

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 settings,
21 $\frac{dx}{dt} = f(x)$, or difference equations in discrete-time settings, $x_t = f(x_{t-1})$, where $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is an
22 arbitrary function describing how the system changes on a moment-to-moment basis (e.g. in the context of
23 communities, f could be Lotka-Volterra, Holling-Type-III or DeAngelis-Beddington functional response).
24 The form of this functional response in real systems is effectively unknown, and some forms are
25 inherently more "forecastable" than others (Beckage *et al.* 2011; Chen *et al.* 2019; Pennekamp *et al.* 2019).
26 The initial success of these forms of models can be traced back to the larger program of ontological
27 reductionism, which became the default approach to modeling in the sciences after its early success in
28 physics, which, by the time ecology was becoming a quantitative science (sometime in the 20th century,
29 depending on who you ask), became the foundation for mathematical models in ecology.

30 However, we run into many problems when aiming to apply this type of model to empirical ecological
31 data. Ecosystems are perhaps the quintessential example of system that cannot be understood by iterative
32 reduction of its components into constituent parts—ecological phenomena are emergent: the product of
33 different mechanisms operating at different spatial, temporal, and organizational scales (Levin 1992).
34 Further this analytical approach to modeling explicitly ignores known realities: ecological dynamics not
35 deterministic and many analytic models in ecology assume long-run equilibrium. Finally, perhaps the
36 biggest challenge in using these models to describe ecological processes is ecosystems consist of more
37 dimensions than the tools of analytic models are suited for. As the number of variables in an analytic
38 model increases, so does the ability of the scientist to discern clear relationships between them given a
39 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 last 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 observation 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 of 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 which together structure the “architecture of
66 biodiversity” (Bascompte & Jordano 2007). The functioning and stability of ecosystems emerge from these
67 interactions, but anthropogenic change threatens to unravel and “rewire” these interaction networks
68 (CaraDonna *et al.* 2017), jeopardizing the persistence of these systems. Plant-pollinator networks face two
69 possible forms of rewiring in response to anthropogenic environmental change: spatial and temporal.
70 Range shifts could cause interacting species to no longer overlap in space, and shifts in phenology could
71 cause interacting species to no longer occur at the same time of year. This chapter uses several years of
72 data on bumblebee-flower phenology and interactions across several field sites, each consisting of several
73 plots across an elevational gradient, combined with spatial records of species occurrence via GBIF to
74 forecast the uncoupling of the plant-pollinator metaweb of Colorado.

75 [Figure 2 about here.]

76 **Data**

77 The data for this chapter is derived from multiple sources that can be split into four categories. (1) Field
78 data from three different field sites across Colorado, each with multiple plots across an elevational
79 gradient, for seven, seven, and three years respectively. This data was collected by Paul CaraDonna and
80 Jane Oglevie (from the Rocky Mountain Biological Laboratory; RMBL) and Julian Resasco (CU Boulder).
81 (2) GBIF spatial occurrence records of each of these species across Colorado, including a metaweb of
82 interactions across all of Colorado taken from GBIF. (3) Remotely sensed data consisting of current and
83 forecasting bioclimatic variables from CHELSA. (4) Phylogenies for both bee and flower species derived
84 from NCBI GenBank barcodes for mitochondrial COI (bumblebees) and chloroplast rbcL (flowers).

85 **Methods**

86 As the data we have is spatially sparse and likely to contain many interaction “false-negatives” (Strydom *et*
87 *al.* 2021b), we begin by predicting a metaweb of interactions across Colorado as they exist *in the present*.
88 We do this using a set of candidate interaction prediction models: relative abundance only, phylogenetic
89 embedding only (a la Strydom *et al.* (2021a)), niche embedding only (Gravel *et al.* 2019), and all pairwise
90 combinations of those constituent models. After validating and selecting the best performing model, we
91 then predict how these distributions of each of these species will change under the CMIP6 consensus
92 climate forecast (Karger *et al.* 2017), and then finally quantify the reduction in spatial between species for
93 which there is a predicted interaction.

94 **Preliminary Results**

95 Here we show the in-progress work on the prerequisites for the analysis outlined above: phylogenies for
96 both plant and bee species (fig. 3) and species distribution models for all species (an example shown in
97 fig. 4).

98 [Figure 3 about here.]

99 [Figure 4 about here.]

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

101 This chapter uses simulation models to investigate the relationship between species relative abundance,
102 sampling effort, and probability of observing an interaction between species, and further proposes a
103 method for optimizing the spatial sampling locations to maximize the probability of detecting an
104 interaction between two species given their distributions. This addresses the optimization of monitoring
105 network part of the flow from data to mitigation at the top of fig. 1, left. As explored in the previous
106 chapter, there are false-negatives in interaction data. However, there is more than one way to observe a
107 false-negative when sampling interactions. fig. 5 shows a taxonomy of false-negatives in occurrence,
108 co-occurrence, and interaction data.

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

110

[Figure 5 about here.]

111 The first result is to compute a null expectation of the probability of an interaction false-negative as a
 112 function number of total observations of individuals of *any species*. This is done by using a log-normal
 113 distribution of relative abundance (Hubbell 2001) and simulating the process of observation on food-webs
 114 generated using the niche model (Williams & Martinez 2000) with connectance parameterized by the
 115 flexible-links model (MacDonald *et al.* 2020). An example of this relation for networks with varying
 116 species richness is shown in fig. 6.

117

[Figure 6 about here.]

118 We then go on to testing some assumptions of this neutral model with empirical data. Primarily that we
 119 analytically show that our neutral model, if anything, underestimates the probability of false-negatives if
 120 there are positive associations between species co-occurrence, and we show these positive associations
 121 exist in two sets of spatially replicated samples of interaction networks (Thompson & Townsend 2000;
 122 Hadfield *et al.* 2014), fig. 7—further I'm planning to add the field data from the previous chapter into this
 123 analysis once available.

124

[Figure 7 about here.]

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

128 **Chapter Three: Optimizing corridor placement against ecological
 129 dynamics**

130 As land-use change has caused many habitats to become fragmented and patchy, promoting landscape
 131 connectivity has become of significant interest to mitigate the effects of this change on Earth's biodiversity.
 132 However, the practical realities of conservation mean that there is a limitation on how much we can
 133 modify landscapes in order to do this. So what is the best place to put a corridor given a constraint on how

134 much surface-area you can change in a landscape? This is the question this chapter seeks to answer.
135 Models for inferring corridor locations have been developed, but are limited in that are not developed
136 around promoting some element of ecosystem function, but instead by trying to find the path of least
137 resistance in an existing landscape from a derived resistance surface (Peterman 2018). This chapter
138 proposes a general algorithm for choosing corridor placement to optimize a measurement of ecosystem
139 functioning derived from simulations run on each proposed landscape modification.

140 [Figure 8 about here.]

141 We propose various landscape modifications which alter the cover of a landscape, represented as a raster.
142 We then compute a new resistance surface based on the proposed landscape modification using
143 Circuitscape (McRae *et al.* 2008), and based on the values of resistance to dispersal between pair of
144 locations we simulate spatially-explicit metapopulation dynamics model (Hanski & Ovaskainen 2000;
145 Ovaskainen *et al.* 2002) to estimate a distribution of time until extinction for each landscape modification.
146 The largest challenge in implementing this algorithm is the space of potential modifications grows as
147 $O(nm)$ for an n by m raster. For most actual landscapes to which we wish to apply this method, the set of
148 possible modifications becomes uncomputably large, so we use simulated annealing to explore the search
149 space of possible modifications to estimate the modification that maximizes the time-until extinction of
150 simulated metapopulation dynamics under that hypothetical modified landscape.
151 The biggest challenge in implementing simulated annealing in this context is defining a proposal function
152 for landscape modifications. This is done by computing the minimum-spanning-tree (MST) of the spatial
153 nodes, and then proposing corridors that connect nodes that are already connected in the MST.

154 **Chapter Four: MetacommunityDynamics.jl: a virtual laboratory for 155 community ecology**

156 The final chapter consists of a collection of modules in the Julia language for different aspects of
157 community ecology, including most of the code used for the preceding chapters. Indeed
158 MetacommunityDynamics.jl (MCD.jl) is the epicenter of this set of tools, but due to the nature of the Julia
159 language, MCD.jl is interoperable with several existing packages within the EcoJulia organization,

160 including several to which I have contributed. A diagram showing the relation between these packages is
161 shown in fig. 9.

162 [Figure 9 about here.]

163 Conclusion

164 References

- 165 Bascompte, J. & Jordano, P. (2007). Plant-Animal Mutualistic Networks: The Architecture of Biodiversity.
166 *Annual Review of Ecology, Evolution, and Systematics*, 38, 567–593.
- 167 Bauer, P., Thorpe, A. & Brunet, G. (2015). The quiet revolution of numerical weather prediction. *Nature*,
168 525, 47–56.
- 169 Beckage, B., Gross, L.J. & Kauffman, S. (2011). The limits to prediction in ecological systems. *Ecosphere*, 2,
170 art125.
- 171 Blanchet, F.G., Cazelles, K. & Gravel, D. (2020). Co-occurrence is not evidence of ecological interactions.
172 *Ecology Letters*, 23, 1050–1063.
- 173 CaraDonna, P.J., Petry, W.K., Brennan, R.M., Cunningham, J.L., Bronstein, J.L., Waser, N.M., et al. (2017).
174 Interaction rewiring and the rapid turnover of plantpollinator networks. *Ecology Letters*, 20, 385–394.
- 175 Chen, Y., Angulo, M.T. & Liu, Y.-Y. (2019). Revealing Complex Ecological Dynamics via Symbolic
176 Regression. *BioEssays*, 41, 1900069.
- 177 Dietze, M.C. (2017). Prediction in ecology: A first-principles framework. *Ecological Applications*, 27,
178 2048–2060.
- 179 Gravel, D., Baiser, B., Dunne, J.A., Kopelke, J.-P., Martinez, N.D., Nyman, T., et al. (2019). Bringing Elton
180 and Grinnell together: A quantitative framework to represent the biogeography of ecological
181 interaction networks. *Ecography*, 42, 401–415.
- 182 Hadfield, J.D., Krasnov, B.R., Poulin, R. & Nakagawa, S. (2014). A Tale of Two Phylogenies: Comparative
183 Analyses of Ecological Interactions. *The American Naturalist*, 183, 174–187.

- 184 Hanski, I. & Ovaskainen, O. (2000). The metapopulation capacity of a fragmented landscape. *Nature*, 404,
185 755–758.
- 186 Hill, C., DeLuca, C., Balaji, Suarez, M. & Da Silva, A. (2004). The architecture of the Earth System
187 Modeling Framework. *Computing in Science Engineering*, 6, 18–28.
- 188 Hubbell, S.P. (2001). *The unified neutral theory of biodiversity and biogeography*. Monographs in
189 population biology. Princeton University Press, Princeton.
- 190 Karger, D.N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., et al. (2017). Climatologies at
191 high resolution for the earth's land surface areas. *Scientific Data*, 4, 170122.
- 192 Levin, S.A. (1992). The Problem of Pattern and Scale in Ecology: The Robert H. MacArthur Award
193 Lecture. *Ecology*, 73, 1943–1967.
- 194 MacDonald, A.A.M., Banville, F. & Poisot, T. (2020). Revisiting the Links-Species Scaling Relationship in
195 Food Webs. *Patterns*, 1.
- 196 Makiola, A., Compson, Z.G., Baird, D.J., Barnes, M.A., Boerlijst, S.P., Bouchez, A., et al. (2020). Key
197 Questions for Next-Generation Biomonitoring. *Frontiers in Environmental Science*, 7.
- 198 McRae, B.H., Dickson, B.G., Keitt, T.H. & Shah, V.B. (2008). Using Circuit Theory to Model Connectivity
199 in Ecology, Evolution, and Conservation. *Ecology*, 89, 2712–2724.
- 200 Ovaskainen, O., Sato, K., Bascompte, J. & Hanski, I. (2002). Metapopulation Models for Extinction
201 Threshold in Spatially Correlated Landscapes. *Journal of Theoretical Biology*, 215, 95–108.
- 202 Ovaskainen, O., Sato, K., Bascompte, J. & Hanski, I. (2002). Metapopulation Models for Extinction
203 Threshold in Spatially Correlated Landscapes. *Journal of Theoretical Biology*, 215, 95–108.
- 204 Pennekamp, F., Iles, A.C., Garland, J., Brennan, G., Brose, U., Gaedke, U., et al. (2019). The intrinsic
205 predictability of ecological time series and its potential to guide forecasting. *Ecological Monographs*, 89,
206 e01359.
- 207 Petchey, O.L., Pontarp, M., Massie, T.M., Kéfi, S., Ozgul, A., Weilenmann, M., et al. (2015). The ecological
208 forecast horizon, and examples of its uses and determinants. *Ecology Letters*, 18, 597–611.
- 209 Peterman, W.E. (2018). ResistanceGA: An R package for the optimization of resistance surfaces using
210 genetic algorithms. *Methods in Ecology and Evolution*, 9, 1638–1647.

- 211 Strydom, T., Bouskila, S., Banville, F., Barros, C., Caron, D., Farrell, M.J., *et al.* (2021a). Food web
212 reconstruction through phylogenetic transfer of low-rank network representation.
- 213 Strydom, T., Catchen, M.D., Banville, F., Caron, D., Dansereau, G., Desjardins-Proulx, P., *et al.* (2021b). *A*
214 *Roadmap Toward Predicting Species Interaction Networks (Across Space and Time)* (Preprint).
- 215 EcoEvoRxiv.
- 216 Thompson, R.M. & Townsend, C.R. (2000). Is resolution the solution?: The effect of taxonomic resolution
217 on the calculated properties of three stream food webs. *Freshwater Biology*, 44, 413–422.
- 218 Urban, M.C., Travis, J.M.J., Zurell, D., Thompson, P.L., Synes, N.W., Scarpa, A., *et al.* (2021). Coding for
219 Life: Designing a Platform for Projecting and Protecting Global Biodiversity. *BioScience*.
- 220 Williams, R.J. & Martinez, N.D. (2000). Simple rules yield complex food webs. *Nature*, 404, 180–183.

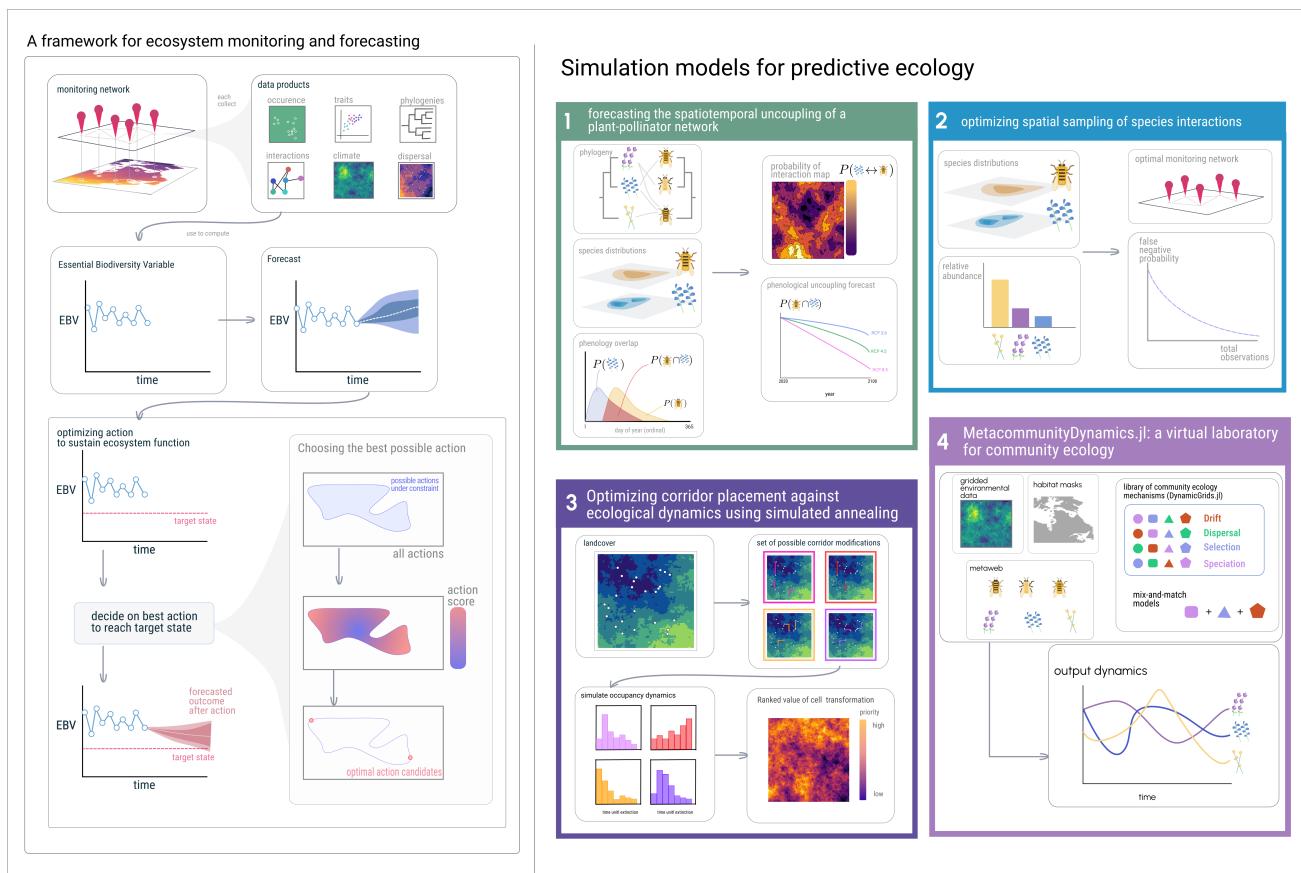


Figure 1: thesis concept

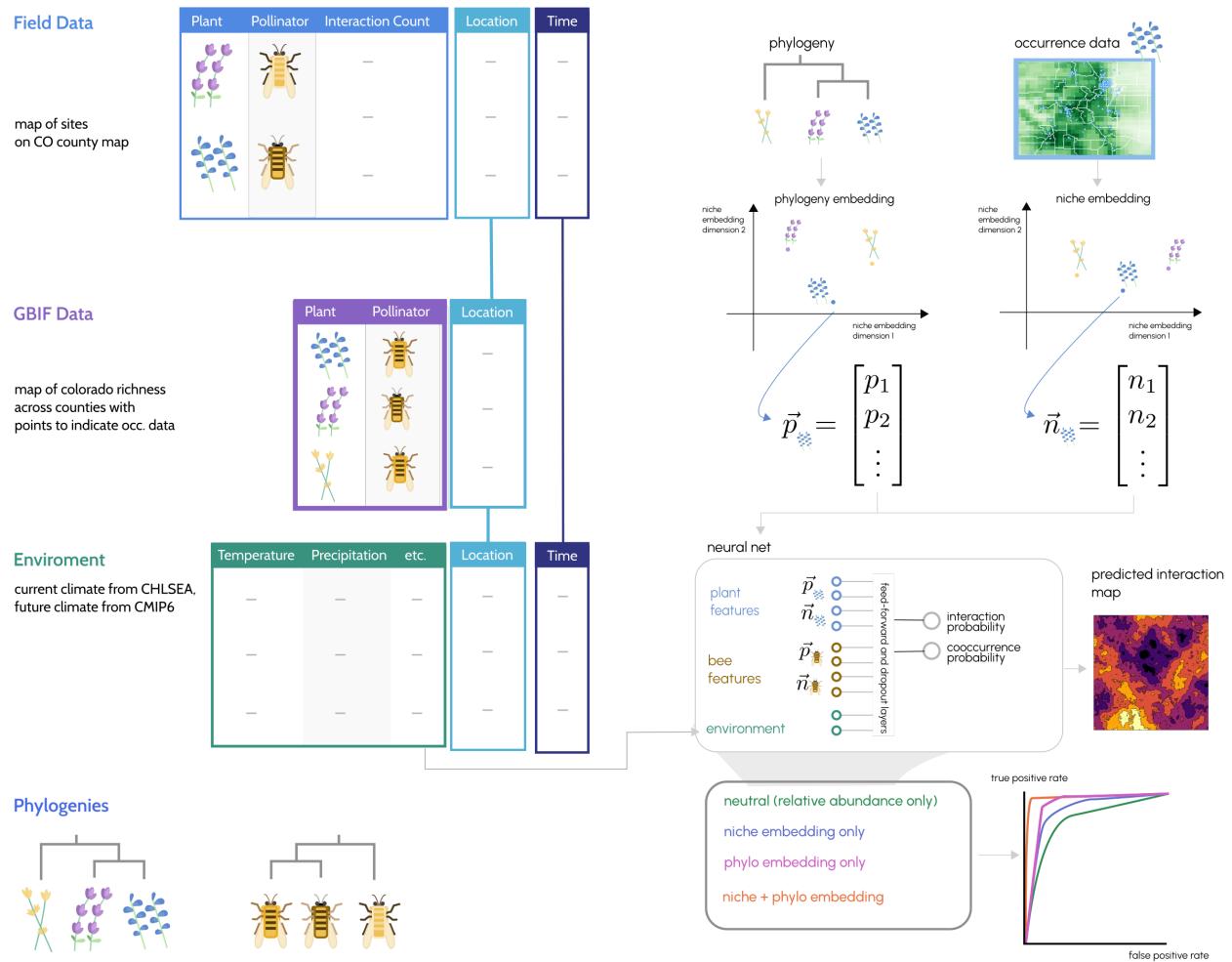


Figure 2: Chapter One conceptual figure. Left: the sources of data and how they can be synthesized. Right: The flow from data to interaction prediction using a few different interaction prediction models.

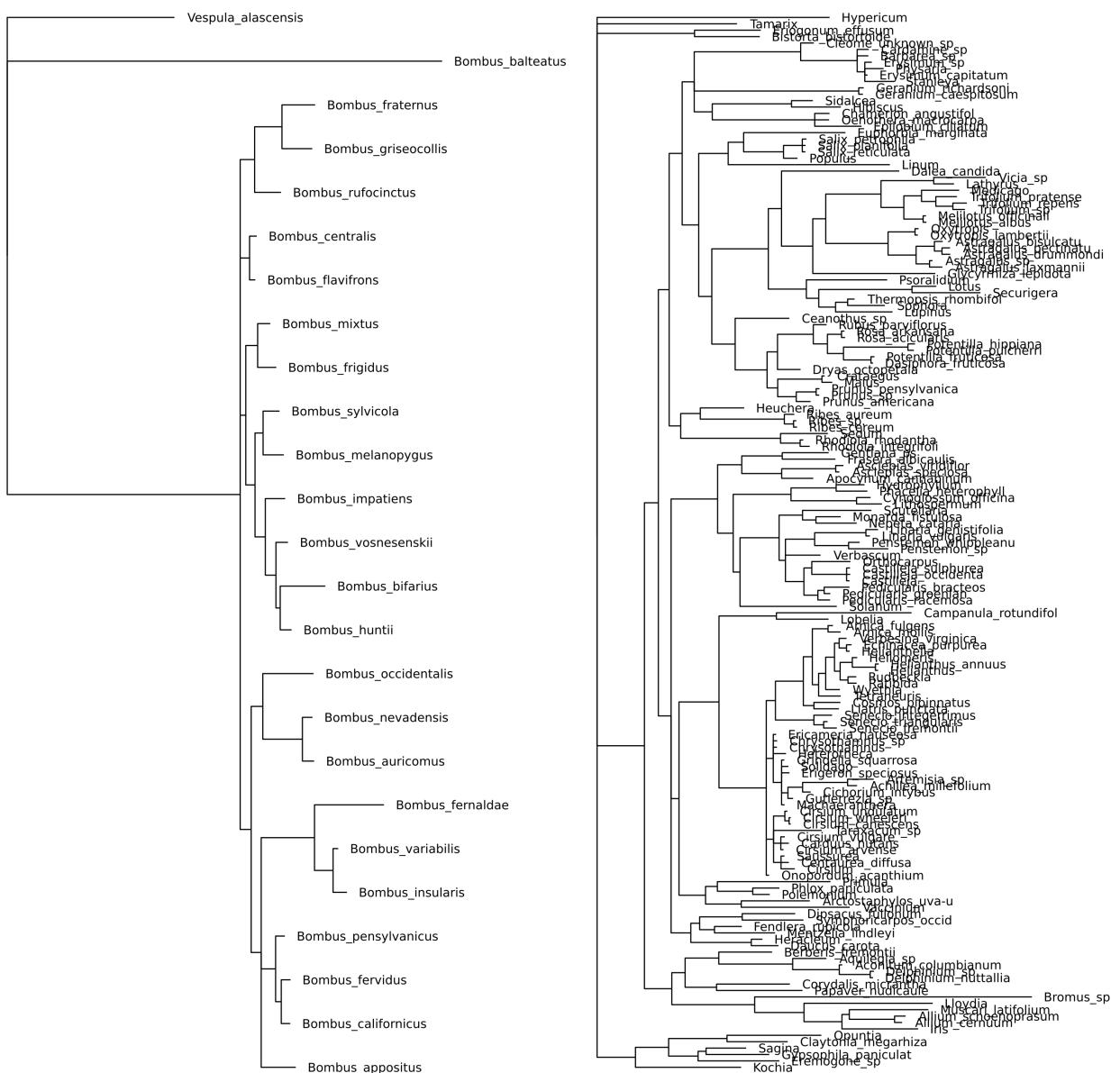


Figure 3: Phylogeny for both bumblebee species (left) and flower species (right)

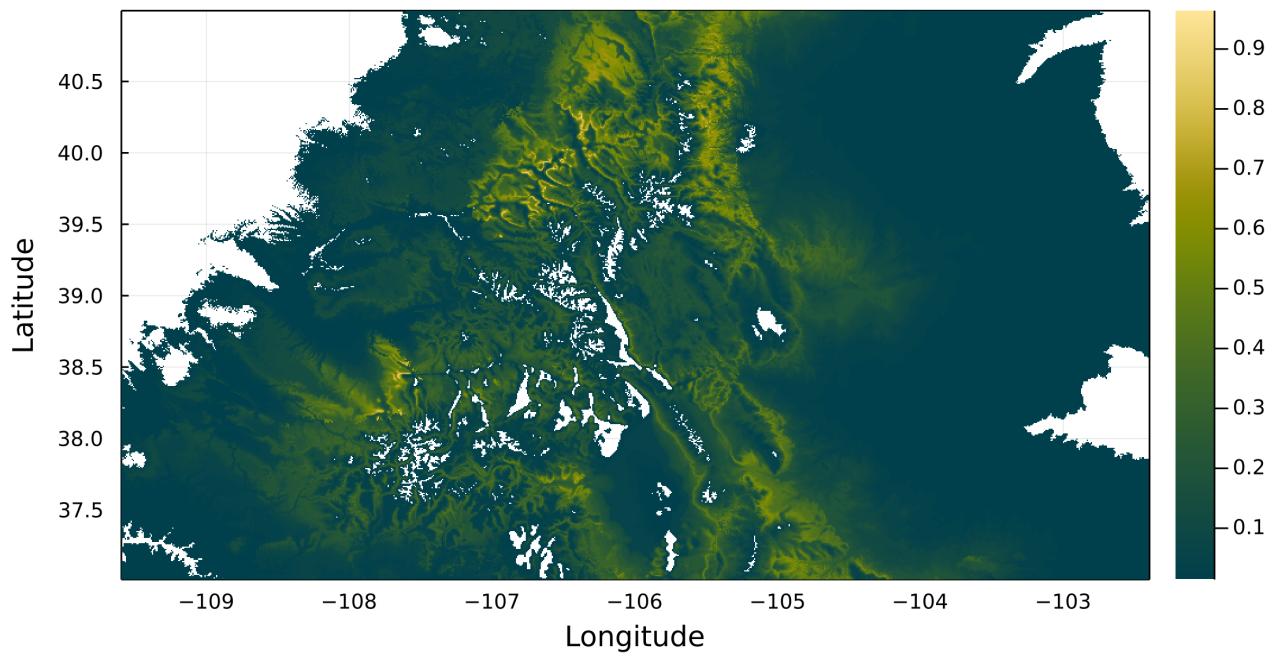


Figure 4: Example SDM for *Achillea millefolium*

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

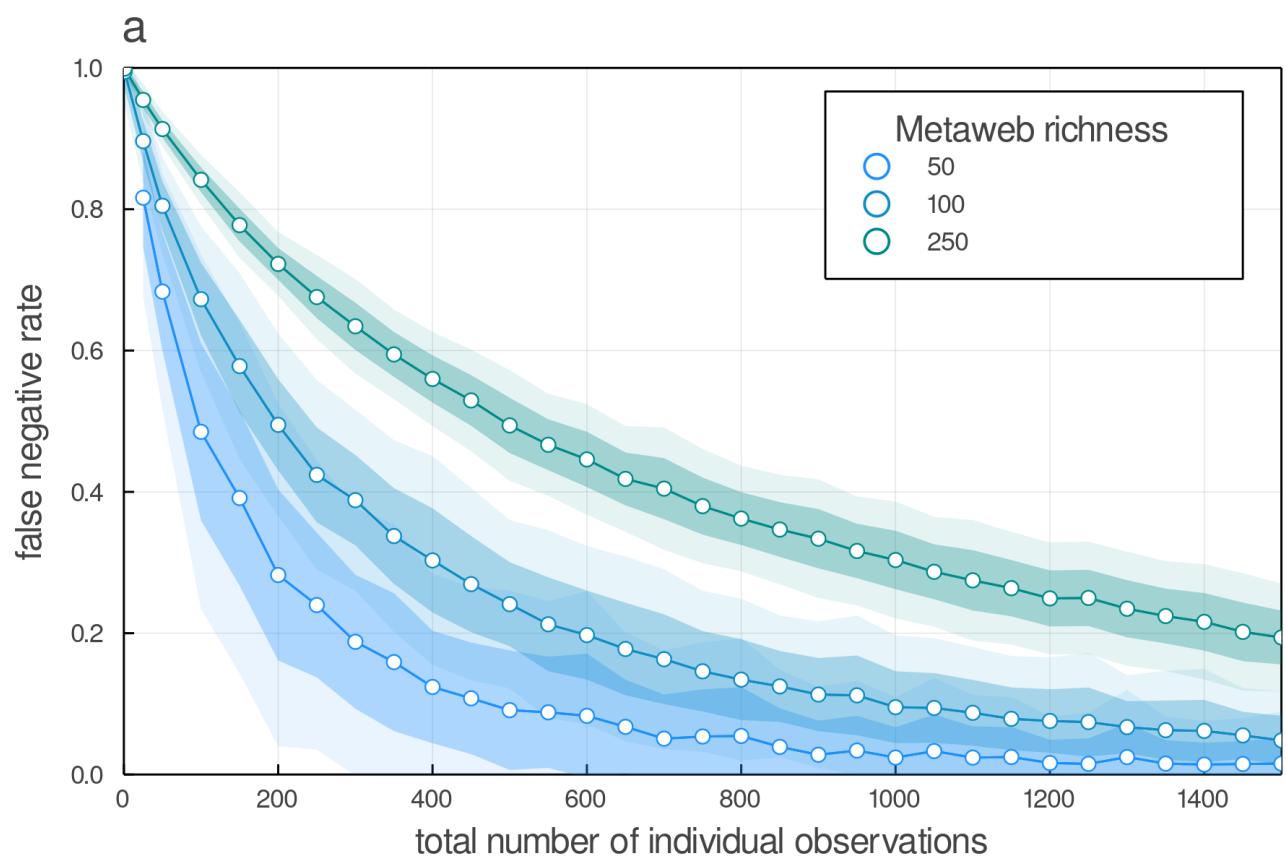


Figure 6: foo

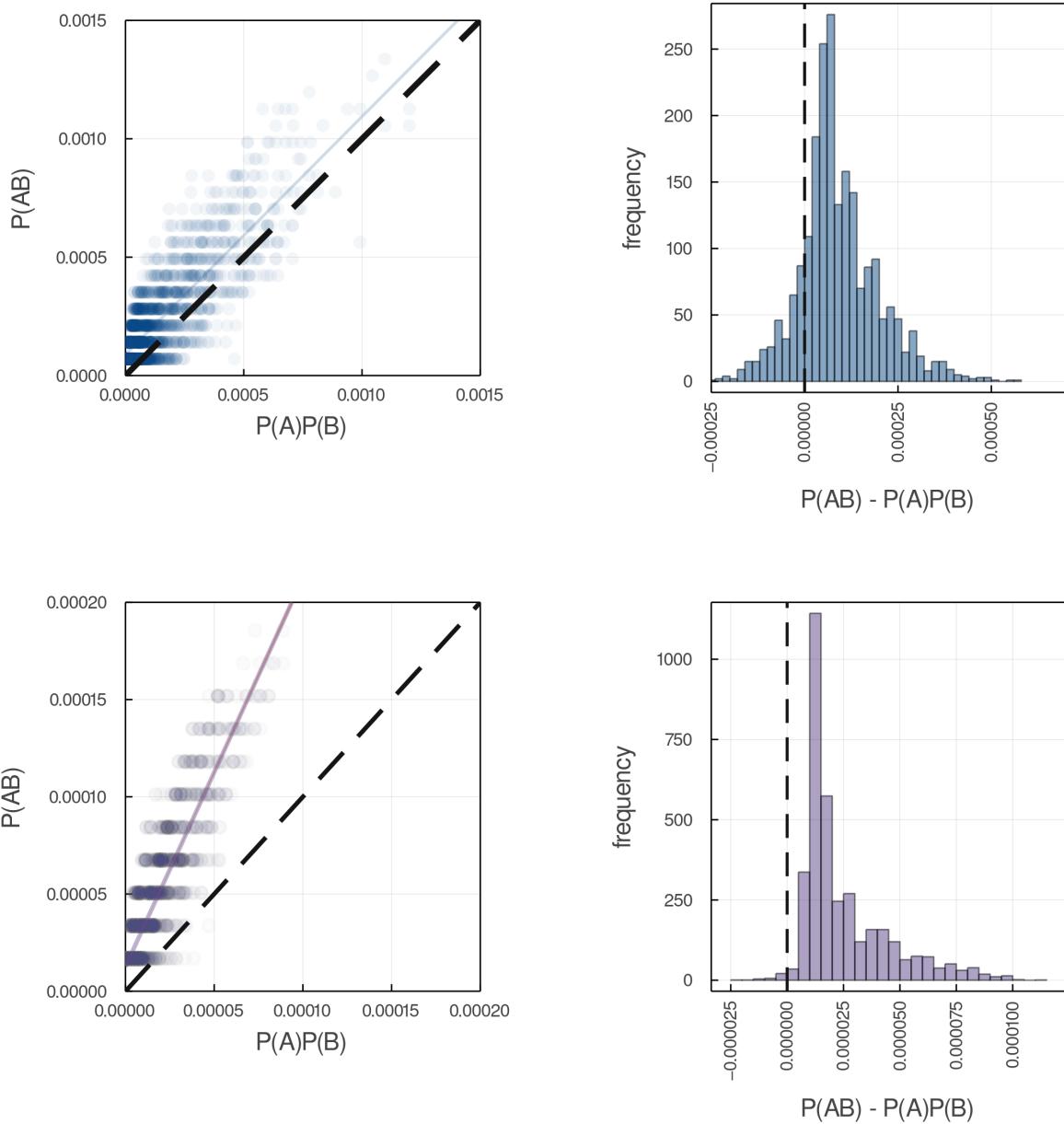


Figure 7: Demonstrates positive associations in co-occurrence

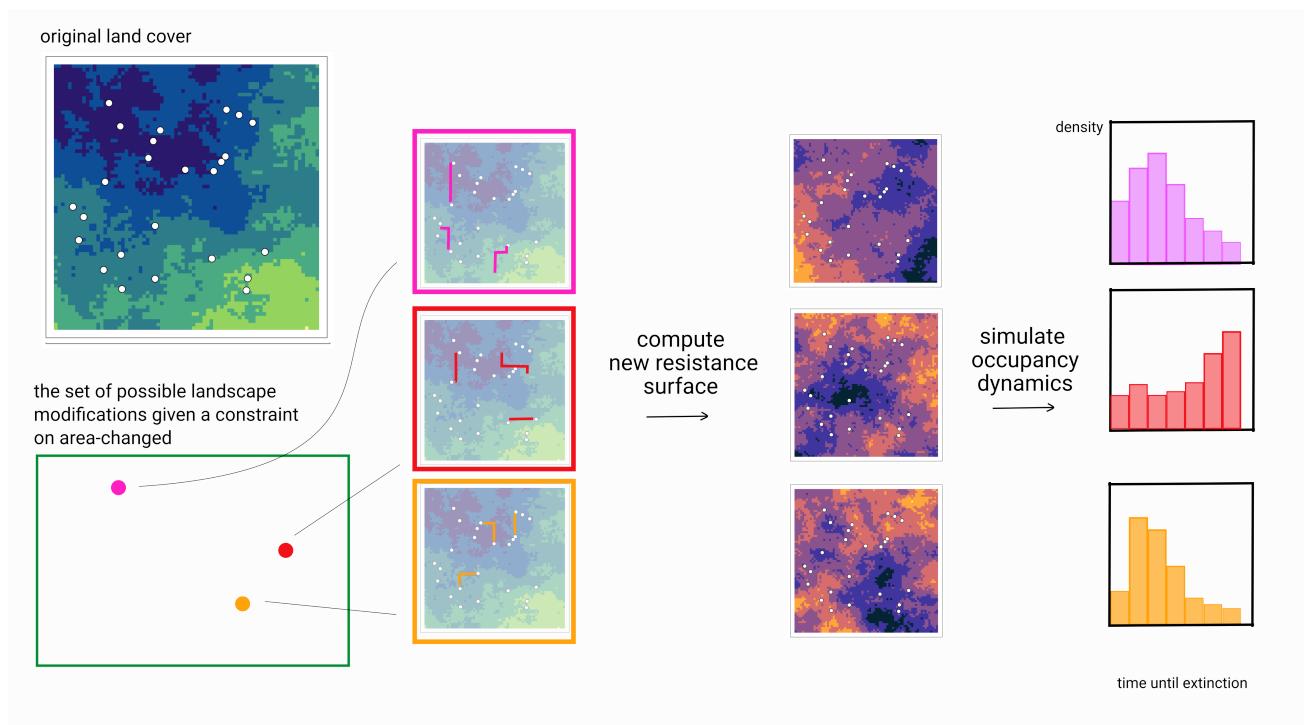


Figure 8: foo

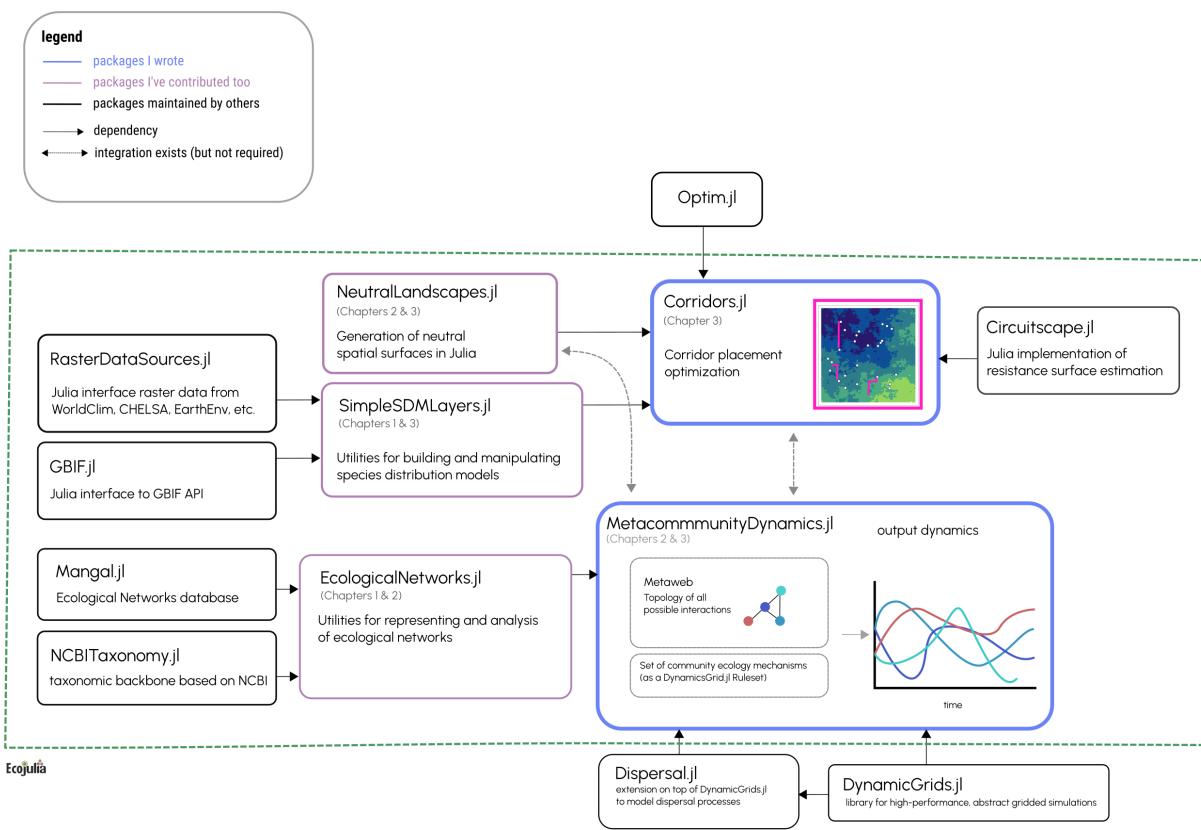


Figure 9: The structure of the software libraries used as part of MCD.jl