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 21, 2021*

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

Introduction

P1

Within the last several hundred years, human activity has induced rapid changes in Earth's atmosphere, oceans, and surface. Greenhouse gas emissions have caused an increase the temperature of both Earth's terrain and oceans, and both agricultural and urban development has rapidly reshaped the Earth's land cover. These the bulk of this change has occurred within the last several hundred years, a geological instant, inducing a sudden shift in conditions to Earth's climate and biosphere. As a result, predicting how ecosystems will change in the future, *ecological forecasting*, and then using these forecasts to make decisions to mitigate the negative consequences of this change on ecosystems, their functioning, and the services they provide to humans has emerged as an imperative for ecology and environmental science (Dietze 2017). However, robust prediction of ecological processes 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, noisey, and sparse in both space and time. The current paucity of ecological data has resulted in much interest in developing global systems for *ecosystem monitoring* (Makiola *et al.* 2020), which would systematize the collection of biodiversity data in manner that makes detecting and predicting change more possible than at the moment (Urban *et al.* 2021).

P2

The second major challenge in ecological forecasting is that the underlying dynamics of most ecological processes are unknown and instead must be inferred from this (sparse) data. 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, 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 initial success of these forms of models can be traced back to the larger program of ontological reductionism, which became the default approach to modeling in the 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.

P3

However, we run into many problems when aiming to apply this type of model to empirical data in ecology. Ecosystems are perhaps the quintessential example of system that cannot be understood by iterative reduction of its components into constituent parts—ecological phenomena are emergent are the product of different mechanisms operating a different spatial, temporal, and organizational scales (Levin 1992). Further, the form of this functional response in real systems is effectively unknown, and some forms are inherently more "forecastable" than others (Beckage *et al.* 2011; Chen *et al.* 2019; Pennekamp *et al.* 2019). Further this analytical approach to modeling explicitly ignores known realities: ecological dynamics not deterministic, many analytic models in ecology assume long-run equilibrium. Finally, perhaps the biggest challenge in using these models to describe ecological processes is ecosystems 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 discern clear relationships between them given a fixed amount of data, the so-called "curse" of dimensionality.

P4

But these problems are not solely unique to ecology. The term *ecological forecasting* implicitly creates an analogy with weather forecasting. Although it has become a trite joke to complain about the weather forecast being wrong, over the least 50 years the field of numerical weather prediction (NWP) has dramatically improved out ability to predict weather across the board (Bauer *et al.* 2015). The success of NWP, and the Earth observations systems that support it (Hill *et al.* 2004), 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. The success of NWP is that, rather than, say, attempt to forecast the weather in Quebec by applying Navier-Stokes to entire province, to instead use simulation models which describe known mechanisms at different scales, and use the availability to increasing computational power to directly simulate many batches of dynamics which directly incorporate stochasticity and uncertainty in parameter estimates via random number generation.

P6

But forecasting is only half the story—if indeed "[ecologists] have hitherto only interpreted the world in various ways; the point is to change it," then once we have a forecast about how an ecosystem will change in the future, what if this forecast predicts a critical ecosystem service will deteriorate? We are still left with the question, what do we in the time being to mitigate the negative consequences a forecast predicts? In this framing, mitigating the consequences of anthropogenic change on ecosystems becomes an optimization problem: given a forecast of the probability. 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.

[Figure 1 about here.]

P7

This dissertation aims to formalize a framework for ecosystem monitoring and forecasting (fig. 1, left), and each chapter address some aspect of this pipeline to data from a monitoring network to forecasts to mitigation strategy (fig. 1, right).

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

Interactions between plants and pollinators form networks of interactions, which structure the "architecture of biodiversity" (**Jordano2007?**). The functioning and stability of ecosystems emerge from these interactions, but antropogenic change threatens to unravel and "rewire" these networks (**CaraDonna2017IntRew?**), threating the persistence of these systems. Plant-pollinator networks face two possible forms of rewiring in response to anthropogenic environmental change: spatial and temporal. Spatially, range shifts could cause interacting species to no longer overlap in space, and shifts in phenology could cause interacting species to no longer overlap in time.

This chapter uses several years of data on bumblebee-flower phenology and interactions across several field sites, each consisting of several plots across an elevational gradient, combined with spatial records of species occurrence via GBIF to forecast this uncoupling. This addresses the EBV to forecast of EBV element of the flow from data to mitigation in fig. 1 (left).

[Figure 2 about here.]

Data

The data for this chapter is derived from multiple souces and can be split into four categories. (1) Field data from three different field sites across Colorado, each with multiple plots across an elevational gradient, for seven, seven, and three years resptively. This data was collected by Paul CaraDonna and Jane Oglevie (from the Rocky Mountain Biological Laboratory) and Julian Resasco (CU Boulder). (2) GBIF spatial occurrence records of each of these species across Colorado, including a metaweb of interactions across all of Colroado taken from GBIF. (3) Remotely sensed data consisting of current and forecasting bioclimatic variables from CHELSA. (4) Phylogeny derived from NCBI GenBank barcodes for mitochondrial COI (bumblebees) and chloroplast rbcL (flowers).

Methods

As the data we have is spatially sparse and likely to contain many interaction "false-negatives" (Strydom *et al.* 2021), we beign by predicting a metaweb of interactions as they exist *in the present*.

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

The process of going from data to forecast can be split into the following parts

1) Building an interaction prediction model (or rather a set of candidate models, relative abundance, phylo embedding relative abundance + phylo embedding) a la Strydom *et al.* (2021)

Reconstructing latent features for each species based on simulating trait evolution on a phylogeny (**Strydom2021FooWeb?**).

2) Make it spatial based on distributions.

3) Forecast distributions based on CMIP6

Preliminary Results

1) we got a tree and SDMs

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

Chapter Two: Optimizing spatial sampling of species interactions

This chapter uses simulation models to investigate the relationship between species relative abundance, sampling effort, and probability of accurately detecting an interaction between species, and further proposes a method for optimizing the spatial sampling locations to maximize the probability of detecting an interaction between two species given their distributions. This addresses the optimization of monitoring network part of the flow from data to mitigation in fig. 1.

As explored in the previous chapter, there are false-negatives in interation data. There is more than one way to observe a false-negative when sampling interactions: (fig. 3). It begins with a conceptual framework for understanding the difference in false-negatives in occurrence, co-occurrence, and interactions. Co-occurrence is not the same thing as interaction (Blanchet *et al.* 2020), but often is used as a proxy.

[Figure 3 about here.]

We use a log-normal distribution as a null model of the relative-abundance distribution (Hubbell 2001) to simulate realized false-negative rate 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 we analytically show that 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 the previous chapter into this anlysis once available.

[Figure 4 about here.]

Finally this chapter proposes a simulated annealing method to optimize the efficacy of interactoin detection given a set of observation points where the dist from observation site decays. optimize set of repeated sampling locations L for a pair of species *known* distributions D_a, D_b.

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 choosing corridor placement to optimize a measurement of ecosystem functioning derived from simulations run on each proposed landscape modification.

[Figure 5 about here.]

Methods

We propose various landscape modifications which alter the cover of a landscape, represented as a raster. 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.

- brief overview of simulated annealing
- describe how you build the proposal function
- optimize landscape optimization

Chapter Four: MetacommunityDynamics.jl: a virtual laboratory for community ecology

This chapter consists of a collection of modules in the Julia language for different aspects of community ecology, including most of the code used for the preceding chapters. Indeed `MetacommunityDynamics.jl` (MCD.jl) is the epicenter of this set of tools, but due to the nature of the Julia language, MCD.jl is interoperable with serveral existing packages within the `EcoJulia` organization, including several to which I have contributed. A diagram showing the relation between these packages is shown in fig. 6.

[Figure 6 about here.]

Conclusion

Appendix

[Figure 7 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.

Blanchet, F.G., Cazelles, K. & Gravel, D. (2020). Co-occurrence is not evidence of ecological interactions. *Ecology Letters*, 23, 1050–1063.

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.

Hill, C., DeLuca, C., Balaji, Suarez, M. & Da Silva, A. (2004). The architecture of the Earth System Modeling Framework. *Computing in Science Engineering*, 6, 18–28.

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

Levin, S.A. (1992). The Problem of Pattern and Scale in Ecology: The Robert H. MacArthur Award Lecture. *Ecology*, 73, 1943–1967.

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.

Pennekamp, F., Iles, A.C., Garland, J., Brennan, G., Brose, U., Gaedke, U., *et al.* (2019). The intrinsic predictability of ecological time series and its potential to guide forecasting. *Ecological Monographs*, 89, e01359.

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.

Strydom, T., Catchen, M.D., Banville, F., Caron, D., Dansereau, G., Desjardins-Proulx, P., *et al.* (2021). *A Roadmap Toward Predicting Species Interaction Networks (Across Space and Time)* (Preprint). EcoEvoRxiv.

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.

Urban, M.C., Travis, J.M.J., Zurell, D., Thompson, P.L., Synes, N.W., Scarpa, A., *et al.* (2021). Coding for Life: Designing a Platform for Projecting and Protecting Global Biodiversity. *BioScience*.

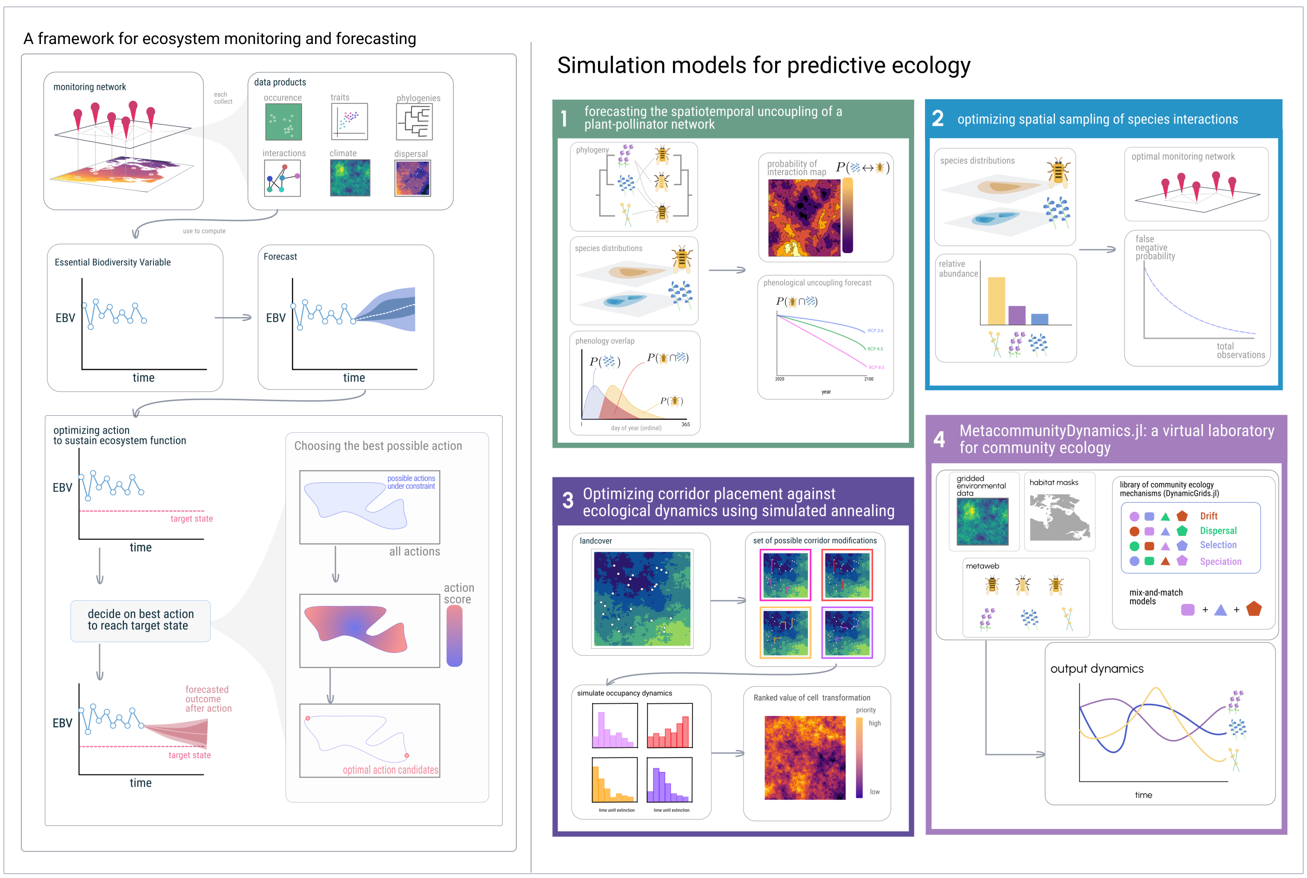

Figure 1: thesis concept

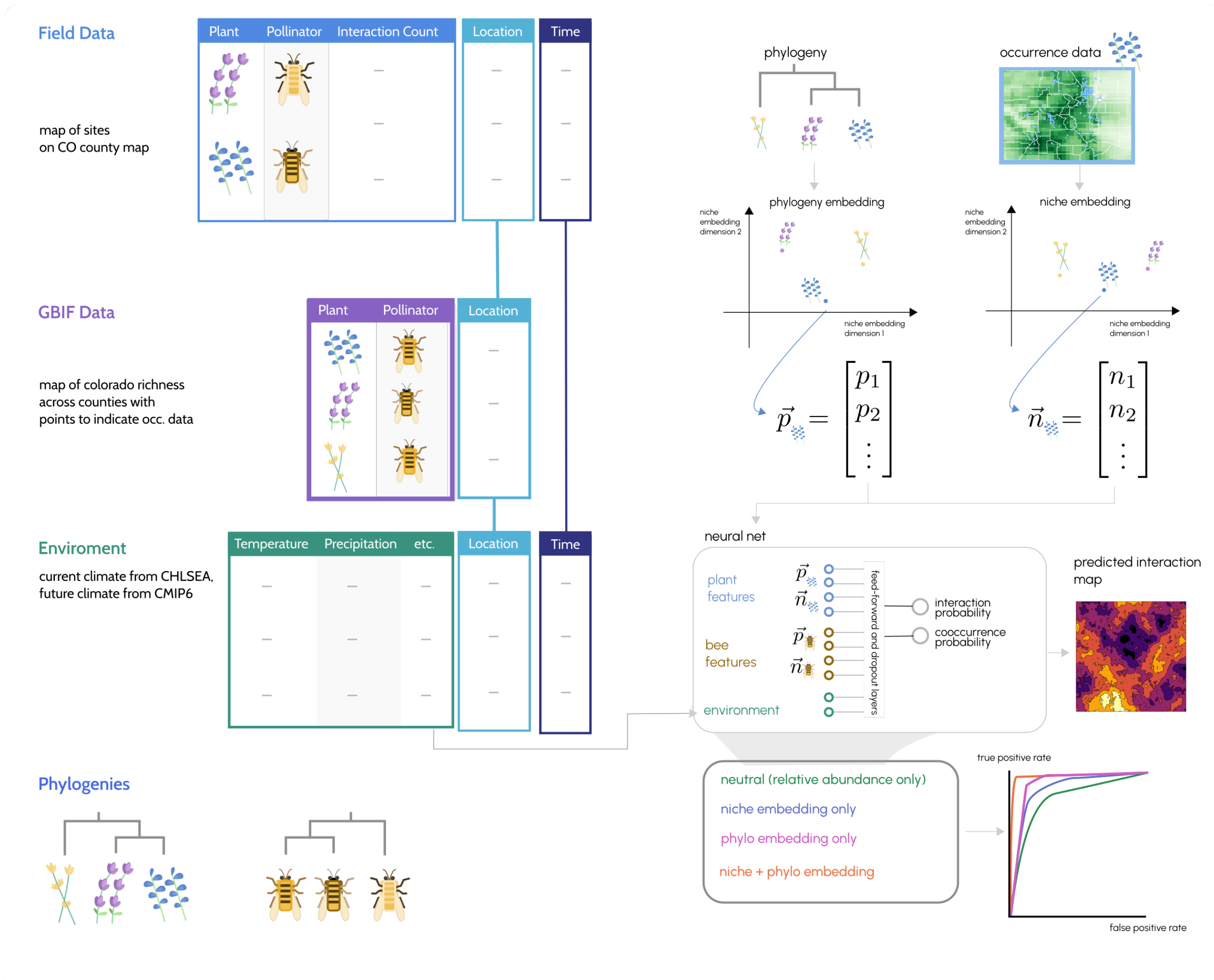

Figure 2: chapter one concept fig

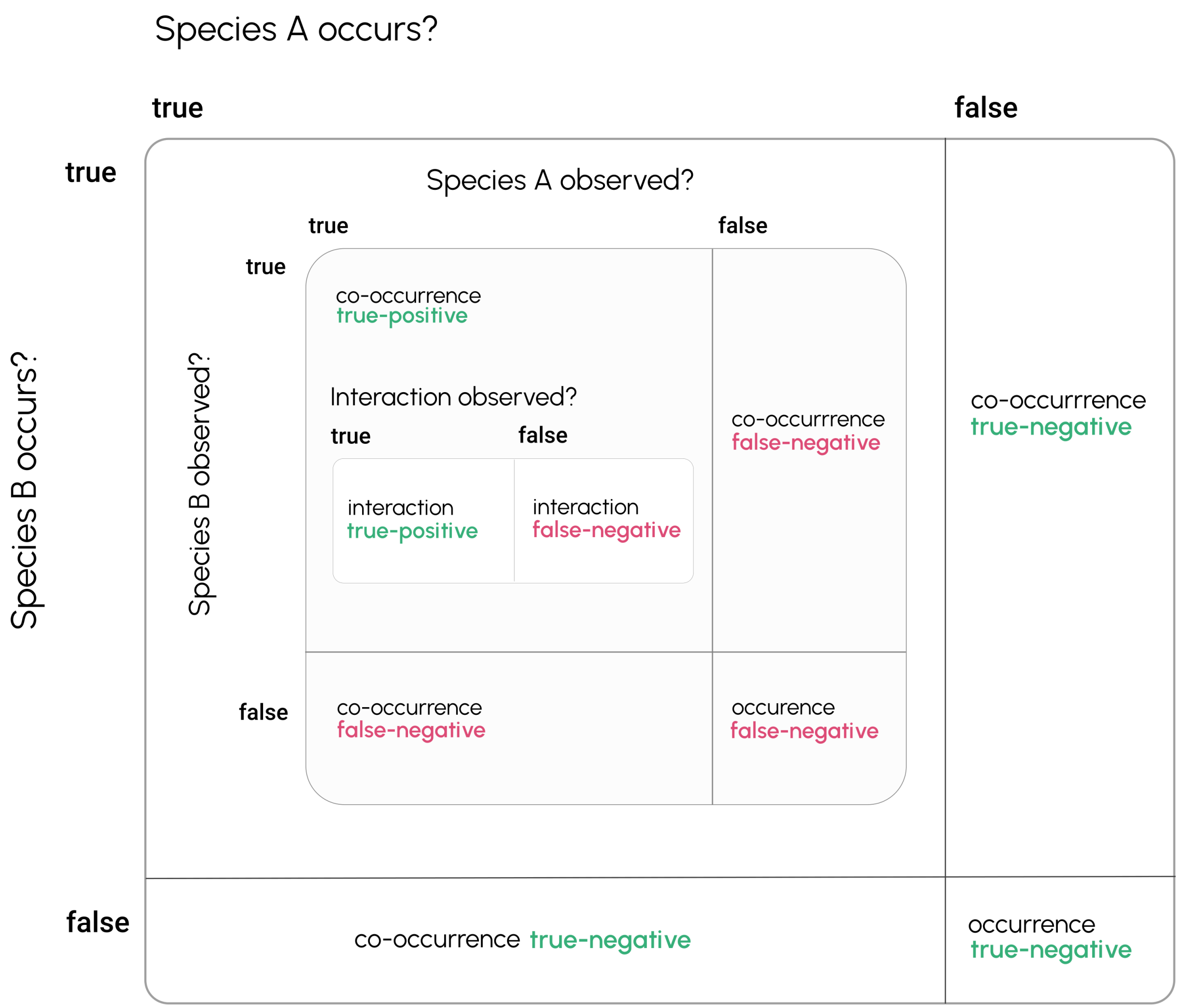

Figure 3: taxonomy of false negatives

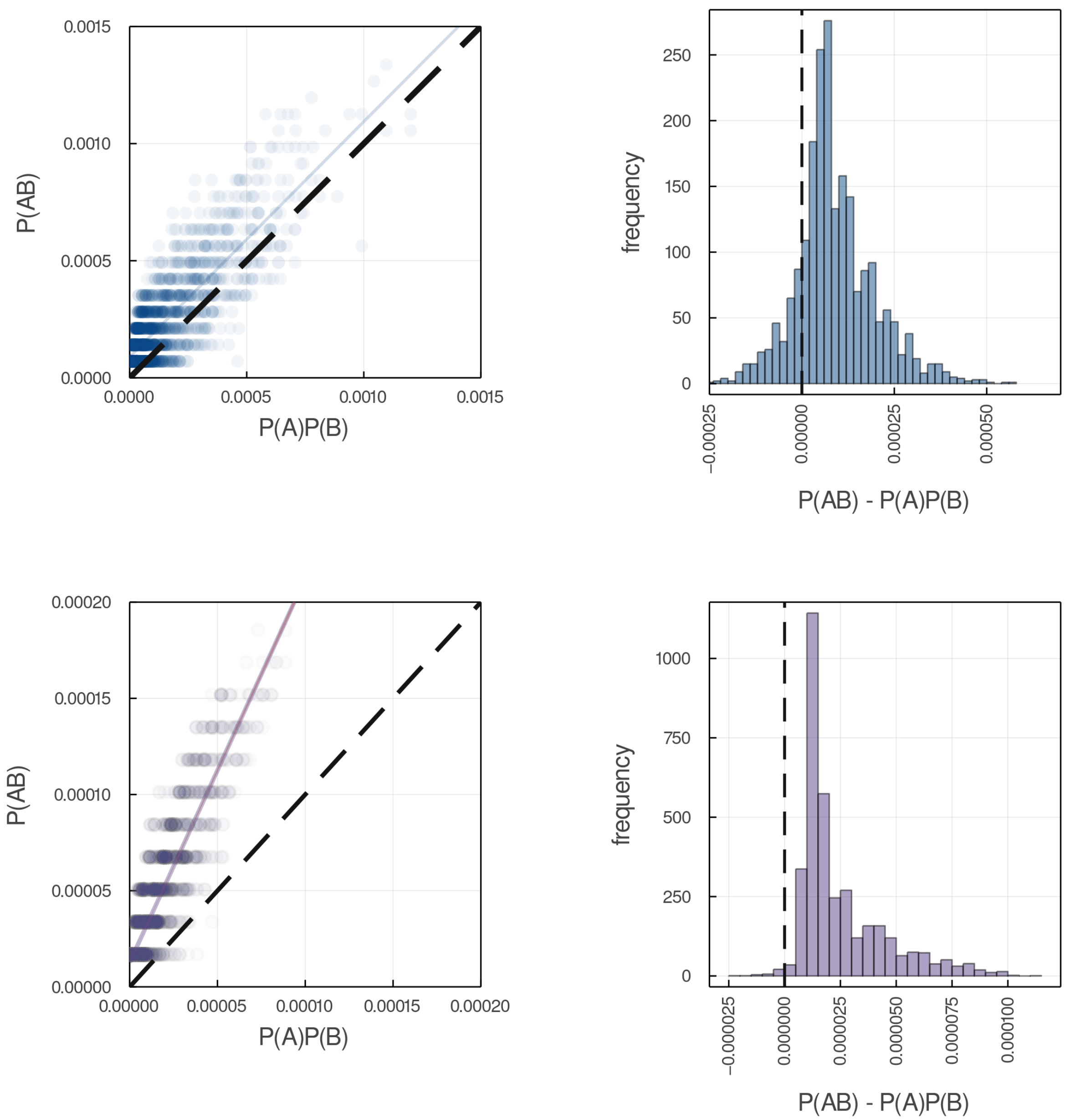

Figure 4: f

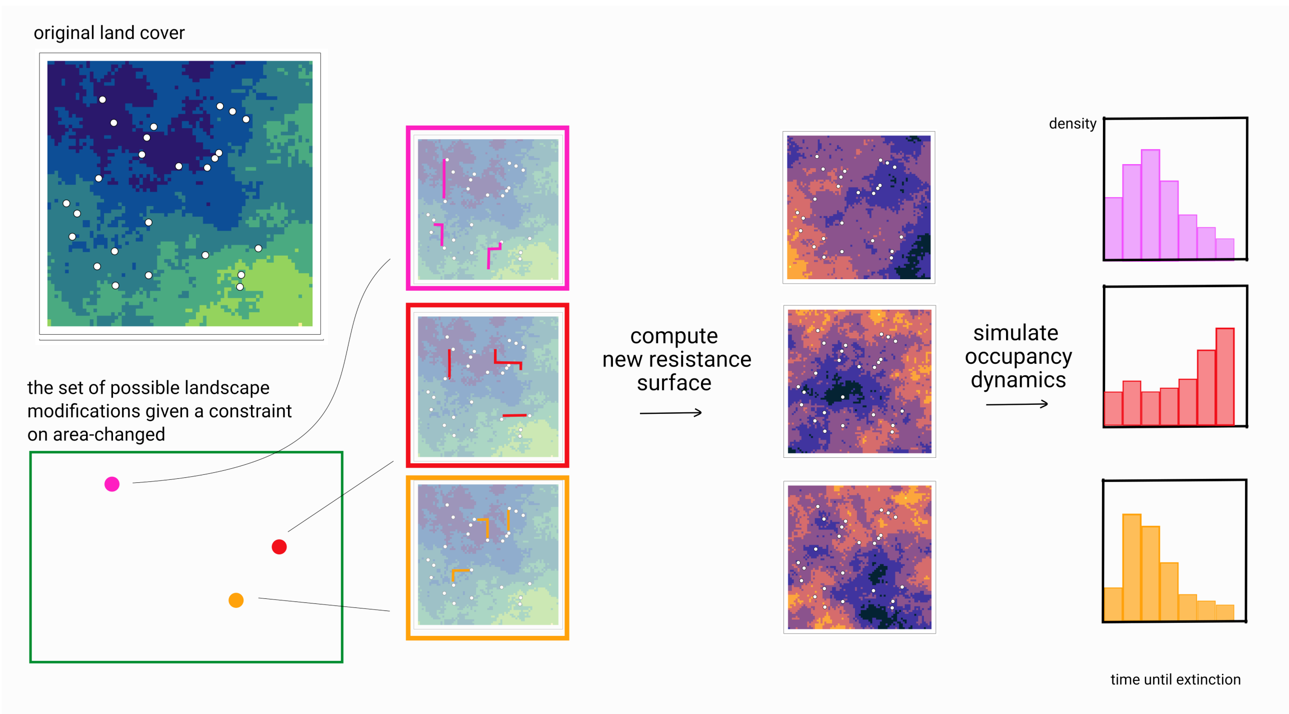
original land cover
the set of possible landscape modifications given a constraint on area-changed
compute new resistance surface
simulate occupancy dynamics
density
time until extinction

Figure 5: fig

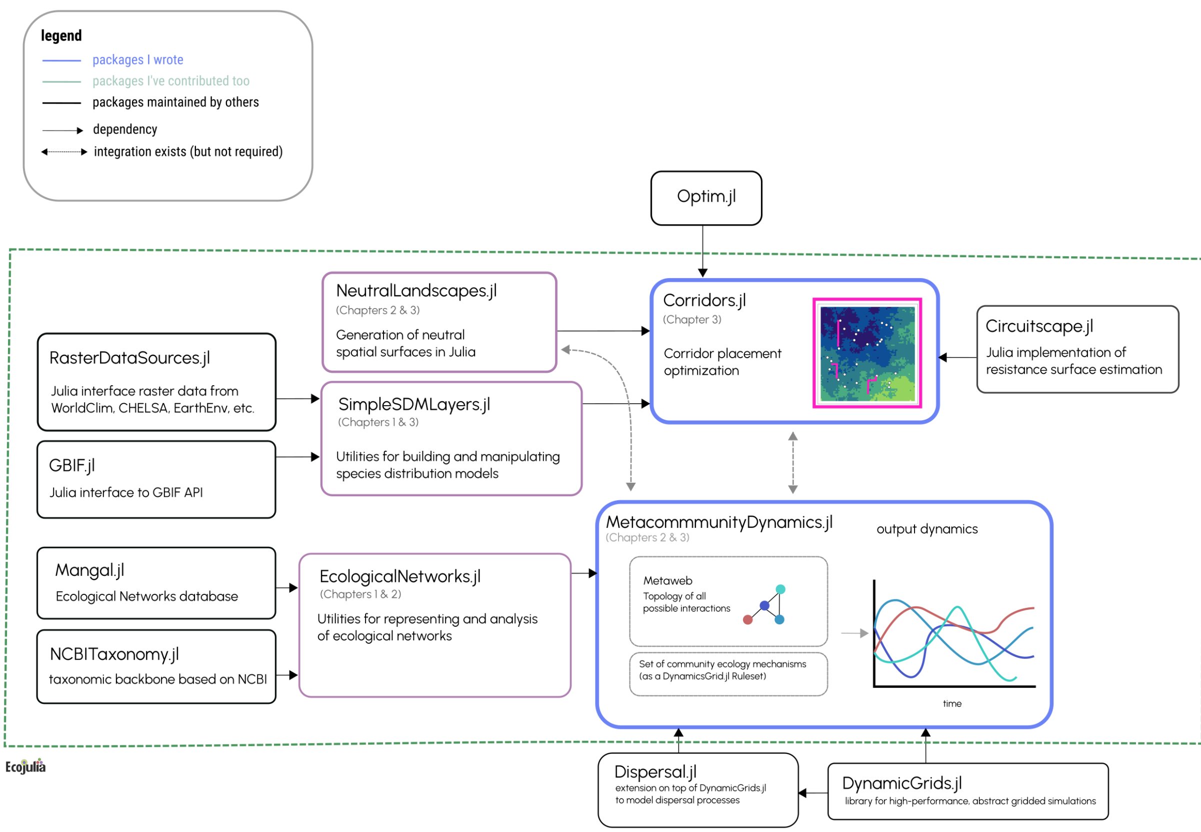

Figure 6: The structure of the software libraries used as part of MCD.jl

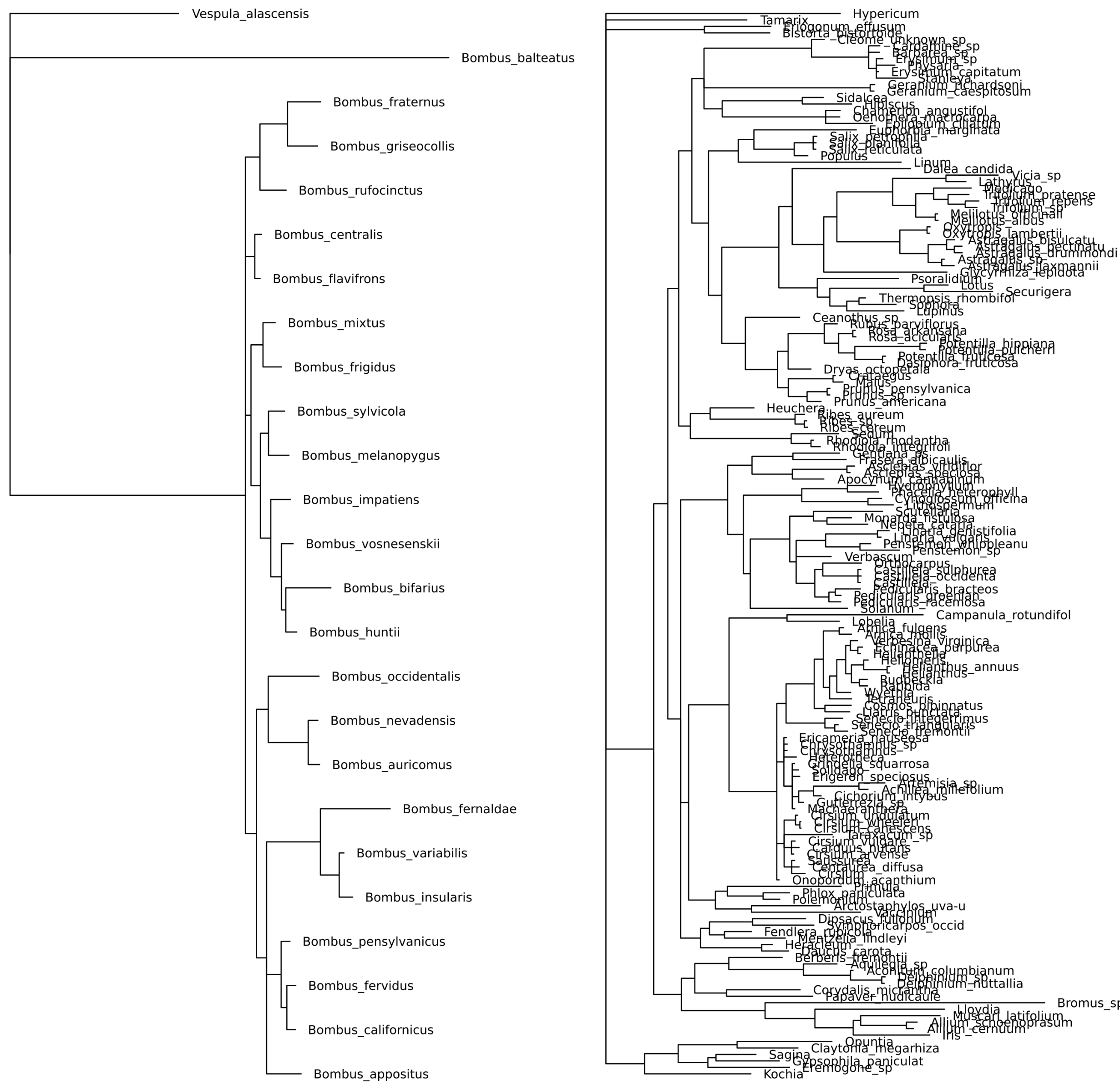

Figure 7: trees