

# 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 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 create networks of interactions, which together form the  
66 “architecture of biodiversity” (Bascompte & Jordano 2007). The functioning and stability of ecosystems  
67 emerge from these interactions, but anthropogenic change threatens to unravel and “rewire” these  
68 interaction networks (CaraDonna *et al.* 2017), threatening the persistence of these systems.

69 Plant-pollinator networks face two possible forms of rewiring in response to anthropogenic environmental  
70 change: spatial and temporal. Spatially, range shifts could cause interacting species to no longer overlap in  
71 space, and shifts in phenology could cause interacting species to no longer occur at the same time of year.  
72 This chapter uses several years of data on bumblebee-flower phenology and interactions across several  
73 field sites, each consisting of several plots across an elevational gradient, combined with spatial records of  
74 species occurrence via GBIF to forecast this uncoupling. This addresses the EBV to forecast element of the  
75 flow from data to mitigation 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 Colorado 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.* 2021b), we begin by predicting a metaweb of interactions across Colorado as they exist *in the present*.  
89 We do this using a set of candidate interaction prediction models: relative abundance only, phylogenetic  
90 embedding only (a la Strydom *et al.* (2021a)), niche embedding only (Gravel *et al.* 2019), and all pairwise  
91 combinations of those constituent models. After validating each model and selecting the best performing  
92 model, we then predict how these distributions of each of these species will change under the CMIP6  
93 consensus climate forecast [@], and then finally quantify the reduction in spatial between species for  
94 which there is a predicted interaction.

95 **Preliminary Results**

96 1) we got a tree and SDMs. See appendix figure.

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

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

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

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

107 interactions. Co-occurrence is not the same thing as interaction (Blanchet *et al.* 2020), but often is used as  
108 a proxy.

109 [Figure 3 about here.]

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

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

117 [Figure 4 about here.]

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

## 121 **Chapter Three: Optimizing corridor placement against ecological 122 dynamics**

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

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

132

[Figure 5 about here.]

133 **Methods**

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

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

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

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

148

[Figure 6 about here.]

149 **Conclusion**

150 **Appendix**

151

[Figure 7 about here.]

152 **References**

- 153 Bascompte, J. & Jordano, P. (2007). Plant-Animal Mutualistic Networks: The Architecture of Biodiversity.  
154 *Annual Review of Ecology, Evolution, and Systematics*, 38, 567–593.
- 155 Bauer, P., Thorpe, A. & Brunet, G. (2015). The quiet revolution of numerical weather prediction. *Nature*,  
156 525, 47–56.
- 157 Beckage, B., Gross, L.J. & Kauffman, S. (2011). The limits to prediction in ecological systems. *Ecosphere*, 2,  
158 art125.
- 159 Blanchet, F.G., Cazelles, K. & Gravel, D. (2020). Co-occurrence is not evidence of ecological interactions.  
160 *Ecology Letters*, 23, 1050–1063.
- 161 CaraDonna, P.J., Petry, W.K., Brennan, R.M., Cunningham, J.L., Bronstein, J.L., Waser, N.M., *et al.* (2017).  
162 Interaction rewiring and the rapid turnover of plantpollinator networks. *Ecology Letters*, 20, 385–394.
- 163 Chen, Y., Angulo, M.T. & Liu, Y.-Y. (2019). Revealing Complex Ecological Dynamics via Symbolic  
164 Regression. *BioEssays*, 41, 1900069.
- 165 Dietze, M.C. (2017). Prediction in ecology: A first-principles framework. *Ecological Applications*, 27,  
166 2048–2060.
- 167 Gravel, D., Baiser, B., Dunne, J.A., Kopelke, J.-P., Martinez, N.D., Nyman, T., *et al.* (2019). Bringing Elton  
168 and Grinnell together: A quantitative framework to represent the biogeography of ecological  
169 interaction networks. *Ecography*, 42, 401–415.
- 170 Hadfield, J.D., Krasnov, B.R., Poulin, R. & Nakagawa, S. (2014). A Tale of Two Phylogenies: Comparative  
171 Analyses of Ecological Interactions. *The American Naturalist*, 183, 174–187.
- 172 Hanski, I. & Ovaskainen, O. (2000). The metapopulation capacity of a fragmented landscape. *Nature*, 404,  
173 755–758.
- 174 Hill, C., DeLuca, C., Balaji, Suarez, M. & Da Silva, A. (2004). The architecture of the Earth System  
175 Modeling Framework. *Computing in Science Engineering*, 6, 18–28.
- 176 Hubbell, S.P. (2001). *The unified neutral theory of biodiversity and biogeography*. Monographs in  
177 population biology. Princeton University Press, Princeton.

- 178 Levin, S.A. (1992). The Problem of Pattern and Scale in Ecology: The Robert H. MacArthur Award  
179 Lecture. *Ecology*, 73, 1943–1967.
- 180 Makiola, A., Compson, Z.G., Baird, D.J., Barnes, M.A., Boerlijst, S.P., Bouchez, A., et al. (2020). Key  
181 Questions for Next-Generation Biomonitoring. *Frontiers in Environmental Science*, 7.
- 182 Ovaskainen, O., Sato, K., Bascompte, J. & Hanski, I. (2002). Metapopulation Models for Extinction  
183 Threshold in Spatially Correlated Landscapes. *Journal of Theoretical Biology*, 215, 95–108.
- 184 Ovaskainen, O., Sato, K., Bascompte, J. & Hanski, I. (2002). Metapopulation Models for Extinction  
185 Threshold in Spatially Correlated Landscapes. *Journal of Theoretical Biology*, 215, 95–108.
- 186 Pennekamp, F., Iles, A.C., Garland, J., Brennan, G., Brose, U., Gaedke, U., et al. (2019). The intrinsic  
187 predictability of ecological time series and its potential to guide forecasting. *Ecological Monographs*, 89,  
188 e01359.
- 189 Petchey, O.L., Pontarp, M., Massie, T.M., Kéfi, S., Ozgul, A., Weilenmann, M., et al. (2015). The ecological  
190 forecast horizon, and examples of its uses and determinants. *Ecology Letters*, 18, 597–611.
- 191 Peterman, W.E. (2018). ResistanceGA: An R package for the optimization of resistance surfaces using  
192 genetic algorithms. *Methods in Ecology and Evolution*, 9, 1638–1647.
- 193 Strydom, T., Bouskila, S., Banville, F., Barros, C., Caron, D., Farrell, M.J., et al. (2021a). Food web  
194 reconstruction through phylogenetic transfer of low-rank network representation.
- 195 Strydom, T., Catchen, M.D., Banville, F., Caron, D., Dansereau, G., Desjardins-Proulx, P., et al. (2021b). A  
196 *Roadmap Toward Predicting Species Interaction Networks (Across Space and Time)* (Preprint).  
197 EcoEvoRxiv.
- 198 Thompson, R.M. & Townsend, C.R. (2000). Is resolution the solution?: The effect of taxonomic resolution  
199 on the calculated properties of three stream food webs. *Freshwater Biology*, 44, 413–422.
- 200 Urban, M.C., Travis, J.M.J., Zurell, D., Thompson, P.L., Synes, N.W., Scarpa, A., et al. (2021). Coding for  
201 Life: Designing a Platform for Projecting and Protecting Global Biodiversity. *BioScience*.

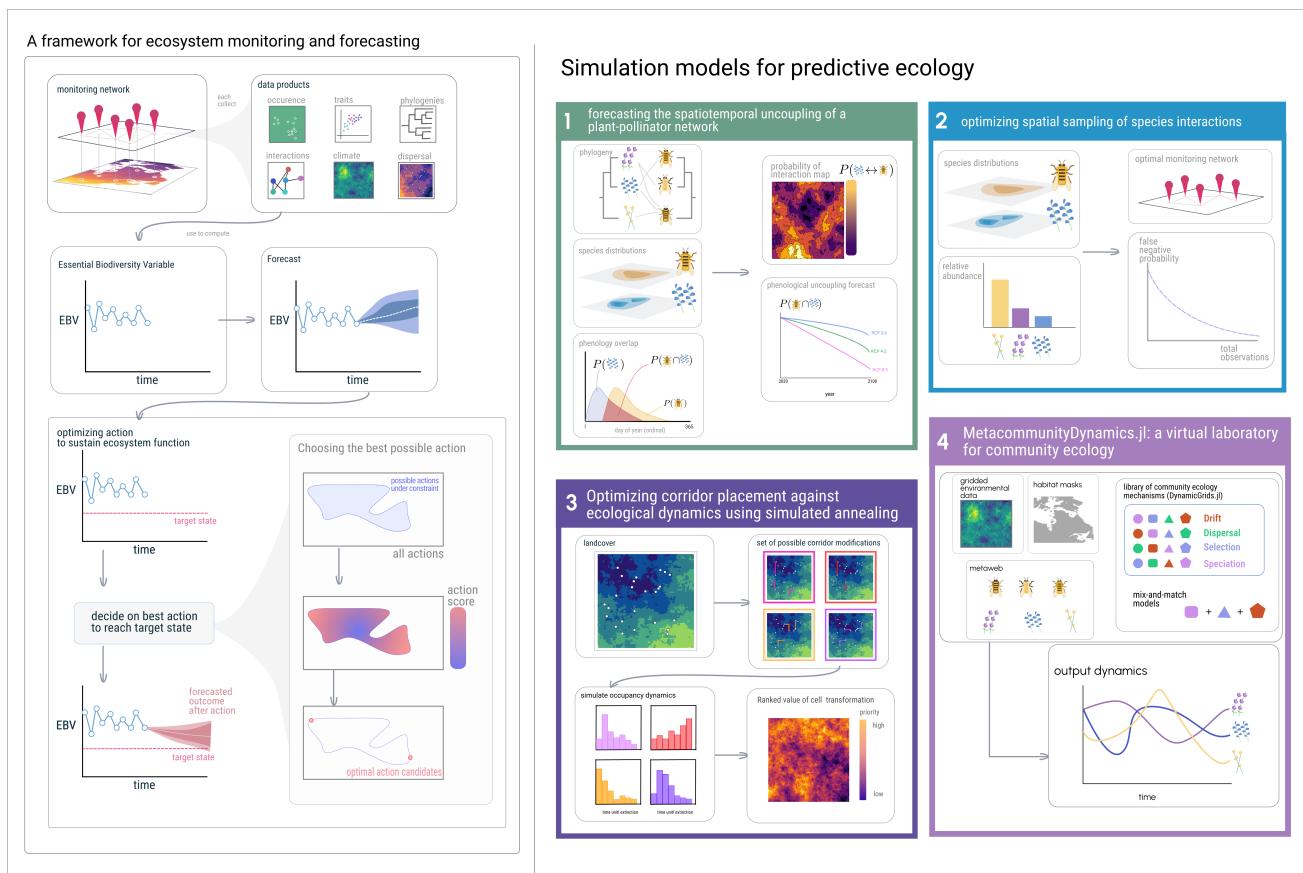


Figure 1: thesis concept

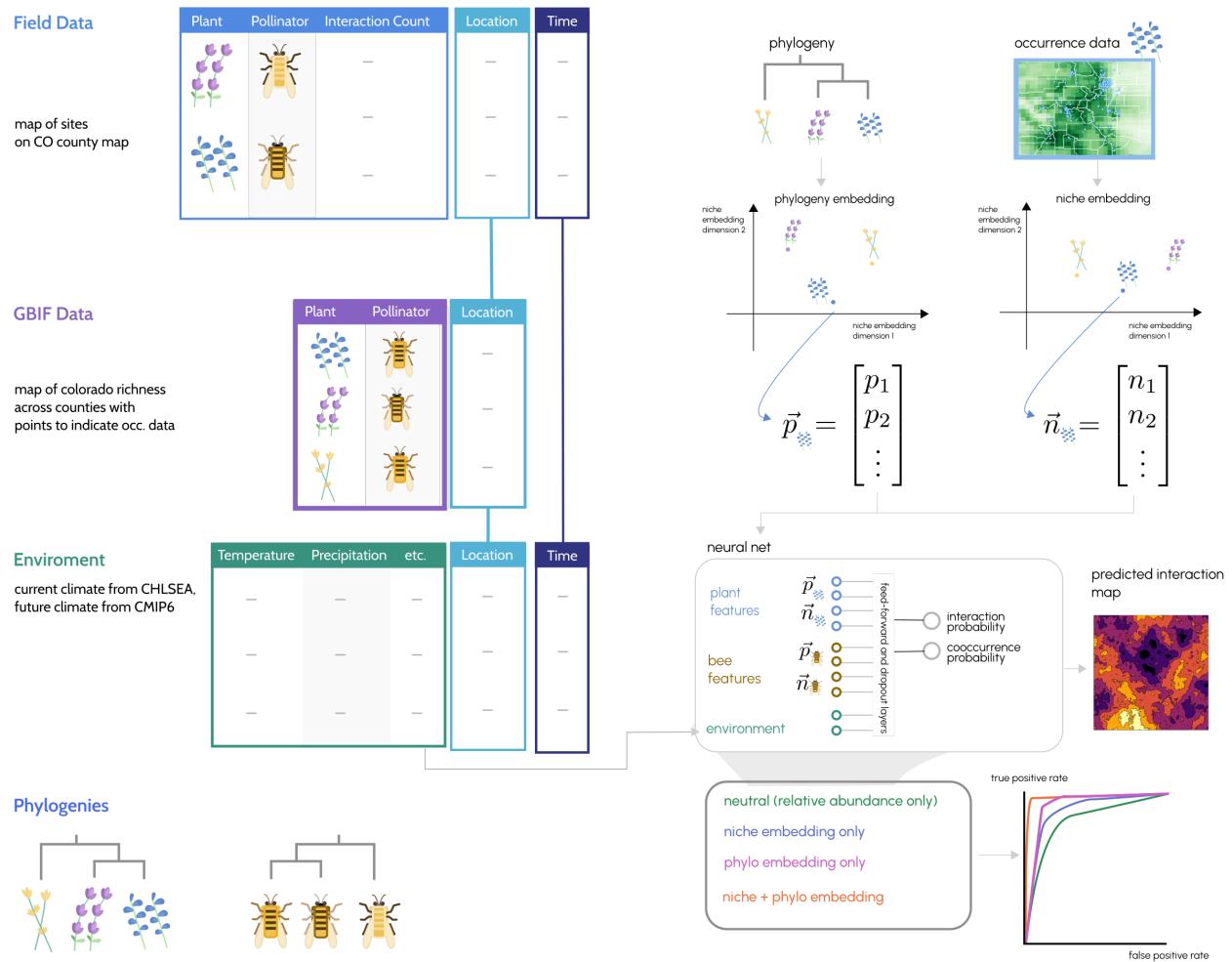


Figure 2: Chapter One conceptual figure. Left: the sources of data and how they can be synthesized. Right: The flow from data to interaction prediction using a few different interaction prediction models.

Species A occurs?

Species B occurs?

		true		false
		true	Species A observed?	
		true	Species A observed? true co-occurrence <b>true-positive</b> Interaction observed? true      false interaction <b>true-positive</b> interaction <b>false-negative</b>	co-occurrence <b>true-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: A taxonomy of occurrence, co-occurrence, and interaction false negatives in data

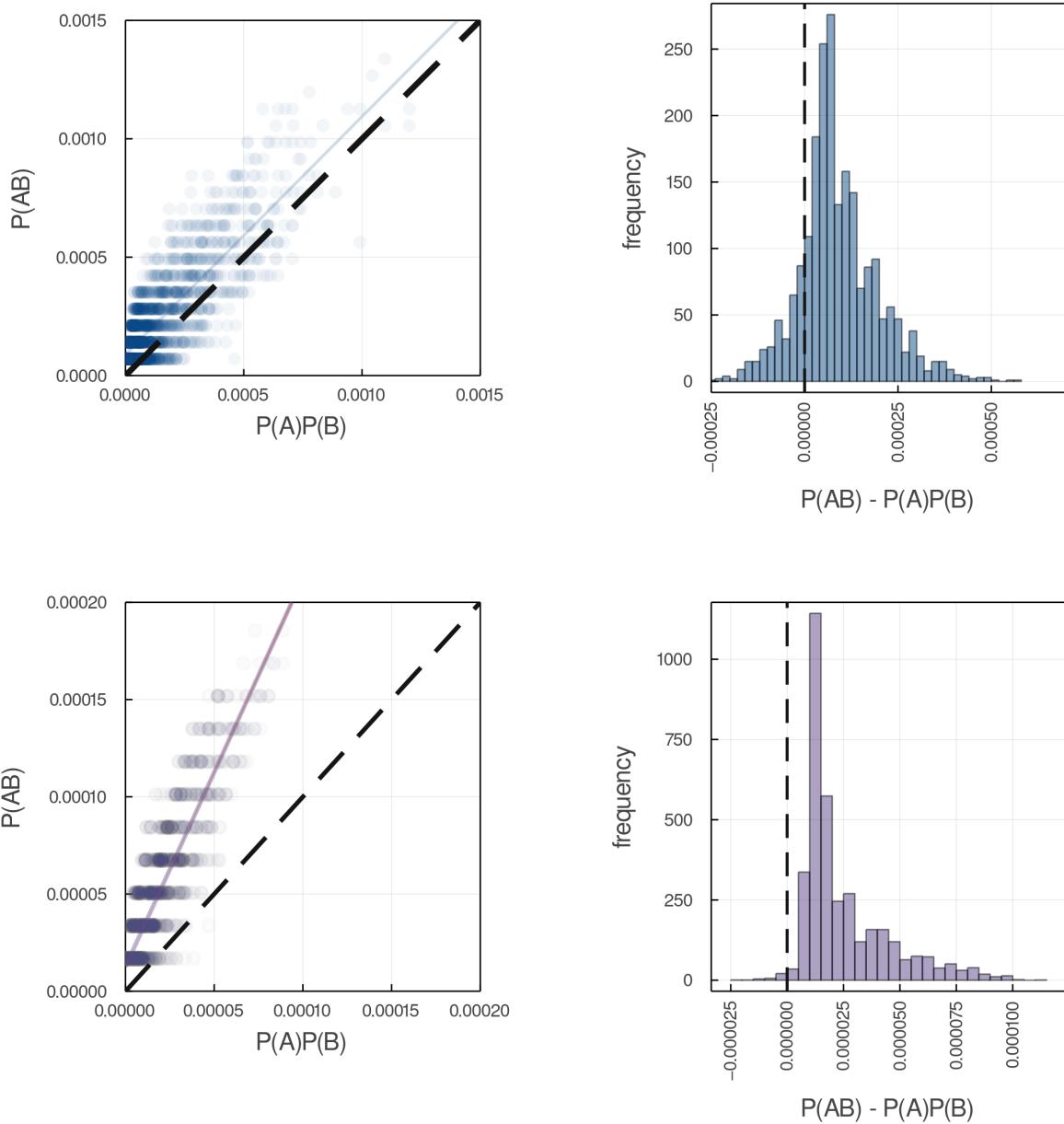


Figure 4: f

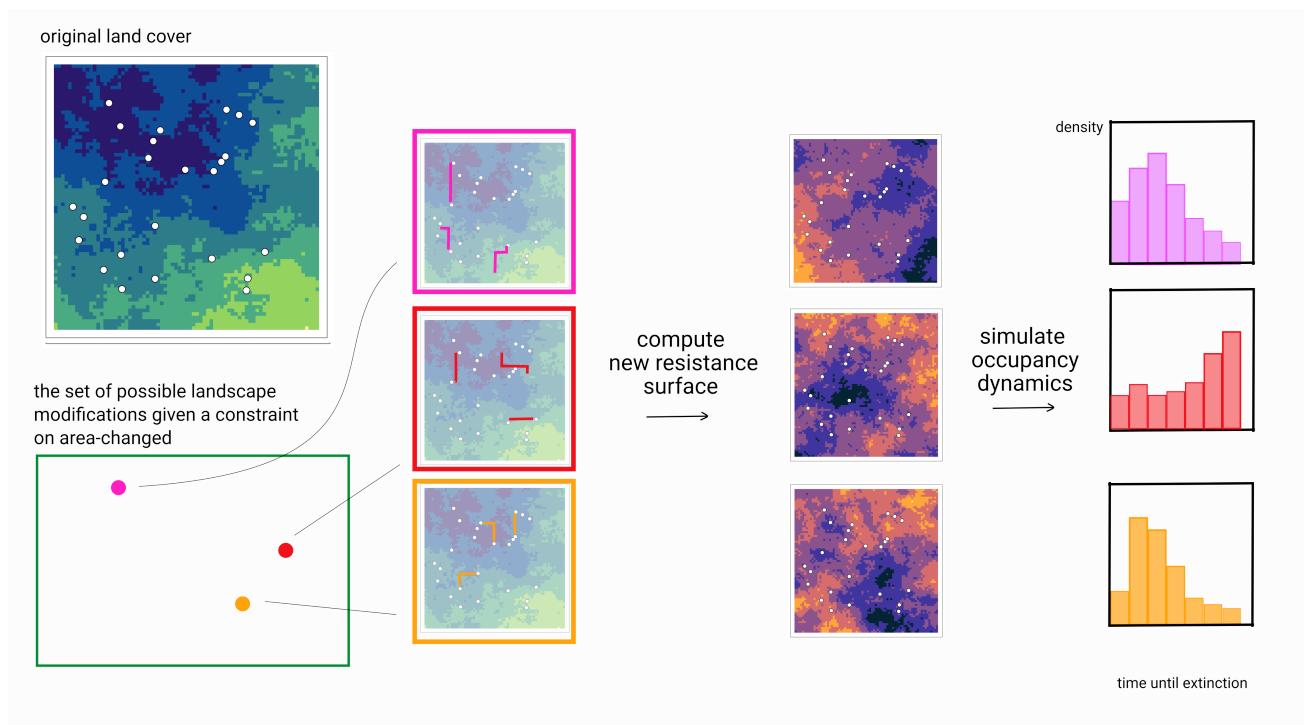


Figure 5: fig

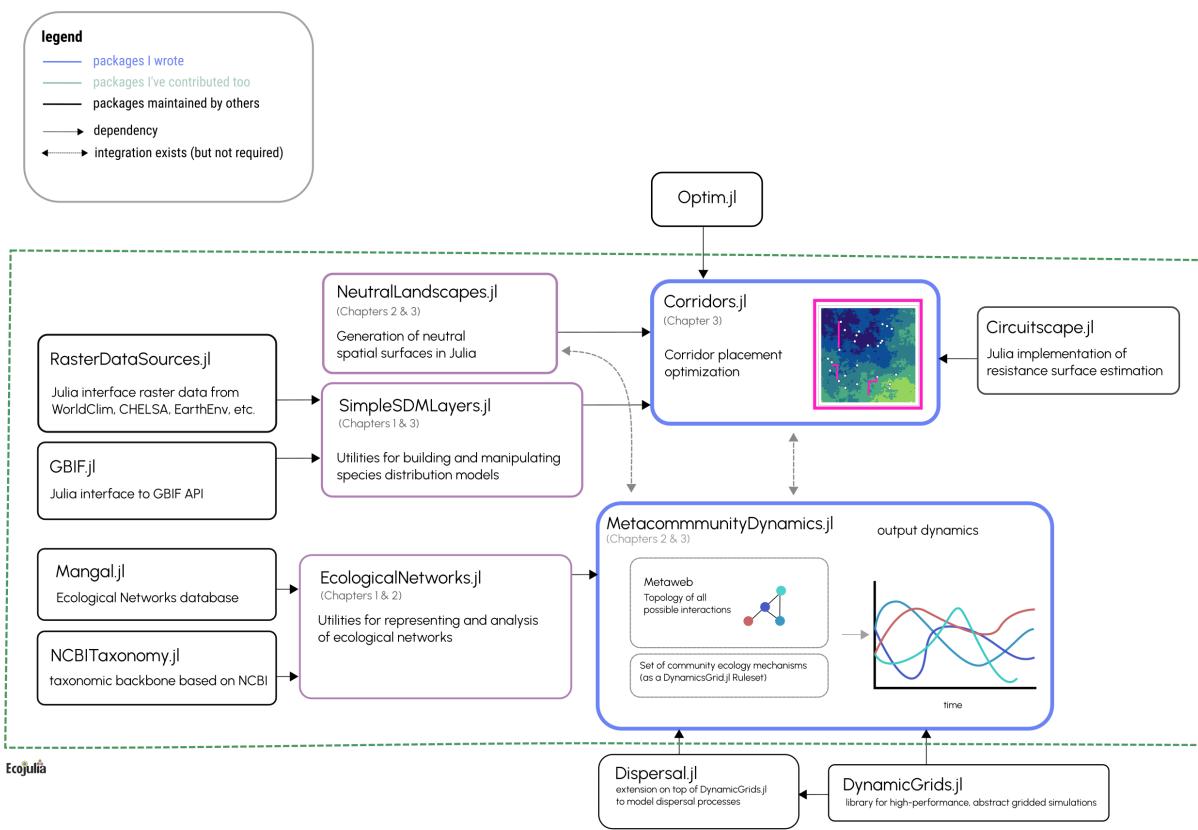


Figure 6: The structure of the software libraries used as part of MCD.jl

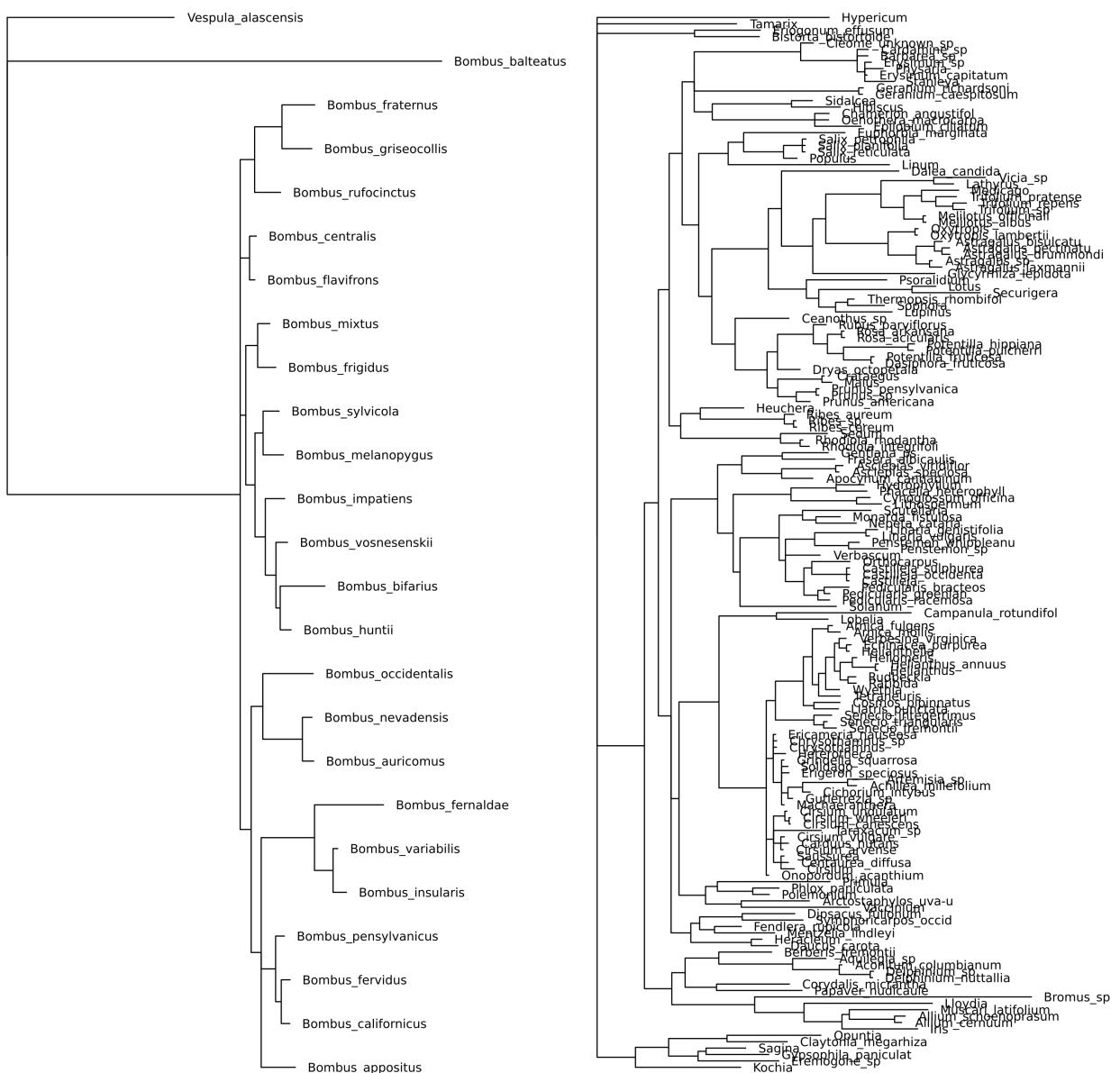


Figure 7: trees