Second Commitee Meeting Report — Septemer 22, 2021

Michael D. Catchen

Summary

This is the annual report for my second committee meeting on September 22, 2021. This document includes a brief introduction to the topic of my thesis, an updated outline of each chapter, what progress had been made on these chapters in the previous year, and a timeline for the next year of work. Additionally, as in the last month the research committee has discussed a revised time-frame and restructure for my PhD, which here I propose a timetable to finish by Fall 2023.

Introduction

Developing a predictive theory of ecology is an imperative, both for our understanding of ecosystem function but also because of the applied need to make robust, actionable forecasts of how ecosystem composition and function will change in the future (Dietze et al. 2018; Dietze 2017)—(fig. 1). Effective prediction has long evaded ecological systems as they are variable, high-dimensional, and the intrinsic dynamics that govern are system are unknown (Chen, Angulo, and Liu 2019). Further the spatiotemporal scale and resolution of a model effects the intrinsic predictability of ecological dynamics (Pennekamp et al. 2019). A primary theme of my proposed dissertation is understanding how simulation tools and methods can aid prediction and forecasting in ecology, much in the way simulated has aided other fields which aim to tackle similarly complex systems.

In particular, the first two chapters discuss use of simulation for sampling, understanding and predicting interactions between species (fig. 2). The third chapter then shifts to apply simulation methods in landscape ecology to optimize corridor placement with respect to a given ecosystem function. The fourth and final chapter is the software (*MetacommunityDynamics.jl*) which enables the rest of the dissertation.

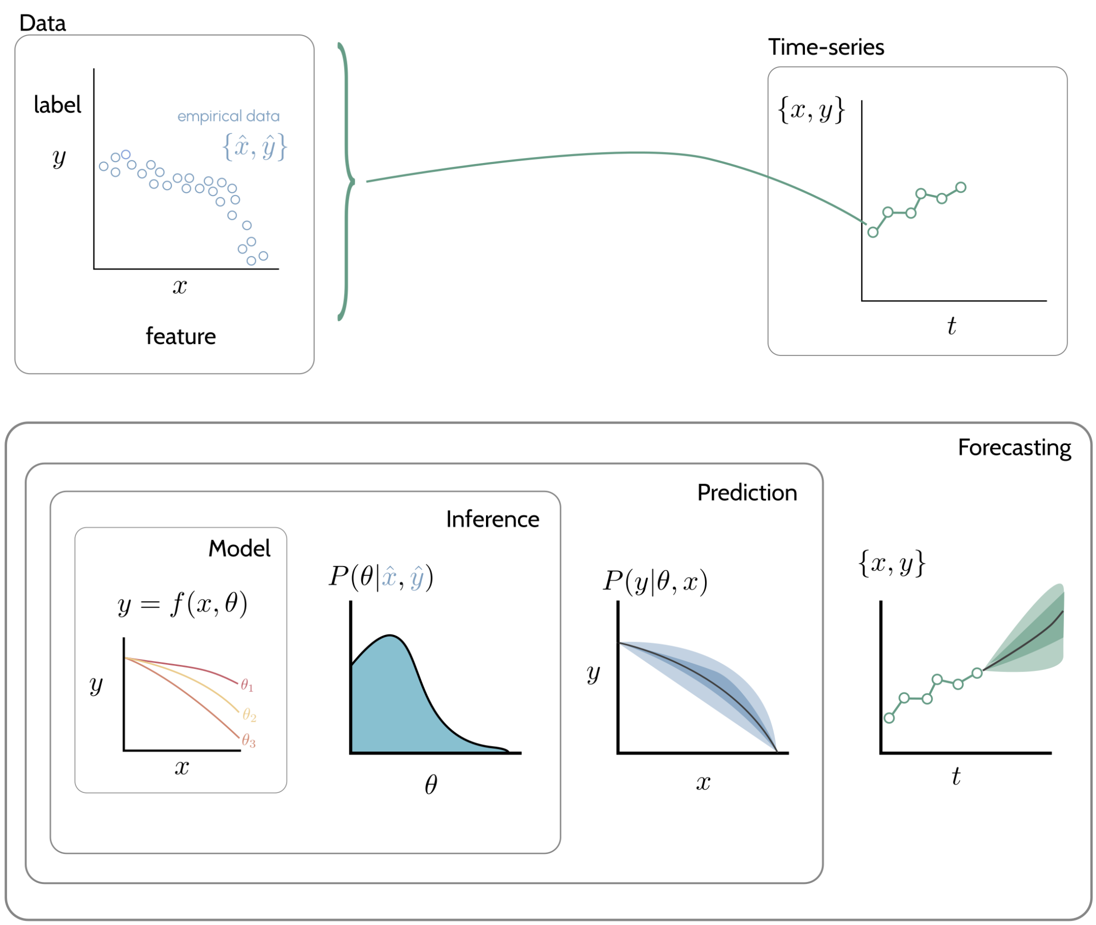

Figure 1: Predictive ecology and its relation to inference and forecasting. Adapted from Strydom, Catchen, et al. (2021)

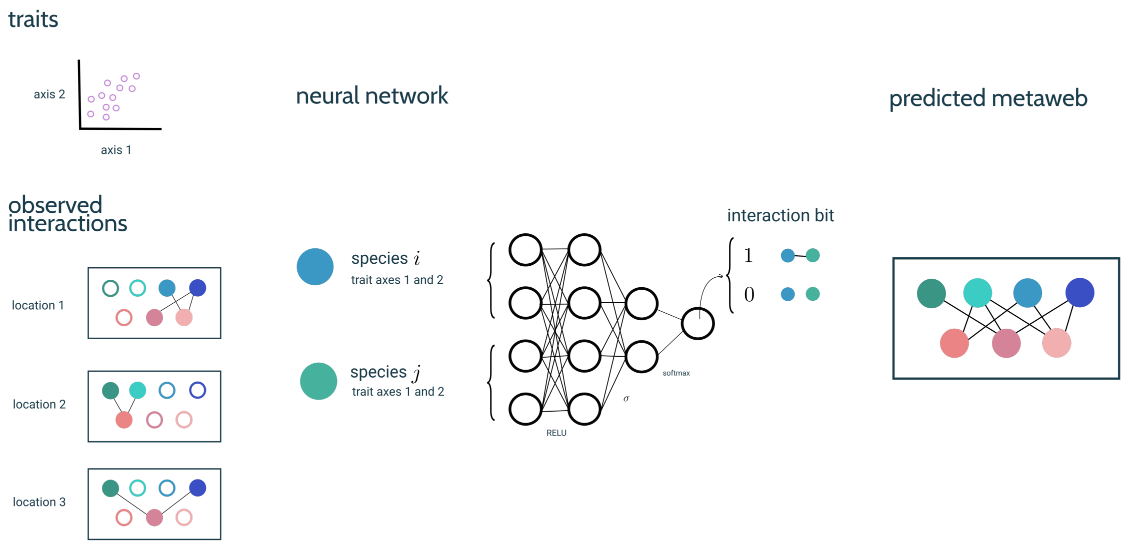

Figure 2: Prediction of interactions between species

Dissertation status

This section briefly describes the status of each chapter, how the structure has changed in the last year. During last years committee meeting, it was suggested that the first chapter be a full review of the spatial/temporal/taxonomic scale at which empirical data on ecological networks exists. In the year since, a couple of papers have effectively done this (Guimarães 2020 ; Resasco, Chacoff, and Vázquez 2021; Schwarz et al. 2020).

The second change is relating to the structure and timeline of the PhD. Initially my proposed dissertation was aimed at building the foundations for a research program which could be competitive for faculty positions. As a collective decision the committee that decided based on my adjusted career interests post-PhD (ideally a semi-permanent data-science/research position at a government sector), the goal is now narrow the scope of the PhD to finish before September 2023, in addition to highlighting the applied potential of my work (primarily in chapter three).

Here I briefly summarize the goal of each chapter and describe in-progress work.

Dissertation Introduction

This is a general introduction to the topics in the dissertation: roughly a 2,000 word history of community ecology, metacommunities, theory and simulation in ecology, and so on. At the moment I have about 1200 words on this, which will likely be adapted into part of the dissertation proposal for my qualifying exams.

Chapter one (*The missing link: differentiating true from false negatives when sampling species interactions*)

The first chapter is a vignette of how simulation can have pragmatic use in ecology, specifically for guiding sampling of interaction based on neutral probabilities of observing interactions due to relative abundance. This is now a paper that we are (hopefully) close to preprinting and submitting, and as a result I won't go too far into the details here.

It begins with a conceptual framework for understanding the difference in false-negatives in occurrence, co-occurrence, and interactions (fig. 3). We use a null model of the relative-abundance distribution (Hubbell 2001) to simulate realized false-negatives as a function of varying sampling effort.

This also goes on to includes testing some assumptions of the model with empirical data (fig. 4), which indicate our neutral model, if anything, underestimates the probability of false-negatives due to positive correlations in co-occurrence in two spatially replicated networks (Hadfield et al. 2014; Thompson and Townsend 2000).

This chapter concludes with a discussion how argument null models, like those presented, in order to plan sampling effort across space and derive estimates of the probability of incorrect interaction detection (which can be implemented into predictive models (Joseph 2020)).

Chapter two (*Generative learning for predictive ecology*)

This chapter further addresses the need to predict interactions between species. As species interaction data is limited (for the reasons explored in the previous chapter) this has limited the models we can use for interaction prediction, as many models we would like to use (e.g. deep neural-nets) require large amounts of data. The central idea of this chapter is that predictive models can be trained on primarily (or entirely) simulated data and still can be used to effectively predict interactions in empirical data (fig. 5).

Many generative models have been proposed to explain food-web structure which fit data reasonably well (Cohen 1985; Williams and Martinez 2000; Allesina, Alonso, and Pascual 2008). The idea for this chapter is to train a

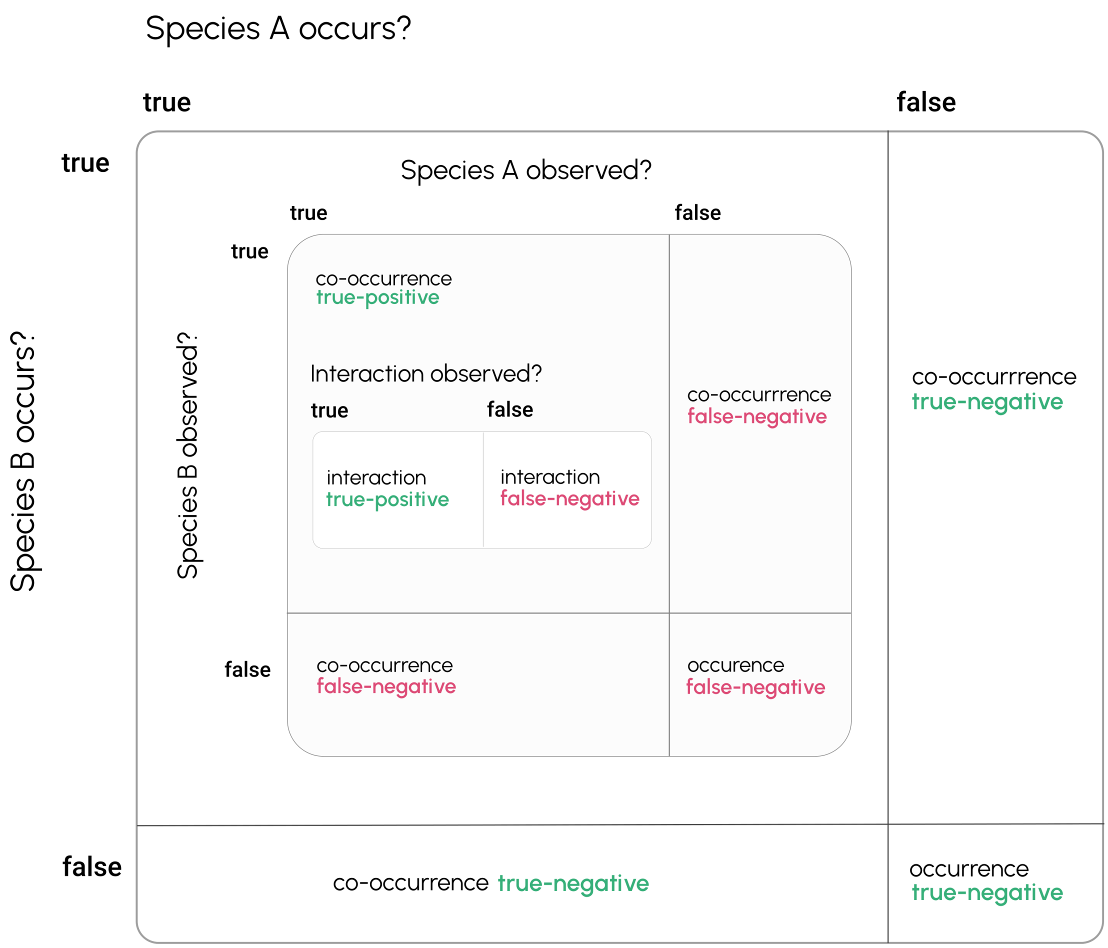

Figure 3: Taxonomy of false negatives

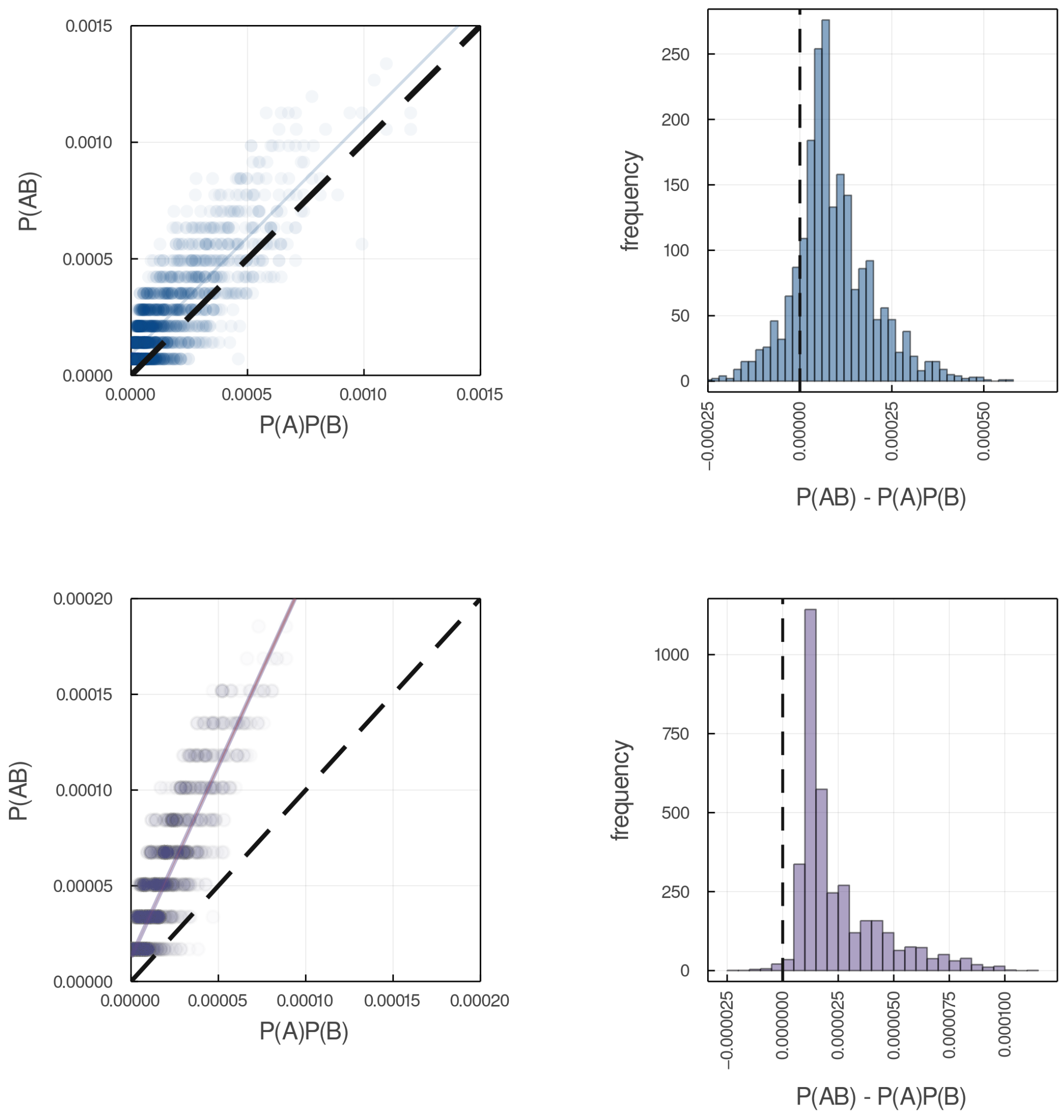

Figure 4: Primary novel result from this chapter

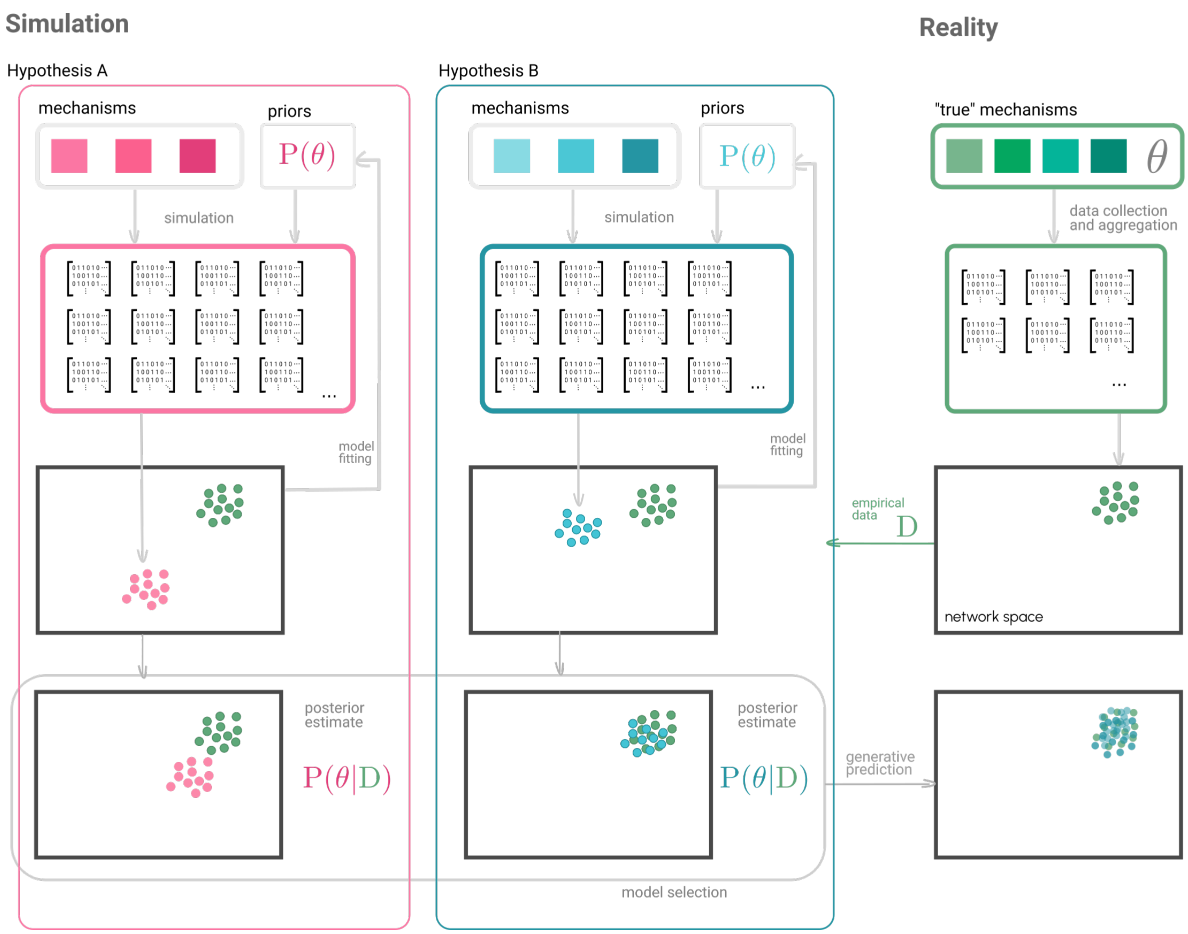

Figure 5: Conceptual overview of generative modeling

neural network to predict interactions (a la (Strydom, Catchen, et al. 2021)) on simulated data from each of these generative food web models. As this predicted model uses inputs for each species (fig. 2), to train this model on generated food-webs, we need summary statistics to go from the network to species-level features. In some code thus far I've considered: trophic level, degree, omnivory index, generality index, centrality, position in left subspace of SVD (Strydom, Bouskila, et al. 2021), and in the future I plan to testing various combinations.

The future work here is drafting some results figures, with a complete first draft done by March 2022.

Chapter three (*Optimizing corridor placement to minimize extinction probability*)

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

This chapter proposes a general algorithm for optimizing corridor placement based on a measurement of ecosystem functioning derived from simulations run on a proposed landscape modification. We propose various landscape modifications which alter the cover of a landscape, represented as a raster (fig. 6, left). We then compute a new resistance surface based on the proposed landscape modification, and based on the values of resistance to dispersal between each location we simulate spatially-explicit metapopulation dynamics model (Ovaskainen et al. 2002; Hanski and Ovaskainen 2000) to estimate a distribution of time until extinction for each landscape modification (fig. 6, right).

We denote the space of landscape modifications M_B, given you have a budget of B cells in the raster which you can change (to the minimum value of resistance across all cells). This space of possible modifications is big— if we fix B as a proportion of the total size of the lattice, it's NP-complete with respect to lattice size. We propose a simulated-annealing algorithm to estimate the optimal landscape modification. Simulated annealing works by defining a Markov-chain $\vec{\pi} = [\pi_1, \pi_2, \dots, \pi_N]$ which has an associated temperature parameter α, where each item in the chain is a proposed landscape modification $\pi_i \in M_B$. For each step

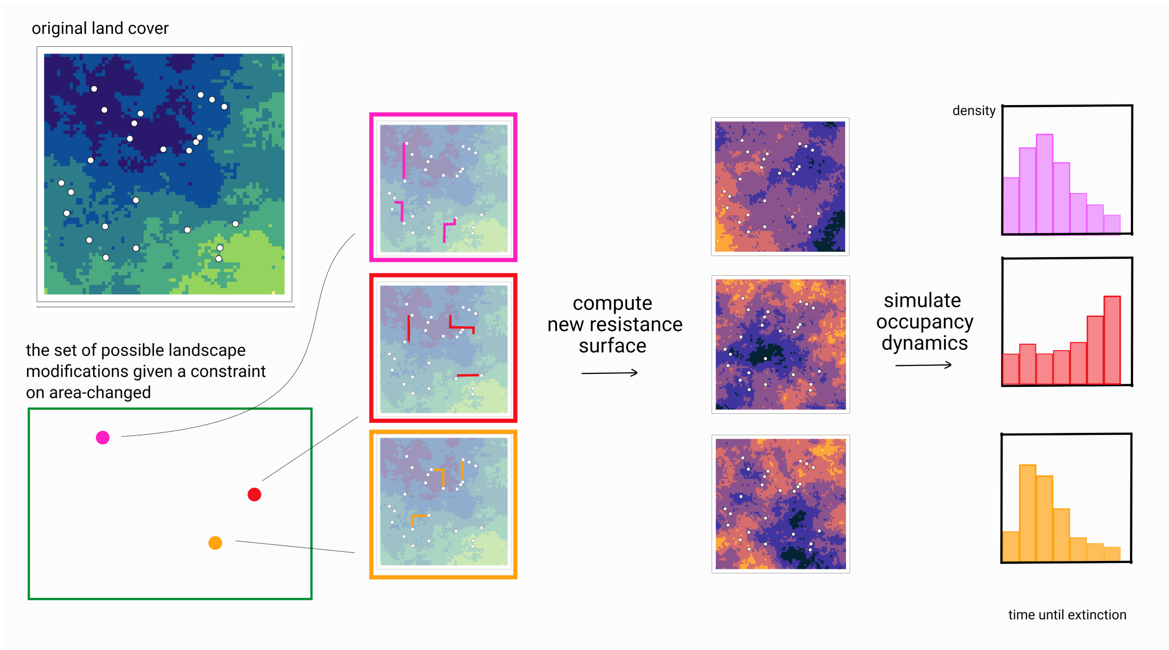

Figure 6: Workflow for measuring the distribution of extinction times for a given landscape modification

$$\pi_{i+1} = \begin{cases} q(\pi_i, x) \\ \pi_i \end{cases}$$

where $q(x, y)$ is a function that describes the probability of transitioning from state x.

The practical difficulties in implement this here are two-fold. 1) Defining a transition probability function q, and 2) defining a proposal algorithm. First, to describe a transition probability function $q(x, y)$, which gives the probability that a chain π will move from modification x to modification y at a given step, we consider a logistic function

$$q(x, y) = \frac{1}{1 + e^{-\alpha f(x,y)}}$$

where $f(x, y)$ is then a function to measure the distance between two candidate modifications x and y. A simple way of defining f would be the difference of the mean extinction time for each modification (fig. **??** right), i.e. $f(x, y) = MTE(y) - MTE(x)$.

In the next year I plan to implement this algorithm and start testing summary stats and chain temperatures. The goal for the first draft of this chapter is November 2022.

Chapter four (*MetacommunityDynamics.jl: a virtual laboratory for community ecology*)

This chapter is the software (*MetacommunityDynamics.jl*) which enables simulation for the previous four chapters. In the end this will hopefully be published as a series of case-studies which use the software, although this may not be finished by the end of my PhD.

The case studies are: 1) occupancy dynamics, 2) food-web dynamics, 3) evolution of plant-pollinator network. The functionality required by the occupancy dynamics case-study will be required for the previous chapter (and already is mostly complete), and the food-web case-study is already nearly functional (fig. 7)

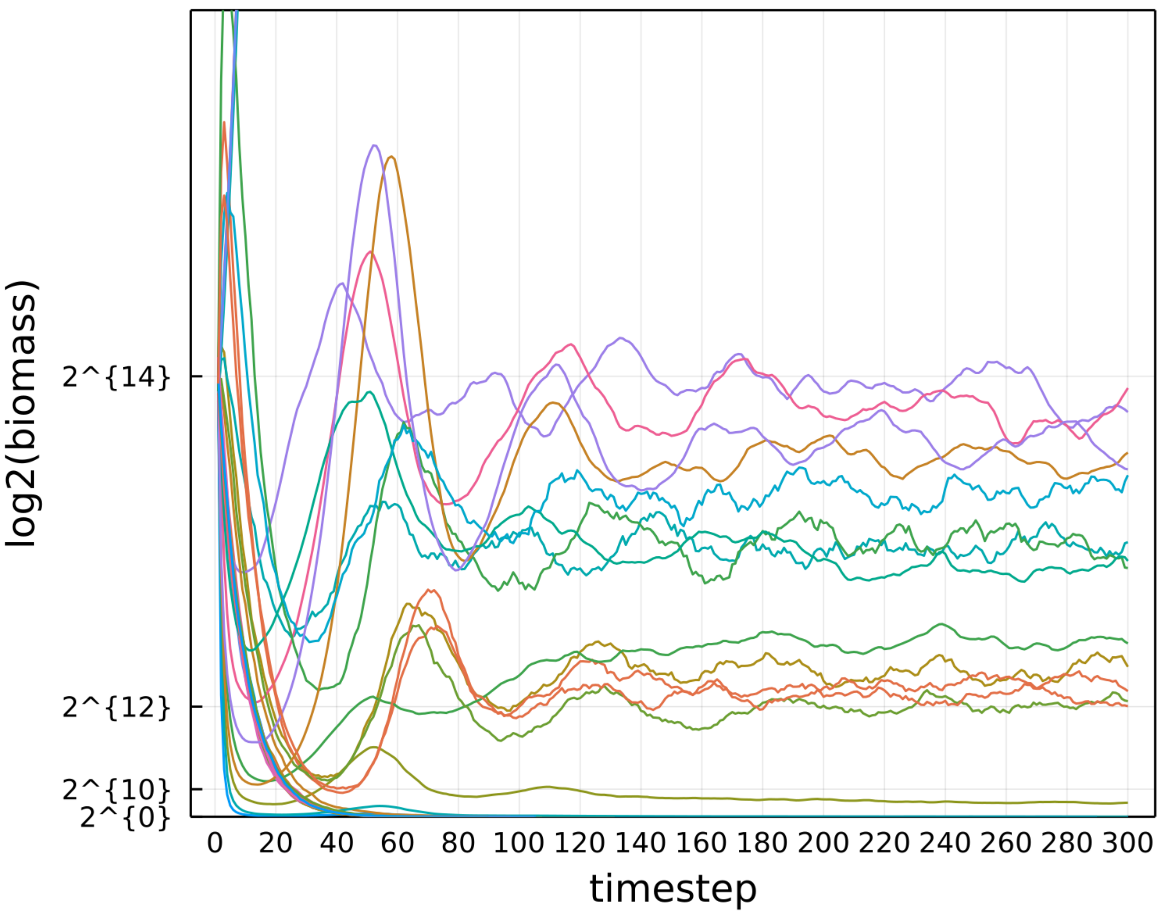

Figure 7: Sample output for the food-web case study

Dissertation Conclusion

The conclusion of the dissertation will be composed of ~3000 words, composed of a summary of the dissertation, how it fits into the larger context of ecological research, and future directions for this work.

Time-table

> You wait for darkness,
> then you wait for day.
>
> You wait for August,
> then you wait for May.
>
> *Built to Spill*

After discussions at the end of summer 2021, my committee agreed that it makes sense for my career interests to aim to complete the remainder of the PhD in the next two years.

Month	Exams	Drafts	Submissions	Courses
October 2021				
November 2021			CH1 Submitted	
December 2021	Quals			
January 2022				
February 2022				
March 2022		CH2 Draft 1		
April 2022				
May 2022				
June 2022		CH2 Draft 2		Bios2 summer school (3cr)
July 2022				
August 2022			CH2 Submitted	
September 2022				Fall 2022 course (3cr)
October 2022				
November 2022		CH3 Draft 1		
December 2022				
January 2023				
February 2023		CH3 Draft 2		
March 2023				
April 2023			CH3 Submitted	
May 2023	Exit Seminar			
June 2023				
July 2023				
August 2023	Defense			

References

Allesina, S., D. Alonso, and M. Pascual. 2008. "A General Model for Food Web Structure." *Science* 320 (5876): 658–61. https://doi.org/10.1126/science.1156269.

Chen, Yize, Marco Tulio Angulo, and Yang-Yu Liu. 2019. "Revealing Complex Ecological Dynamics via Symbolic Regression." *BioEssays* 41 (12): 1900069. https://doi.org/10.1002/bies.201900069.

Cohen, Newman, J. E. 1985. "A Stochastic Theory of Community Food Webs i. Models and Aggregated Data." *Proceedings of the Royal Society of London. Series B. Biological Sciences* 224 (1237): 421–48. https://doi.org/10.1098/rspb.1985.0042.

Dietze, Michael C. 2017. "Prediction in Ecology: A First-Principles Framework." *Ecological Applications* 27 (7): 2048–60. https://doi.org/10.1002/eap.1589.

Dietze, Michael C., Andrew Fox, Lindsay M. Beck-Johnson, Julio L. Betancourt, Mevin B. Hooten, Catherine S. Jarnevich, Timothy H. Keitt, et al. 2018. "Iterative Near-Term Ecological Forecasting: Needs, Opportunities, and Challenges." *Proceedings of the National Academy of Sciences* 115 (7): 1424–32. https://doi.org/10.1073/pnas.1710231115.

Guimarães, Paulo R. 2020. "The Structure of Ecological Networks Across Levels of Organization." *Annual Review of Ecology, Evolution, and Systematics* 51 (1): 433–60. https://doi.org/10.1146/annurev-ecolsys-012220-120819.

Hadfield, Jarrod D., Boris R. Krasnov, Robert Poulin, and Shinichi Nakagawa. 2014. "A Tale of Two Phylogenies: Comparative Analyses of Ecological Interactions." *The American Naturalist* 183 (2): 174–87. https://doi.org/10.1086/674445.

Hanski, Ilkka, and Otso Ovaskainen. 2000. "The Metapopulation Capacity of a Fragmented Landscape." *Nature* 404 (6779): 755–58. https://doi.org/10.1038/35008063.

Hubbell, Stephen P. 2001. *The Unified Neutral Theory of Biodiversity and Biogeography*. Monographs in Population Biology 32. Princeton: Princeton University Press.

Joseph, Maxwell B. 2020. "Neural Hierarchical Models of Ecological Populations." *Ecology Letters* 23 (4): 734–47. https://doi.org/10.1111/ele.13462.

Ovaskainen, Otso, Kazunori Sato, Jordi Bascompte, and Ilkka Hanski. 2002. "Metapopulation Models for Extinction Threshold in Spatially Correlated Landscapes." *Journal of Theoretical Biology* 215 (1): 95–108. https://doi.org/10.1006/jtbi.2001.2502.

Pennekamp, Frank, Alison C. Iles, Joshua Garland, Georgina Brennan, Ulrich Brose, Ursula Gaedke, Ute Jacob, et al. 2019. "The Intrinsic Predictability of Ecological Time Series and Its Potential to Guide Forecasting."

Ecological Monographs 89 (2): e01359. https://doi.org/10.1002/ecm.1359.

Peterman, William E. 2018. "ResistanceGA: An R Package for the Optimization of Resistance Surfaces Using Genetic Algorithms." *Methods in Ecology and Evolution* 9 (6): 1638–47. https://doi.org/10.1111/2041-210X.12984.

Resasco, Julian, Natacha P. Chacoff, and Diego P. Vázquez. 2021. "Plantpollinator Interactions Between Generalists Persist over Time and Space." *Ecology* 102 (6): e03359. https://doi.org/10.1002/ecy.3359.

Schwarz, Benjamin, Diego P. Vázquez, Paul J. CaraDonna, Tiffany M. Knight, Gita Benadi, Carsten F. Dormann, Benoit Gauzens, et al. 2020. "Temporal Scale-Dependence of Plantpollinator Networks." *Oikos* 129 (9): 1289–1302. https://doi.org/10.1111/oik.07303.

Strydom, Tanya, Salomé Bouskila, Francis Banville, Cerres Barros, Dominique Caron, Maxwell J Farrell, Marie-Josée Fortin, et al. 2021. "Food Web Reconstruction Through Phylogenetic Transfer of Low-Rank Network Representation," August. https://doi.org/10.32942/osf.io/y7sdz.

Strydom, Tanya, Michael David Catchen, Francis Banville, Dominique Caron, Gabriel Dansereau, Philippe Desjardins-Proulx, Norma Forero, et al. 2021. "A Roadmap Toward Predicting Species Interaction Networks (Across Space and Time)." Preprint. EcoEvoRxiv. https://doi.org/10.32942/osf.io/eu7k3.

Thompson, Ross M., and Colin R. Townsend. 2000. "Is Resolution the Solution?: The Effect of Taxonomic Resolution on the Calculated Properties of Three Stream Food Webs." *Freshwater Biology* 44 (3): 413–22. https://doi.org/10.1046/j.1365-2427.2000.00579.x.

Williams, Richard J., and Neo D. Martinez. 2000. "Simple Rules Yield Complex Food Webs." *Nature* 404 (6774): 180–83. https://doi.org/10.1038/35004572.