

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 23, 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 *ecological*
7 *forecasting*—TK DEFINITION—and then using these forecasts to make decisions to mitigate the negative
8 consequences of this change on ecosystems, their functioning, and the services they provide to humans
9 has emerged as an imperative for ecology and environmental science (Dietze 2017). However, robust
10 prediction of ecological processes is, to say the least, quite difficult (Beckage *et al.* 2011; Petchey *et al.*
11 2015). This difficultly is compounded by a few factors, the first being that sampling ecosystems is not easy.
12 Ecological data is often biased, noisy, and sparse in both space and time. The current paucity of
13 ecological data has resulted in much interest in developing global systems for *ecosystem monitoring*
14 (Makiola *et al.* 2020), which would systematize the collection of biodiversity data in manner that makes
15 detecting and predicting change more possible than at the moment (Urban *et al.* 2021).

16 **TK add mention of interaction prediction here as they effect dynamics** The second major challenge in
17 ecological forecasting is that the underlying dynamics of most ecological processes are unknown and
18 instead must be inferred from this (sparse) data. Much of the history of quantitatively modeling
19 ecosystems have been done in the language of dynamical systems, describing how the value of an
20 observable state of the system, represented by a vector of numbers $[x_1, x_2, \dots, x_n]^T = \vec{x}$ changes as over
21 time, yielding models in the form of differential equations in continuous-time settings, $\frac{dx}{dt} = f(x)$, or
22 difference equations in discrete-time settings, $x_t = f(x_{t-1})$, where $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is an arbitrary function
23 describing how the system changes on a moment-to-moment basis (e.g. in the context of communities, f
24 could be Lotka-Volterra, Holling-Type-III or DeAngelis-Beddington functional response). The form of this
25 functional response in real systems is effectively unknown, and some forms are inherently more
26 “forecastable” than others (Beckage *et al.* 2011; Chen *et al.* 2019; Pennekamp *et al.* 2019). The initial
27 success of these forms of models can be traced back to the larger program of ontological reductionism,
28 which became the default approach to modeling in the sciences after its early success in physics, which, by
29 the time ecology was becoming a quantitative science (sometime in the 20th century, depending on who

30 you ask), became the foundation for mathematical models in ecology.

31 However, we run into many problems when aiming to apply this type of model to empirical ecological
32 data. Ecosystems are perhaps the quintessential example of system that cannot be understood by iterative
33 reduction of its components into constituent parts—ecological phenomena are emergent: the product of
34 different mechanisms operating at different spatial, temporal, and organizational scales (Levin 1992).

35 Further this analytical approach to modeling explicitly ignores known realities: ecological dynamics not
36 deterministic and many analytic models in ecology assume long-run equilibrium. Finally, perhaps the
37 biggest challenge in using these models to describe ecological processes is ecosystems consist of more
38 dimensions than the tools of analytic models are suited for. As the number of variables in an analytic
39 model increases, so does the ability of the scientist to discern clear relationships between them given a
40 fixed amount of data, the so-called “curse” of dimensionality.

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

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

60 probability the system approaches our “goal” state. This dissertation aims to this framework for ecosystem
61 monitoring and forecasting (fig. 1, left), and each chapter address some aspect of this pipeline to data from
62 a monitoring network to forecasts to mitigation strategy (fig. 1, right).

63 [Figure 1 about here.]

64 **TK final paragraph** > Set up the key research challenges in this field so that the reader understands
65 where your thesis fits in. > Finish the introduction with a short summary of what you will do across your
66 chapters. This is the bridging paragraph to your chapters.

67 The primary reserach challenges this thesis addresses: how do we design ecological samples? Data
68 assimilation pipeline. How do we propgate uncertainty from these samples into forecasts? More details on
69 pipeline flow

70 **Chapter One: Optimizing spatial sampling of species interactions**

71 **Objective**

72 This chapter uses simulation models to investigate the relationship between species relative abundance,
73 sampling effort, and probability of observing an interaction between species in order to aid in the design of
74 samples of ecological interactions, and to provide a null expectation of false-negative probability for a
75 dataset of a given size. Further it then proposes a method for optimizing the spatial sampling locations to
76 maximize the probability of detecting an interaction between two species given a fixed number of total of
77 observations, and the distributions of each species. This addresses the optimization of monitoring
78 network part of the flow from data to mitigation at the top of fig. 1, left. As explored in the previous
79 chapter, there are false-negatives in interaction data. However, there is more than one way to observe a
80 false-negative when sampling interactions. fig. 2 shows a taxonomy of false-negatives in occurrence,
81 co-occurrence, and interaction data.

82 [Figure 2 about here.]

83 **Methods**

84 The first result is to compute a null expectation of the probability of an interaction false-negative as a
85 function number of total observations of individuals of *any species*. This is done by simulating the process
86 of observation, where the probability of observing a given species is its relative abundance. We use a
87 log-normal distribution of relative abundance (Hubbell 2001) and simulating the process of observation on
88 food-webs generated using the niche model (Williams & Martinez 2000) with connectance parameterized
89 by the flexible-links model (MacDonald *et al.* 2020).

90 An example of this relation for networks with varying species richness is shown in fig. 3.

91 We then go on to testing some assumptions of this neutral model with empirical data. Primarily that we
92 analytically show that our neutral model, if anything, underestimates the probability of false-negatives if
93 there are positive associations between species co-occurrence, and we show these positive associations
94 exist in two sets of spatially replicated samples of interaction networks (Thompson & Townsend 2000;
95 Hadfield *et al.* 2014), fig. 4—further I'm planning to add the field data from the previous chapter into this
96 analysis once available.

97 **Results**

98 [Figure 3 about here.]

99 [Figure 4 about here.]

100 **Discussion**

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

104 **Progress**

105 This chapter is mostly complete. The only remaining work is the implementation of simulated annealing
106 optimization process.

107 **Chapter Two: Forecasting the spatial uncoupling of a plant-pollinator**
108 **network**

109 Interactions between plants and pollinators form networks which together structure the “architecture of
110 biodiversity” (Bascompte & Jordano 2007). The functioning and stability of ecosystems emerge from these
111 interactions, but anthropogenic change threatens to unravel and “rewire” these interaction networks
112 (CaraDonna *et al.* 2017), jeopardizing the persistence of these systems. Plant-pollinator networks face two
113 possible forms of rewiring in response to anthropogenic environmental change: spatial and temporal.
114 Range shifts could cause interacting species to no longer overlap in space, and shifts in phenology could
115 cause interacting species to no longer occur at the same time of year. This chapter uses several years of
116 data on bumblebee-flower phenology and interactions across several field sites, each consisting of several
117 plots across an elevational gradient, combined with spatial records of species occurrence via GBIF to
118 forecast the uncoupling of the plant-pollinator metaweb of Colorado.

119 [Figure 5 about here.]

120 **Methods**

121 The data for this chapter is derived from multiple sources that can be split into four categories. (1) Field
122 data from three different field sites across Colorado, each with multiple plots across an elevational
123 gradient, for seven, seven, and three years respectively. This data was collected by Paul CaraDonna and
124 Jane Oglevie (from the Rocky Mountain Biological Laboratory; RMBL) and Julian Resasco (CU Boulder).
125 (2) GBIF spatial occurrence records of each of these species across Colorado, including a metaweb of
126 interactions across all of Colorado taken from GBIF. (3) Remotely sensed data consisting of current and
127 forecasting bioclimatic variables from CHELSA. (4) Phylogenies for both bee and flower species derived
128 from NCBI GenBank barcodes for mitochondrial COI (bumblebees) and chloroplast rbcL (flowers).
129 As the data we have is spatially sparse and likely to contain many interaction “false-negatives” (Strydom *et*
130 *al.* 2021b), we begin by predicting a metaweb of interactions across Colorado as they exist *in the present*.
131 We do this using a set of candidate interaction prediction models: relative abundance only, phylogenetic
132 embedding only (a la Strydom *et al.* (2021a)), niche embedding only (Gravel *et al.* 2019), and all pairwise
133 combinations of those constituent models. After validating and selecting the best performing model, we

134 then predict how these distributions of each of these species will change under the CMIP6 consensus
135 climate forecast (Karger *et al.* 2017), and then finally quantify the reduction in spatial between species for
136 which there is a predicted interaction.

137 **Results**

138 Here we show the in-progress work on the prerequisites for the analysis outlined above: phylogenies for
139 both plant and bee species (fig. 6) and species distribution models for all species (an example shown in
140 fig. 7).

141 [Figure 6 about here.]

142 [Figure 7 about here.]

143 **Progress**

144 I have a phylogeny and SDMs for all the species. Waiting on data agreement with RMBL.

145 **Chapter Three: Optimizing corridor placement against ecological 146 dynamics**

147 As land-use change has caused many habitats to become fragmented and patchy, promoting landscape
148 connectivity has become of significant interest to mitigate the effects of this change on Earth's biodiversity.
149 However, the practical realities of conservation mean that there is a limitation on how much we can
150 modify landscapes in order to do this. So what is the best place to put a corridor given a constraint on how
151 much surface-area you can change in a landscape? This is the question this chapter seeks to answer.
152 Models for inferring corridor locations have been developed, but are limited in that they are not developed
153 around promoting some element of ecosystem function, but instead by trying to find the path of least
154 resistance in an existing landscape from a derived resistance surface (Peterman 2018). This chapter
155 proposes a general algorithm for choosing corridor placement to optimize a measurement of ecosystem
156 functioning derived from simulations run on each proposed landscape modification.

158 We propose various landscape modifications which alter the cover of a landscape, represented as a raster.
 159 We then compute a new resistance surface based on the proposed landscape modification using
 160 Circuitscape (McRae *et al.* 2008), and based on the values of resistance to dispersal between pair of
 161 locations we simulate spatially-explicit metapopulation dynamics model (Hanski & Ovaskainen 2000;
 162 Ovaskainen *et al.* 2002) to estimate a distribution of time until extinction for each landscape modification.
 163 The largest challenge in implementing this algorithm is the space of potential modifications grows as
 164 $O((nm)!)$ for an n by m raster. For most actual landscapes to which we wish to apply this method, the set
 165 of possible modifications becomes uncomputably large, so we use simulated annealing to explore the
 166 search space of possible modifications to estimate the modification that maximizes the time-until
 167 extinction of simulated metapopulation dynamics under that hypothetical modified landscape.
 168 The biggest challenge in implementing simulated annealing in this context is defining a proposal function
 169 for landscape modifications. This is done by computing the minimum-spanning-tree (MST) of the spatial
 170 nodes, and then proposing corridors that connect nodes that are already connected in the MST.

171 **Chapter Four: MetacommunityDynamics.jl: a virtual laboratory for 172 community ecology**

173 **Objective**

174 The final chapter consists of a collection of modules in the Julia language for different aspects of
 175 community ecology, including most of the code used for the preceding chapters. Indeed
 176 MetacommunityDynamics.jl (MCD.jl) is the epicenter of this set of tools, but due to the nature of the Julia
 177 language, MCD.jl is interoperable with several existing packages within the EcoJulia organization,
 178 including several to which I have contributed. A diagram showing the relation between these packages is
 179 shown in fig. 10.

180 We need tools to generate synthetic data from a *known* set of mechanisms and parameters to test our
 181 methods for parameter inference and forecasting on this *known* system to assess the effectiveness of these
 182 inference and forecasting methods.

183 **Methods**

- 184 Software is structured based on DG.jl (**cite?**) and Dispersal.jl (**cite?**).
185 Uses methods from EN.jl to generate metawebs, or can use empirical networks from Mangal.jl (**cite?**).
186 Framework based on (**Velland2010ConSyn?**), processes divided into four categories: selection, dispersal,
187 drift, speciation.

188 **Results**

189 [Figure 9 about here.]

190 **Progress**

- 191 The software as it exists is capable of simulating the biomass dynamics of arbitrarily large food-webs using
192 Lotka-Volterra, Holling Type-II, or Holling Type-III functional responses. It currently has methods to
193 implement Gaussian drift, and various forms of dispersal via Dispersal.jl. Also occupancy dynamics for
194 Levins metapopulations (**levins1967?**), and spatially explicit Hanski metapopulations (**hanski2001?**).
195 This is most of what needs to exist for the preceding chapters.
196 Selection on arbitrary environmental variables in progress, as well as traits.

197 [Figure 10 about here.]

198 **Discussion**

- 199 Describing expected/anticipated contributions of the thesis. Very important for QE. This
200 should be at least half a page.

201 **References**

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

- 204 Bauer, P., Thorpe, A. & Brunet, G. (2015). The quiet revolution of numerical weather prediction. *Nature*,
205 525, 47–56.
- 206 Beckage, B., Gross, L.J. & Kauffman, S. (2011). The limits to prediction in ecological systems. *Ecosphere*, 2,
207 art125.
- 208 CaraDonna, P.J., Petry, W.K., Brennan, R.M., Cunningham, J.L., Bronstein, J.L., Waser, N.M., *et al.* (2017).
209 Interaction rewiring and the rapid turnover of plantpollinator networks. *Ecology Letters*, 20, 385–394.
- 210 Chen, Y., Angulo, M.T. & Liu, Y.-Y. (2019). Revealing Complex Ecological Dynamics via Symbolic
211 Regression. *BioEssays*, 41, 1900069.
- 212 Dietze, M.C. (2017). Prediction in ecology: A first-principles framework. *Ecological Applications*, 27,
213 2048–2060.
- 214 Gravel, D., Baiser, B., Dunne, J.A., Kopalke, J.-P., Martinez, N.D., Nyman, T., *et al.* (2019). Bringing Elton
215 and Grinnell together: A quantitative framework to represent the biogeography of ecological
216 interaction networks. *Ecography*, 42, 401–415.
- 217 Hadfield, J.D., Krasnov, B.R., Poulin, R. & Nakagawa, S. (2014). A Tale of Two Phylogenies: Comparative
218 Analyses of Ecological Interactions. *The American Naturalist*, 183, 174–187.
- 219 Hanski, I. & Ovaskainen, O. (2000). The metapopulation capacity of a fragmented landscape. *Nature*, 404,
220 755–758.
- 221 Hill, C., DeLuca, C., Balaji, Suarez, M. & Da Silva, A. (2004). The architecture of the Earth System
222 Modeling Framework. *Computing in Science Engineering*, 6, 18–28.
- 223 Hubbell, S.P. (2001). *The unified neutral theory of biodiversity and biogeography*. Monographs in
224 population biology. Princeton University Press, Princeton.
- 225 Karger, D.N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., *et al.* (2017). Climatologies at
226 high resolution for the earth's land surface areas. *Scientific Data*, 4, 170122.
- 227 Levin, S.A. (1992). The Problem of Pattern and Scale in Ecology: The Robert H. MacArthur Award
228 Lecture. *Ecology*, 73, 1943–1967.
- 229 MacDonald, A.A.M., Banville, F. & Poisot, T. (2020). Revisiting the Links-Species Scaling Relationship in
230 Food Webs. *Patterns*, 1.

- 231 Makiola, A., Compson, Z.G., Baird, D.J., Barnes, M.A., Boerlijst, S.P., Bouchez, A., *et al.* (2020). Key
232 Questions for Next-Generation Biomonitoring. *Frontiers in Environmental Science*, 7.
- 233 McRae, B.H., Dickson, B.G., Keitt, T.H. & Shah, V.B. (2008). Using Circuit Theory to Model Connectivity
234 in Ecology, Evolution, and Conservation. *Ecology*, 89, 2712–2724.
- 235 Ovaskainen, O., Sato, K., Bascompte, J. & Hanski, I. (2002). Metapopulation Models for Extinction
236 Threshold in Spatially Correlated Landscapes. *Journal of Theoretical Biology*, 215, 95–108.
- 237 Ovaskainen, O., Sato, K., Bascompte, J. & Hanski, I. (2002). Metapopulation Models for Extinction
238 Threshold in Spatially Correlated Landscapes. *Journal of Theoretical Biology*, 215, 95–108.
- 239 Pennekamp, F., Iles, A.C., Garland, J., Brennan, G., Brose, U., Gaedke, U., *et al.* (2019). The intrinsic
240 predictability of ecological time series and its potential to guide forecasting. *Ecological Monographs*, 89,
241 e01359.
- 242 Petchey, O.L., Pontarp, M., Massie, T.M., Kéfi, S., Ozgul, A., Weilenmann, M., *et al.* (2015). The ecological
243 forecast horizon, and examples of its uses and determinants. *Ecology Letters*, 18, 597–611.
- 244 Peterman, W.E. (2018). ResistanceGA: An R package for the optimization of resistance surfaces using
245 genetic algorithms. *Methods in Ecology and Evolution*, 9, 1638–1647.
- 246 Strydom, T., Bouskila, S., Banville, F., Barros, C., Caron, D., Farrell, M.J., *et al.* (2021a). Food web
247 reconstruction through phylogenetic transfer of low-rank network representation.
- 248 Strydom, T., Catchen, M.D., Banville, F., Caron, D., Dansereau, G., Desjardins-Proulx, P., *et al.* (2021b). A
249 *Roadmap Toward Predicting Species Interaction Networks (Across Space and Time)* (Preprint).
- 250 EcoEvoRxiv.
- 251 Thompson, R.M. & Townsend, C.R. (2000). Is resolution the solution?: The effect of taxonomic resolution
252 on the calculated properties of three stream food webs. *Freshwater Biology*, 44, 413–422.
- 253 Urban, M.C., Travis, J.M.J., Zurell, D., Thompson, P.L., Synes, N.W., Scarpa, A., *et al.* (2021). Coding for
254 Life: Designing a Platform for Projecting and Protecting Global Biodiversity. *BioScience*.
- 255 Williams, R.J. & Martinez, N.D. (2000). Simple rules yield complex food webs. *Nature*, 404, 180–183.

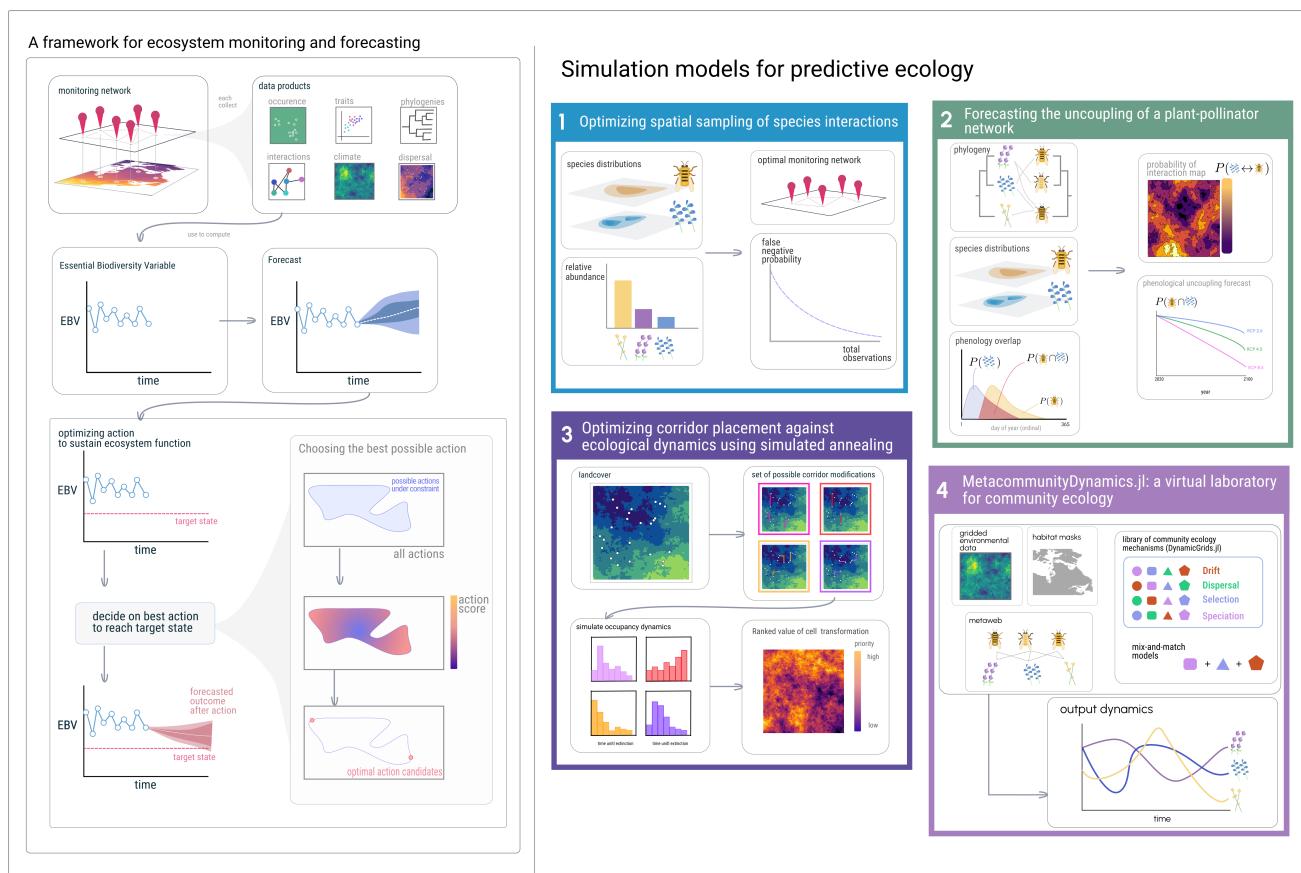


Figure 1: thesis concept

Species A occurs?

Species B occurs?

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

Figure 2: A taxonomy of occurrence, co-occurrence, and interaction false negatives in data

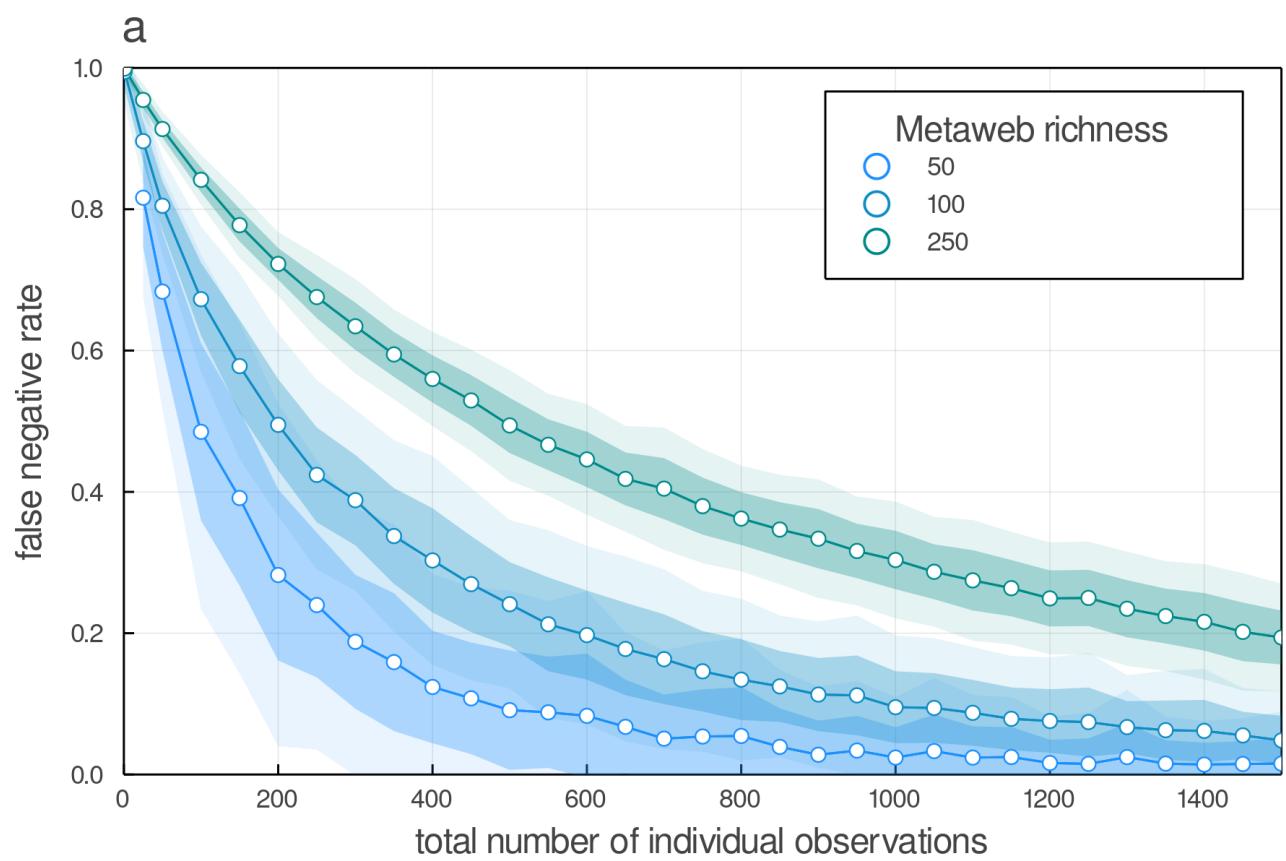


Figure 3: foo

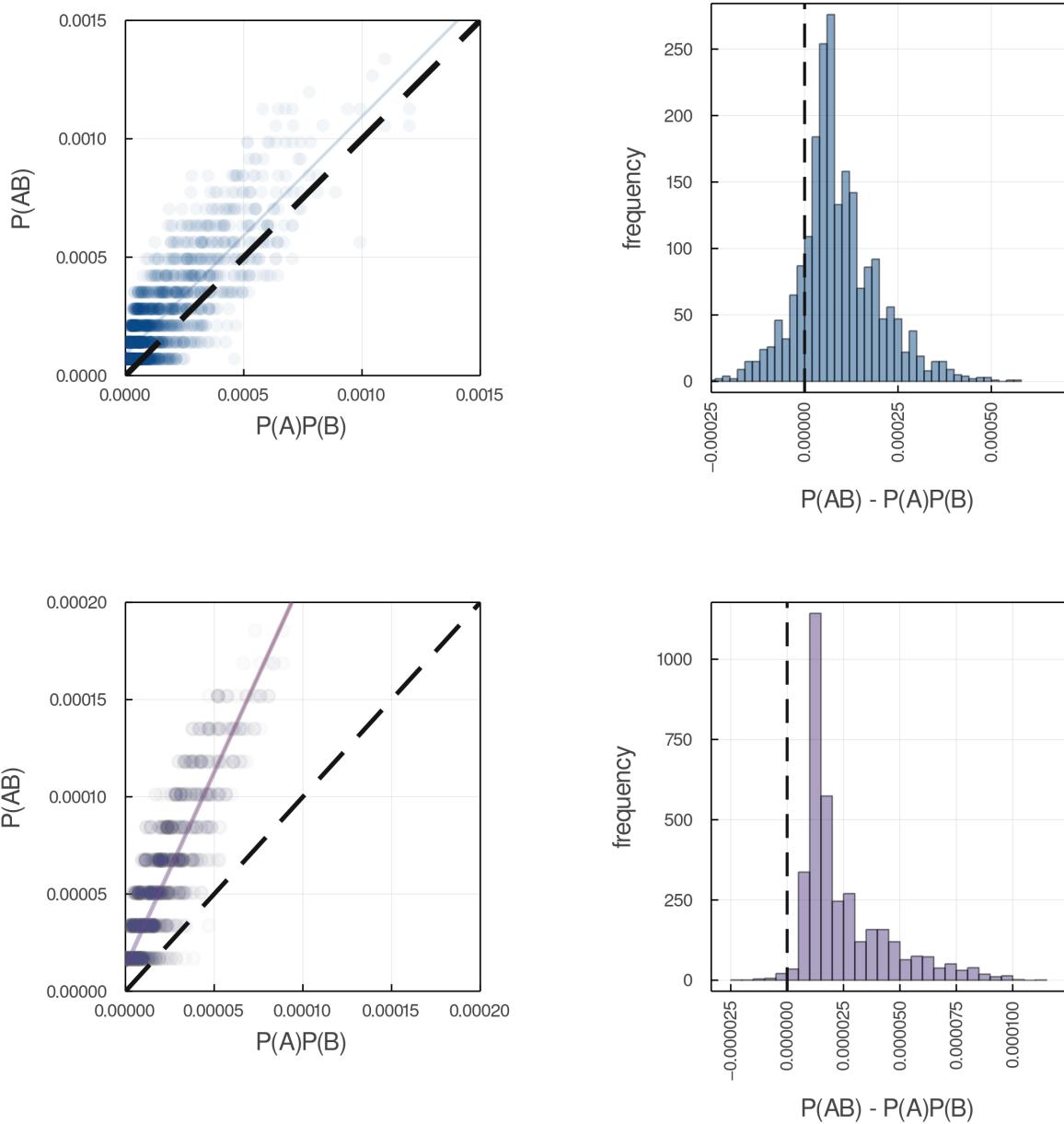


Figure 4: Demonstrates positive associations in co-occurrence

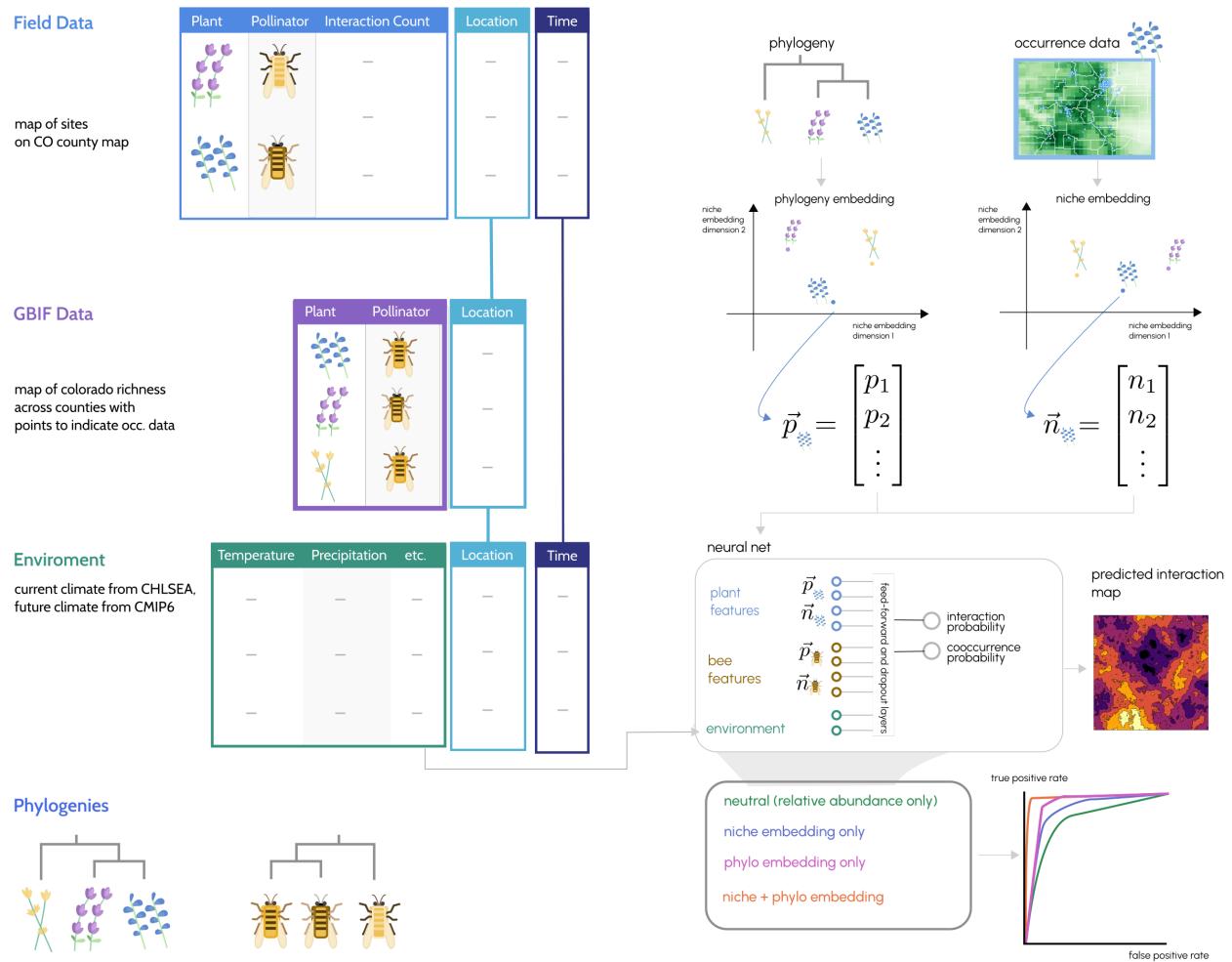


Figure 5: Chapter One conceptual figure. Left: the sources of data and how they can be synthesized. Right: The flow from data to interaction prediction using a few different interaction prediction models.

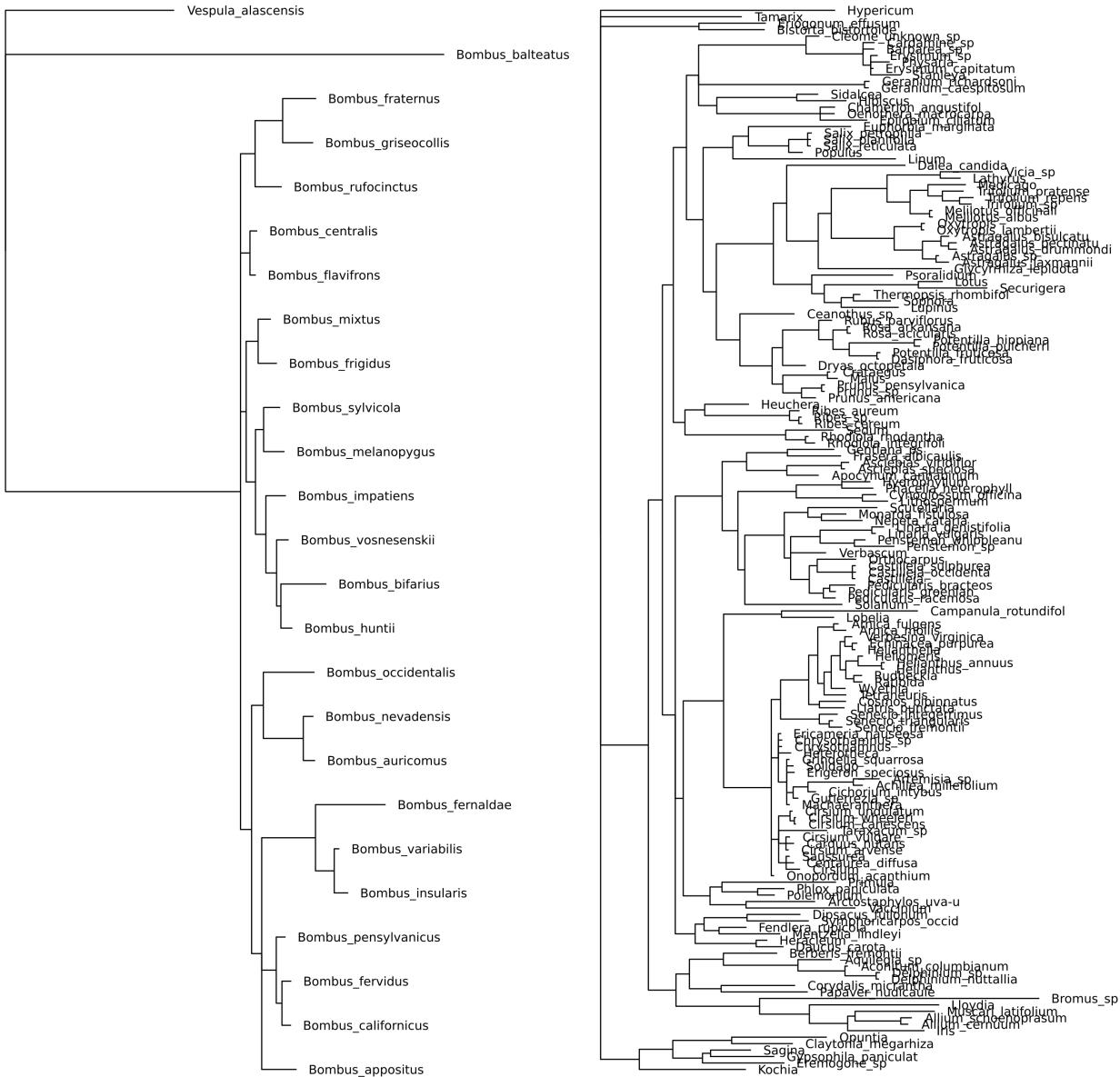


Figure 6: Phylogeny for both bumblebee species (left) and flower species (right)

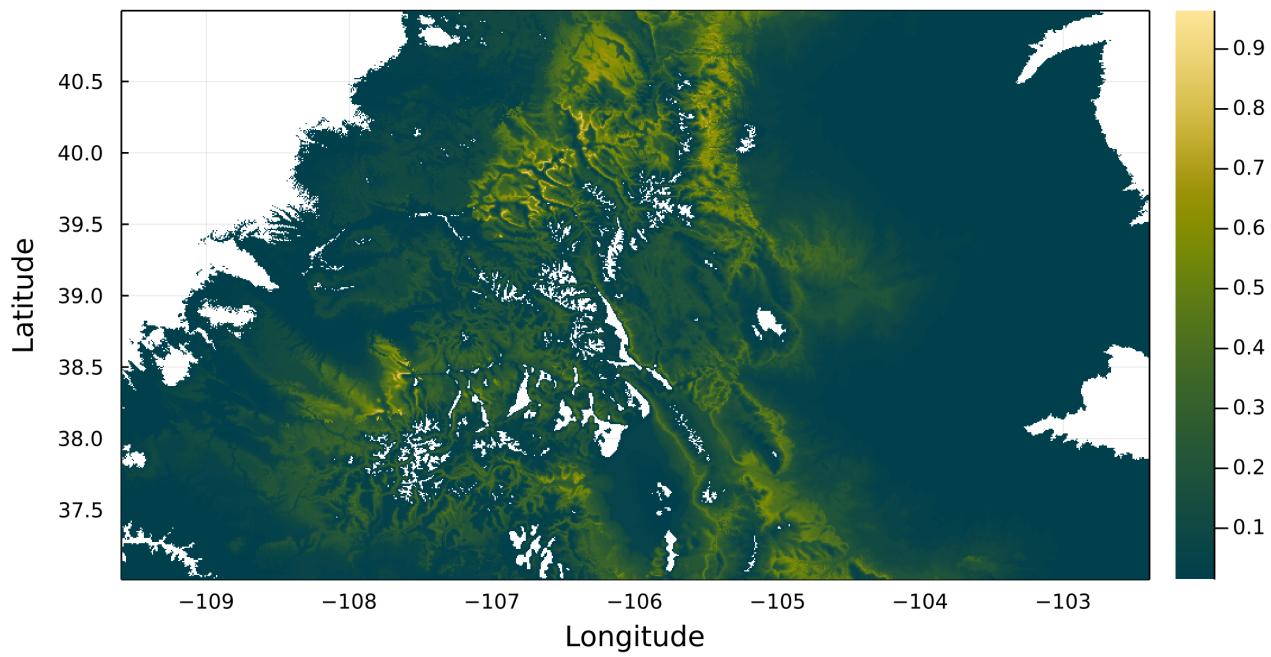


Figure 7: Example SDM for *Achillea millefolium*

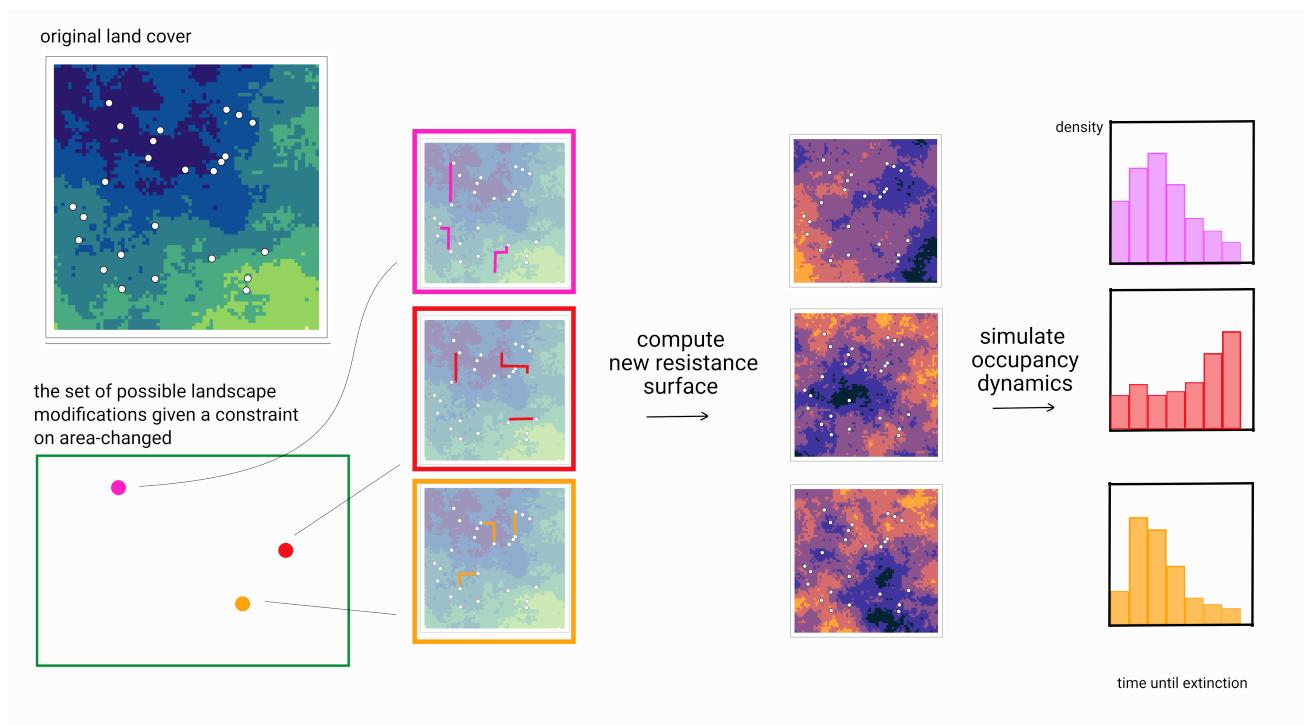


Figure 8: foo

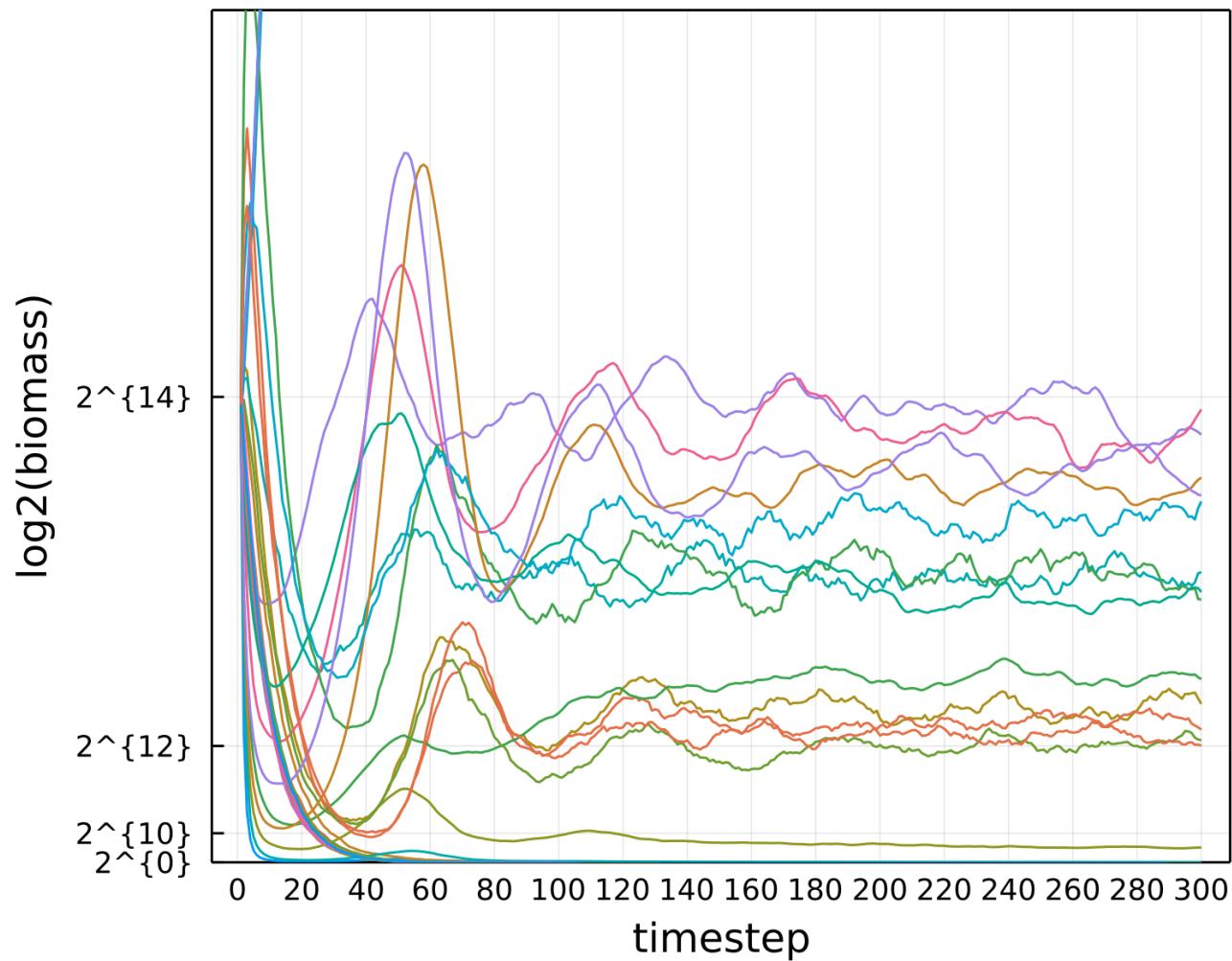


Figure 9: Sample output of simulated food web dynamics from MetacommunityDynamics.jl

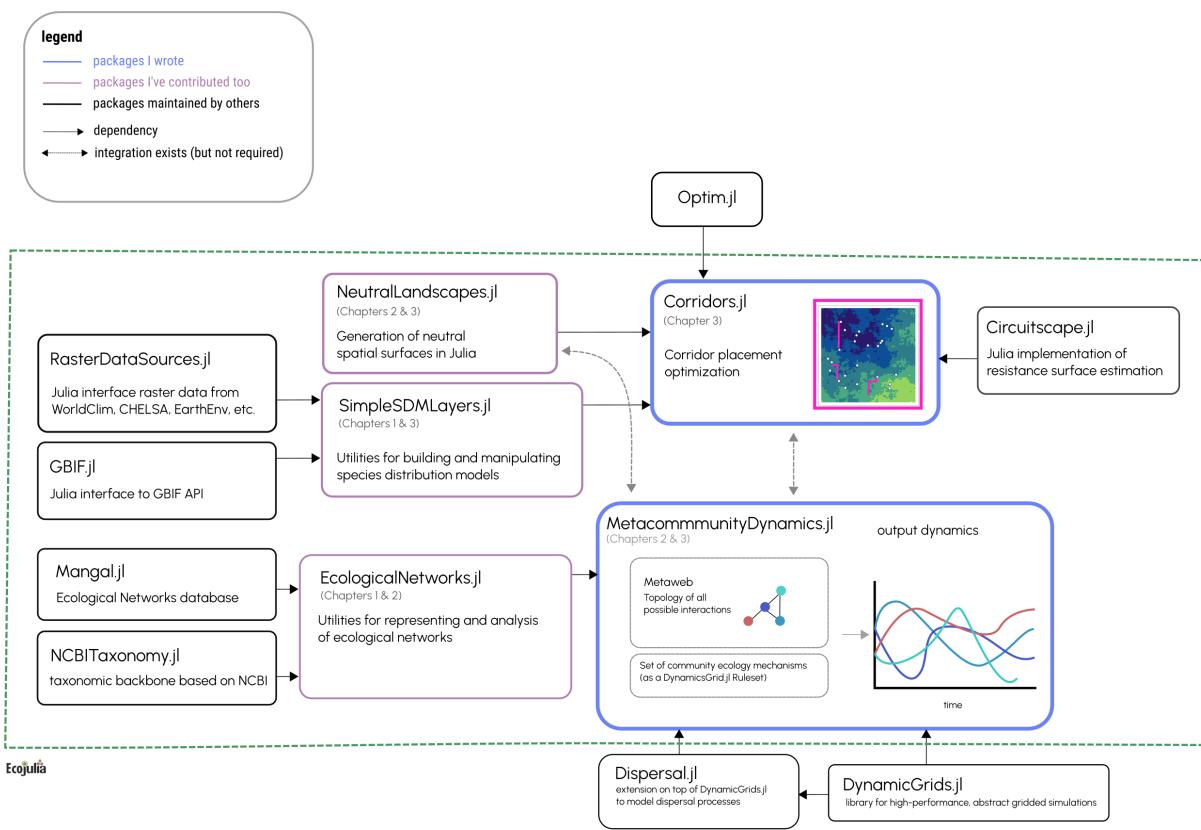


Figure 10: The structure of the software libraries used as part of MCD.jl