

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

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

30 becoming a quantitative science (sometime in the 20th century, depending on who you ask), became the
31 foundation for mathematical models in ecology.

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

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

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

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

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

64 [Figure 1 about here.]

65 The primary research challenges this thesis addresses: how do we design ecological samples to? How do
66 we build the software infrastructure to assimilate data from a variety of sources? How do we propagate
67 uncertainty from data to forecasts? The flow of chapters follows the flow in fig. 1 (left), from data
68 collection via a monitoring network, to forecasting an essential biodiversity variable (EBV), to optimizing
69 mitigation strategy based on constraints. In chapter one, we discuss how simulation can aid in the design
70 of ecological samples and monitoring network design. In chapter two we use data to forecast the
71 uncoupling of a plant-pollinator network. In chapter three, we apply simulation methods in landscape
72 ecology to optimize corridor placement with respect maximize the time-until-extinction of a
73 metapopulation. The fourth and final chapter is the software (*MetacommunityDynamics.jl*) which enables
74 the rest of the dissertation.

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

76 **Objective**

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

87

[Figure 2 about here.]

88 Methods

89 We begin by proposing a method to compute a null expectation of the probability of an interaction
 90 false-negative as a function number of total observations of individuals of *all species in the species pool*.
 91 This is done by simulating the process of observation, where the probability of observing a given species is
 92 its relative abundance. We show that the realized false-negative rate can be quite high, simply as a
 93 byproduct of the distribution of relative abundance in communities. We use a log-normal distribution of
 94 relative abundance (Hubbell 2001) and simulating the process of observation on food-webs generated
 95 using the niche model (Williams & Martinez 2000) with connectance parameterized by the flexible-links
 96 model (MacDonald *et al.* 2020). An example of this relation for networks with varying species richness is
 97 shown in fig. 4.

98

[Figure 3 about here.]

99 We then go on to testing some assumptions of this neutral model with empirical data. Primarily that we
 100 analytically show that our neutral model, if anything, underestimates the probability of false-negatives if
 101 there are positive associations between species co-occurrence, and we show these positive associations
 102 exist in two sets of spatially replicated samples of interaction networks (Thompson & Townsend 2000;
 103 Hadfield *et al.* 2014), fig. 5—further I'm planning to add the field data from the previous chapter into this
 104 analysis once available. Finally this chapter proposes a simulated annealing method to optimize the a set
 105 of n points in space to maximize the probability of detecting an interaction between two species a and b
 106 with *known* distributions D_a, D_b .

107 Results

108 The first major result is using the simulation of the observation process described above to generate
 109 expectations of interaction false-negative rate (FNR) as a function of total number of observations, with
 110 the goal being for this estimate to be used as correction for detection error when fitting an interaction
 111 prediction model. This relationship varies with the total richness of the metaweb fig. 4.

112

[Figure 4 about here.]

113 The second major result is that we analytically show that the this simulated observation model, by
114 assuming that there is no association between observing two species given that they interact, actually
115 under predicts the realized false-negative interaction rate. We then demonstrate that this positive
116 association association exists in two empirical systems fig. 5.

117

[Figure 5 about here.]

118 **Progress**

119 This chapter is mostly complete. The only remaining work is the implementation of simulated annealing
120 optimization process. This will be done by using a proposal function which takes a set of coordinates in
121 space and proposes a new location for each point based on a distance-decaying kernel.

122 **Chapter Two: Forecasting the spatial uncoupling of a plant-pollinator 123 network**

124 **Objective**

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

135

[Figure 6 about here.]

136 **Methods**

137 The data for this chapter is derived from multiple sources that can be split into four categories. (1) Field
138 data from three different field sites across Colorado, each with multiple plots across an elevational
139 gradient, for seven, seven, and three years respectively. This data was collected by Paul CaraDonna and
140 Jane Oglevie (from the Rocky Mountain Biological Laboratory; RMBL) and Julian Resasco (CU Boulder).

141 (2) GBIF spatial occurrence records of each of these species across Colorado, including a metaweb of
142 interactions across all of Colorado taken from GBIF. (3) Remotely sensed data consisting of current and
143 forecasting bioclimatic variables from CHELSA. (4) Phylogenies for both bee and flower species derived
144 from NCBI GenBank barcodes for mitochondrial COI (bumblebees) and chloroplast rbcL (flowers).

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

147 We do this using a set of candidate interaction prediction models: relative abundance only, phylogenetic
148 embedding only (a la Strydom *et al.* (2021a)), niche embedding only (Gravel *et al.* 2019), and all pairwise
149 combinations of those constituent models. After validating and selecting the best performing model, we
150 then predict how these distributions of each of these species will change under the CMIP6 consensus
151 climate forecast (Karger *et al.* 2017), and then finally quantify the reduction in spatial between species for
152 which there is a predicted interaction.

153 **Results**

154 The in-progress results are the prerequisites for the analysis outlined above: phylogenies for both plant
155 and bee species (appendix figure one) and species distribution models for all species (an example shown in
156 appendix figure two).

157 **Progress**

158 At the moment, we have derived phylogenies and SDMs for all the species present in the Colorado GBIF
159 metaweb (appendix figures one and two). I’ve also been exploring the data available from Julian Resasco.
160 The primary constraint on further progress is that we are waiting on the finalization of a data sharing
161 agreement with RMBL.

162 **Chapter Three: Optimizing corridor placement against ecological**
163 **dynamics**

164 **Objective**

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

175 **Methods**

176 [Figure 7 about here.]

177 We propose various landscape modifications which alter the cover of a landscape, represented as a raster.
178 We then compute a new resistance surface based on the proposed landscape modification using
179 Circuitscape (McRae *et al.* 2008), and based on the values of resistance to dispersal between pairs of
180 locations we simulate spatially-explicit metapopulation dynamics model (Hanski & Ovaskainen 2000;
181 Ovaskainen *et al.* 2002) to estimate a distribution of time until extinction for each landscape modification.
182 The largest challenge in implementing this algorithm is the space of potential modifications grows as
183 $O((nm)!)$ for an n by m raster. For most actual landscapes to which we wish to apply this method, the set
184 of possible modifications becomes uncomputably large, so we use simulated annealing to explore the
185 search space of possible modifications to estimate the modification that maximizes the time-until
186 extinction of simulated metapopulation dynamics under that hypothetical modified landscape.
187 The biggest challenge in implementing simulated annealing in this context is defining a proposal function

188 for landscape modifications. At the moment this is done by computing the minimum-spanning-tree
189 (MST) of the spatial nodes, and then proposing corridors that connect nodes that are already connected in
190 the MST. The primary reason for doing this is the cut down the size of the solution space to enable quicker
191 convergence, although the final software that implements this algorithm will enable alternative methods
192 of proposing modifications.

193 The goal output of this chapter is not only to provide a set of discrete corridor options, but also to rank the
194 cells in the raster by priority based on how many times they are converted in the distribution of “good”
195 corridors after simulated annealing has reached a pseudoequilibrium. Further, the final component of this
196 chapter is measuring the effect of land-use change on the robustness of the optimized corridor by
197 simulating various neutral models of urban and agricultural sprawl, and determining if the proposed
198 modifications still maximize time-until-extinction when the landcover in the landscape is not static.

199 **Progress**

200 The current progress is that I have an algorithm for proposing landscape modifications and a simple
201 implementation of simulated annealing. The only gap left is implementing Circuitscape estimation of
202 resistance surfaces.

203 **Chapter Four: MetacommunityDynamics.jl: a virtual laboratory for 204 community ecology**

205 **Objective**

206 The final chapter consists of a collection of modules in the Julia language for different aspects of
207 community ecology, including most of the code used for the preceding chapters. Indeed
208 MetacommunityDynamics.jl (MCD.jl) is the epicenter of this set of tools, but due to the nature of the Julia
209 language, MCD.jl is interoperable with several existing packages within the EcoJulia organization,
210 including several to which I have contributed. We need a software library like this to generate synthetic
211 data from a *known* set of mechanisms and parameters to test our methods for parameter inference and
212 forecasting on this *known* system to assess the effectiveness of these inference and forecasting methods.

213

[Figure 8 about here.]

214 **Methods**

215 A diagram showing the relation between these packages is shown in fig. 8. `MetacommunityDynamics.jl` is
 216 built on `DynamicGrids.jl`, a library for high-performance gridded simulations in the Julia language, and
 217 `Dispersal.jl` (Maino *et al.* 2021), and extension of `DynamicGrids.jl` specifically for modeling organism
 218 dispersal. It also contains integrations with `EcologicalNetworks.jl` (Poisot *et al.* 2019) to generate
 219 metawebs, or can use empirical networks from `Mangal.jl` (Banville *et al.* 2021). It implements the general
 220 framework for community dynamics proposed by Vellend (2010), where all community processes can
 221 divided into four categories: selection, dispersal, drift, and speciation.

222 **Results**

223 Below (fig. 9) is a sample output of simulated food-web dynamics for a metaweb of 100 species generated
 224 using the minimum-potential-niche model with connectance $C = 0.05$ and forbidden-link probability of
 225 0.5. The dynamics change according to a Lotka-Volterra functional response, dispersal (with dispersal
 226 distance inverse proportional to trophic-level), linear mortality, and logistic growth for any species at the
 227 producer trophic-level.

228

[Figure 9 about here.]

229 **Progress**

230 The software as it exists is capable of simulating the biomass dynamics of arbitrarily large food-webs using
 231 Lotka-Volterra, Holling Type-II, or Holling Type-III functional responses. It currently has methods to
 232 implement Gaussian drift, and various forms of dispersal via `Dispersal.jl`. Also occupancy dynamics for
 233 Levins metapopulations (Levins 1969), and spatially explicit Hanski-Ovaskainen metapopulations
 234 (Hanski & Ovaskainen 2000; Ovaskainen *et al.* 2002). This is most of what needs to exist for the preceding
 235 chapters. In-progress functionality includes selection (which affects growth-rate) on arbitrary
 236 environmental variables in progress, as well as traits.

237 **Discussion**

238 Developing a system for global biodiversity monitoring is an imperative to mitigate biodiversity loss and
239 its impacts on humanity. In my thesis I hope to provide a template for the digital infrastructure that
240 enables the pipeline from data collection, to forecast, to actionable information, both through software
241 that can be used to solve these problems (chapters one, three, four), and vignettes of how these software
242 can be applied (chapters one, two).

243 Biodiversity science would be wise to use the success of numerical-weather prediction and the Earth
244 monitoring system that supports it as a template. Further, we should embrace methodological advances in
245 computational statistics that can enable more robust prediction of the dynamics of complex systems than
246 is possible using the tools of analytic models—not just because they can provide more efficacious
247 predictions, but also because they enable a more direct representation of uncertainty, which is crucial for
248 developing ecological forecasting models which provide actionable information to stakeholders.

249 **References**

- 250 Banville, F., Vissault, S. & Poisot, T. (2021). Mangal.jl and EcologicalNetworks.jl: Two complementary
251 packages for analyzing ecological networks in Julia. *Journal of Open Source Software*, 6, 2721.
- 252 Bascompte, J. & Jordano, P. (2007). Plant-Animal Mutualistic Networks: The Architecture of Biodiversity.
253 *Annual Review of Ecology, Evolution, and Systematics*, 38, 567–593.
- 254 Bauer, P., Thorpe, A. & Brunet, G. (2015). The quiet revolution of numerical weather prediction. *Nature*,
255 525, 47–56.
- 256 Beckage, B., Gross, L.J. & Kauffman, S. (2011). The limits to prediction in ecological systems. *Ecosphere*, 2,
257 art125.
- 258 CaraDonna, P.J., Petry, W.K., Brennan, R.M., Cunningham, J.L., Bronstein, J.L., Waser, N.M., *et al.* (2017).
259 Interaction rewiring and the rapid turnover of plantpollinator networks. *Ecology Letters*, 20, 385–394.
- 260 Chen, Y., Angulo, M.T. & Liu, Y.Y. (2019). Revealing Complex Ecological Dynamics via Symbolic
261 Regression. *BioEssays*, 41, 1900069.

- 262 Dietze, M.C. (2017). Prediction in ecology: A first-principles framework. *Ecological Applications*, 27,
263 2048–2060.
- 264 Gravel, D., Baiser, B., Dunne, J.A., Kopalke, J.-P., Martinez, N.D., Nyman, T., *et al.* (2019). Bringing Elton
265 and Grinnell together: A quantitative framework to represent the biogeography of ecological
266 interaction networks. *Ecography*, 42, 401–415.
- 267 Hadfield, J.D., Krasnov, B.R., Poulin, R. & Nakagawa, S. (2014). A Tale of Two Phylogenies: Comparative
268 Analyses of Ecological Interactions. *The American Naturalist*, 183, 174–187.
- 269 Hanski, I. & Ovaskainen, O. (2000). The metapopulation capacity of a fragmented landscape. *Nature*, 404,
270 755–758.
- 271 Hill, C., DeLuca, C., Balaji, Suarez, M. & Da Silva, A. (2004). The architecture of the Earth System
272 Modeling Framework. *Computing in Science Engineering*, 6, 18–28.
- 273 Hubbell, S.P. (2001). *The unified neutral theory of biodiversity and biogeography*. Monographs in
274 population biology. Princeton University Press, Princeton.
- 275 Karger, D.N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., *et al.* (2017). Climatologies at
276 high resolution for the earth's land surface areas. *Scientific Data*, 4, 170122.
- 277 Levin, S.A. (1992). The Problem of Pattern and Scale in Ecology: The Robert H. MacArthur Award
278 Lecture. *Ecology*, 73, 1943–1967.
- 279 Levins, R. (1969). Some Demographic and Genetic Consequences of Environmental Heterogeneity for
280 Biological Control. *Bulletin of the Entomological Society of America*, 15, 237–240.
- 281 MacDonald, A.A.M., Banville, F. & Poisot, T. (2020). Revisiting the Links-Species Scaling Relationship in
282 Food Webs. *Patterns*, 1.
- 283 Maino, J.L., Schouten, R. & Umina, P. (2021). Predicting the global invasion of *Drosophila suzukii* to
284 improve Australian biosecurity preparedness. *Journal of Applied Ecology*, 58, 789–800.
- 285 Makiola, A., Compson, Z.G., Baird, D.J., Barnes, M.A., Boerlijst, S.P., Bouchez, A., *et al.* (2020). Key
286 Questions for Next-Generation Biomonitoring. *Frontiers in Environmental Science*, 7.
- 287 McRae, B.H., Dickson, B.G., Keitt, T.H. & Shah, V.B. (2008). Using Circuit Theory to Model Connectivity
288 in Ecology, Evolution, and Conservation. *Ecology*, 89, 2712–2724.

- 289 Ovaskainen, O., Sato, K., Bascompte, J. & Hanski, I. (2002). Metapopulation Models for Extinction
290 Threshold in Spatially Correlated Landscapes. *Journal of Theoretical Biology*, 215, 95–108.
- 291 Ovaskainen, O., Sato, K., Bascompte, J. & Hanski, I. (2002). Metapopulation Models for Extinction
292 Threshold in Spatially Correlated Landscapes. *Journal of Theoretical Biology*, 215, 95–108.
- 293 Pennekamp, F., Iles, A.C., Garland, J., Brennan, G., Brose, U., Gaedke, U., *et al.* (2019). The intrinsic
294 predictability of ecological time series and its potential to guide forecasting. *Ecological Monographs*, 89,
295 e01359.
- 296 Petchey, O.L., Pontarp, M., Massie, T.M., Kéfi, S., Ozgul, A., Weilenmann, M., *et al.* (2015). The ecological
297 forecast horizon, and examples of its uses and determinants. *Ecology Letters*, 18, 597–611.
- 298 Peterman, W.E. (2018). ResistanceGA: An R package for the optimization of resistance surfaces using
299 genetic algorithms. *Methods in Ecology and Evolution*, 9, 1638–1647.
- 300 Poisot, T., Bélisle, Z., Hoebelke, L., Stock, M. & Szefer, P. (2019). EcologicalNetworks.jl: Analysing
301 ecological networks of species interactions. *Ecography*, 42, 1850–1861.
- 302 Strydom, T., Bouskila, S., Banville, F., Barros, C., Caron, D., Farrell, M.J., *et al.* (2021a). Food web
303 reconstruction through phylogenetic transfer of low-rank network representation.
- 304 Strydom, T., Catchen, M.D., Banville, F., Caron, D., Dansereau, G., Desjardins-Proulx, P., *et al.* (2021b). A
305 *Roadmap Toward Predicting Species Interaction Networks (Across Space and Time)* (Preprint).
306 EcoEvoRxiv.
- 307 Thompson, R.M. & Townsend, C.R. (2000). Is resolution the solution?: The effect of taxonomic resolution
308 on the calculated properties of three stream food webs. *Freshwater Biology*, 44, 413–422.
- 309 Urban, M.C., Travis, J.M.J., Zurell, D., Thompson, P.L., Synes, N.W., Scarpa, A., *et al.* (2021). Coding for
310 Life: Designing a Platform for Projecting and Protecting Global Biodiversity. *BioScience*.
- 311 Vellend, M. (2010). Conceptual Synthesis in Community Ecology. *The Quarterly Review of Biology*, 85,
312 183–206.
- 313 Williams, R.J. & Martinez, N.D. (2000). Simple rules yield complex food webs. *Nature*, 404, 180–183.

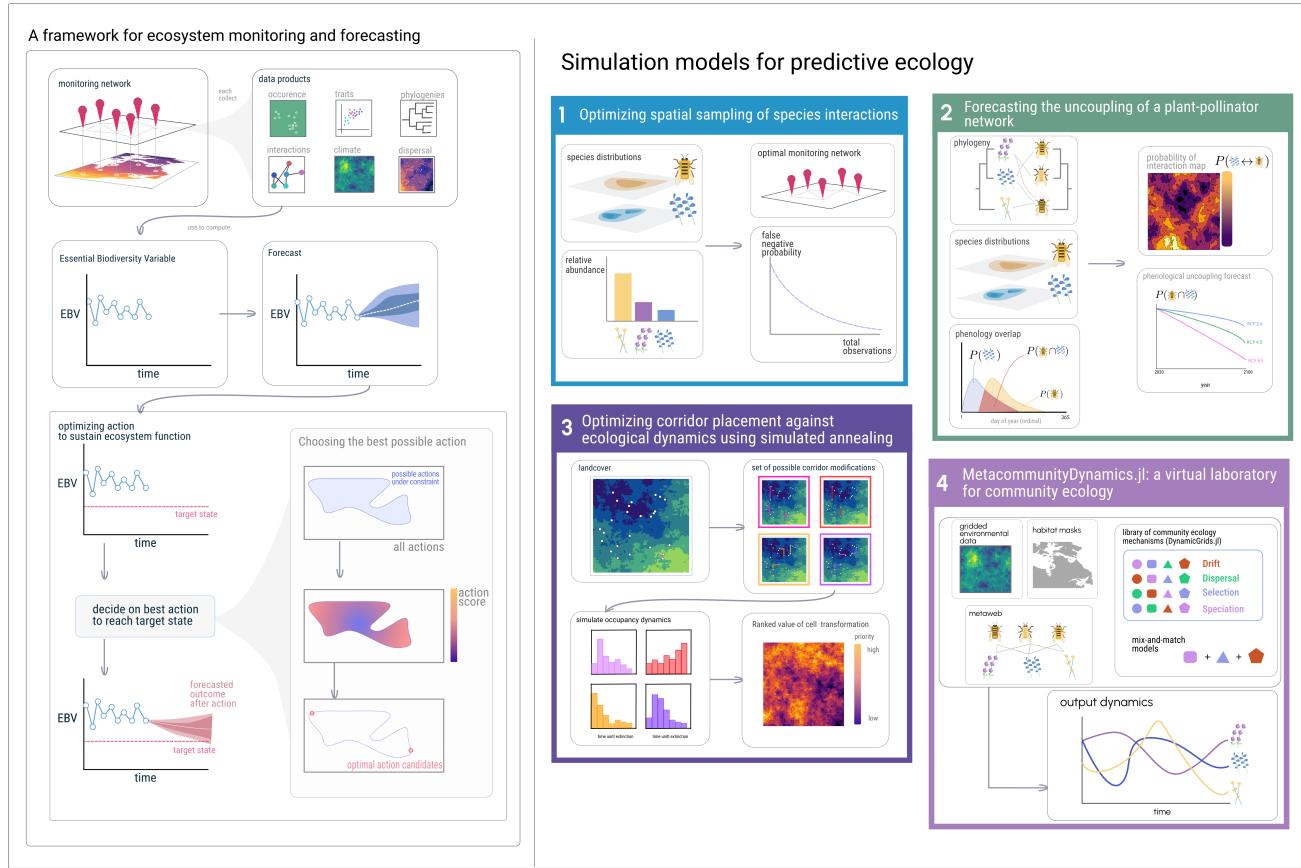


Figure 1: Left: a framework for ecosystem monitoring, forecasting, and mitigation. From the top, a set of biodiversity observatories which form a monitoring network. Each collect various biodiversity data products. From this raw data, we derive essential biodiversity variables (EBVs), and forecast how they change over time. Based on this forecast, we wish to choose the best possible mitigation strategy to maximize the chance the realized future outcomes of the EBV approaches the target state. Right: Each panel represents a chapter of the thesis, which follows the flow of the framework on the left.

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

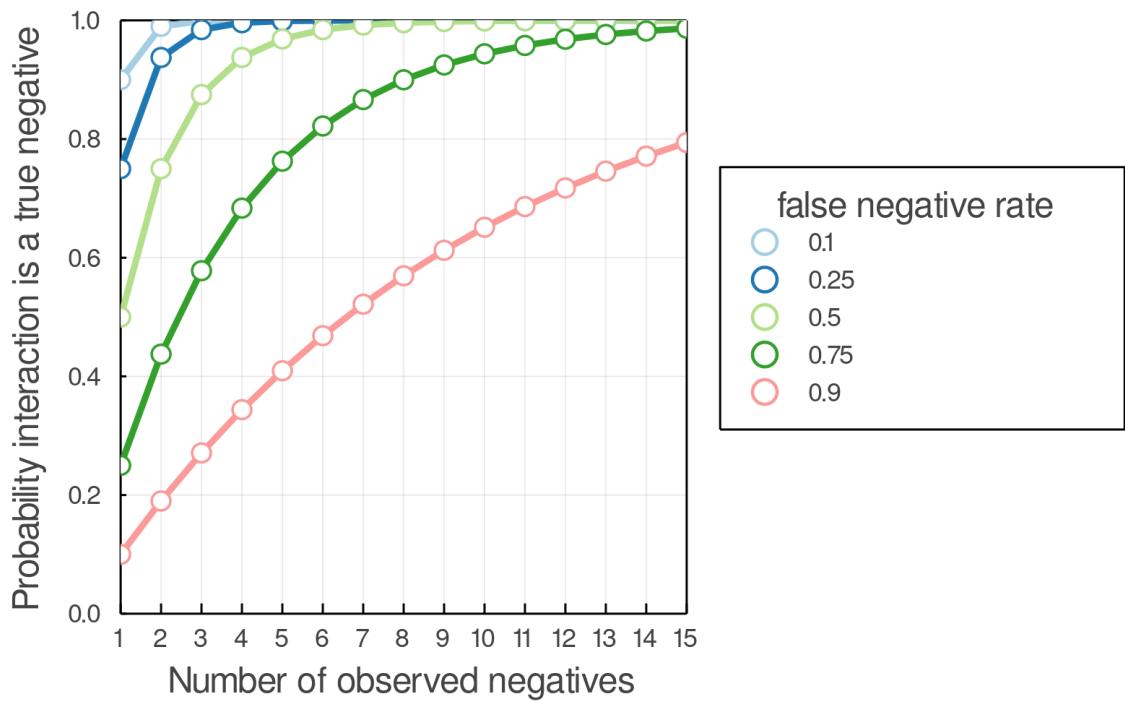


Figure 3: Relationship between total number of observed interactions negatives (x-axis) and the probability that an interaction is a true-negative as a function of different realized false-negative rates (colors), assuming each observed negative is independent—the same as the negative-bernoulli distribution.

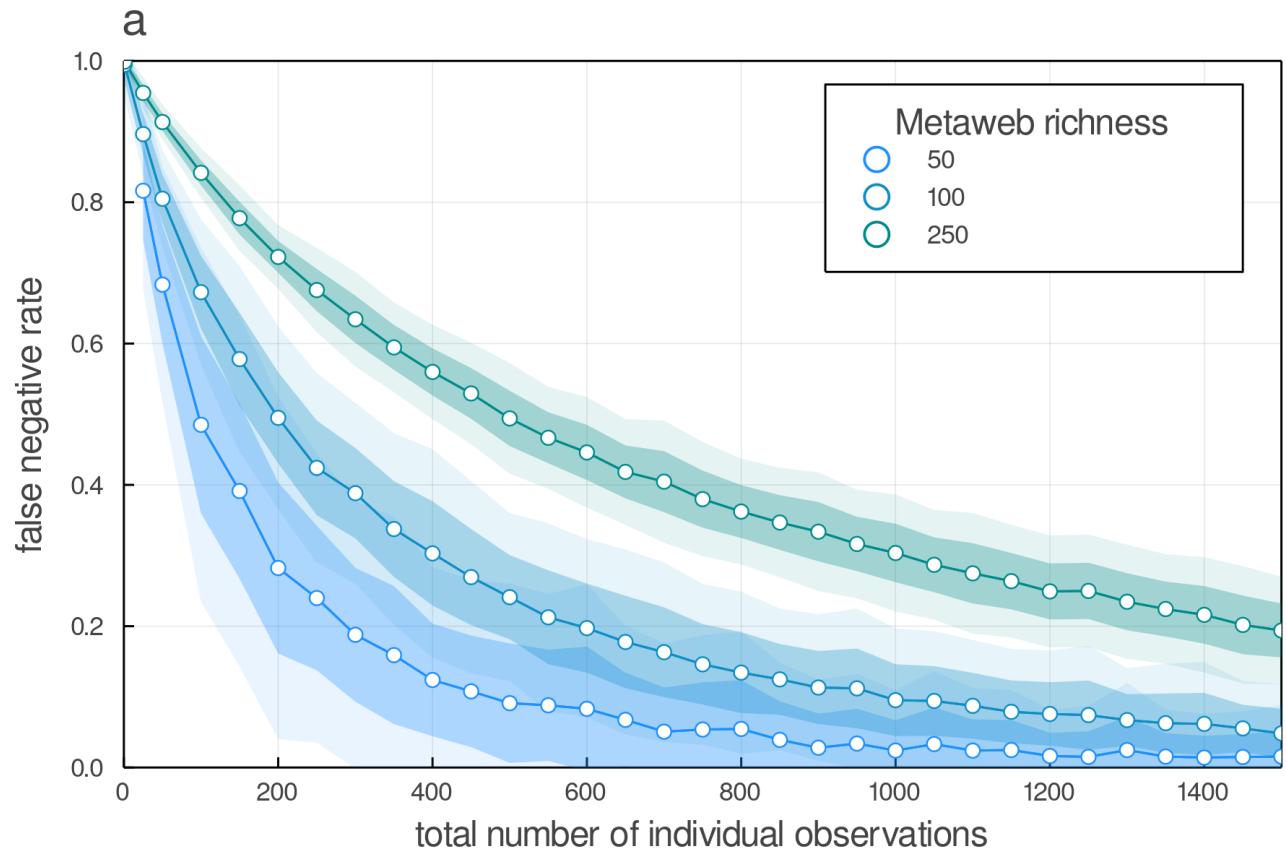


Figure 4: The realized false-negative-rate of interaction detection (y-axis) as a function of the total number of observations of all species in the species pool (different richnesses in different colors). Each point is the mean of 50 replicates, with one standard-deviation in the first shade, and two standard deviations in the second shade.

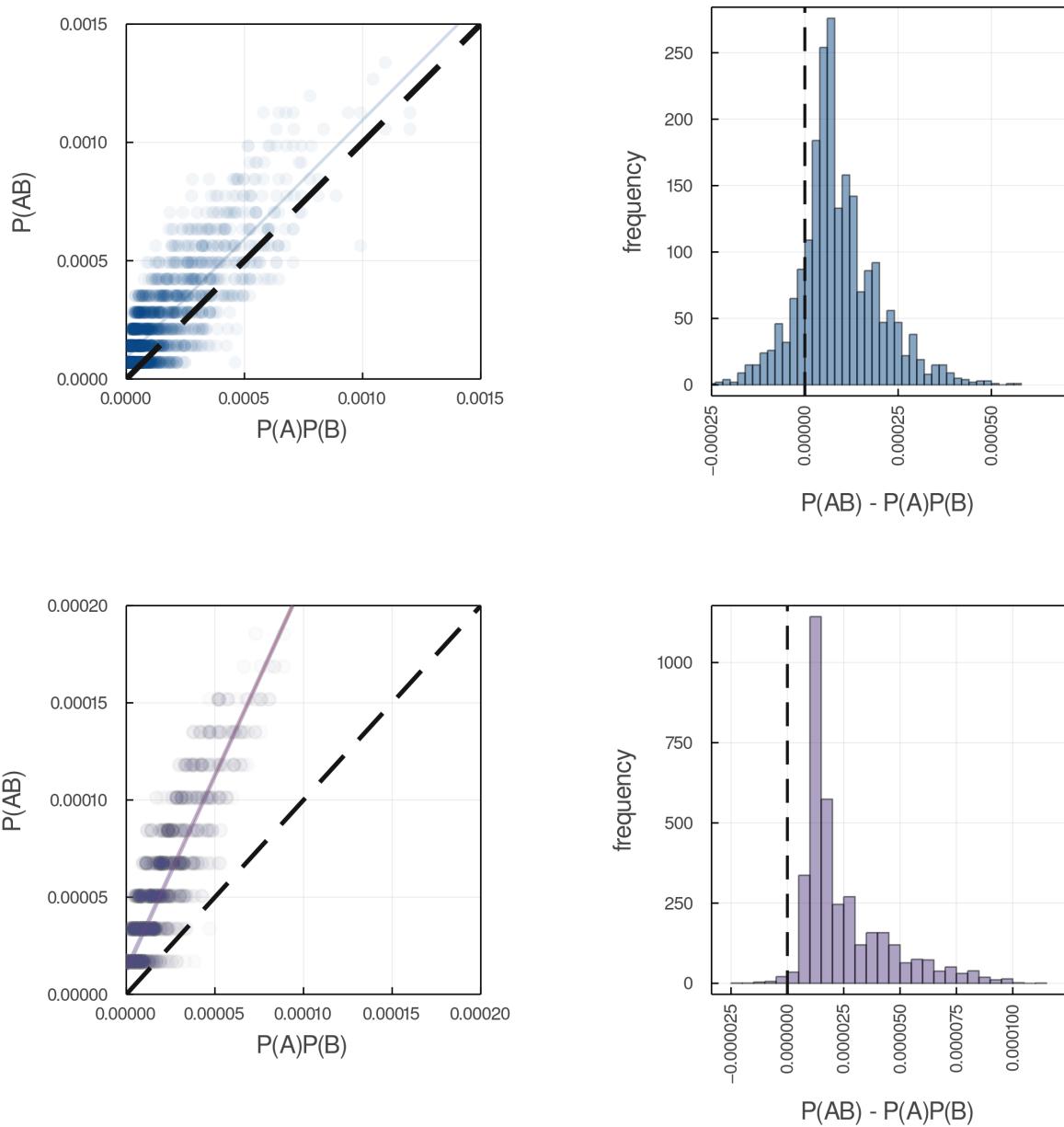


Figure 5: Demonstrates positive associations in co-occurrence. Left: the product of the marginal probability of observing two species (A and B) in a sample (x-axis) against the computed joint probability of observing these species together (y-axis). Dashed line indicates $y = x$, meaning no association between the two. Each point is an observed interaction between two species. Right: the distribution of the difference between these joint and marginal probabilities. Both are non-zero with $p < 10^{-50}$ via a t-test.

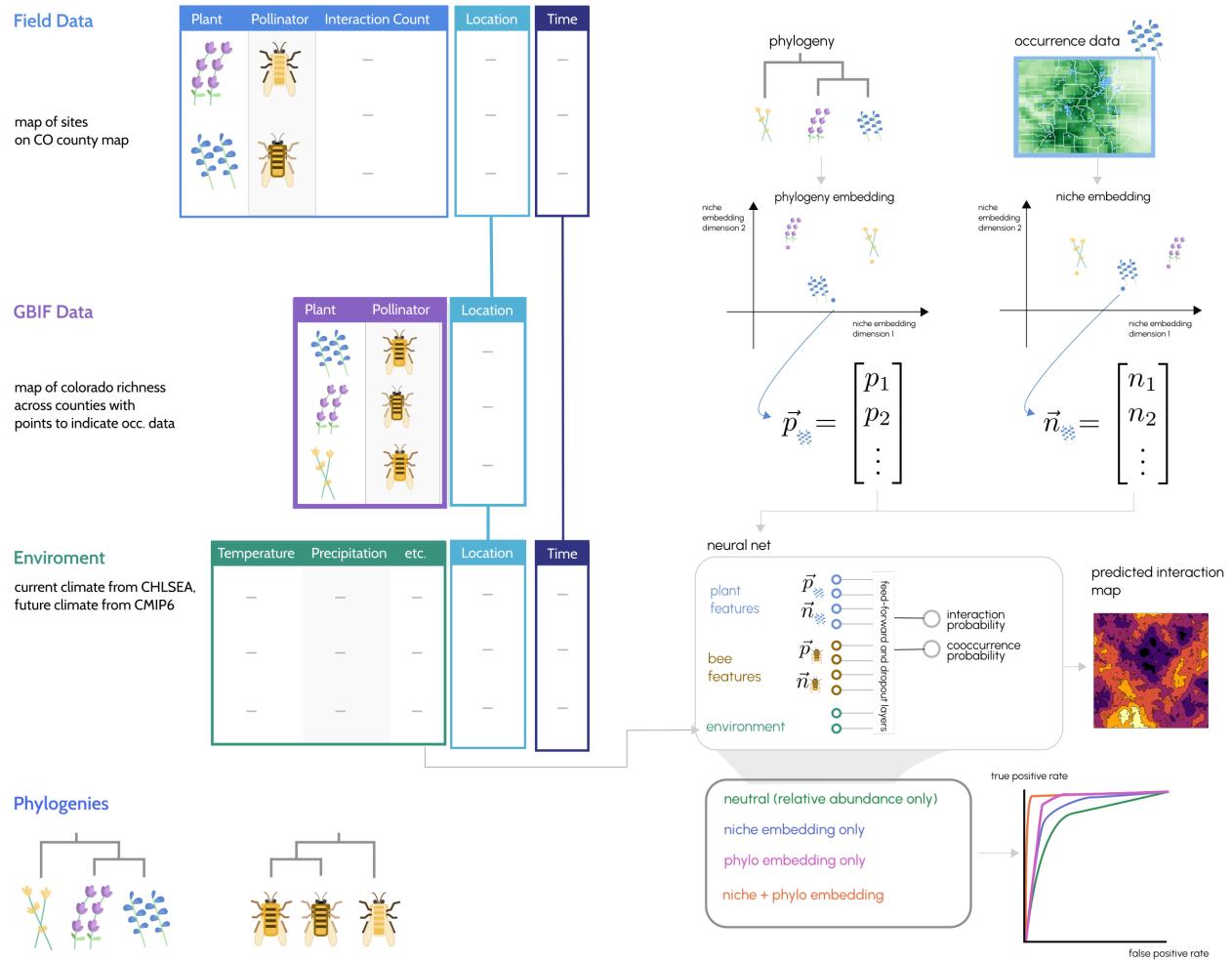


Figure 6: Chapter One conceptual figure. Left: the sources of data and how they can be combined. Right: The flow from data to interaction prediction using a few different interaction prediction models.

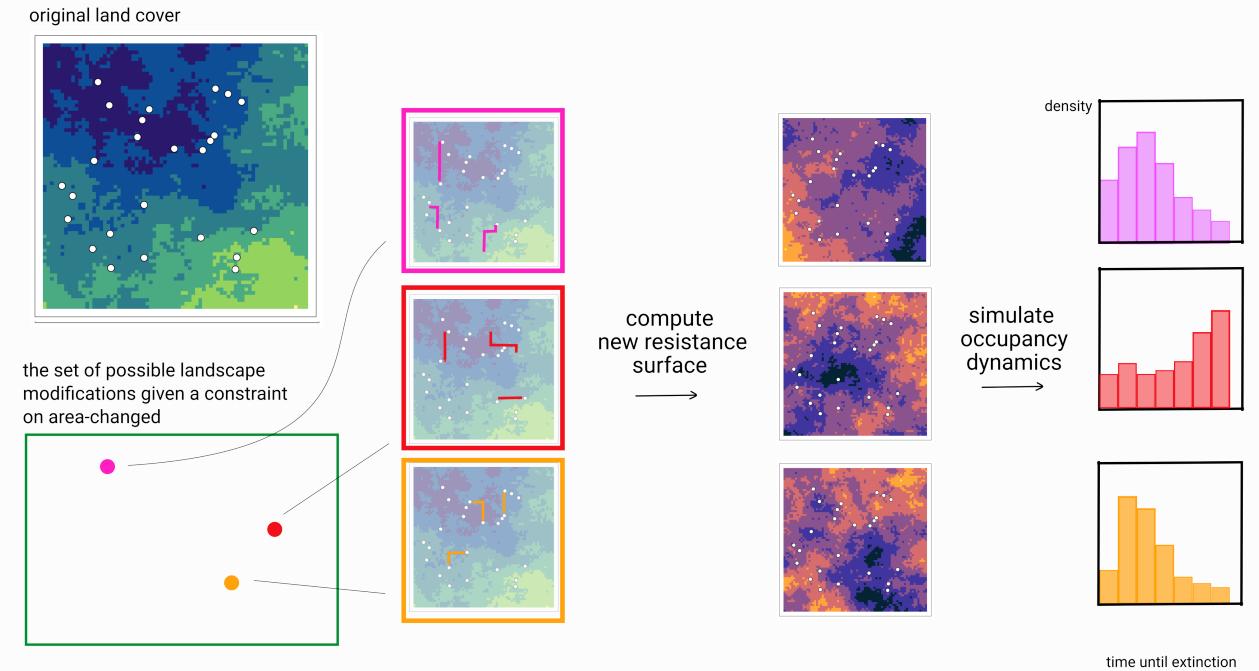


Figure 7: A conceptual example of how we go from a map of landcover with a set of points where occurrence of a species has been recorded (top left), to the set of all possible landscape modifications (green box, where each point in the green box is a unique landscape modification, with three examples shown as pink, red and orange dots), to computing resistance surfaces based on proposed landscape modifications (center) and then simulate the distribution of extinction times for a metapopulation in this new landscape (right).

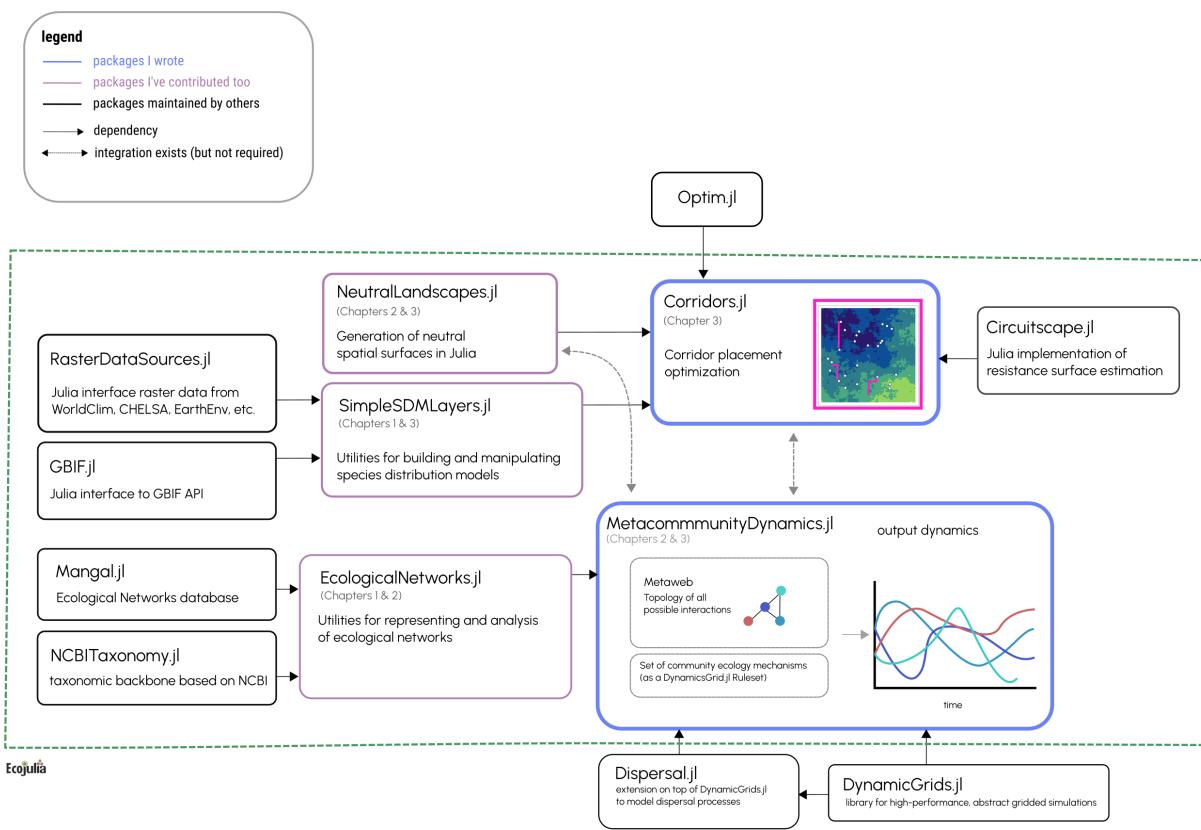


Figure 8: The structure of the software libraries used as part of MCD.jl

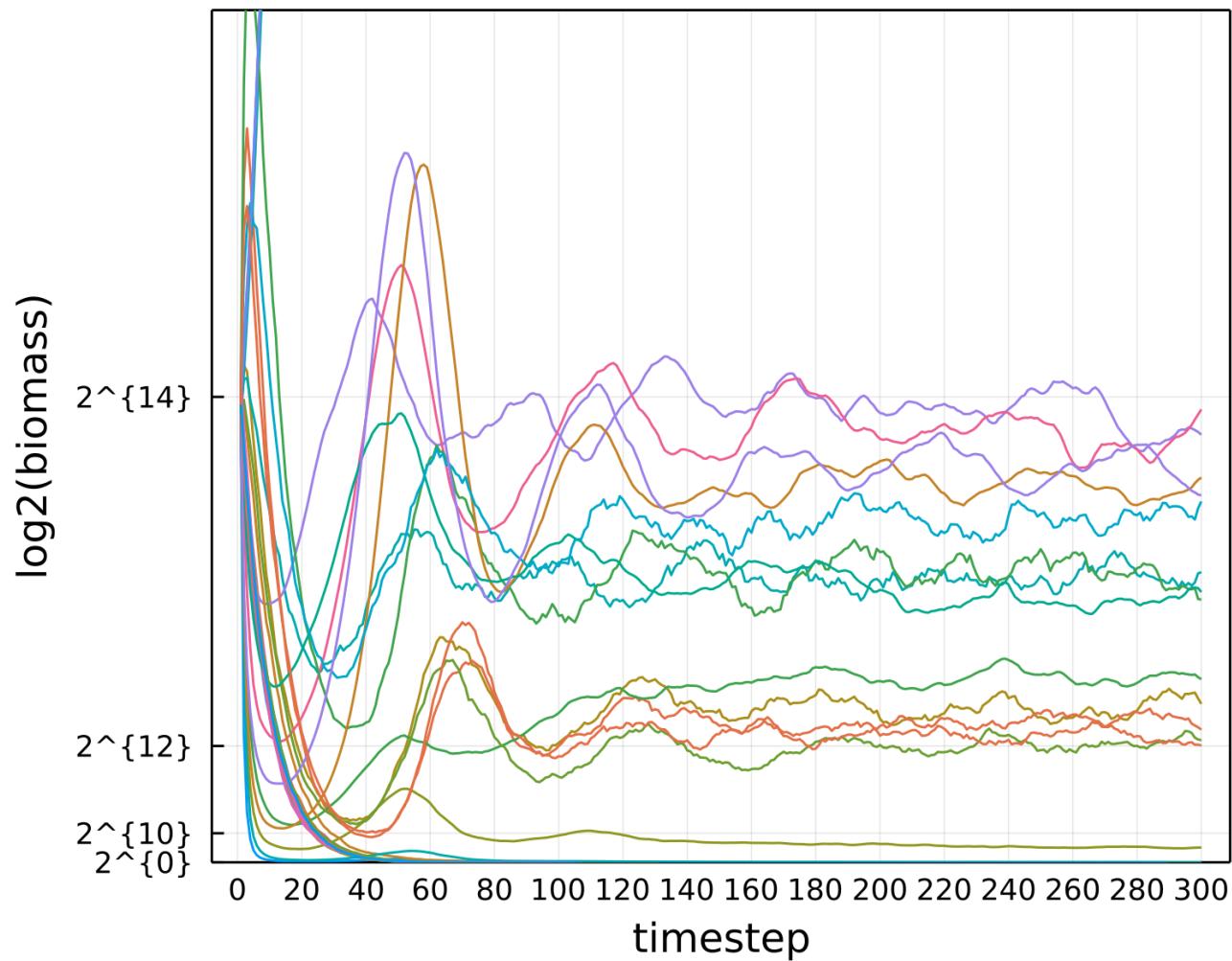


Figure 9: Sample output of simulated food web dynamics from MetacommunityDynamics.jl