

# Thesis proposal

Michael D. Catchen<sup>1,2</sup>

<sup>1</sup> McGill University    <sup>2</sup> Québec Centre for Biodiversity Sciences

## Correspondance to:

Michael D. Catchen — [michael.catchen@mail.mcgill.ca](mailto: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 P1**

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

### **17 P2**

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

30 **P3**

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

43 **P4**

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

56 **P6**

57 But forecasting is only half the story—if indeed “[ecologists] have hitherto only interpreted the world in  
58 various ways; the point is to change it,” then once we have a forecast about how an ecosystem will change

59 in the future, what if this forecast predicts a critical ecosystem service will deteriorate? We are still left  
60 with the question, what do we in the time being to mitigate the negative consequences a forecast predicts?  
61 In this framing, mitigating the consequences of anthropogenic change on ecosystems becomes an  
62 optimization problem: given a forecast of the probability. We have some goal state for the future, and some  
63 estimate of what the state of the world will be given a set of actions. Frame optimization problem  
64 mathematically an introduce concept of solution-space and constraint.

65 [Figure 1 about here.]

66 **P7**

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

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

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

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

83 [Figure 2 about here.]

84 **Data**

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

93 **Methods**

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

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

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

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

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

104 2) Make it spatial based on distributions.

105 3) Forecast distributions based on CMIP6

106 **Preliminary Results**

107 1) we got a tree and SDMs

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

## 109 **Chapter Two: Optimizing spatial sampling of species interactions**

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

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

119 [Figure 3 about here.]

120 We use a log-normal distribution as a null model of the relative-abundance distribution (Hubbell 2001) to  
121 simulate realized false-negative rate as a function of varying sampling effort.

122 This also goes on to includes testing some assumptions of the model with empirical data fig. 4, which we  
123 analytically show that our neutral model, if anything, underestimates the probability of false-negatives  
124 due to positive correlations in co-occurrence in two spatially replicated networks (Thompson & Townsend  
125 2000; Hadfield *et al.* 2014)—further I'm planning to add the field data from the previous chapter into this  
126 analysis once available.

127 [Figure 4 about here.]

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

131 **Chapter Three: Optimizing corridor placement against ecological**  
132 **dynamics**

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

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

142 [Figure 5 about here.]

143 **Methods**

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

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

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

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

159 [Figure 6 about here.]

160 **Conclusion**

161 **Appendix**

162 [Figure 7 about here.]

163 **References**

- 164 Bauer, P., Thorpe, A. & Brunet, G. (2015). The quiet revolution of numerical weather prediction. *Nature*,  
165 525, 47–56.
- 166 Beckage, B., Gross, L.J. & Kauffman, S. (2011). The limits to prediction in ecological systems. *Ecosphere*, 2,  
167 art125.
- 168 Blanchet, F.G., Cazelles, K. & Gravel, D. (2020). Co-occurrence is not evidence of ecological interactions.  
169 *Ecology Letters*, 23, 1050–1063.
- 170 Chen, Y., Angulo, M.T. & Liu, Y.-Y. (2019). Revealing Complex Ecological Dynamics via Symbolic  
171 Regression. *BioEssays*, 41, 1900069.
- 172 Dietze, M.C. (2017). Prediction in ecology: A first-principles framework. *Ecological Applications*, 27,  
173 2048–2060.

- 174 Hadfield, J.D., Krasnov, B.R., Poulin, R. & Nakagawa, S. (2014). A Tale of Two Phylogenies: Comparative  
175 Analyses of Ecological Interactions. *The American Naturalist*, 183, 174–187.
- 176 Hanski, I. & Ovaskainen, O. (2000). The metapopulation capacity of a fragmented landscape. *Nature*, 404,  
177 755–758.
- 178 Hill, C., DeLuca, C., Balaji, Suarez, M. & Da Silva, A. (2004). The architecture of the Earth System  
179 Modeling Framework. *Computing in Science Engineering*, 6, 18–28.
- 180 Hubbell, S.P. (2001). *The unified neutral theory of biodiversity and biogeography*. Monographs in  
181 population biology. Princeton University Press, Princeton.
- 182 Levin, S.A. (1992). The Problem of Pattern and Scale in Ecology: The Robert H. MacArthur Award  
183 Lecture. *Ecology*, 73, 1943–1967.
- 184 Makiola, A., Compson, Z.G., Baird, D.J., Barnes, M.A., Boerlijst, S.P., Bouchez, A., et al. (2020). Key  
185 Questions for Next-Generation Biomonitoring. *Frontiers in Environmental Science*, 7.
- 186 Ovaskainen, O., Sato, K., Bascompte, J. & Hanski, I. (2002). Metapopulation Models for Extinction  
187 Threshold in Spatially Correlated Landscapes. *Journal of Theoretical Biology*, 215, 95–108.
- 188 Ovaskainen, O., Sato, K., Bascompte, J. & Hanski, I. (2002). Metapopulation Models for Extinction  
189 Threshold in Spatially Correlated Landscapes. *Journal of Theoretical Biology*, 215, 95–108.
- 190 Pennekamp, F., Iles, A.C., Garland, J., Brennan, G., Brose, U., Gaedke, U., et al. (2019). The intrinsic  
191 predictability of ecological time series and its potential to guide forecasting. *Ecological Monographs*, 89,  
192 e01359.
- 193 Petchey, O.L., Pontarp, M., Massie, T.M., Kéfi, S., Ozgul, A., Weilenmann, M., et al. (2015). The ecological  
194 forecast horizon, and examples of its uses and determinants. *Ecology Letters*, 18, 597–611.
- 195 Peterman, W.E. (2018). ResistanceGA: An R package for the optimization of resistance surfaces using  
196 genetic algorithms. *Methods in Ecology and Evolution*, 9, 1638–1647.
- 197 Strydom, T., Catchen, M.D., Banville, F., Caron, D., Dansereau, G., Desjardins-Proulx, P., et al. (2021). *A  
198 Roadmap Toward Predicting Species Interaction Networks (Across Space and Time)* (Preprint).  
199 EcoEvoRxiv.
- 200 Thompson, R.M. & Townsend, C.R. (2000). Is resolution the solution?: The effect of taxonomic resolution  
201 on the calculated properties of three stream food webs. *Freshwater Biology*, 44, 413–422.

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

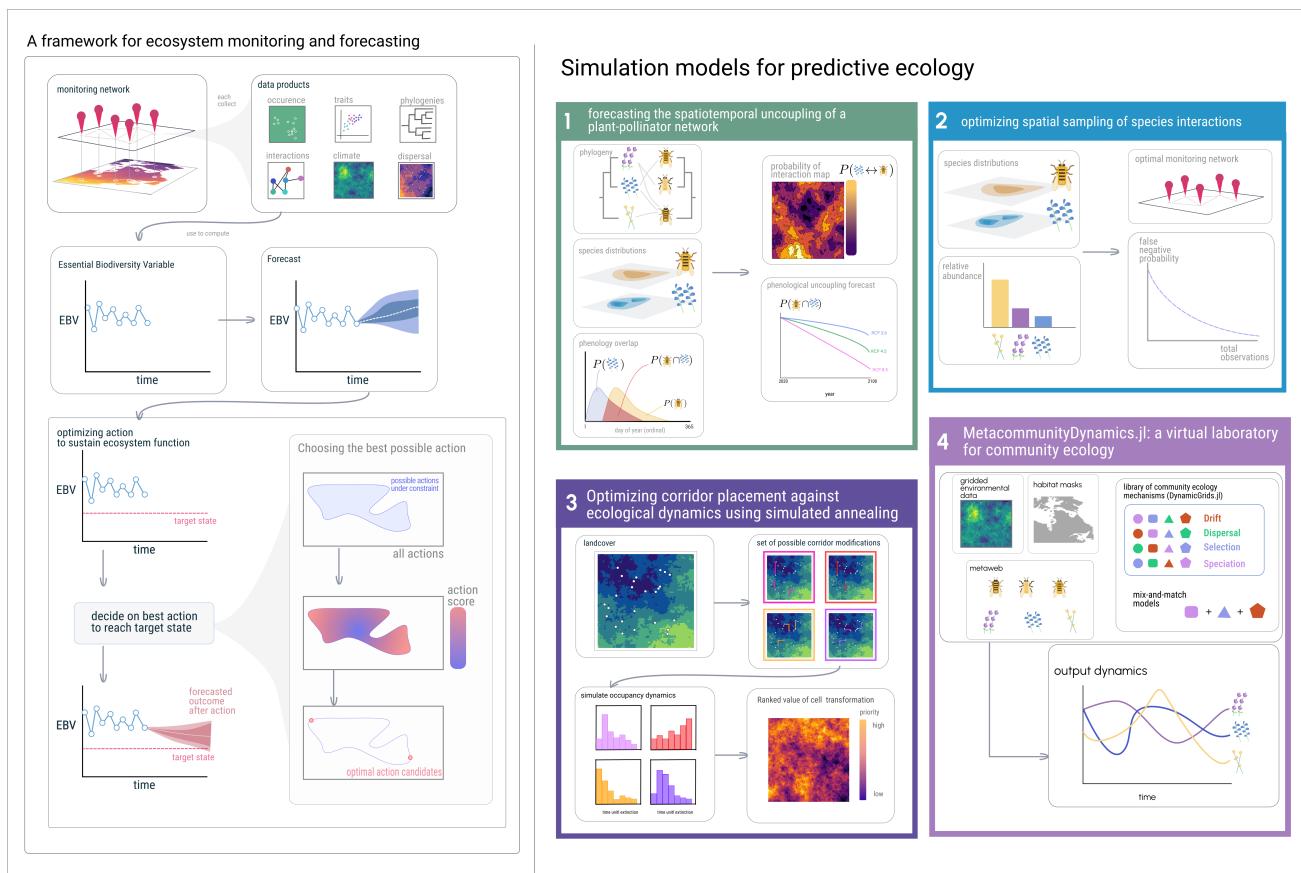


Figure 1: thesis concept

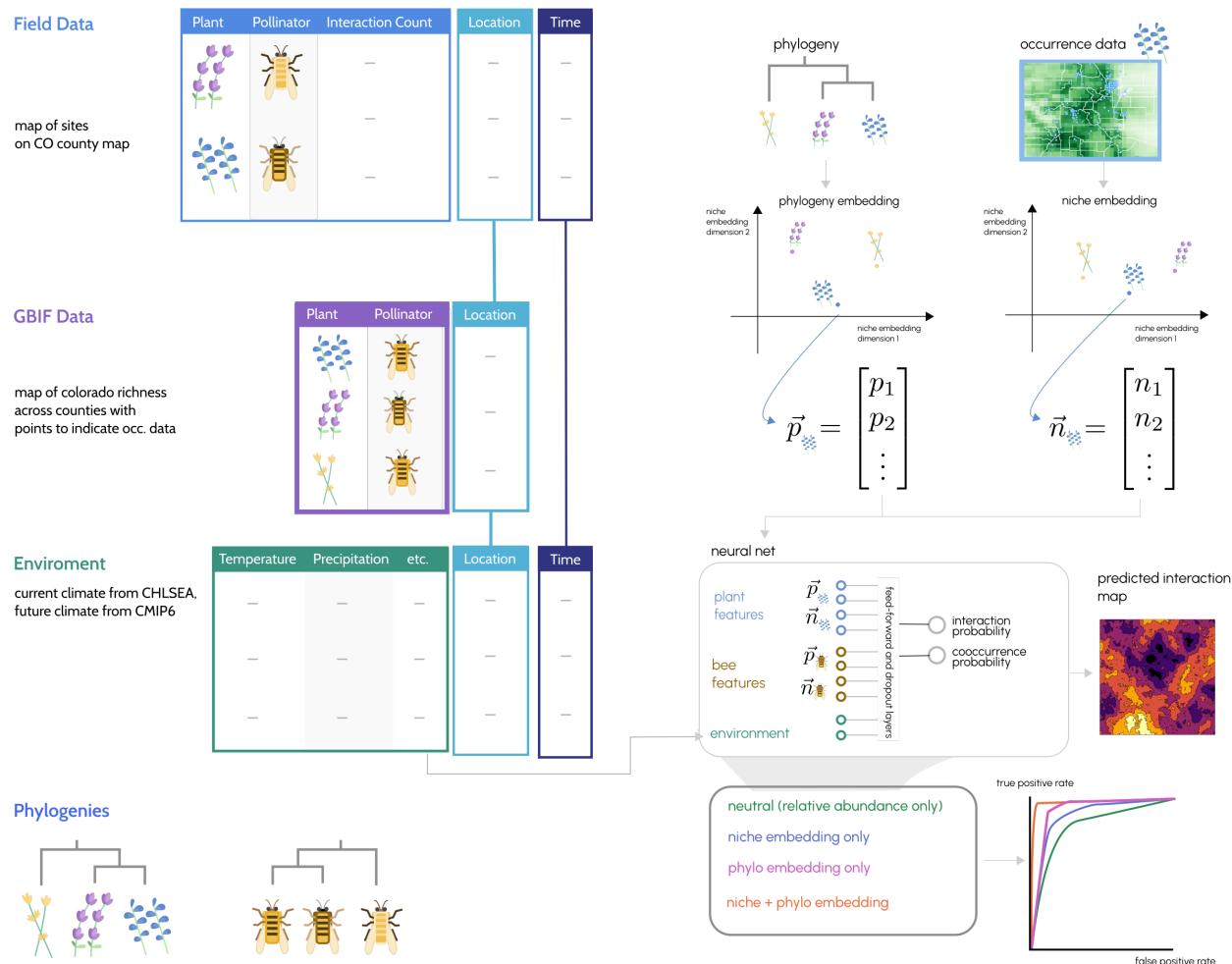


Figure 2: chapter one concept fig

Species A occurs?

		true		false
		true	Species A observed?	
		true	Species B observed? true co-occurrence <b>true-positive</b> Interaction observed? true      false interaction <b>true-positive</b> interaction <b>false-negative</b>	co-occurrence <b>false-negative</b>
		false	co-occurrence <b>false-negative</b>	occurrence <b>false-negative</b>
false		co-occurrence <b>true-negative</b>		occurrence <b>true-negative</b>

Figure 3: taxonomy of false negatives

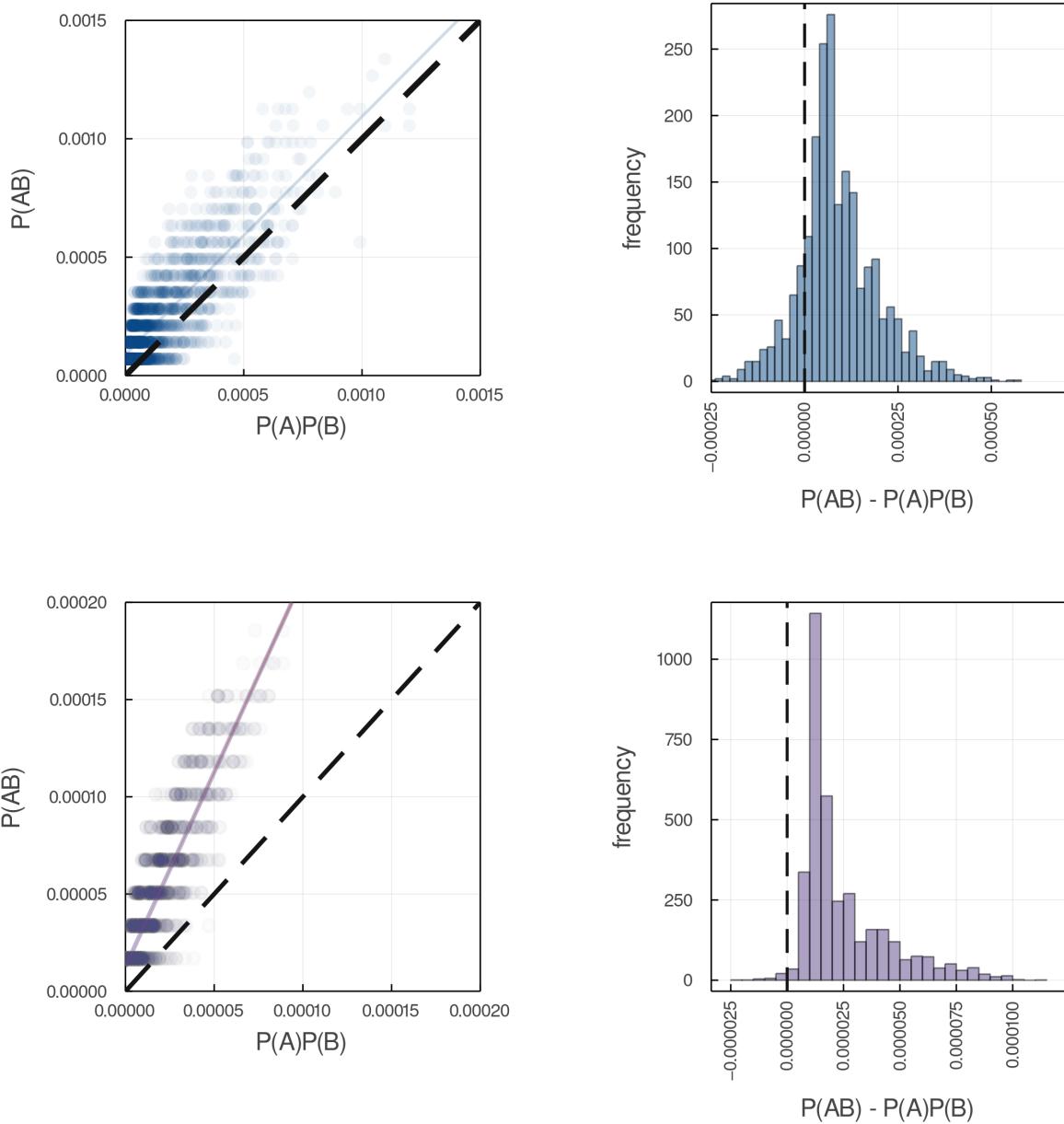


Figure 4: f

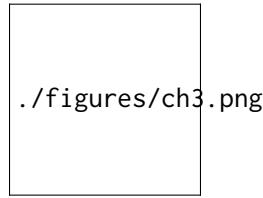


Figure 5: fig

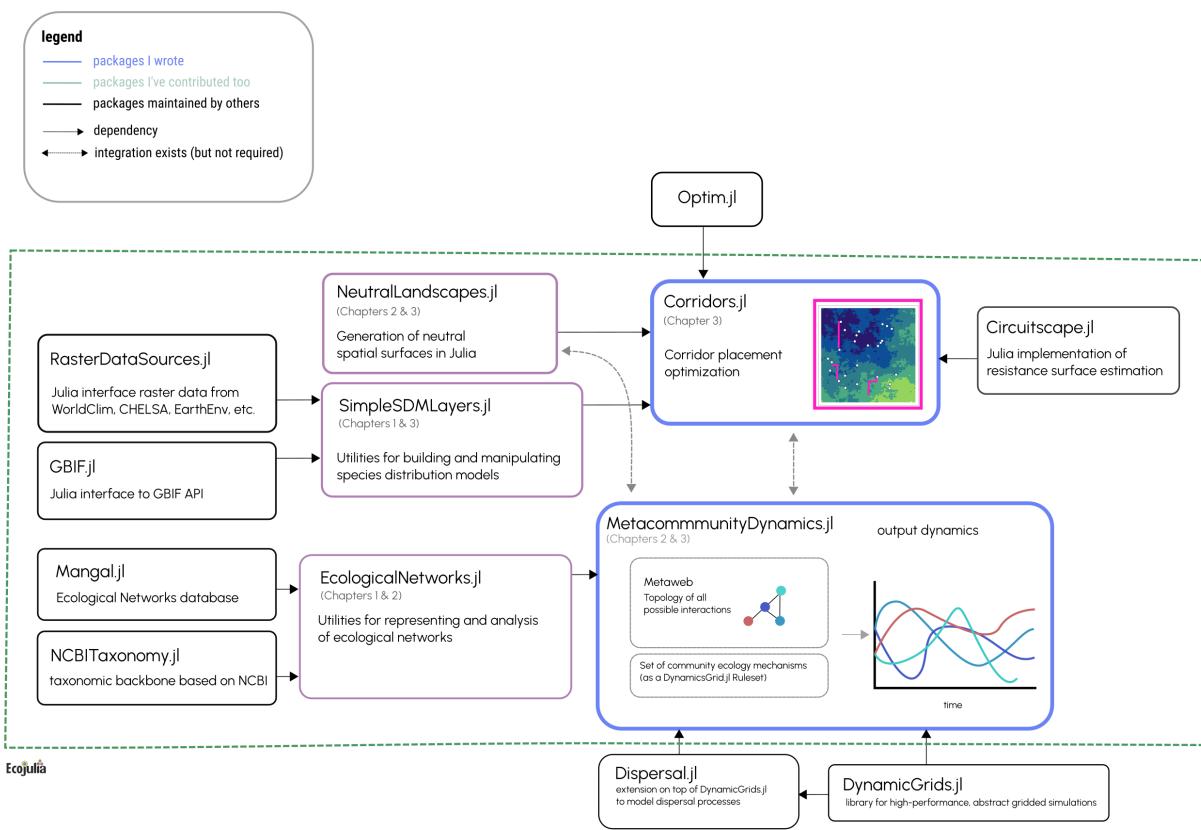


Figure 6: The structure of the software libraries used as part of MCD.jl

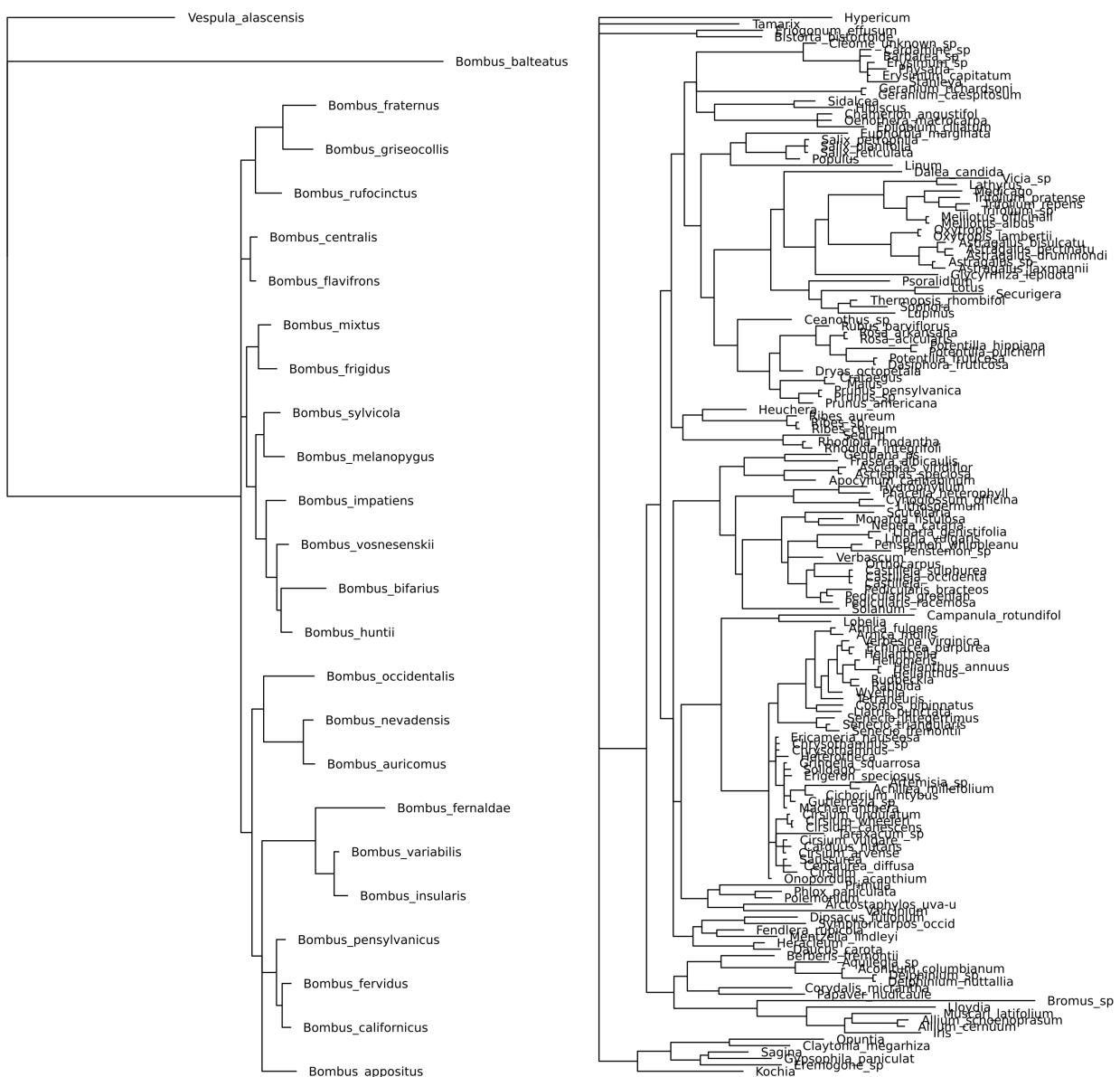


Figure 7: trees