

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 20, 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. Marx’s most well known quote that “philosophers have hitherto only
58 interpreted the world in various ways; the point is to change it.” Indeed, once we have a forecast about

59 how an ecosystem will change in the future, what if this forecast predicts a critical ecosystem service will
60 deteriorate? We are still left with the question, what do we in the time being to mitigate the negative
61 consequences a forecast predicts? In this framing, mitigating the consequences of anthropogenic change
62 on ecosystems becomes an optimization problem: given a forecast of the probability. We have some goal
63 state for the future, and some estimate of what the state of the world will be given a set of actions. Frame
64 optimization problem 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 Plants and pollinators form interaction networks, called the “architecture of biodiversity” (**Jordano2007?**).

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

76 The issue is that we don’t really know what interactions are like now. So not only do we need to predict
77 with data that is spatially and temporally sparse and likely to contain many interaction “false-negatives”
78 (Strydom *et al.* 2021).

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

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

85

[Figure 2 about here.]

86 **Data**

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

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

92 **Methods**

93 Split the process into parts.

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

96 **Preliminary Results**

97 1) we got a tree

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

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

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

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

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

105

[Figure 3 about here.]

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

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

113 [Figure 4 about here.]

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

117 **Results**

- 118 • nonrandom association figure sampling effort under neutral model

119 **Chapter Three: Optimizing corridor placement against ecological 120 dynamics**

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

128 This chapter proposes a general algorithm for optimizing corridor placement based on a measurement of
129 ecosystem functioning derived from simulations run on a proposed landscape modification. We propose

130 various landscape modifications which alter the cover of a landscape, represented as a raster (fig. 6, left).
131 We then compute a new resistance surface based on the proposed landscape modification, and based on
132 the values of resistance to dispersal between each location we simulate spatially-explicit metapopulation
133 dynamics model (Hanski & Ovaskainen 2000; Ovaskainen *et al.* 2002) to estimate a distribution of time
134 until extinction for each landscape modification (fig. 6, right).

135 **Methods**

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

138 **CH4 a software note on the resulting packages.**

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

142 [Figure 5 about here.]

143 **Conclusion**

144 **Appendix**

145 [Figure 6 about here.]

146 **References**

- 147 Bauer, P., Thorpe, A. & Brunet, G. (2015). The quiet revolution of numerical weather prediction. *Nature*,
148 525, 47–56.
149 Beckage, B., Gross, L.J. & Kauffman, S. (2011). The limits to prediction in ecological systems. *Ecosphere*, 2,
150 art125.

- 151 Chen, Y., Angulo, M.T. & Liu, Y.-Y. (2019). Revealing Complex Ecological Dynamics via Symbolic
152 Regression. *BioEssays*, 41, 1900069.
- 153 Dietze, M.C. (2017). Prediction in ecology: A first-principles framework. *Ecological Applications*, 27,
154 2048–2060.
- 155 Hadfield, J.D., Krasnov, B.R., Poulin, R. & Nakagawa, S. (2014). A Tale of Two Phylogenies: Comparative
156 Analyses of Ecological Interactions. *The American Naturalist*, 183, 174–187.
- 157 Hanski, I. & Ovaskainen, O. (2000). The metapopulation capacity of a fragmented landscape. *Nature*, 404,
158 755–758.
- 159 Hill, C., DeLuca, C., Balaji, Suarez, M. & Da Silva, A. (2004). The architecture of the Earth System
160 Modeling Framework. *Computing in Science Engineering*, 6, 18–28.
- 161 Hubbell, S.P. (2001). *The unified neutral theory of biodiversity and biogeography*. Monographs in
162 population biology. Princeton University Press, Princeton.
- 163 Levin, S.A. (1992). The Problem of Pattern and Scale in Ecology: The Robert H. MacArthur Award
164 Lecture. *Ecology*, 73, 1943–1967.
- 165 Makiola, A., Compson, Z.G., Baird, D.J., Barnes, M.A., Boerlijst, S.P., Bouchez, A., et al. (2020). Key
166 Questions for Next-Generation Biomonitoring. *Frontiers in Environmental Science*, 7.
- 167 Ovaskainen, O., Sato, K., Bascompte, J. & Hanski, I. (2002). Metapopulation Models for Extinction
168 Threshold in Spatially Correlated Landscapes. *Journal of Theoretical Biology*, 215, 95–108.
- 169 Ovaskainen, O., Sato, K., Bascompte, J. & Hanski, I. (2002). Metapopulation Models for Extinction
170 Threshold in Spatially Correlated Landscapes. *Journal of Theoretical Biology*, 215, 95–108.
- 171 Pennekamp, F., Iles, A.C., Garland, J., Brennan, G., Brose, U., Gaedke, U., et al. (2019). The intrinsic
172 predictability of ecological time series and its potential to guide forecasting. *Ecological Monographs*, 89,
173 e01359.
- 174 Petchey, O.L., Pontarp, M., Massie, T.M., Kéfi, S., Ozgul, A., Weilenmann, M., et al. (2015). The ecological
175 forecast horizon, and examples of its uses and determinants. *Ecology Letters*, 18, 597–611.
- 176 Peterman, W.E. (2018). ResistanceGA: An R package for the optimization of resistance surfaces using
177 genetic algorithms. *Methods in Ecology and Evolution*, 9, 1638–1647.

- 178 Strydom, T., Catchen, M.D., Banville, F., Caron, D., Dansereau, G., Desjardins-Proulx, P., *et al.* (2021). *A*
179 *Roadmap Toward Predicting Species Interaction Networks (Across Space and Time)* (Preprint).
180 EcoEvoRxiv.
- 181 Thompson, R.M. & Townsend, C.R. (2000). Is resolution the solution?: The effect of taxonomic resolution
182 on the calculated properties of three stream food webs. *Freshwater Biology*, 44, 413–422.
- 183 Urban, M.C., Travis, J.M.J., Zurell, D., Thompson, P.L., Synes, N.W., Scarpa, A., *et al.* (2021). Coding for
184 Life: Designing a Platform for Projecting and Protecting Global Biodiversity. *BioScience*.

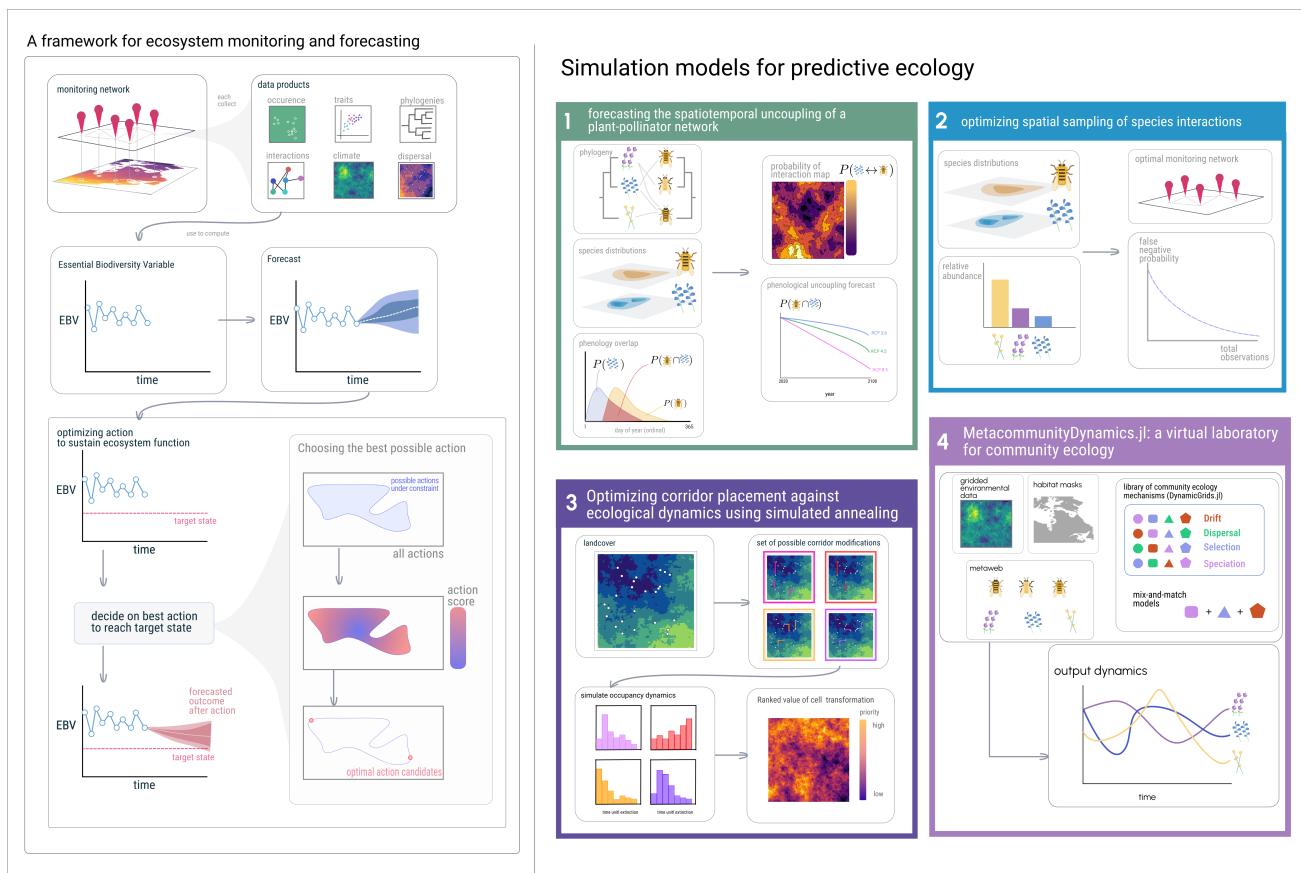


Figure 1: thesis concept

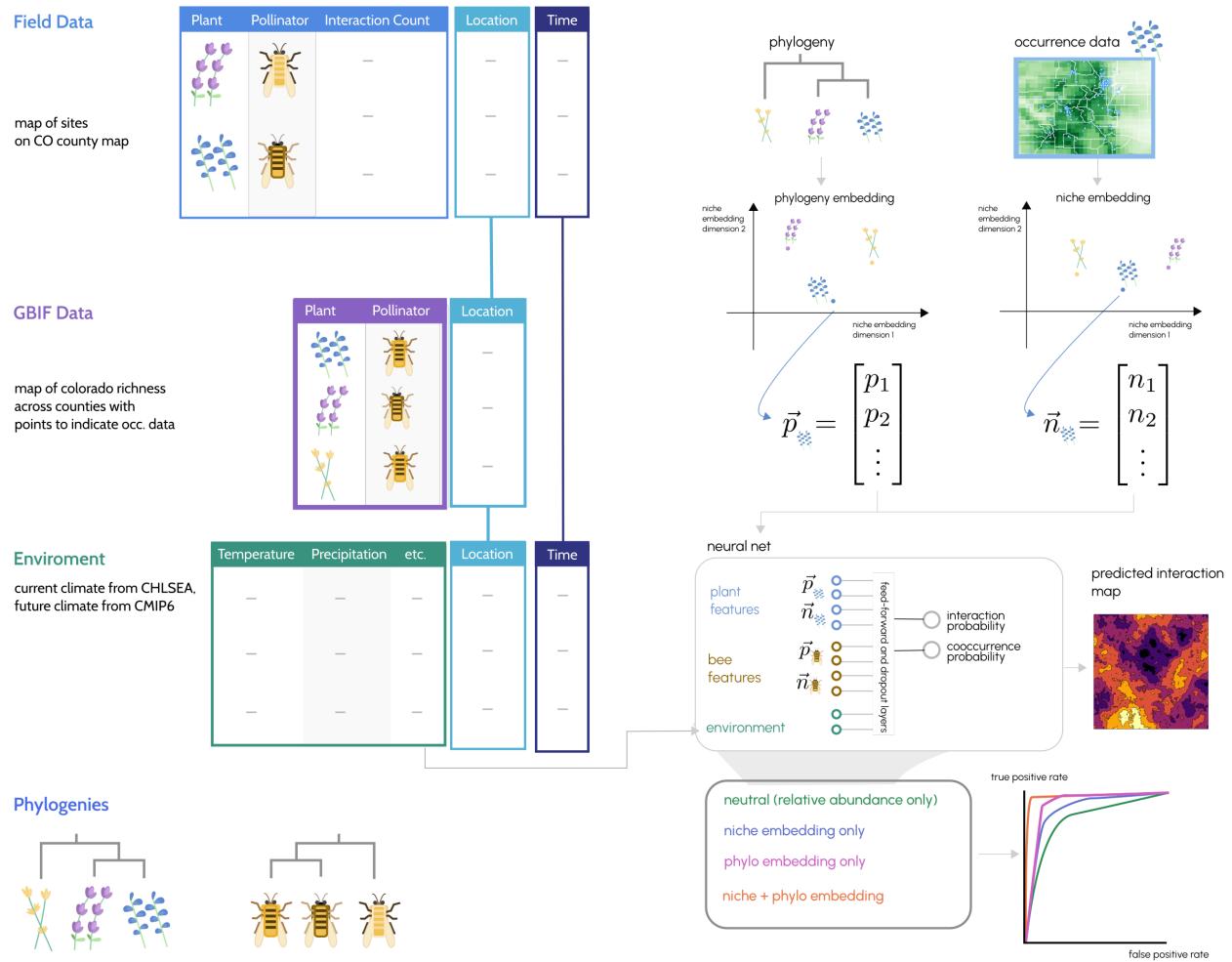


Figure 2: chapter one concept fig

Species A occurs?

		true		false								
		true	Species A observed?									
		true	Species B observed? <table border="1"> <tr> <td>true</td><td>co-occurrence true-positive</td><td>Interaction observed? true false</td><td>co-occurrence false-negative</td></tr> <tr> <td>false</td><td>co-occurrence false-negative</td><td>interaction true-positive interaction false-negative</td><td>occurrence false-negative</td></tr> </table> <th></th>	true	co-occurrence true-positive	Interaction observed? true false	co-occurrence false-negative	false	co-occurrence false-negative	interaction true-positive interaction false-negative	occurrence false-negative	
true	co-occurrence true-positive	Interaction observed? true false	co-occurrence false-negative									
false	co-occurrence false-negative	interaction true-positive interaction false-negative	occurrence false-negative									
Species B occurs?	true	co-occurrence true-negative		occurrence true-negative								
false												

Figure 3: taxonomy of false negatives

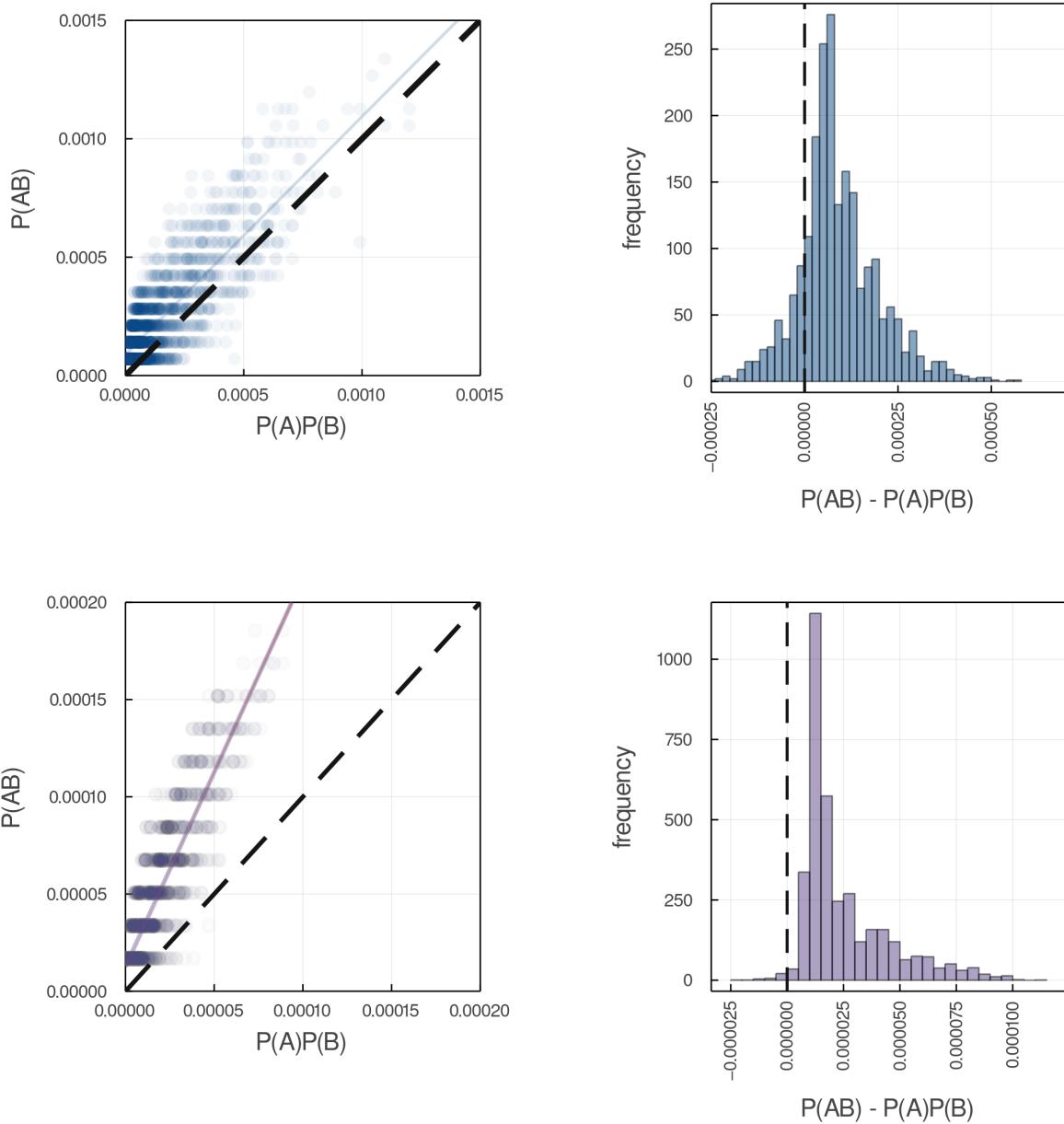


Figure 4: f

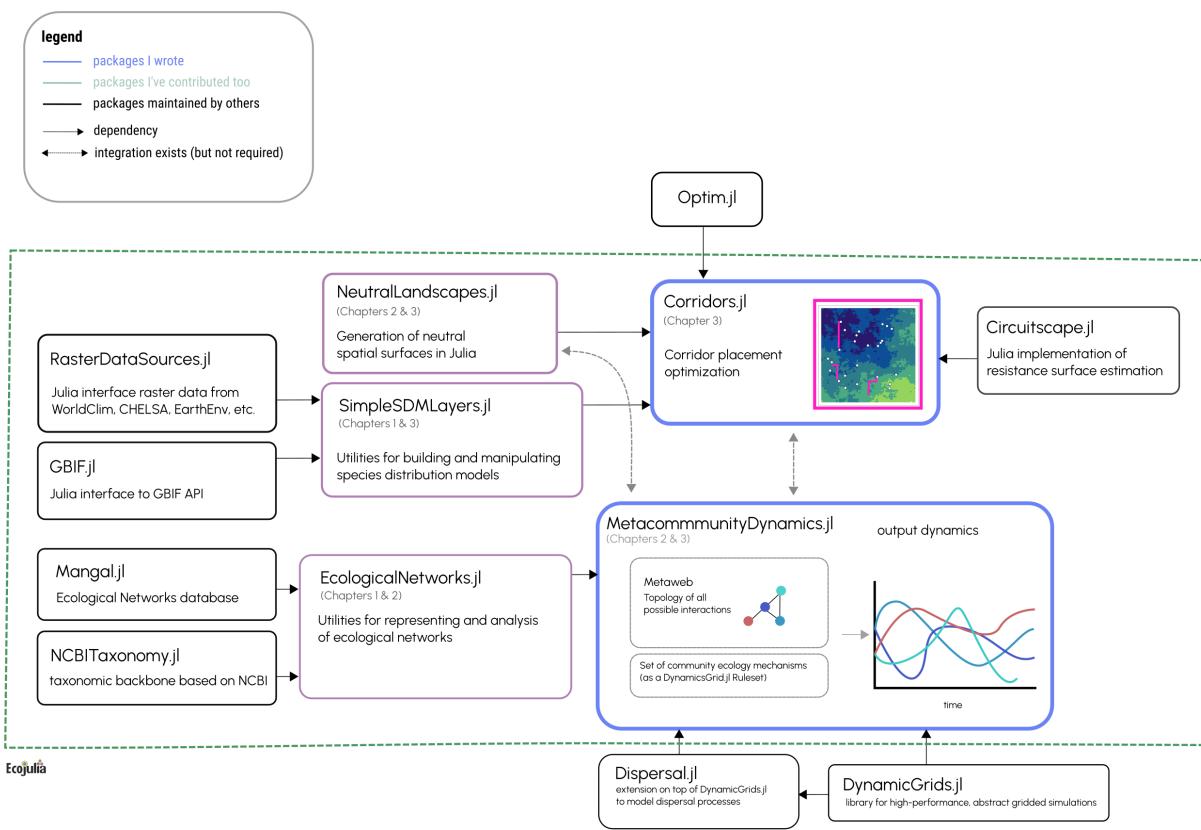


Figure 5: todo

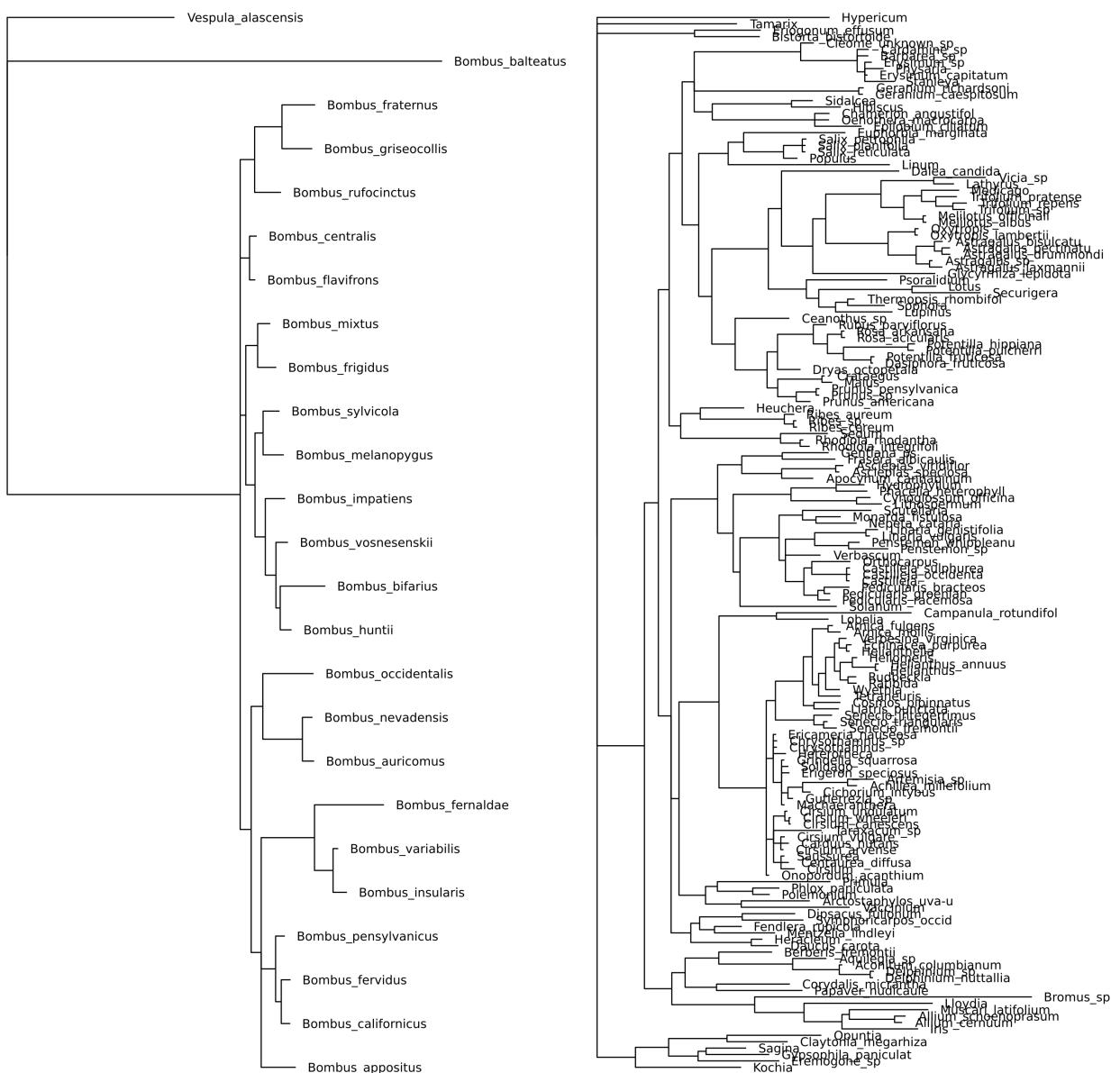


Figure 6: trees