

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*

1 Introduction

2 Within the last several hundred years, human activity has induced rapid changes in Earth's atmosphere,
3 oceans, and surface. Greenhouse gas emissions have caused an increase the temperature of both Earth's
4 terrain and oceans, and both agricultural and urban development has rapidly reshaped the Earth's land
5 cover. These the bulk of this change has occurred within the last several hundred years, a geological
6 instant, inducing a sudden shift in conditions to Earth's climate and biosphere. As a result, predicting how
7 ecosystems will change in the future, *ecological forecasting*, and then using these forecasts to make
8 decisions to mitigate the negative consequences of this change on ecosystems, their functioning, and the
9 services they provide to humans has emerged as an imperative for ecology and environmental science
10 (Dietze 2017). However, robust prediction of ecological processes is, to say the least, quite difficult
11 (Beckage *et al.* 2011; Petchey *et al.* 2015). This difficultly is compounded by a few factors, the first being
12 that sampling ecosystems is not easy. Ecological data is often biased, noisy, and sparse in both space and
13 time. The current paucity of ecological data has resulted in much interest in developing global systems for
14 *ecosystem monitoring* (Makiola *et al.* 2020), which would systematize the collection of biodiversity data in
15 manner that makes detecting and predicting change more possible than at the moment (Urban *et al.* 2021).

16 The second major challenge in ecological forecasting is that the underlying dynamics of most ecological
17 processes are unknown and instead must be inferred from this (sparse) data. Much of the history of
18 quantitatively modeling ecosystems have been done in the language of dynamical systems, describing how
19 the value of an observable state of the system, represented by a vector of numbers $[x_1, x_2, \dots, x_n]^T = \vec{x}$
20 changes as over time, yielding models in the form of differential equations in continuous-time
21 settings $\frac{dx}{dt} = f(x)$ – or difference equations in discrete-time settings $x_t = f(x_{t-1})$ – where $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is
22 an arbitrary function describing how the system changes on a moment-to-moment basis (e.g. in the
23 context of communities, f could be Lotka-Volterra, Holling-Type-III or DeAngelis-Beddington functional
24 response). The initial success of these forms of models can be traced back to the larger program of
25 ontological reductionism, which became the default approach to modeling in the sciences after its early
26 success in physics, which, by the time ecology was becoming a quantitative science (sometime in the 20th
27 century, depending on who you ask), became the foundation for early quantitative models in ecology.

28 However, we run into many problems when aiming to apply this type of model to empirical data in
29 ecology. Ecosystems are perhaps the quintessential example of system that cannot be understood by

30 iterative reduction of its components into constituent parts—ecological phenomena are emergent are the
31 product of different mechanisms operating a different spatial, temporal, and organizational scales (Levin
32 1992). Further, the form of this functional response in real systems is effectively unknown, and some
33 forms are inherently more “forecastable” than others (Beckage *et al.* 2011; Chen *et al.* 2019; Pennekamp *et*
34 *al.* 2019). Further this analytical approach to modeling explicitly ignores known realities: ecological
35 dynamics not deterministic, many analytic models in ecology assume long-run equilibrium. Finally,
36 perhaps the biggest challenge in using these models to describe ecological processes is ecosystems vary
37 across more variables than the tools of analytic models are suited for. As the number of variables in an
38 analytic model increases, so does the ability of the scientist to discern clear relationships between them
39 given a fixed amount of data, the so-called “curse” of dimensionality.

40 But these problems are not solely unique to ecology. The term *ecological forecasting* implicitly creates an
41 analogy with weather forecasting. Although it has become a trite joke to complain about the weather
42 forecast being wrong, over the least 50 years the field of numerical weather prediction (NWP) has
43 dramatically improved our ability to predict weather across the board (Bauer *et al.* 2015). The success of
44 NWP, and the Earth observations systems that support it (Hill *et al.* 2004), should serve as a template for
45 development of a system for monitoring Earth’s biodiversity. Much like ecology, NWP is faced with
46 high-dimensional systems that are governed by different mechanisms at different scales. The success of
47 NWP is that, rather than, say, attempt to forecast the weather in Quebec by applying Navier-Stokes to
48 entire province, to instead use simulation models which describe known mechanisms at different scales,
49 and use the availability to increasing computational power to directly simulate many batches of dynamics
50 which directly incorporate stochasticity and uncertainty in parameter estimates via random number
51 generation.

52 But forecasting is only half the story—if indeed “[ecologists] have hitherto only interpreted the world in
53 various ways; the point is to change it,” then once we have a forecast about how an ecosystem will change
54 in the future, what if this forecast predicts a critical ecosystem service will deteriorate? We are still left
55 with the question, what do we in the time being to mitigate the potentially negative consequences a
56 forecast predicts? In this framing, mitigating the consequences of anthropogenic change on ecosystems
57 becomes an optimization problem: given a forecast of the future state of the system, and some “goal” state
58 for the future, the problem is then to optimize our intervention into the system to maximize the
59 probability the system approaches our “goal” state. This dissertation aims to this framework for ecosystem

60 monitoring and forecasting (fig. 1, left), and each chapter address some aspect of this pipeline to data from
61 a monitoring network to forecasts to mitigation strategy (fig. 1, right).

62 [Figure 1 about here.]

63 **Chapter One: Forecasting the spatial uncoupling of a plant-pollinator 64 network**

65 Interactions between plants and pollinators form networks of interactions, which together create the
66 “architecture of biodiversity” (**Jordano2007?**). The functioning and stability of ecosystems emerge from
67 these interactions, but anthropogenic change threatens to unravel and “rewire” these interaction networks
68 (**CaraDonna2017IntRew?**), threatening the persistence of these systems. Plant-pollinator networks face
69 two possible forms of rewiring in response to anthropogenic environmental change: spatial and temporal.
70 Spatially, range shifts could cause interacting species to no longer overlap in space, and shifts in phenology
71 could cause interacting species to no longer occur at the same time of year. This chapter uses several years
72 of data on bumblebee-flower phenology and interactions across several field sites, each consisting of
73 several plots across an elevational gradient, combined with spatial records of species occurrence via GBIF
74 to forecast this uncoupling. This addresses the EBV to forecast element of the flow from data to mitigation
75 in fig. 1 (left).

76 [Figure 2 about here.]

77 **Data**

78 The data for this chapter is derived from multiple sources that can be split into four categories. (1) Field
79 data from three different field sites across Colorado, each with multiple plots across an elevational
80 gradient, for seven, seven, and three years respectively. This data was collected by Paul CaraDonna and
81 Jane Oglevie (from the Rocky Mountain Biological Laboratory; RMBL) and Julian Resasco (CU Boulder).
82 (2) GBIF spatial occurrence records of each of these species across Colorado, including a metaweb of
83 interactions across all of Colroado taken from GBIF. (3) Remotely sensed data consisting of current and

84 forecasting bioclimatic variables from CHELSA. (4) Phylogenies for both bee and flower species derived
85 from NCBI GenBank barcodes for mitochondrial COI (bumblebees) and chloroplast rbcL (flowers).

86 **Methods**

87 As the data we have is spatially sparse and likely to contain many interaction “false-negatives” (Strydom *et*
88 *al.* 2021), we begin by predicting a metaweb of interactions across Colorado as they exist *in the present*. We
89 do this using a set of candidate interaction prediction models: relative abundance only, phylogenetic
90 embedding only (a la (**Strydom2021FooWeb?**)), niche embedding only (**Gravel2019BriElt?**), and all
91 pairwise combinations of those constituent models. After validating each model and selecting the best
92 performing model, we then predict how these distributions of each of these species will change under the
93 CMIP6 consensus climate forecast [@], and then finally quantify the reduction in spatial between species
94 for which there is a predicted interaction.

95 **Preliminary Results**

96 1) we got a tree and SDMs

97 **Next steps**

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

99 **Chapter Two: Optimizing spatial sampling of species interactions**

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

105 As explored in the previous chapter, there are false-negatives in interaction data. However, there is more
106 than one way to observe a false-negative when sampling interactions: (fig. 3). It begins with a conceptual

107 framework for understanding the difference in false-negatives in occurrence, co-occurrence, and
108 interactions. Co-occurrence is not the same thing as interaction (Blanchet *et al.* 2020), but often is used as
109 a proxy.

110 [Figure 3 about here.]

111 We use a log-normal distribution as a null model of the relative-abundance distribution (Hubbell 2001)
112 and then simulate realized false-negative rate as a function of varying sampling effort.
113 This also goes on to includes testing some assumptions of the model with empirical data fig. 4, which we
114 analytically show that our neutral model, if anything, underestimates the probability of false-negatives
115 due to positive correlations in co-occurrence in two sets of spatially replicated samples of interaction
116 networks (Thompson & Townsend 2000; Hadfield *et al.* 2014)—further I'm planning to add the field data
117 from the previous chapter into this analysis once available.

118 [Figure 4 about here.]

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

122 **Chapter Three: Optimizing corridor placement against ecological 123 dynamics**

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

131 This chapter proposes a general algorithm for choosing corridor placement to optimize a measurement of
132 ecosystem functioning derived from simulations run on each proposed landscape modification.

133 [Figure 5 about here.]

134 **Methods**

135 We propose various landscape modifications which alter the cover of a landscape, represented as a raster.
136 We then compute a new resistance surface based on the proposed landscape modification, and based on
137 the values of resistance to dispersal between each location we simulate spatially-explicit metapopulation
138 dynamics model (Hanski & Ovaskainen 2000; Ovaskainen *et al.* 2002) to estimate a distribution of time
139 until extinction for each landscape modification.

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

142 **Chapter Four: MetacommunityDynamics.jl: a virtual laboratory for** 143 **community ecology**

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

149 [Figure 6 about here.]

150 **Conclusion**

151 **Appendix**

152 [Figure 7 about here.]

153 **References**

- 154 Bauer, P., Thorpe, A. & Brunet, G. (2015). The quiet revolution of numerical weather prediction. *Nature*,
155 525, 47–56.
- 156 Beckage, B., Gross, L.J. & Kauffman, S. (2011). The limits to prediction in ecological systems. *Ecosphere*, 2,
157 art125.
- 158 Blanchet, F.G., Cazelles, K. & Gravel, D. (2020). Co-occurrence is not evidence of ecological interactions.
159 *Ecology Letters*, 23, 1050–1063.
- 160 Chen, Y., Angulo, M.T. & Liu, Y.-Y. (2019). Revealing Complex Ecological Dynamics via Symbolic
161 Regression. *BioEssays*, 41, 1900069.
- 162 Dietze, M.C. (2017). Prediction in ecology: A first-principles framework. *Ecological Applications*, 27,
163 2048–2060.
- 164 Hadfield, J.D., Krasnov, B.R., Poulin, R. & Nakagawa, S. (2014). A Tale of Two Phylogenies: Comparative
165 Analyses of Ecological Interactions. *The American Naturalist*, 183, 174–187.
- 166 Hanski, I. & Ovaskainen, O. (2000). The metapopulation capacity of a fragmented landscape. *Nature*, 404,
167 755–758.
- 168 Hill, C., DeLuca, C., Balaji, Suarez, M. & Da Silva, A. (2004). The architecture of the Earth System
169 Modeling Framework. *Computing in Science Engineering*, 6, 18–28.
- 170 Hubbell, S.P. (2001). *The unified neutral theory of biodiversity and biogeography*. Monographs in
171 population biology. Princeton University Press, Princeton.
- 172 Levin, S.A. (1992). The Problem of Pattern and Scale in Ecology: The Robert H. MacArthur Award
173 Lecture. *Ecology*, 73, 1943–1967.
- 174 Makiola, A., Compson, Z.G., Baird, D.J., Barnes, M.A., Boerlijst, S.P., Bouchez, A., et al. (2020). Key
175 Questions for Next-Generation Biomonitoring. *Frontiers in Environmental Science*, 7.
- 176 Ovaskainen, O., Sato, K., Bascompte, J. & Hanski, I. (2002). Metapopulation Models for Extinction
177 Threshold in Spatially Correlated Landscapes. *Journal of Theoretical Biology*, 215, 95–108.
- 178 Ovaskainen, O., Sato, K., Bascompte, J. & Hanski, I. (2002). Metapopulation Models for Extinction
179 Threshold in Spatially Correlated Landscapes. *Journal of Theoretical Biology*, 215, 95–108.

- 180 Pennekamp, F., Iles, A.C., Garland, J., Brennan, G., Brose, U., Gaedke, U., *et al.* (2019). The intrinsic
181 predictability of ecological time series and its potential to guide forecasting. *Ecological Monographs*, 89,
182 e01359.
- 183 Petchey, O.L., Pontarp, M., Massie, T.M., Kéfi, S., Ozgul, A., Weilenmann, M., *et al.* (2015). The ecological
184 forecast horizon, and examples of its uses and determinants. *Ecology Letters*, 18, 597–611.
- 185 Peterman, W.E. (2018). ResistanceGA: An R package for the optimization of resistance surfaces using
186 genetic algorithms. *Methods in Ecology and Evolution*, 9, 1638–1647.
- 187 Strydom, T., Catchen, M.D., Banville, F., Caron, D., Dansereau, G., Desjardins-Proulx, P., *et al.* (2021). *A
188 Roadmap Toward Predicting Species Interaction Networks (Across Space and Time)* (Preprint).
189 EcoEvoRxiv.
- 190 Thompson, R.M. & Townsend, C.R. (2000). Is resolution the solution?: The effect of taxonomic resolution
191 on the calculated properties of three stream food webs. *Freshwater Biology*, 44, 413–422.
- 192 Urban, M.C., Travis, J.M.J., Zurell, D., Thompson, P.L., Synes, N.W., Scarpa, A., *et al.* (2021). Coding for
193 Life: Designing a Platform for Projecting and Protecting Global Biodiversity. *BioScience*.

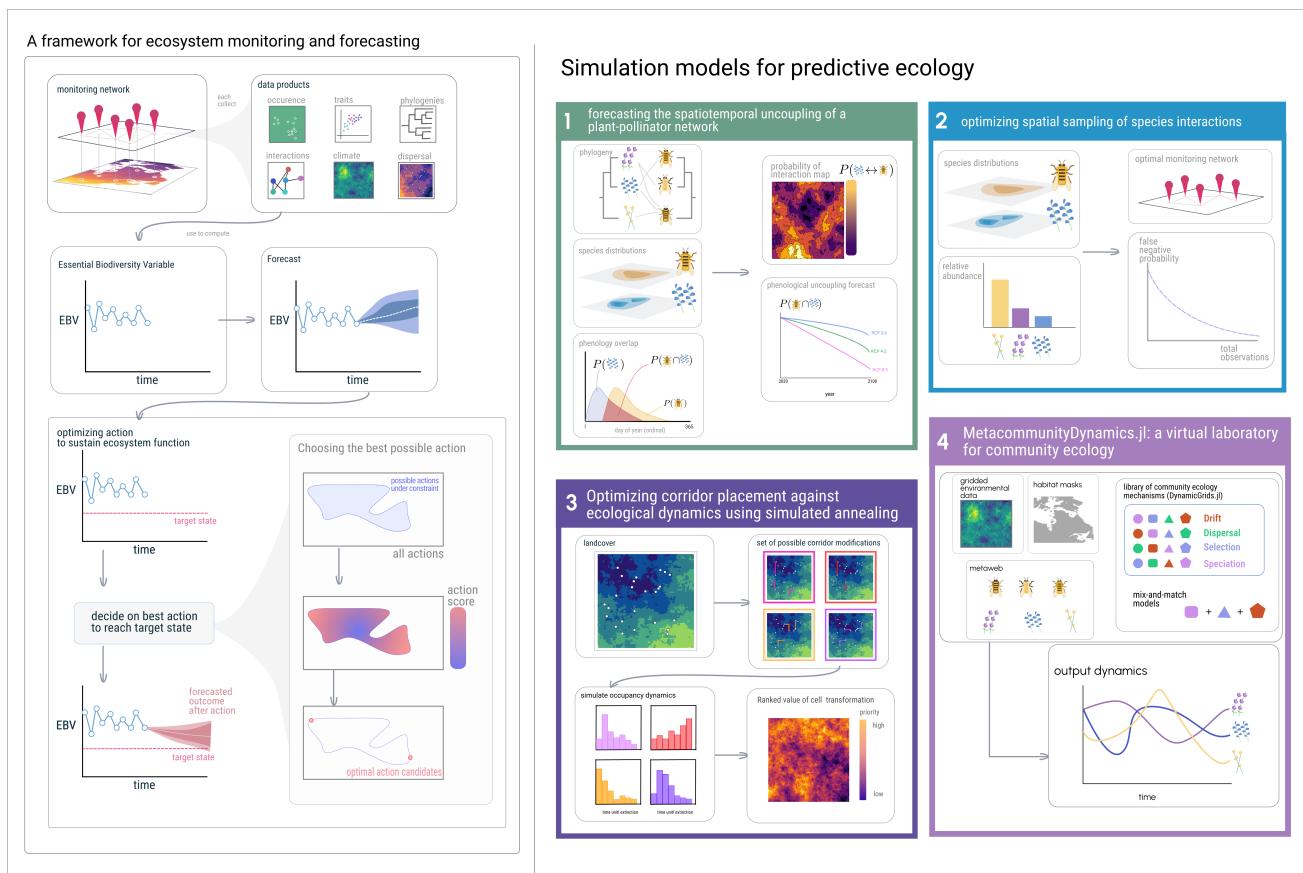


Figure 1: thesis concept

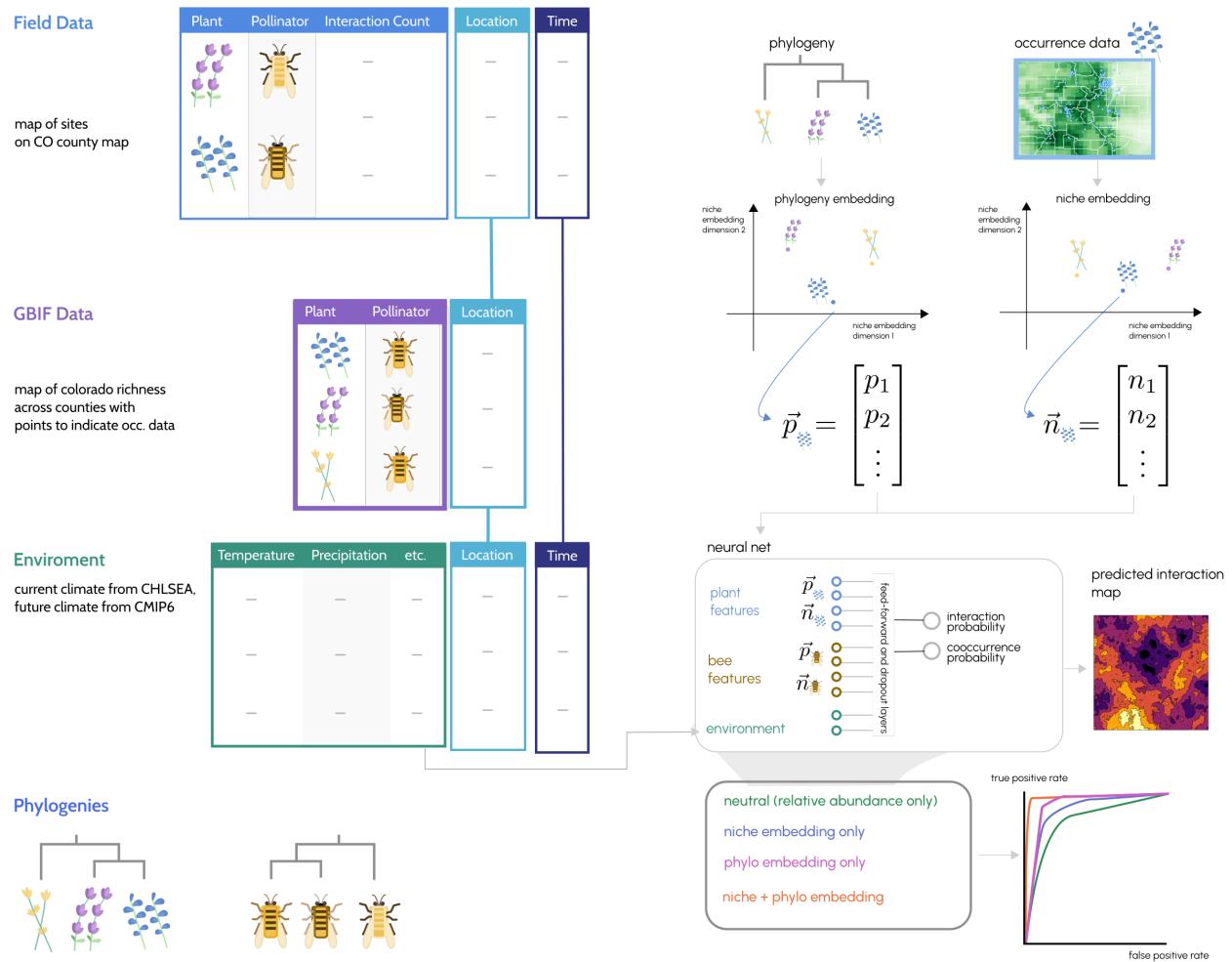


Figure 2: Chapter One conceptual figure. Left: the sources of data and how they can be synthesized. Right: The flow from data to interaction prediction using a few different interaction prediction models.

Species A occurs?

Species B occurs?

		true		false
		true	Species A observed?	
		true	Species A observed? true co-occurrence true-positive Interaction observed? true false interaction true-positive interaction false-negative	co-occurrence true-negative
		false	co-occurrence false-negative	occurrence false-negative
false		co-occurrence true-negative		occurrence true-negative

Figure 3: A taxonomy of occurrence, co-occurrence, and interaction false negatives in data

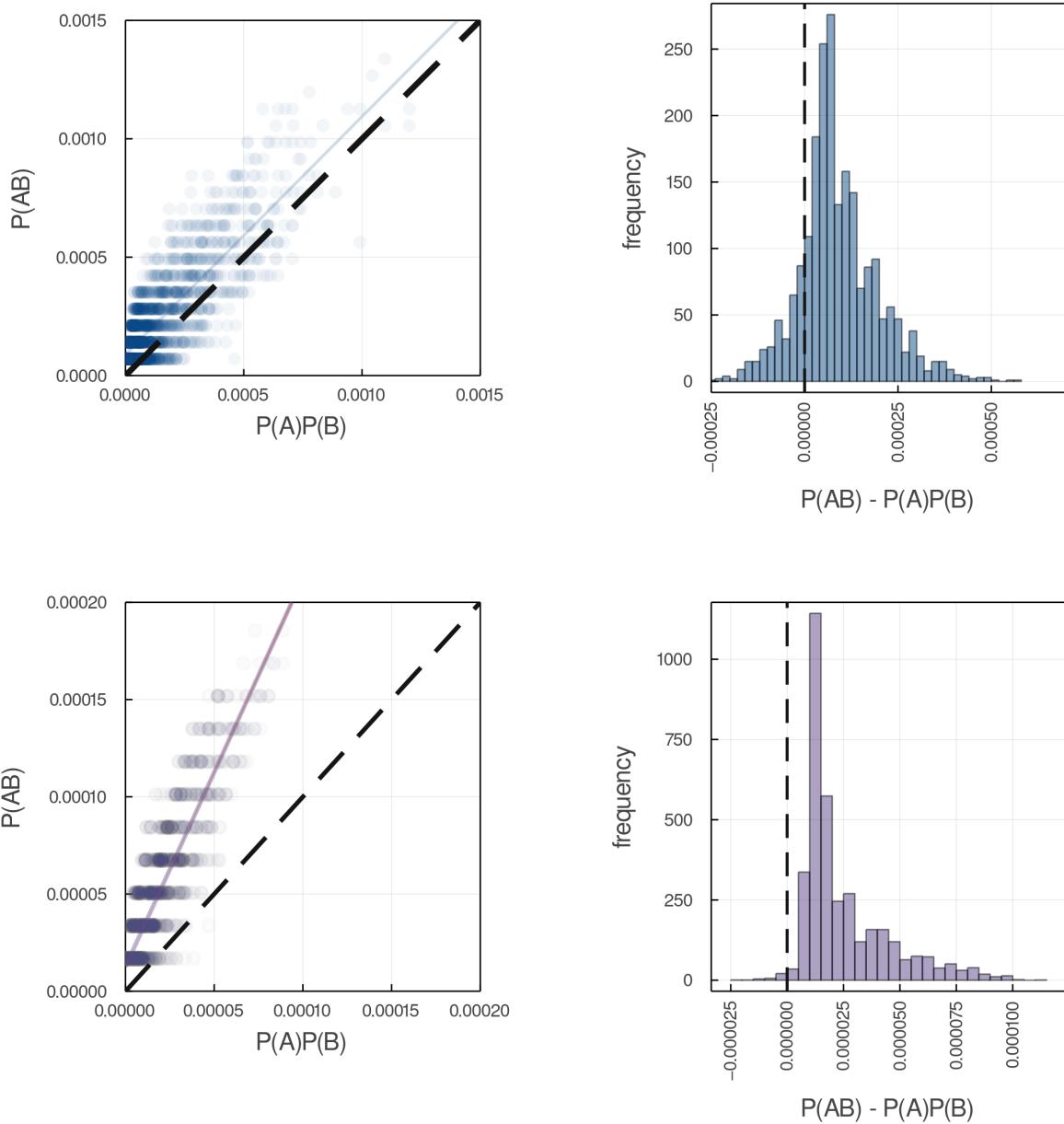


Figure 4: f

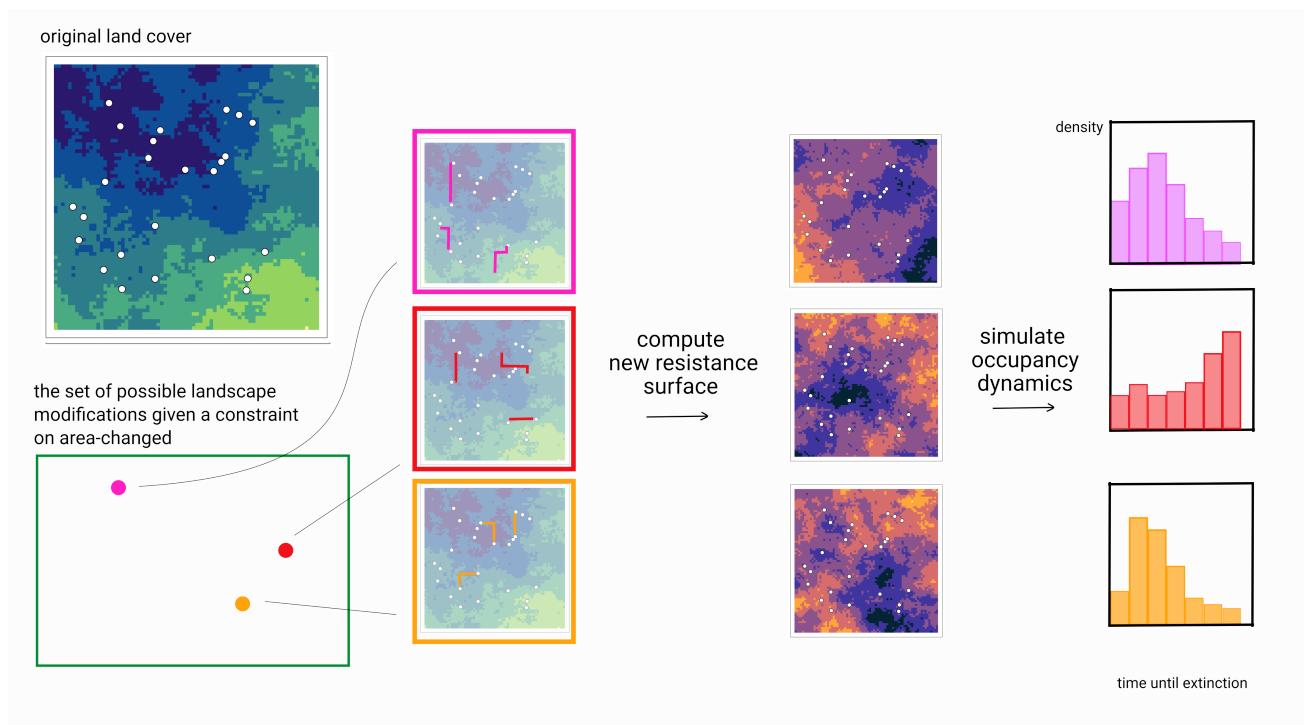


Figure 5: fig

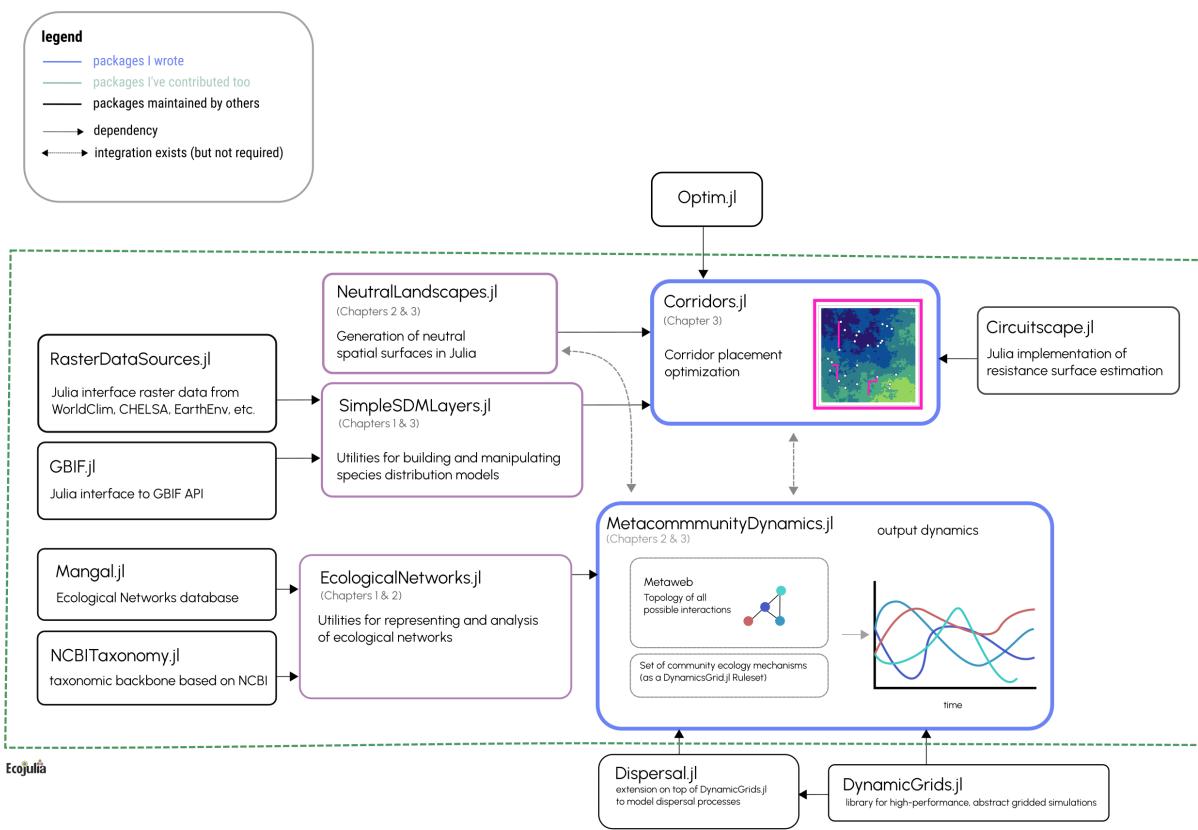


Figure 6: The structure of the software libraries used as part of MCD.jl

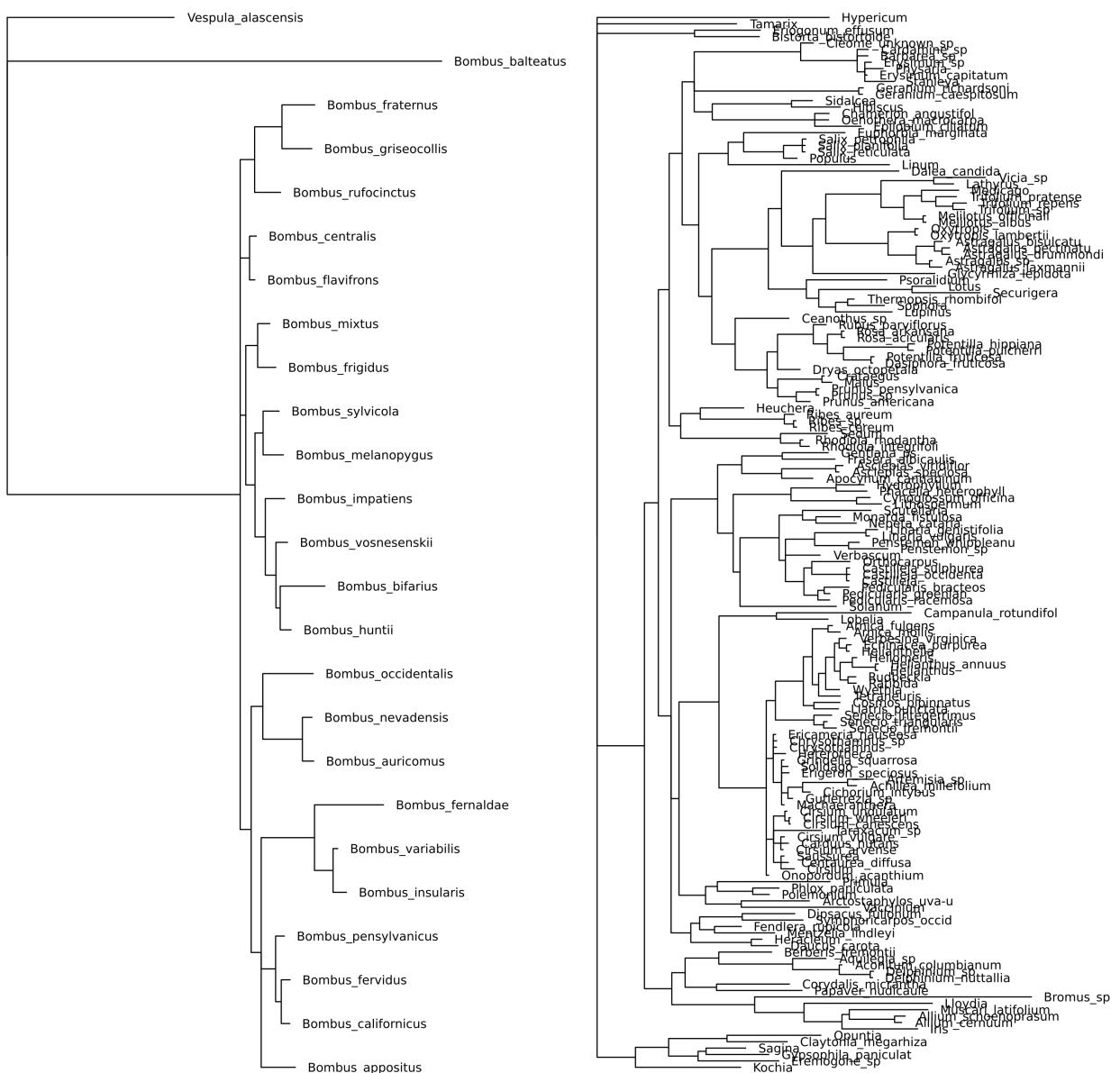


Figure 7: trees