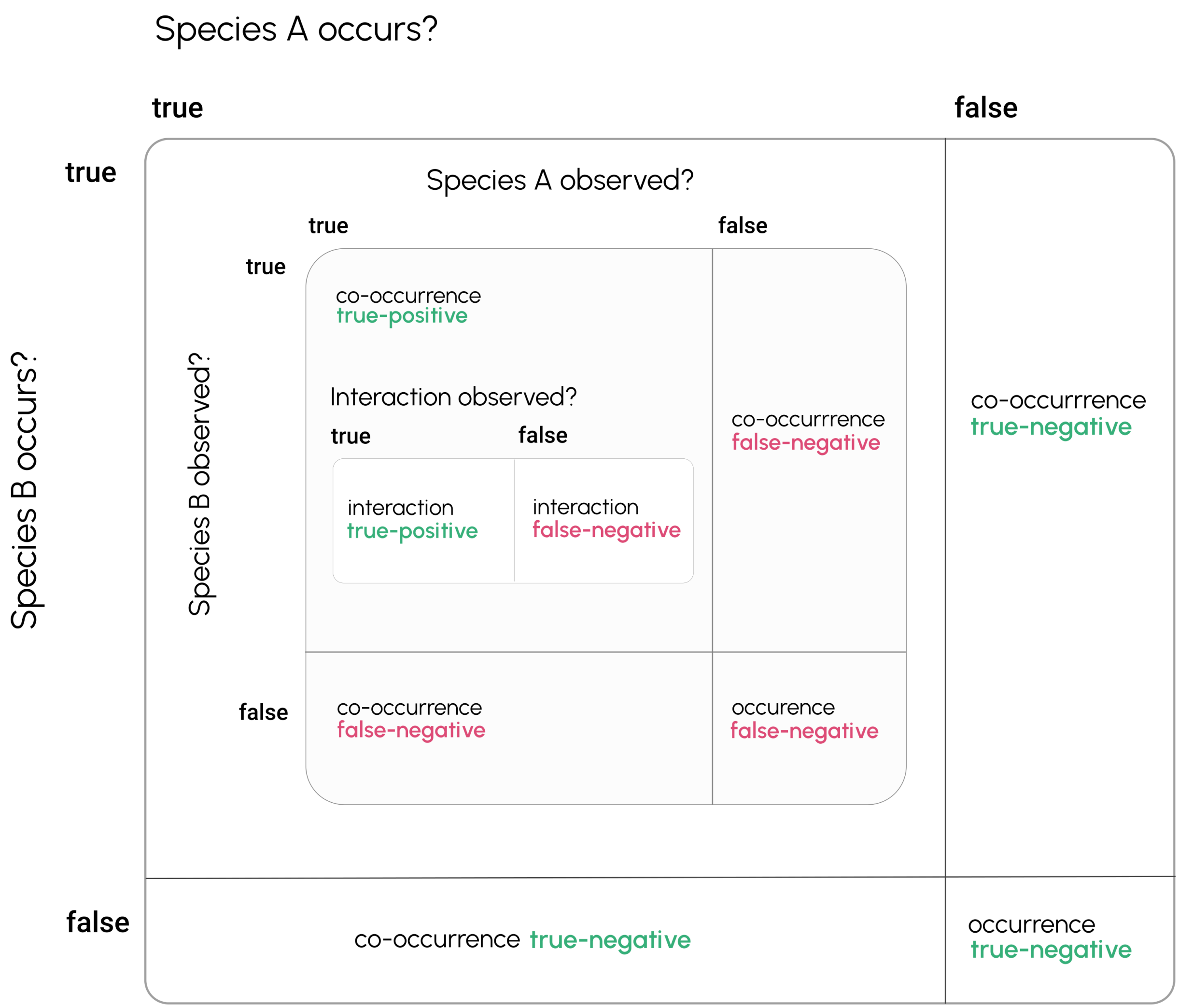


Figure 3: taxonomy of false negatives

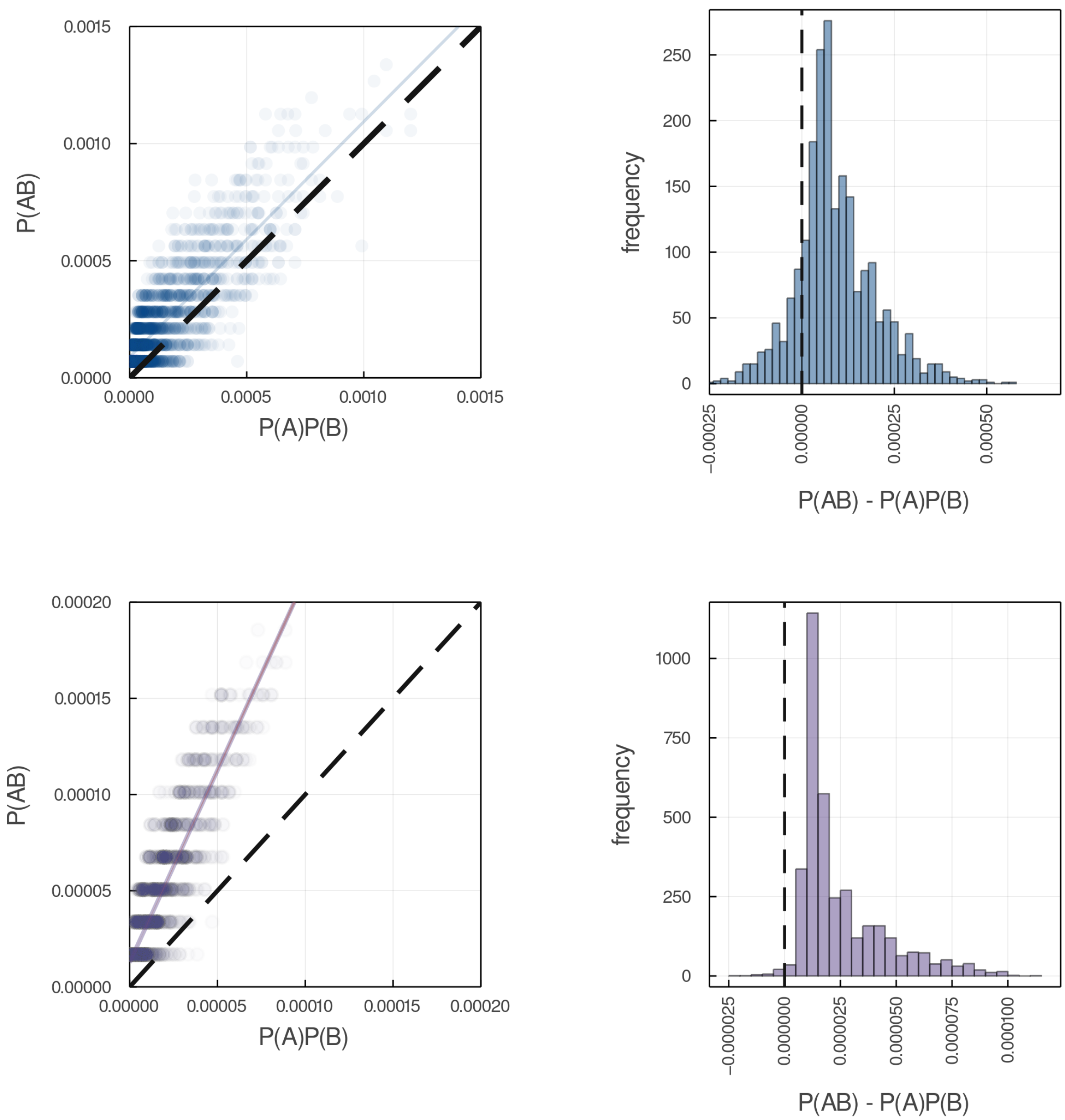
P(AB)
P(A)P(B)
0.0000
0.0005
0.0010
0.0015
frequency
P(AB) - P(A)P(B)
0
50
100
150
200
250
−0.00025
0.00000
0.00025
0.00050
0.00000
0.00005
0.00010
0.00015
0.00020
500
750
1000
−0.000025
0.000000
0.000025
0.000050
0.000075
0.000100

Figure 4: f

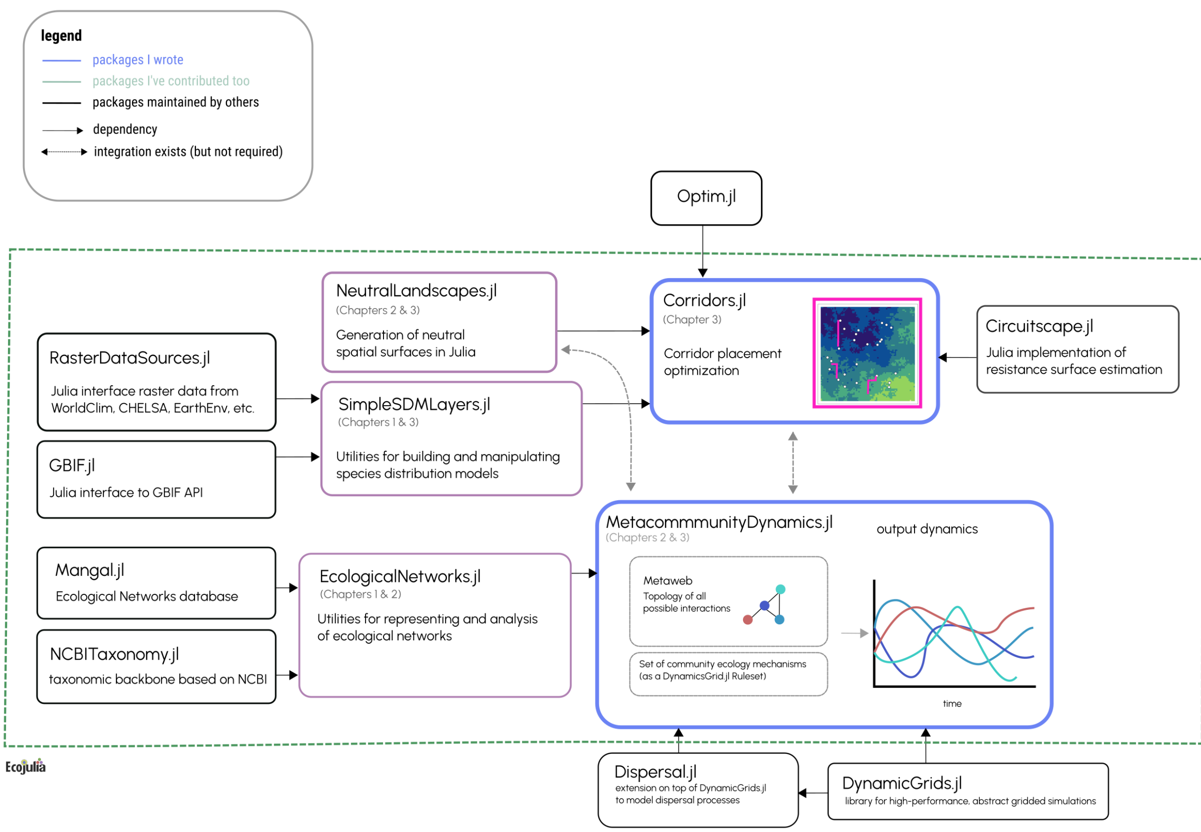


Figure 5: todo

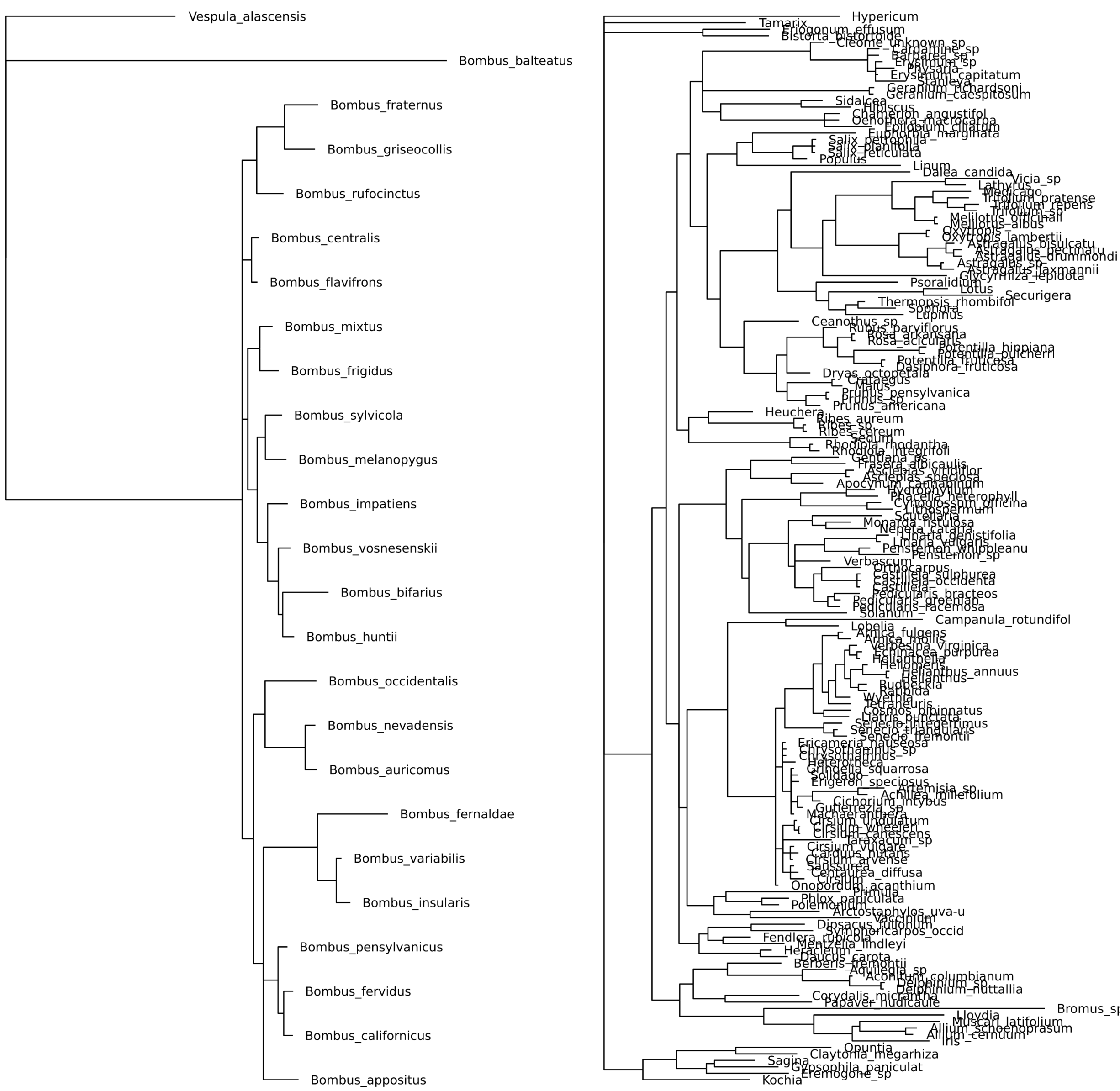


Figure 6: trees