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`

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

Last revision: *November 21, 2021*

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

Introduction

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).

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 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). 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 mathematical models in ecology.

However, we run into many problems when aiming to apply this type of model to empirical ecological data. 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: the product of different mechanisms operating a different spatial, temporal, and organizational scales (Levin 1992). Further this analytical approach to modeling explicitly ignores known realities: ecological dynamics not deterministic and many analytic models in ecology assume long-run equilibrium. Finally, perhaps the biggest challenge in using these models to describe ecological processes is ecosystems consist of more dimensions 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.

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.

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 potentially 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 future state of the system, and some "goal" state for the future, the problem is then to optimize our intervention into the system to maximize the probability the system approaches our "goal" state. This dissertation aims to this 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).

[Figure 1 about here.]

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

Interactions between plants and pollinators form networks which together structure the "architecture of biodiversity" (Bascompte & Jordano 2007). The functioning and stability of ecosystems emerge from these interactions, but anthropogenic change threatens to unravel and "rewire" these interaction networks (CaraDonna *et al.* 2017), jeopardizing the persistence of these systems. Plant-pollinator networks face two possible forms of rewiring in response to anthropogenic environmental change: spatial and temporal. Range shifts could cause interacting species to no longer overlap in space, and shifts in phenology could cause interacting species to no longer occur at the same time of year. 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 the uncoupling of the plant-pollinator metaweb of Colorado.

[Figure 2 about here.]

Data

The data for this chapter is derived from multiple sources that 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 respectively. This data was collected by Paul CaraDonna and Jane Oglevie (from the Rocky Mountain Biological Laboratory; RMBL) 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 Colorado taken from GBIF. (3) Remotely sensed data consisting of current and forecasting bioclimatic variables from CHELSA. (4) Phylogenies for both bee and flower species 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.* 2021b), we begin by predicting a metaweb of interactions across Colorado as they exist *in the present*. We do this using a set of candidate interaction prediction models: relative abundance only, phylogenetic embedding only (a la Strydom *et al.* (2021a)), niche embedding only (Gravel *et al.* 2019), and all pairwise combinations of those constituent models. After validating and selecting the best performing model, we then predict how these distributions of each of these species will change under the CMIP6 consensus climate forecast (Karger *et al.* 2017), and then finally quantify the reduction in spatial between species for which there is a predicted interaction.

Preliminary Results

Here we show the in-progress work on the prerequisites for the analysis outlined above: phylogenies for both plant and bee species (fig. 3) and species distribution models for all species (an example shown in fig. 4).

[Figure 3 about here.]

[Figure 4 about here.]

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 observing 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 at the top of fig. 1, left. As explored in the previous chapter, there are false-negatives in interaction data. However, there is more than one way to observe a false-negative when sampling interactions. fig. 5 shows a taxonomy of false-negatives in occurrence, co-occurrence, and interaction data.

[Figure 5 about here.]

The first result is to compute a null expectation of the probability of an interaction false-negative as a function number of total observations of individuals of *any species*. This is done by using a log-normal distribution of relative abundance (Hubbell 2001) and simulating the process of observation on food-webs generated using the niche model (Williams & Martinez 2000) with connectance parameterized by the flexible-links model (MacDonald *et al.* 2020). An example of this relation for networks with varying spceies richness is shown in fig. 6.

[Figure 6 about here.]

We then go on to testing some assumptions of this neutral model with empirical data. Primarily that we analytically show that our neutral model, if anything, underestimates the probability of false-negatives if there are positive associations between species co-occurrence, and we show these positive associations exist in two sets of spatially replicated samples of interaction networks (Thompson & Townsend 2000; Hadfield *et al.* 2014), fig. 7—further I'm planning to add the field data from the previous chapter into this analysis once available.

[Figure 7 about here.]

Finally this chapter proposes a simulated annealing method to optimize the a set of n points in space to maximize the probability of detecting an interaction between two species a and b with *known* distributions D_a, D_b.

Chapter Three: Optimizing corridor placement against ecological dynamics

As land-use change has caused many habitats to become fragmented and patchy, promoting landscape connectivity has become of significant interest to mitigate the effects of this 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 inferring 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 in an existing landscape from a derived 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 8 about here.]

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 using Circuitscape (McRae *et al.* 2008), and based on the values of resistance to dispersal between pair of locations 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. The largest challenge in implementing this algorithm is the space of potential modifications grows as $O((nm)!)$ for an n by m raster. For most actual landscapes to which we wish to apply this method, the set of possible modifications becomes uncomputably large, so we use simulated annealing to explore the search space of possible modifications to estimate the modification that maximizes the time-until extinction of simulated metapopulation dynamics under that hypothetical modified landscape.

The biggest challenge in implementing simulated annealing in this context is defining a proposal function for landscape modifications. This is done by computing the minimum-spanning-tree (MST) of the spatial nodes, and then proposing corridors that connect nodes that are already connected in the MST.

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

The final 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. 9.

[Figure 9 about here.]

Conclusion

References

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

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.

CaraDonna, P.J., Petry, W.K., Brennan, R.M., Cunningham, J.L., Bronstein, J.L., Waser, N.M., *et al.* (2017). Interaction rewiring and the rapid turnover of plantpollinator networks. *Ecology Letters*, 20, 385–394.

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.

Gravel, D., Baiser, B., Dunne, J.A., Kopelke, J.-P., Martinez, N.D., Nyman, T., *et al.* (2019). Bringing Elton and Grinnell together: A quantitative framework to represent the biogeography of ecological interaction networks. *Ecography*, 42, 401–415.

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.

Karger, D.N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., *et al.* (2017). Climatologies at high resolution for the earth's land surface areas. *Scientific Data*, 4, 170122.

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

MacDonald, A.A.M., Banville, F. & Poisot, T. (2020). Revisiting the Links-Species Scaling Relationship in Food Webs. *Patterns*, 1.

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.

McRae, B.H., Dickson, B.G., Keitt, T.H. & Shah, V.B. (2008). Using Circuit Theory to Model Connectivity in Ecology, Evolution, and Conservation. *Ecology*, 89, 2712–2724.

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., Bouskila, S., Banville, F., Barros, C., Caron, D., Farrell, M.J., *et al.* (2021a). Food web reconstruction through phylogenetic transfer of low-rank network representation.

Strydom, T., Catchen, M.D., Banville, F., Caron, D., Dansereau, G., Desjardins-Proulx, P., *et al.* (2021b). *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*.

Williams, R.J. & Martinez, N.D. (2000). Simple rules yield complex food webs. *Nature*, 404, 180–183.

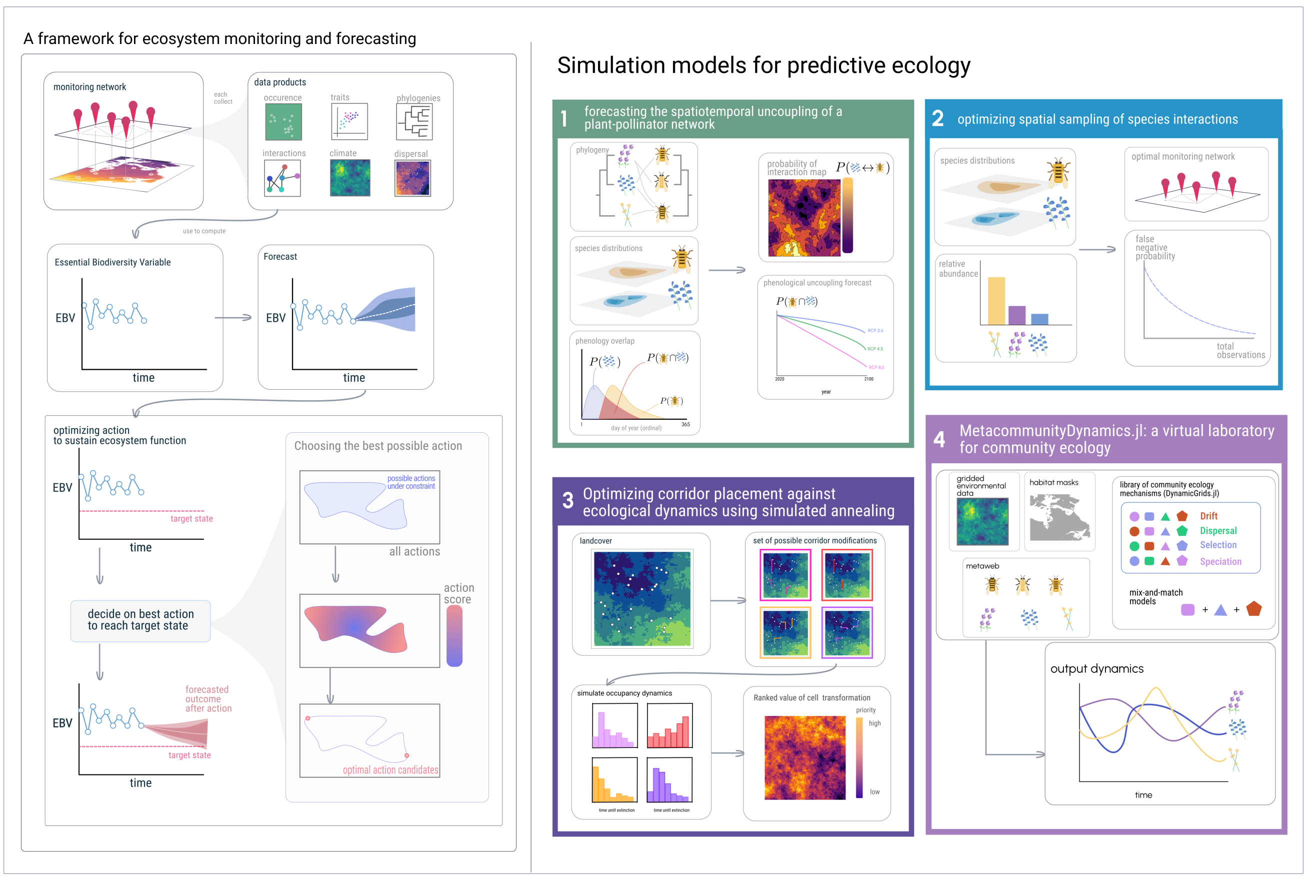

Figure 1: thesis concept

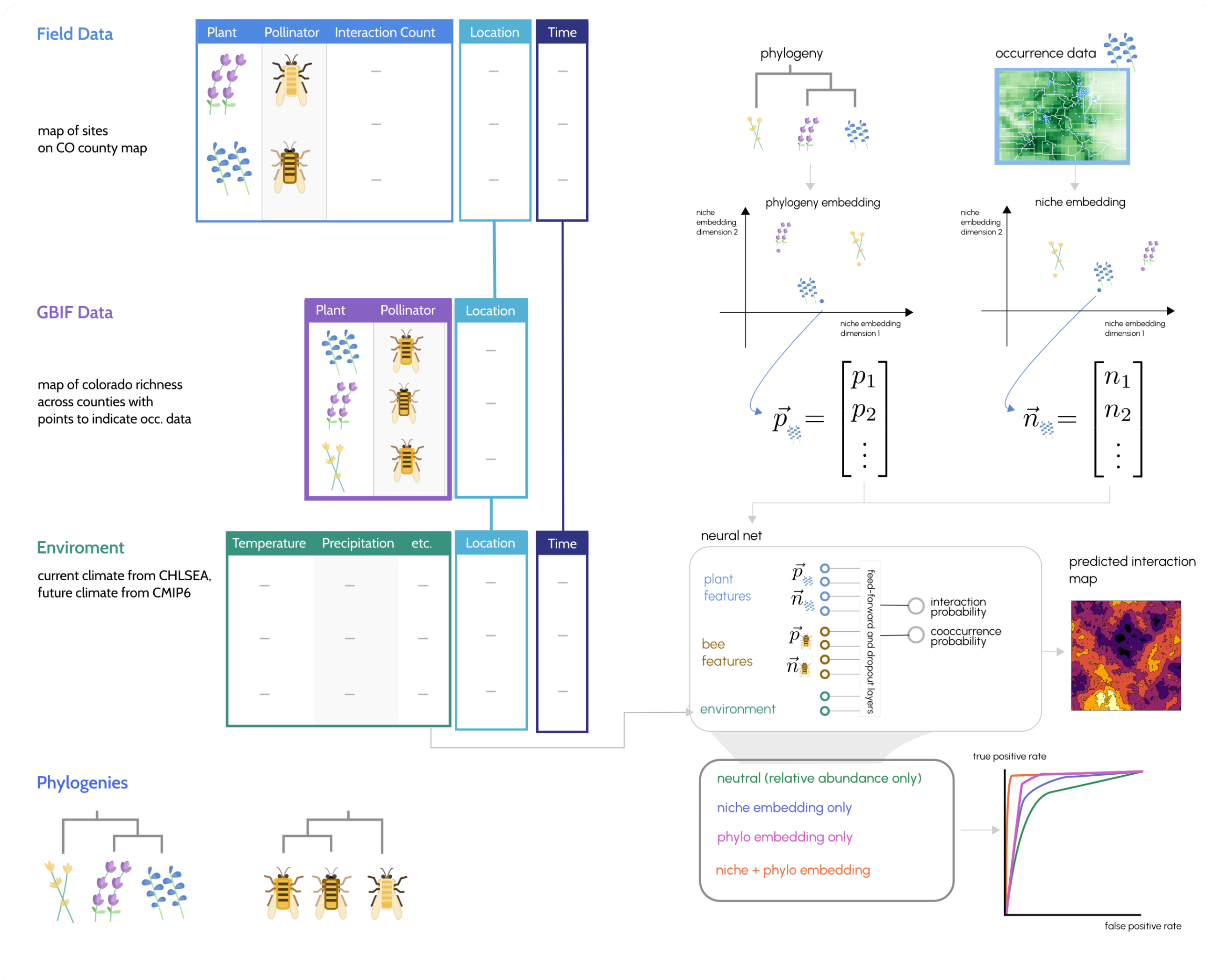

Figure 2: Chapter One conceptual figure. Left: the sources of data and how they can be synthesized. Right: The flow from data to interaction prediction using a few different interaction prediction models.

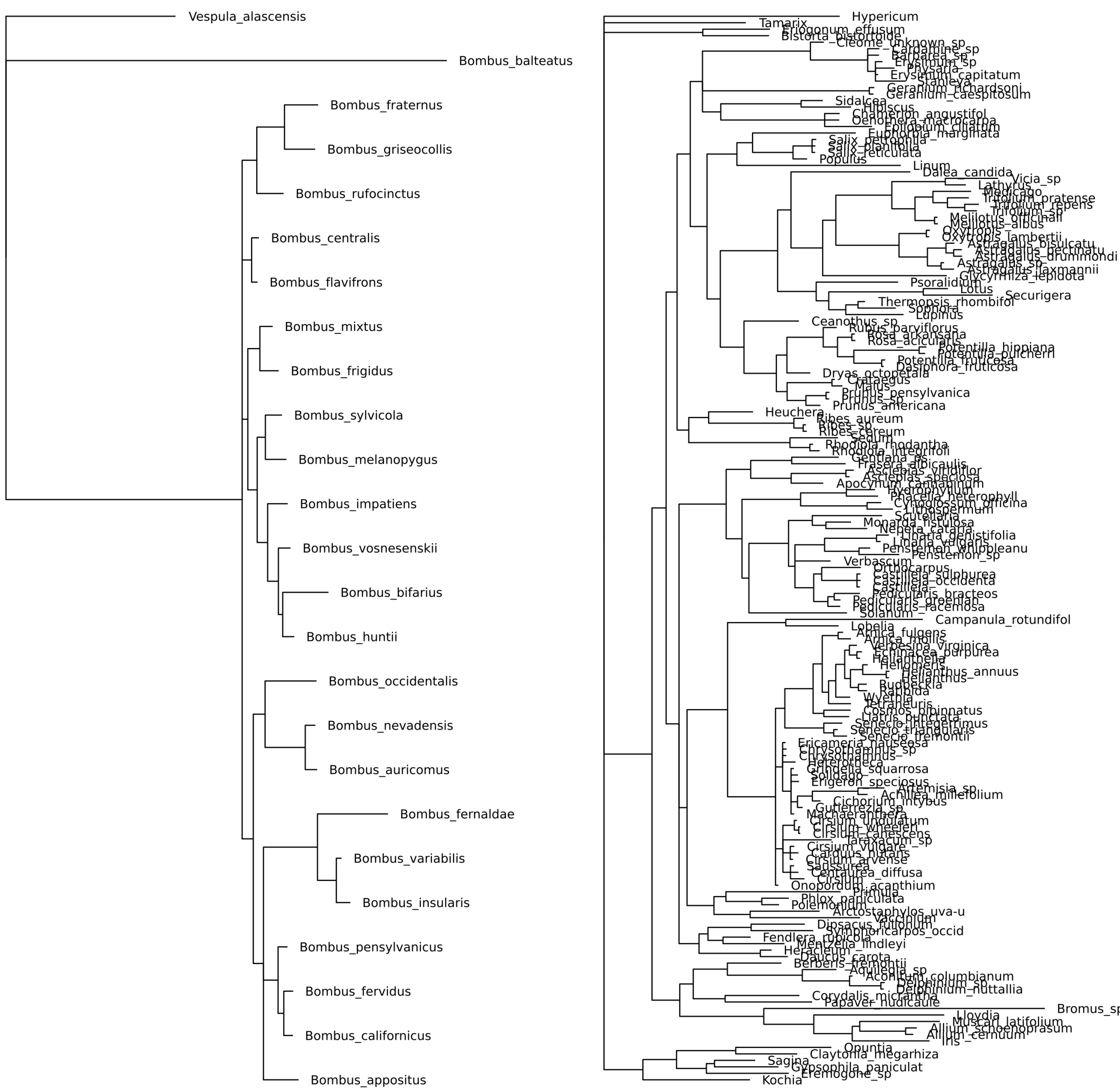

Figure 3: Phylogeny for both bumblebee species (left) and flower species (right)

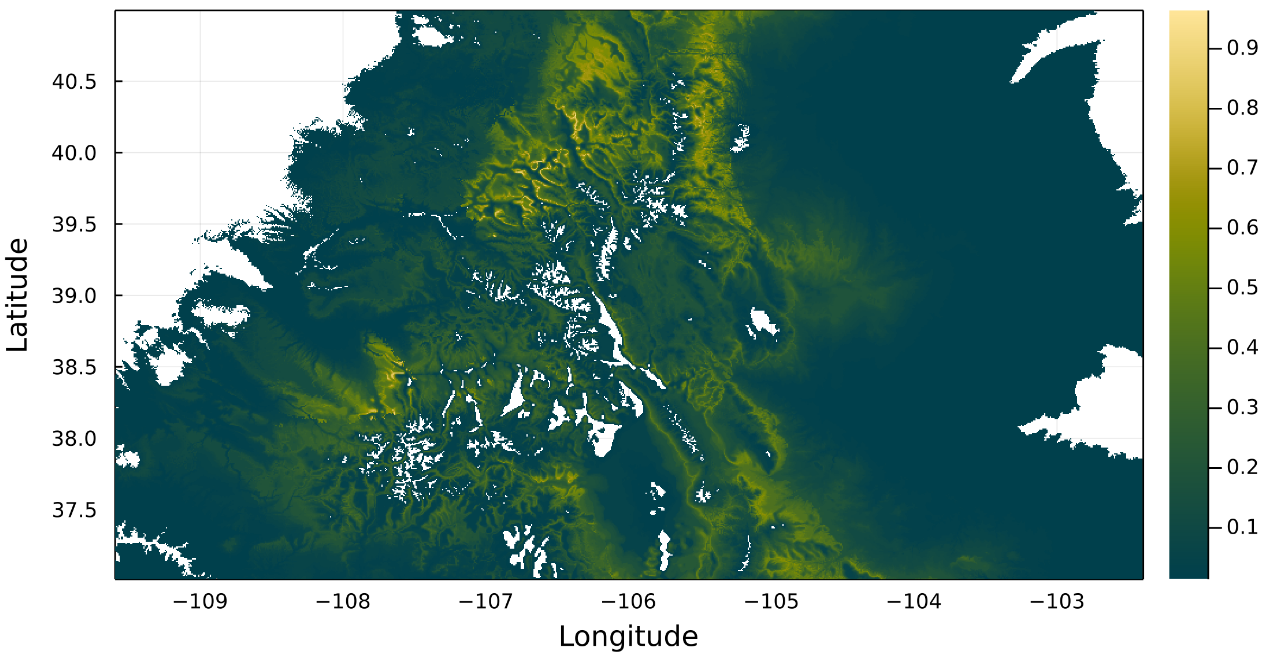

Figure 4: Example SDM for *Archillea millefolium*

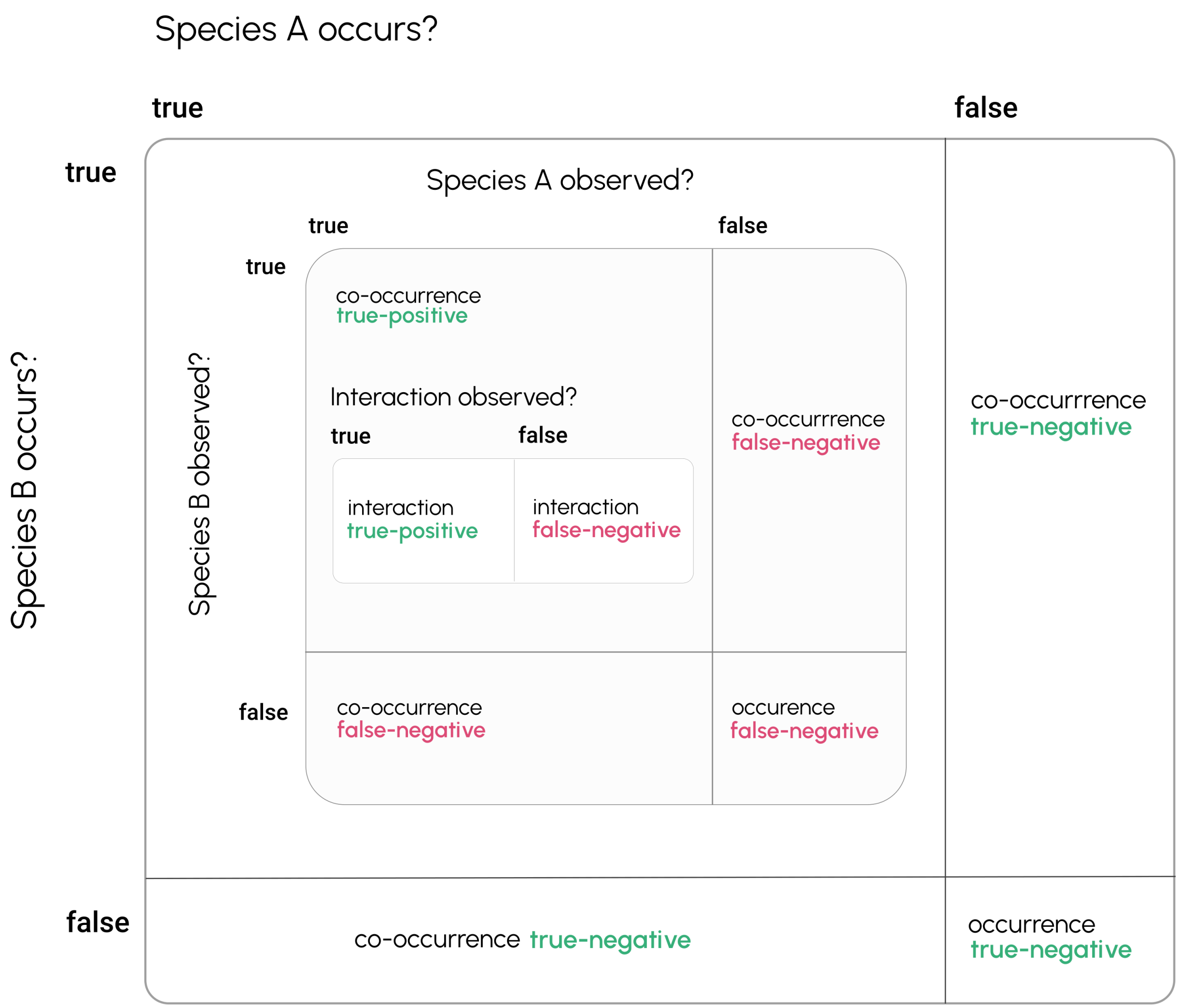

Figure 5: A taxonomy of occurrence, co-occurrence, and interaction false negatives in data

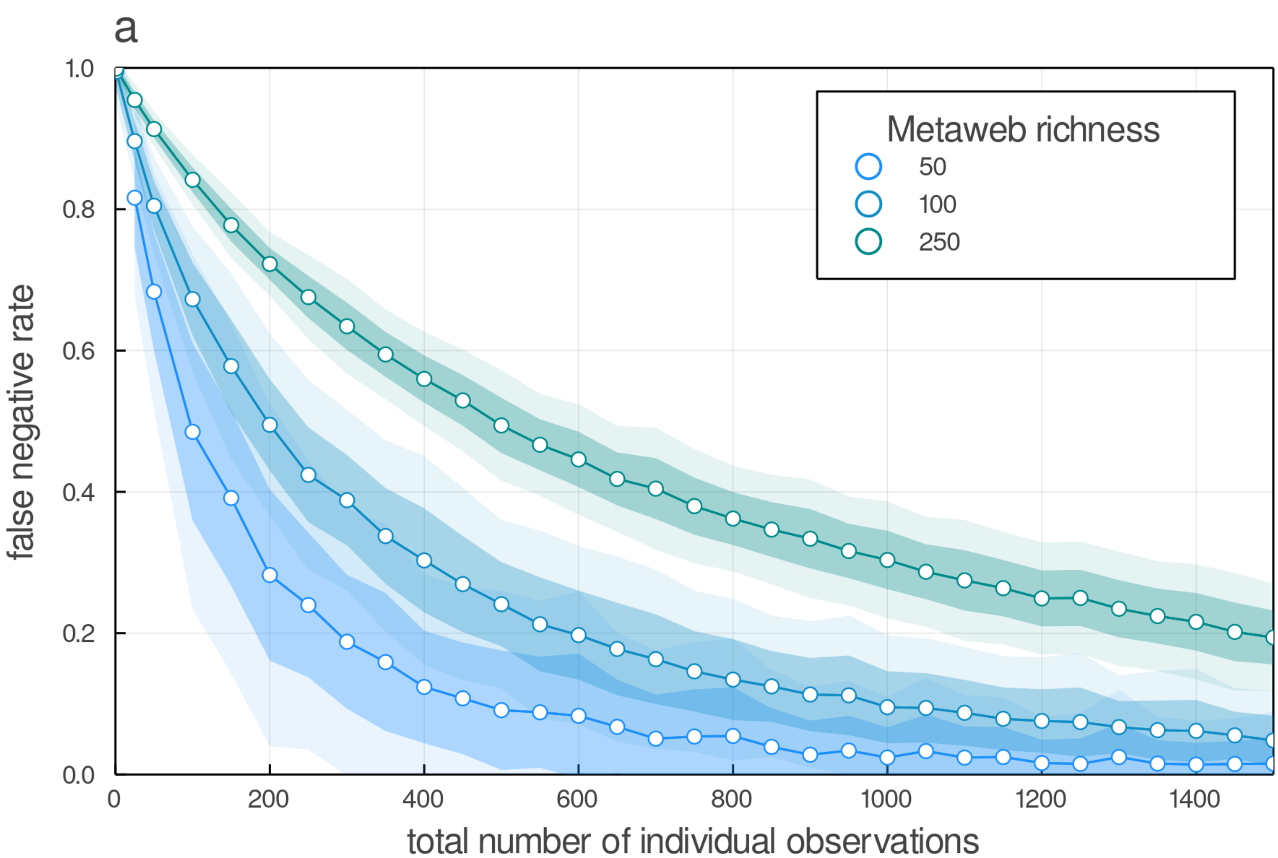
a
Metaweb richness
50
100
250
false negative rate
1.0
0.8
0.6
0.4
0.2
0.0
0
200
400
600
800
1000
1200
1400
total number of individual observations

Figure 6: foo

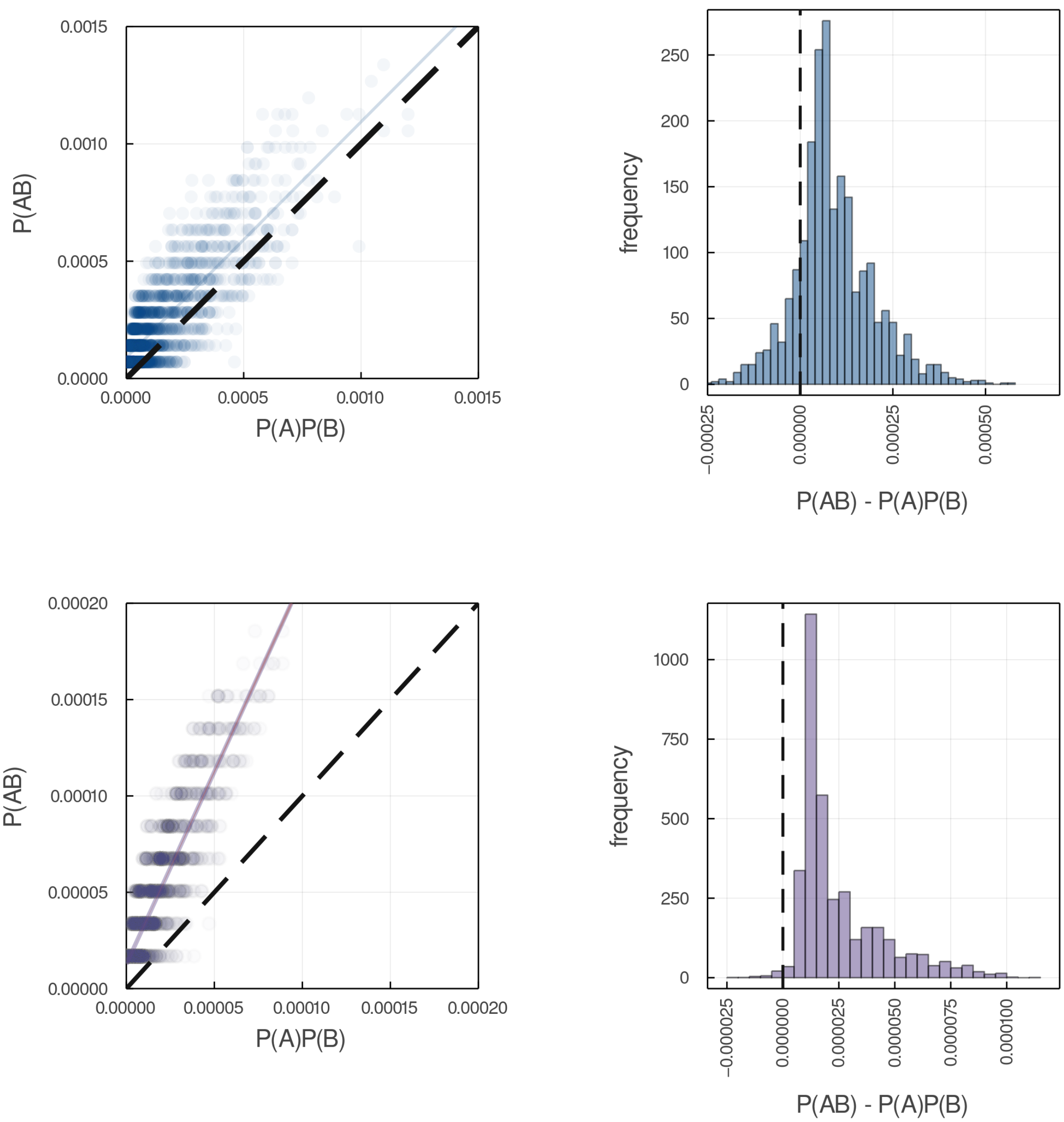

Figure 7: Demonstrates positive associations in co-occurrence

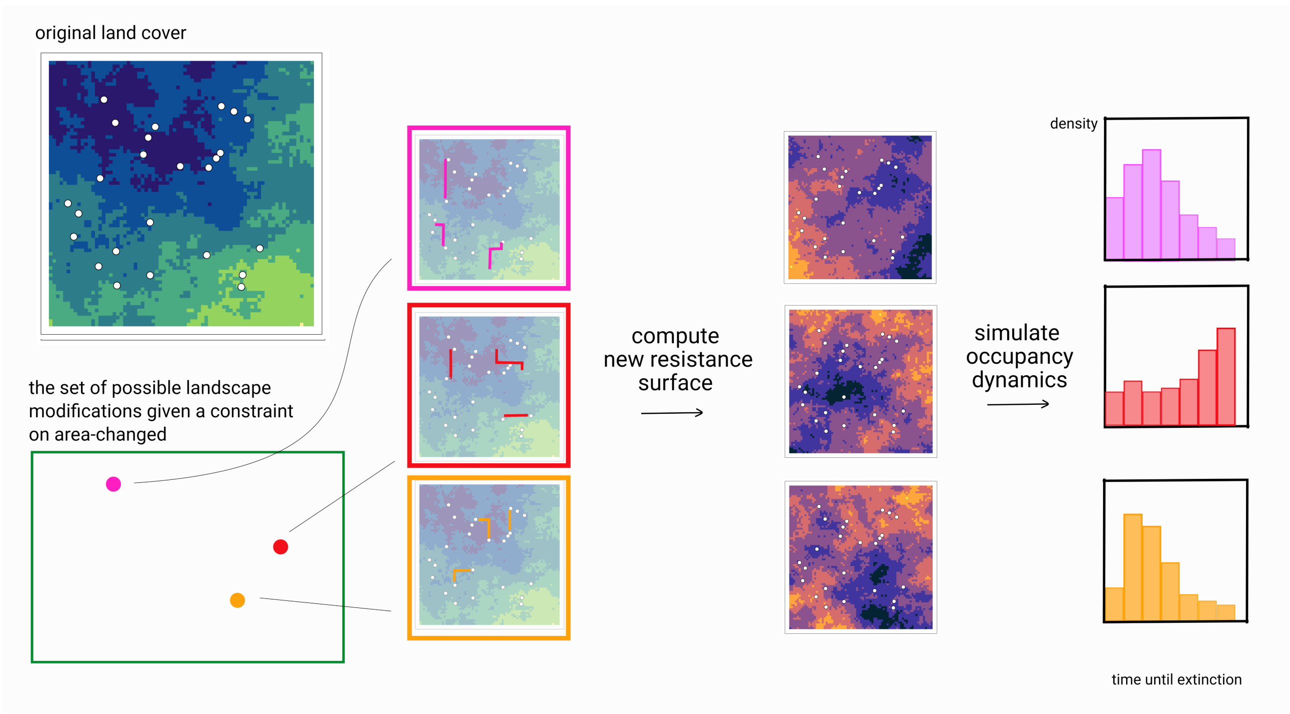
original land cover
the set of possible landscape modifications given a constraint on area-changed
compute new resistance surface
simulate occupancy dynamics
density
time until extinction

Figure 8: foo

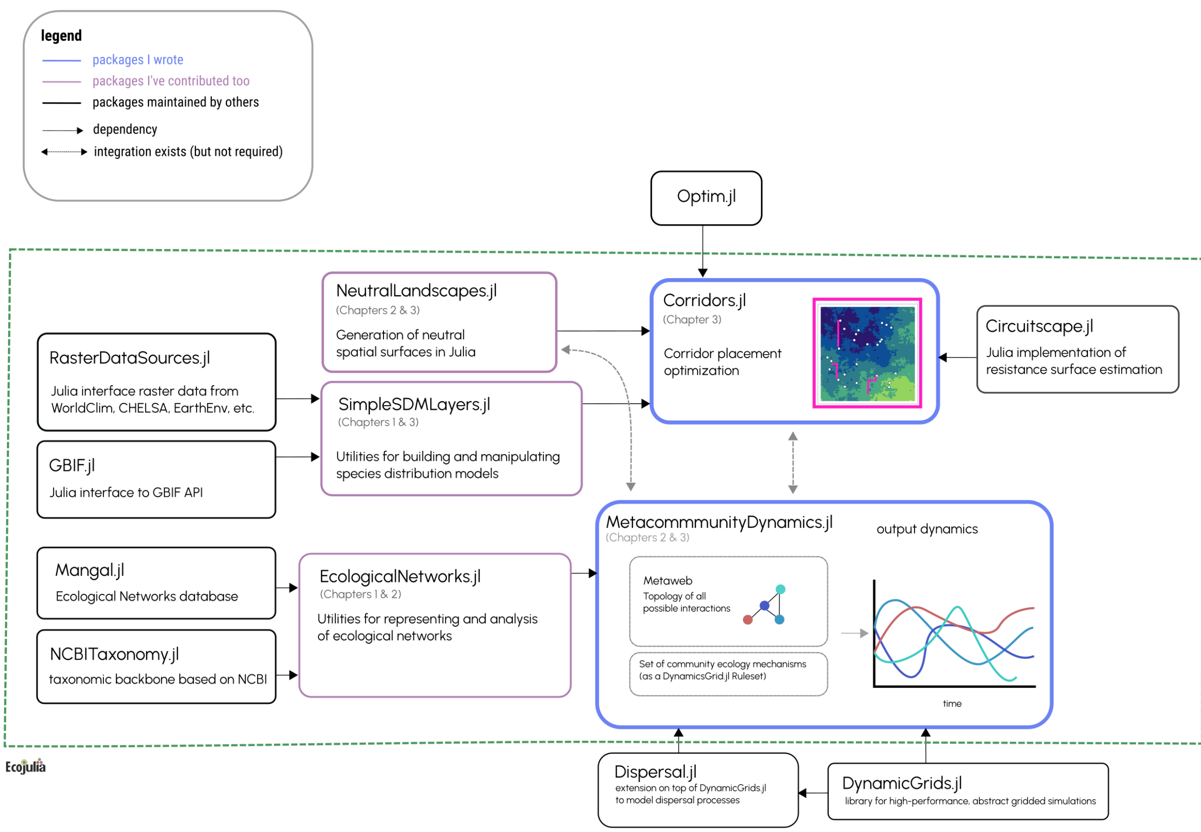

Figure 9: The structure of the software libraries used as part of MCD.jl