

Thesis proposal

Michael D. Catchen

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 Earth's land cover.
5 The bulk of this change has occurred within the last several hundred years, a geological instant, inducing
6 a sudden shift in conditions to Earth's climate and biosphere. As a result *ecological forecasting*—modeling
7 how ecosystems and their services will change in the future—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 difficulty 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 a
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 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 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-II or DeAngelis-Beddington functional response).
24 The form of this functional response in real systems, and whether it is meaningfully non-zero for a given
25 species interaction, is effectively unknown and must be predicted (Strydom *et al.* 2021b), and some forms
26 of these dynamics are inherently more "forecastable" than others (Beckage *et al.* 2011; Chen *et al.* 2019;
27 Pennekamp *et al.* 2019). The initial success of these forms of models can be traced back to the larger
28 program of ontological reductionism, which became the default approach to modeling in the sciences after
29 its early success in physics, which, by the time ecology was becoming a quantitative science (sometime in

30 the 20th century, depending on who 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 a model
39 increases, so does the ability of the scientist to discern clear relationships between them given a fixed
40 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 observation 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 Québec 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 to 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 provide a framework for
61 ecosystem monitoring and forecasting (fig. 1, left), and each chapter address some aspect of this pipeline
62 to data from a monitoring network to forecasts to mitigation strategy (fig. 1, right).

63 [Figure 1 about here.]

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

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

75 **Objective**

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

[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 that an observation is of a
 92 given species is that species' relative abundance. We show that the realized false-negative rate can be quite
 93 high, simply as a byproduct of the distribution of relative abundance in communities. We use a log-normal
 94 distribution of relative abundance (Hubbell 2001) and simulating the process of observation on food-webs
 95 generated using the niche model (Williams & Martinez 2000) with connectance parameterized by the
 96 flexible-links model (MacDonald *et al.* 2020). An example of this relation for networks with varying
 97 species richness is shown in fig. 4.

[Figure 3 about here.]

99 We then go on to test some assumptions of this neutral model with empirical data. Primarily 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 chapter two into this analysis
 104 once available. Finally this chapter proposes a simulated annealing method to optimize the a set of n
 105 points in space to maximize the probability of detecting an interaction between two species a and b with
 106 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 (see 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 forecasted 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, we begin by
146 predicting a metaweb of interactions across Colorado as they exist *in the present*. We do this using a set of
147 candidate interaction prediction models: relative abundance only, phylogenetic embedding only (a la
148 Strydom *et al.* (2021a)), niche embedding only (encompassing both phenology and environmental niche,
149 similar to Gravel *et al.* (2019)), and all pairwise combinations of those constituent models. After validating
150 and selecting the best performing model, we then predict how these distributions of each of these species
151 will change under the CMIP6 consensus climate forecast (Karger *et al.* 2017), and then finally quantify the
152 reduction in spatial overlap between species for 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 (e.g. Peterman 2018). This chapter
173 proposes a general algorithm for optimizing spatial restoration effort to move a measurement of ecosystem
174 functioning toward a target state.

175 **Methods**

176 [Figure 7 about here.]

177 We propose various landscape modifications that alter the cover of a landscape, represented as a raster. We
178 then compute a new resistance surface based on the proposed landscape modification using Circuitscape
179 (McRae *et al.* 2008), and based on the values of resistance to dispersal between pairs of locations we
180 simulate spatially-explicit metapopulation dynamics (Hanski & Ovaskainen 2000; Ovaskainen *et al.* 2002)
181 to estimate a distribution of time-until-extinction for each landscape modification. The largest challenge
182 in implementing this algorithm is the space of potential modifications grows as $O((nm)!)$ for an n by m
183 raster. For most actual landscapes to which we wish to apply this method, the set of possible modifications
184 becomes uncomputably large, so we use simulated annealing to explore the search space of possible
185 modifications to estimate the modification that maximizes the time-until extinction of simulated
186 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 (locations where occurrence has been observed), and then proposing corridors
190 that only connect nodes that are already connected in the MST. The primary reason for doing this is the
191 cut down the size of the solution space to enable quicker convergence, although the final software that
192 implements this algorithm will enable alternative methods 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 included in the distribution of “good”
195 corridors after simulated annealing has converged. Further, the final component of this chapter is
196 measuring the effect of land-use change on the robustness of the optimized corridor by simulating various
197 neutral models of urban and agricultural sprawl, and determining if the proposed modifications still
198 maximize time-until-extinction when the landcover in the landscape is not static.

199 **Progress**

200 Currently I have an algorithm for proposing landscape modifications and a simple implementation of
201 simulated annealing. The only gap left is implementing Circuitscape estimation of resistance surfaces and
202 running benchmarking tests for the resulting chapter.

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 at the center of this set of tools, but due to the nature of the Julia
209 language, MCD.jl is inter-operable 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 In fig. 9 we see a sample output of simulated food-web dynamics for a metaweb of 100 species generated
 224 using the minimum-potential-niche model (Allesina *et al.* 2008) with connectance $C = 0.05$ and
 225 forbidden-link probability of 0.5. The dynamics change according to a Lotka-Volterra functional response,
 226 dispersal with distance inversely proportional to trophic-level, linear mortality, and logistic growth for any
 227 species at the 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 functional are
 233 occupancy dynamics for Levins metapopulations (Levins 1969), and spatially explicit Hanski-Ovaskainen
 234 metapopulations (Hanski & Ovaskainen 2000; Ovaskainen *et al.* 2002). This is most of what needs to exist
 235 for the preceding chapters. In-progress functionality includes selection (which affects growth-rate) on
 236 arbitrary 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, and to
249 find a way for humans to live sustainably on Earth, the only planet on which life has ever been known to
250 occur.

251 **References**

- 252 Allesina, S., Alonso, D. & Pascual, M. (2008). A General Model for Food Web Structure. *Science*, 320,
253 658–661.
- 254 Banville, F., Vissault, S. & Poisot, T. (2021). Mangal.jl and EcologicalNetworks.jl: Two complementary
255 packages for analyzing ecological networks in Julia. *Journal of Open Source Software*, 6, 2721.
- 256 Bascompte, J. & Jordano, P. (2007). Plant-Animal Mutualistic Networks: The Architecture of Biodiversity.
257 *Annual Review of Ecology, Evolution, and Systematics*, 38, 567–593.
- 258 Bauer, P., Thorpe, A. & Brunet, G. (2015). The quiet revolution of numerical weather prediction. *Nature*,
259 525, 47–56.
- 260 Beckage, B., Gross, L.J. & Kauffman, S. (2011). The limits to prediction in ecological systems. *Ecosphere*, 2,
261 art125.
- 262 CaraDonna, P.J., Petry, W.K., Brennan, R.M., Cunningham, J.L., Bronstein, J.L., Waser, N.M., *et al.* (2017).

- 263 Interaction rewiring and the rapid turnover of plantpollinator networks. *Ecology Letters*, 20, 385–394.
- 264 Chen, Y., Angulo, M.T. & Liu, Y.-Y. (2019). Revealing Complex Ecological Dynamics via Symbolic
265 Regression. *BioEssays*, 41, 1900069.
- 266 Dietze, M.C. (2017). Prediction in ecology: A first-principles framework. *Ecological Applications*, 27,
267 2048–2060.
- 268 Gravel, D., Baiser, B., Dunne, J.A., Kopelke, J.-P., Martinez, N.D., Nyman, T., *et al.* (2019). Bringing Elton
269 and Grinnell together: A quantitative framework to represent the biogeography of ecological
270 interaction networks. *Ecography*, 42, 401–415.
- 271 Hadfield, J.D., Krasnov, B.R., Poulin, R. & Nakagawa, S. (2014). A Tale of Two Phylogenies: Comparative
272 Analyses of Ecological Interactions. *The American Naturalist*, 183, 174–187.
- 273 Hanski, I. & Ovaskainen, O. (2000). The metapopulation capacity of a fragmented landscape. *Nature*, 404,
274 755–758.
- 275 Hill, C., DeLuca, C., Balaji, Suarez, M. & Da Silva, A. (2004). The architecture of the Earth System
276 Modeling Framework. *Computing in Science Engineering*, 6, 18–28.
- 277 Hubbell, S.P. (2001). *The unified neutral theory of biodiversity and biogeography*. Monographs in
278 population biology. Princeton University Press, Princeton.
- 279 Karger, D.N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., *et al.* (2017). Climatologies at
280 high resolution for the earth's land surface areas. *Scientific Data*, 4, 170122.
- 281 Levin, S.A. (1992). The Problem of Pattern and Scale in Ecology: The Robert H. MacArthur Award
282 Lecture. *Ecology*, 73, 1943–1967.
- 283 Levins, R. (1969). Some Demographic and Genetic Consequences of Environmental Heterogeneity for
284 Biological Control. *Bulletin of the Entomological Society of America*, 15, 237–240.
- 285 MacDonald, A.A.M., Banville, F. & Poisot, T. (2020). Revisiting the Links-Species Scaling Relationship in
286 Food Webs. *Patterns*, 1.
- 287 Maino, J.L., Schouten, R. & Umina, P. (2021). Predicting the global invasion of *Drosophila suzukii* to
288 improve Australian biosecurity preparedness. *Journal of Applied Ecology*, 58, 789–800.
- 289 Makiola, A., Compson, Z.G., Baird, D.J., Barnes, M.A., Boerlijst, S.P., Bouchez, A., *et al.* (2020). Key

- 290 Questions for Next-Generation Biomonitoring. *Frontiers in Environmental Science*, 7.
- 291 McRae, B.H., Dickson, B.G., Keitt, T.H. & Shah, V.B. (2008). Using Circuit Theory to Model Connectivity
292 in Ecology, Evolution, and Conservation. *Ecology*, 89, 2712–2724.
- 293 Ovaskainen, O., Sato, K., Bascompte, J. & Hanski, I. (2002). Metapopulation Models for Extinction
294 Threshold in Spatially Correlated Landscapes. *Journal of Theoretical Biology*, 215, 95–108.
- 295 Ovaskainen, O., Sato, K., Bascompte, J. & Hanski, I. (2002). Metapopulation Models for Extinction
296 Threshold in Spatially Correlated Landscapes. *Journal of Theoretical Biology*, 215, 95–108.
- 297 Pennekamp, F., Iles, A.C., Garland, J., Brennan, G., Brose, U., Gaedke, U., et al. (2019). The intrinsic
298 predictability of ecological time series and its potential to guide forecasting. *Ecological Monographs*, 89,
299 e01359.
- 300 Petchey, O.L., Pontarp, M., Massie, T.M., Kéfi, S., Ozgul, A., Weilenmann, M., et al. (2015). The ecological
301 forecast horizon, and examples of its uses and determinants. *Ecology Letters*, 18, 597–611.
- 302 Peterman, W.E. (2018). ResistanceGA: An R package for the optimization of resistance surfaces using
303 genetic algorithms. *Methods in Ecology and Evolution*, 9, 1638–1647.
- 304 Poisot, T., Bélisle, Z., Hoebelke, L., Stock, M. & Szefer, P. (2019). EcologicalNetworks.jl: Analysing
305 ecological networks of species interactions. *Ecography*, 42, 1850–1861.
- 306 Strydom, T., Bouskila, S., Banville, F., Barros, C., Caron, D., Farrell, M.J., et al. (2021a). Food web
307 reconstruction through phylogenetic transfer of low-rank network representation.
- 308 Strydom, T., Catchen, M.D., Banville, F., Caron, D., Dansereau, G., Desjardins-Proulx, P., et al. (2021b). *A
309 Roadmap Toward Predicting Species Interaction Networks (Across Space and Time)* (Preprint).
- 310 EcoEvoRxiv.
- 311 Thompson, R.M. & Townsend, C.R. (2000). Is resolution the solution?: The effect of taxonomic resolution
312 on the calculated properties of three stream food webs. *Freshwater Biology*, 44, 413–422.
- 313 Urban, M.C., Travis, J.M.J., Zurell, D., Thompson, P.L., Synes, N.W., Scarpa, A., et al. (2021). Coding for
314 Life: Designing a Platform for Projecting and Protecting Global Biodiversity. *BioScience*.
- 315 Vellend, M. (2010). Conceptual Synthesis in Community Ecology. *The Quarterly Review of Biology*, 85,
316 183–206.

³¹⁷ Williams, R.J. & Martinez, N.D. (2000). Simple rules yield complex food webs. *Nature*, 404, 180–183.

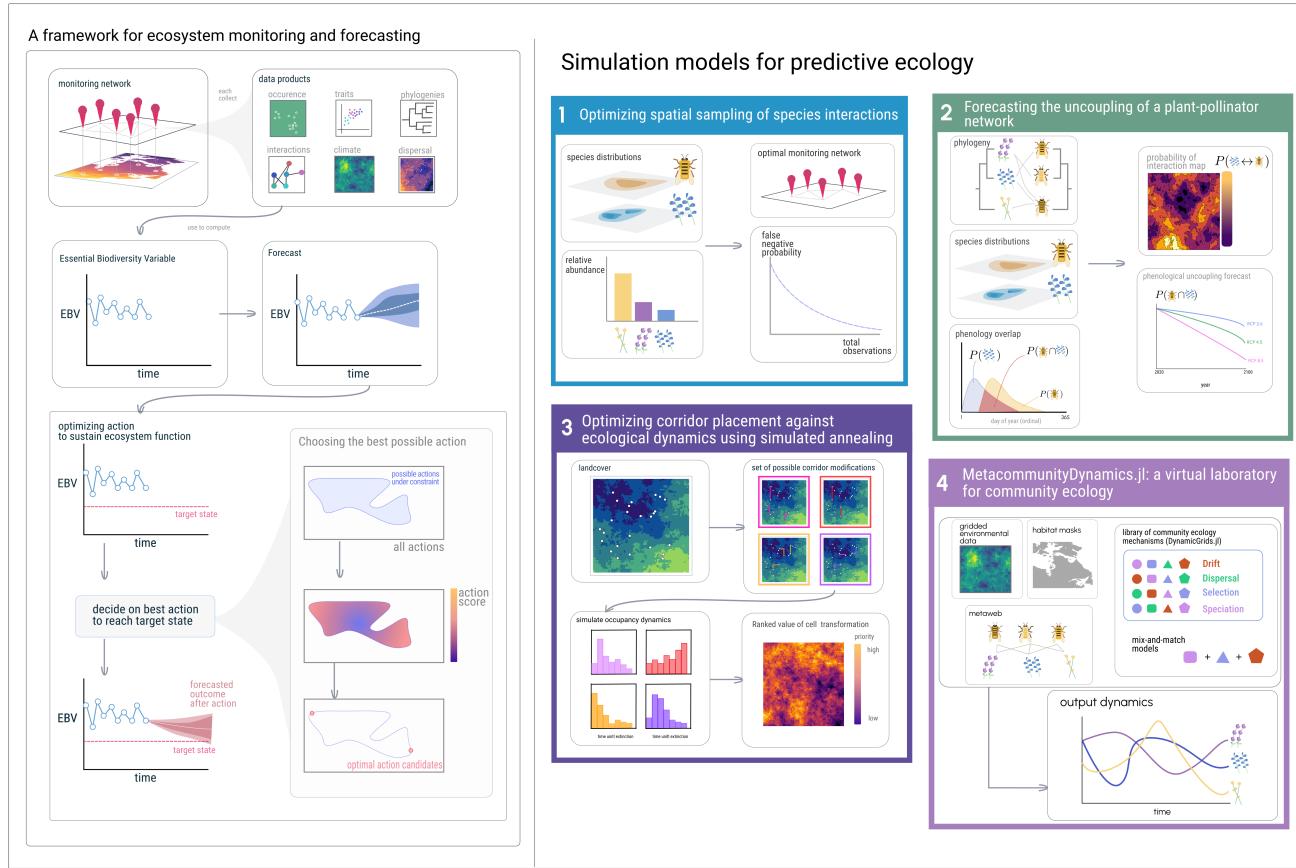


Figure 1: Left: a framework for ecosystem monitoring, forecasting, and mitigation. From the top, a set of biodiversity observatories which form a monitoring network. Each collect various biodiversity data products. From this raw data, we derive essential biodiversity variables (EBVs), and forecast how 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?

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

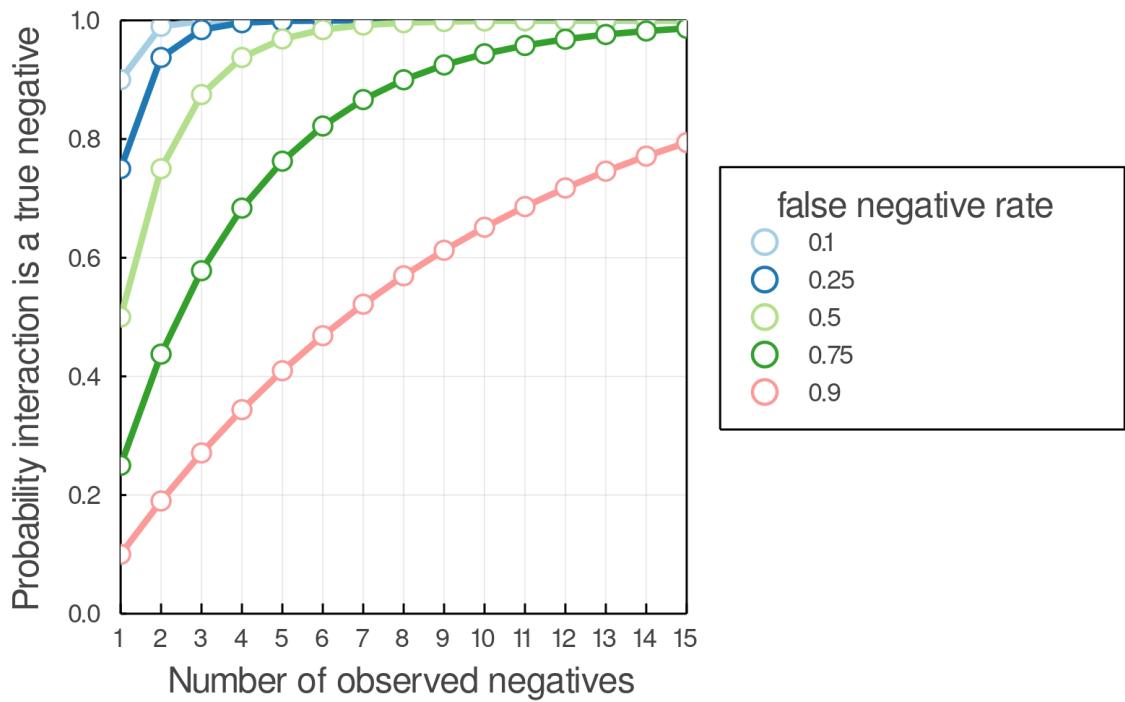


Figure 3: Relationship between total number of observed interactions negatives (x-axis) and the probability that an interaction is a true-negative as a function of different realized false-negative rates (colors), assuming each observed negative is independent—the same as the negative-bernoulli distribution.

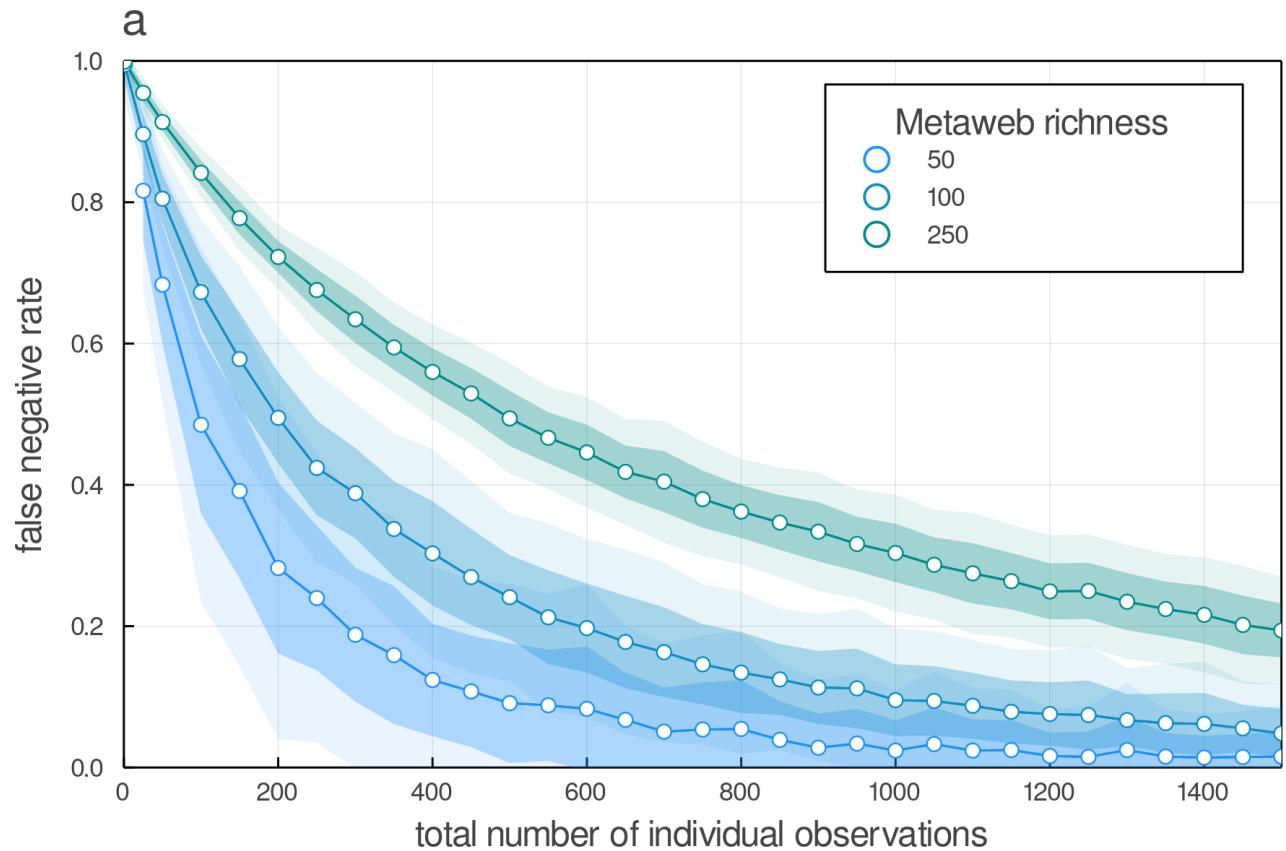


Figure 4: The realized false-negative-rate of interaction detection (y-axis) as a function of the total number of observations of all species in the species pool (different richnesses in different colors). Each point is the mean of 50 replicates, with one standard-deviation in the first shade, and two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

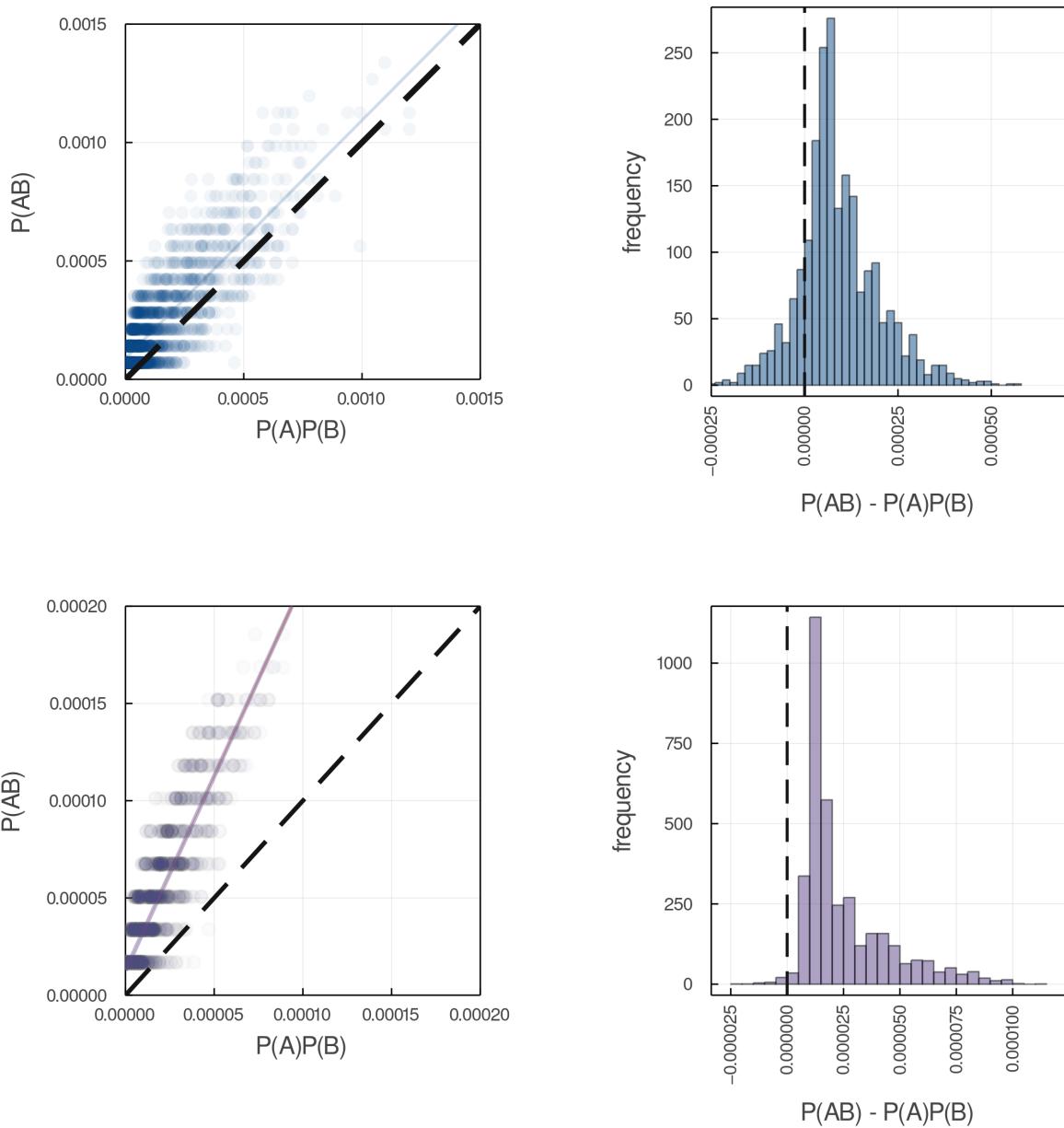


Figure 5: Demonstrates positive associations in co-occurrence. Left: the product of the marginal probability of observing two species (A and B) in a sample (x-axis) against the computed joint probability of observing these species together (y-axis). Dashed line indicates $y = x$, meaning no association 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.

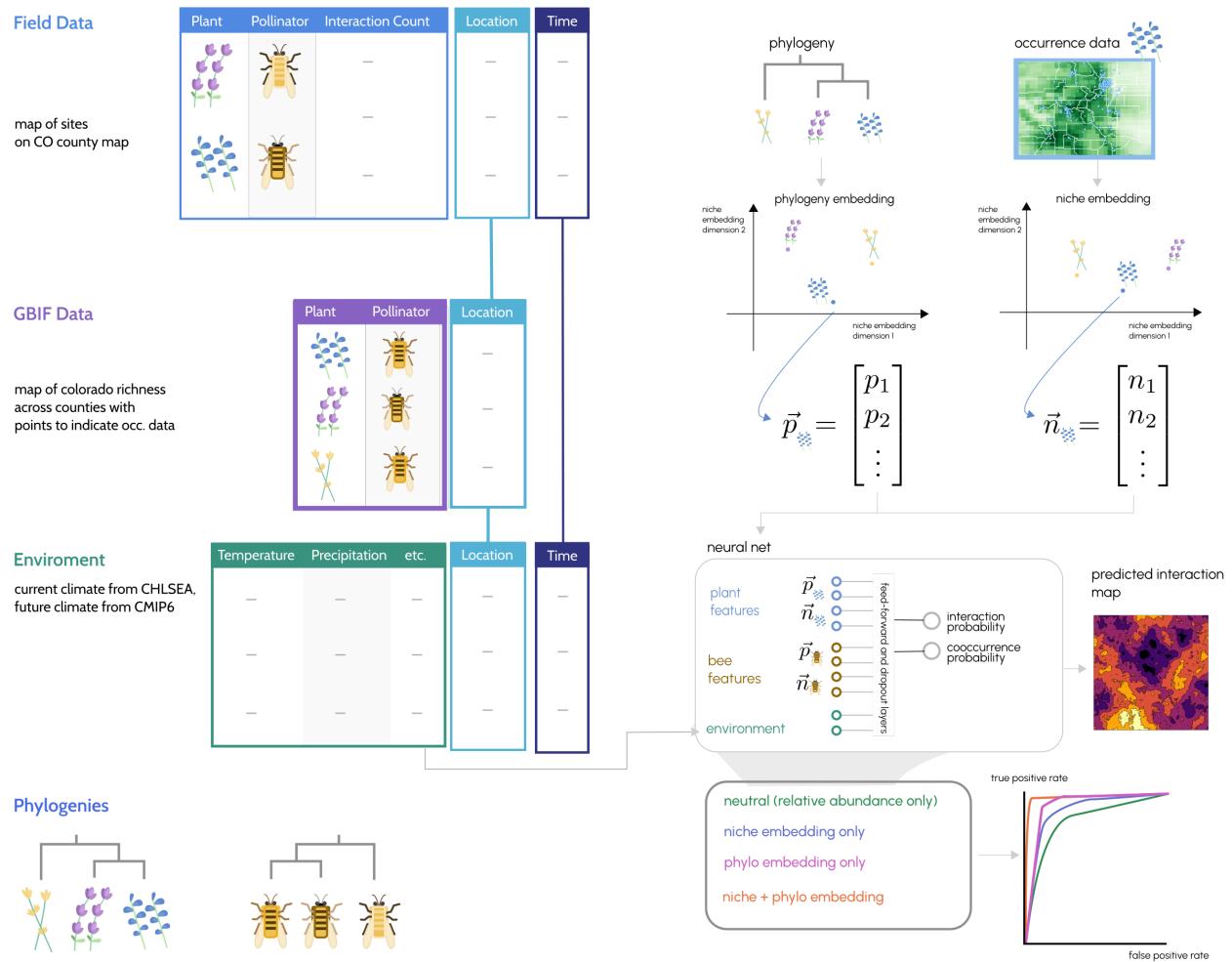


Figure 6: Chapter One conceptual figure. Left: the sources of data and how they can be combined. Right: The flow from data to interaction prediction using a few different interaction prediction models.

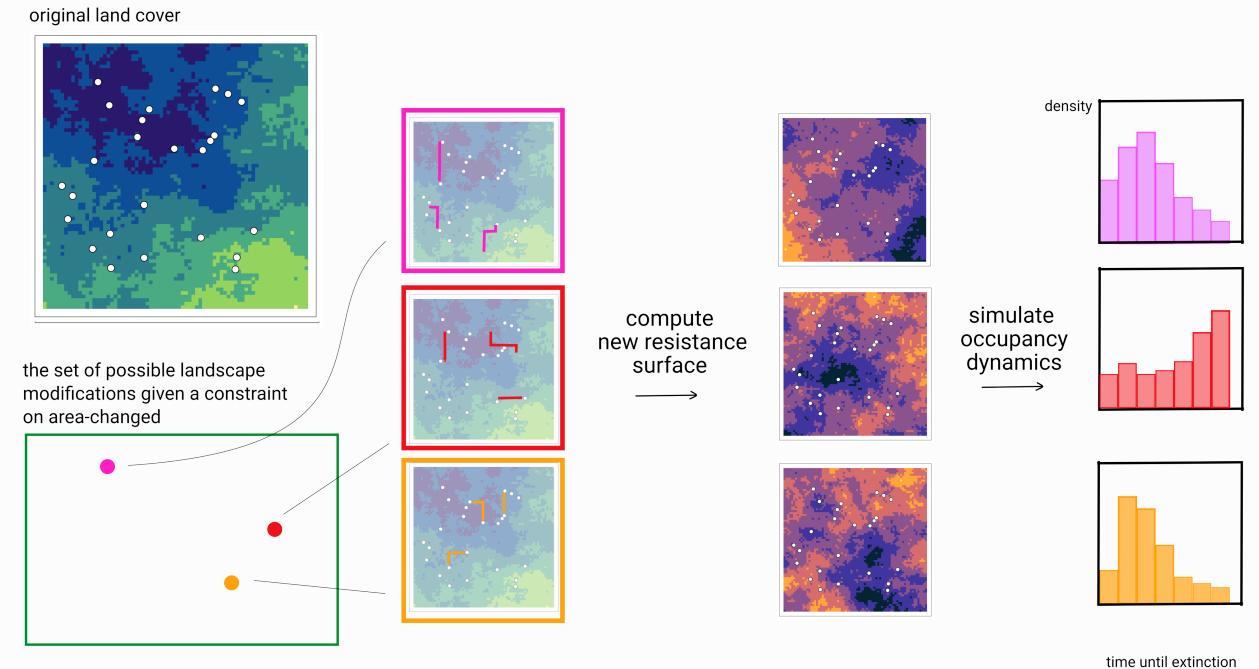


Figure 7: A conceptual example of how we go from a map of landcover with a set of points where occurrence of a species has been recorded (top left), to the set of all possible landscape modifications (green box, where each point in the green box is a unique landscape modification, with three examples 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).

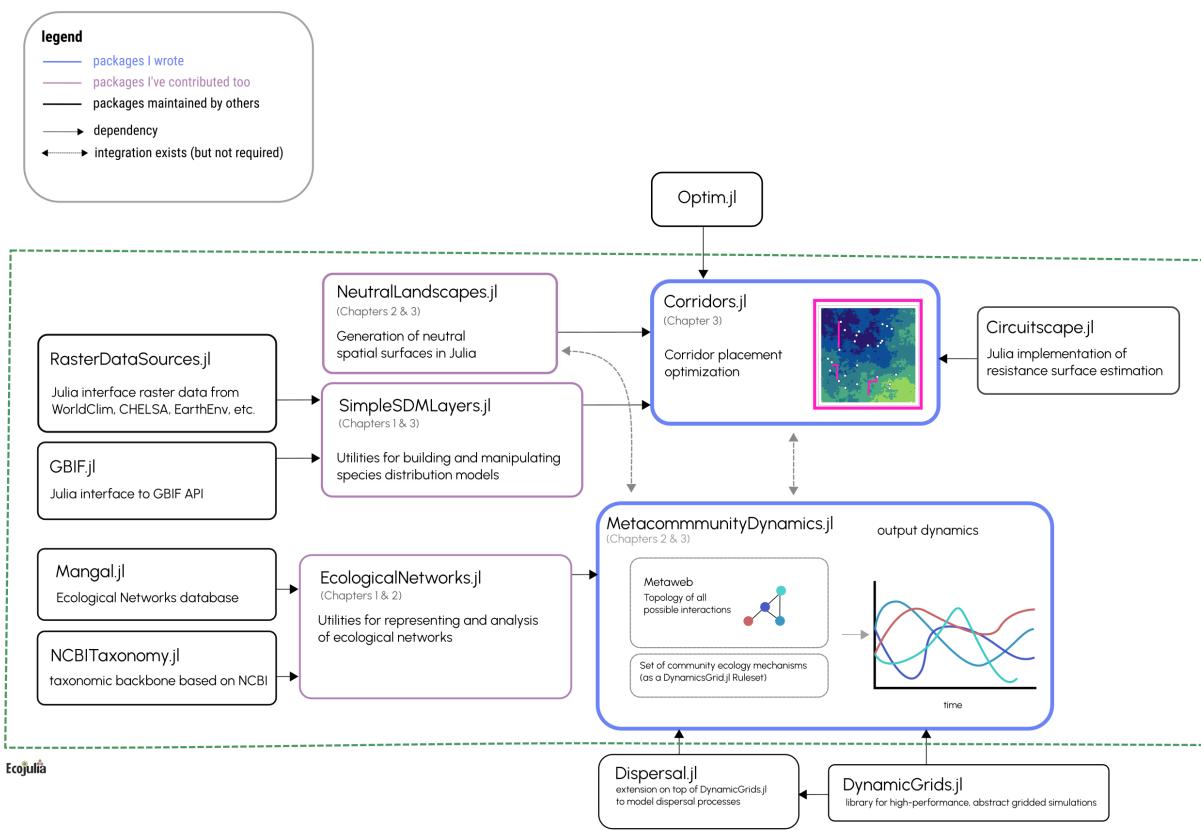


Figure 8: The structure of the software libraries used as part of MCD.jl

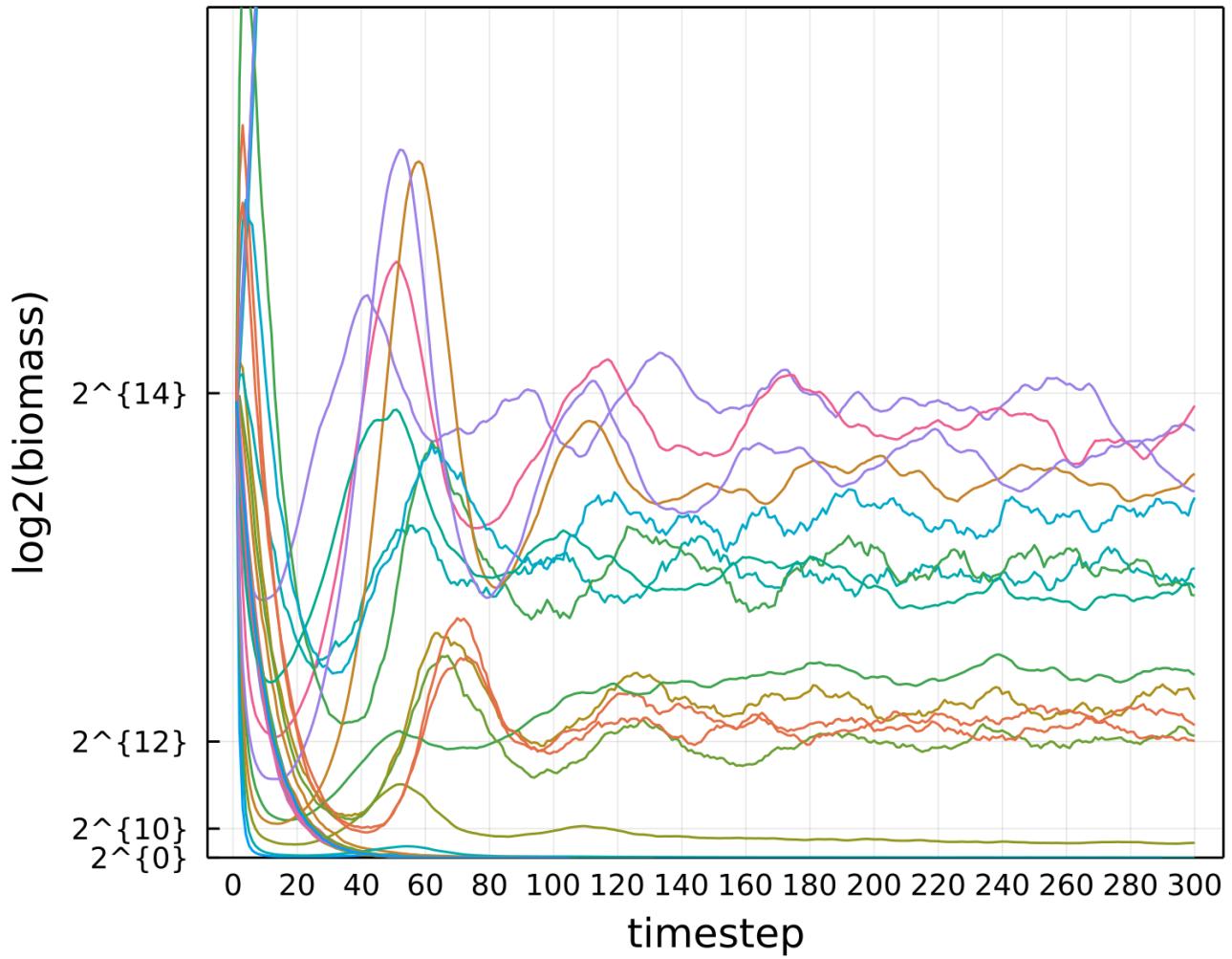


Figure 9: Sample output of simulated food web dynamics from MetacommunityDynamics.jl. Timestep (x-axis), and biomass of each species (y-axis).