

Thesis proposal

Michael D. Catchen^{1,2}

The proposal for my thesis, *Simulation models for predictive ecology*

1

Introduction

P1

Within the last several hundred years, human activity has induced rapid changes in Earth's atmosphere, oceans, and surface. Greenhouse gas emissions have caused an increase in the temperature of both Earth's terrain and oceans, and both agricultural and urban development has rapidly reshaped the Earth's land cover. The bulk of this change has occurred within the last several hundred years, a geological instant, inducing a sudden shift in conditions to Earth's climate and biosphere. As a result, predicting how ecosystems will change in the future, *ecological forecasting*, and then using these forecasts to make decisions to mitigate the negative consequences of this change on ecosystems, their functioning, and the services they provide to humans has emerged as an imperative for ecology and environmental science (Dietze 2017). However, robust prediction of ecological processes is, to say the least, quite difficult (Beckage *et al.* 2011; Petchey *et al.* 2015). This difficulty is compounded by a few factors, the first being that sampling ecosystems is not easy. Ecological data is often biased, noisy, and sparse in both space and time. The current paucity of ecological data has resulted in much interest in developing global systems for *ecosystem monitoring* (Makiola *et al.* 2020), which would systematize the collection of biodiversity data in manner that makes detecting and predicting change more possible than at the moment (Urban *et al.* 2021).

P2

The second major challenge in ecological forecasting is that the underlying dynamics of most ecological processes are unknown and instead must be inferred from this (sparse) data. Much of the history of quantitatively modeling ecosystems have been done in the language of dynamical systems, describing how the value of an observable state of the system, represented by a vector of numbers $[x_1, x_2, \dots, x_n]^T = \vec{x}$ changes over time, yielding models in the form of differential equations in continuous-time settings— $\frac{dx}{dt} = f(x)$ —or difference equations in discrete-time settings— $x_t = f(x_{t-1})$ —where $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is an arbitrary function describing how the system changes on a moment-to-moment basis (e.g. in the context of communities, f could be Lotka-Volterra, Holling-Type-III or DeAngelis-Beddington functional response). The initial success of these forms of models can be traced back to the larger program of ontological reductionism, which became the default approach to modeling in the sciences after its early success in physics, which, by the time ecology was becoming a quantitative science (sometime in the 20th century, depending on who you ask), became the foundation for early quantitative models in ecology.

P3

However, we run into many problems when aiming to apply this type of model to empirical data in ecology. Ecosystems are perhaps the quintessential example of system that cannot be understood by iterative reduction of its components into constituent parts—ecological phenomena are emergent are the product of different mechanisms operating a different spatial, temporal, and organizational scales (Levin 1992). Further, the form of this functional response in real systems is effectively unknown, and

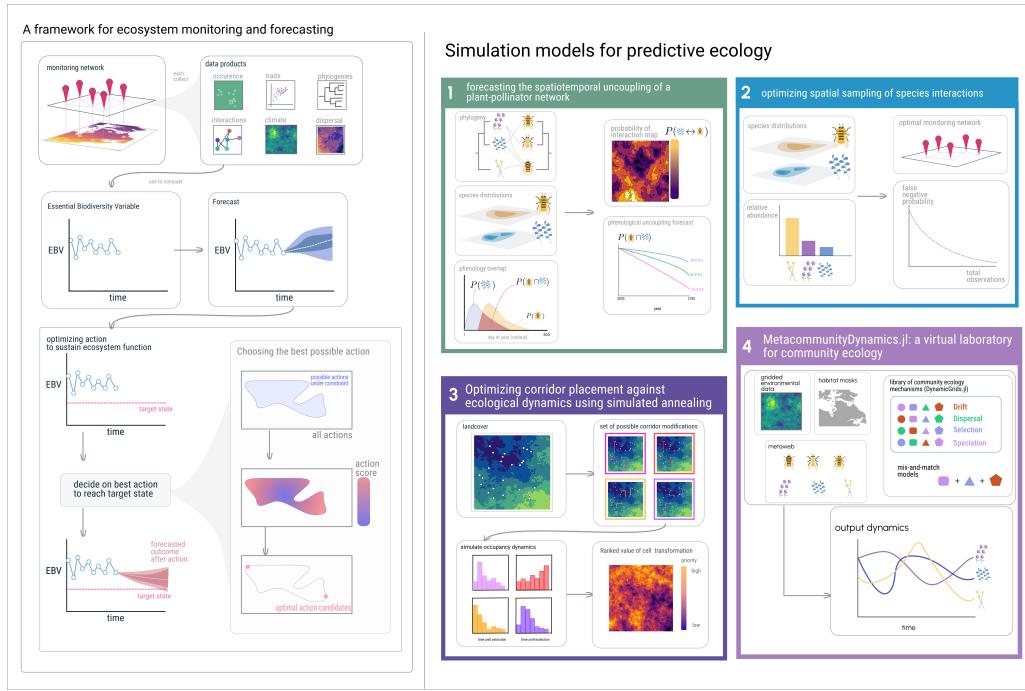


Figure 1 thesis concept

some forms are inherently more “forecastable” than others (Beckage *et al.* 2011; Chen *et al.* 2019; Pennekamp *et al.* 2019). Further this analytical approach to modeling explicitly ignores known realities: ecological dynamics not deterministic, many analytic models in ecology assume long-run equilibrium. Finally, perhaps the biggest challenge in using these models to describe ecological processes is ecosystems vary across more variables than the tools of analytic models are suited for. As the number of variables in an analytic model increases, so does the ability of the scientist to discern clear relationships between them given a fixed amount of data, the so-called “curse” of dimensionality.

P4

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

P6

But forecasting is only half the story. Marx’s most well known quote that “philosophers have hitherto only interpreted the world in various ways; the point is to change it.” Indeed, once we have a forecast about how an ecosystem will change in the future, what if this forecast predicts a critical ecosystem service will deteriorate? We are still left with the question, what do we in the time being to mitigate the negative consequences a forecast predicts? In this framing, mitigating the consequences of anthropogenic change on ecosystems becomes an optimization problem: given a forecast of the probability. We have some goal state for the future, and some estimate of what the state of the world will be given a set of actions. Frame optimization problem mathematically an introduce concept of solution-space and constraint.

P7

This dissertation aims to formalize a framework for ecosystem monitoring and forecasting (fig. 1, left),

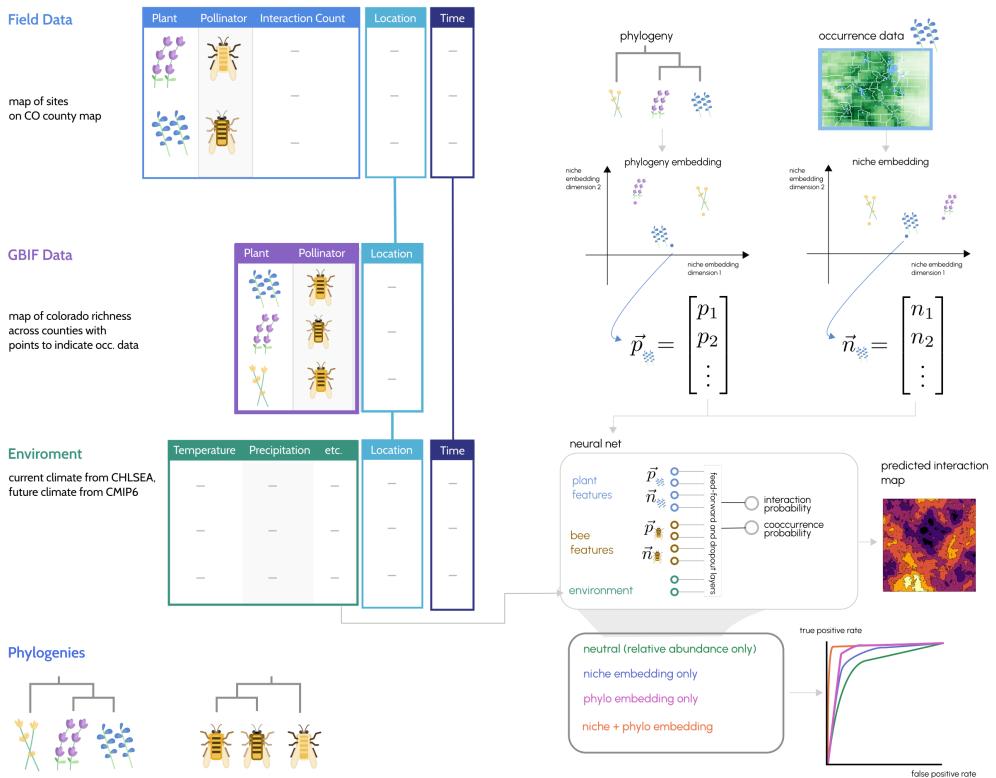


Figure 2 chapter one concept fig

and each chapter address some aspect of this pipeline to data from a monitoring network to forecasts to mitigation strategy (fig. 1, right).

2

Chapter One: Forecasting the spatial uncoupling of a plant-pollinator network

Plants and pollinators form interaction networks, called the “architecture of biodiversity” (**Jordano2007?**).

The stability, function, and persistence of ecosystems relies on the structure of these interactions. Anthropogenic change threatens to unravel these networks. Two aspects to this change: spatial and temporal. Spatially, range shifts along elevational gradient, and temporally, phenological shifts.

The issue is that we don't really know what interactions are like now. So not only do we need to predict with data that is spatially and temporally sparse and likely to contain many interaction “false-negatives” (Strydom *et al.* 2021).

This chapter uses several years of data on bee-flower phenology and interactions, combined with spatial records of species occurrence via GBIF, to forecast how much overlap there will be between plants/pollinators in space/time.

In stages, (1) take data from multiple sites to predict a spatial metaweb of *Bombus*-flower interactions across Colorado. (2) Predict how these spatial distributions will change under CMIP6. and (3) quantify the lack of overlap between species for which there is a predicted

2.1. Data The data for this chapter is derived from multiple sources and can be split into three categories. (1) Field data from three different locations across Colorado. All field sites have multiple plots across an elevational gradient.

System description: lots of data on *Bombus* (bumblebees) and wildflowers. Three different sites, (7/7/3) years each, each covering an elevational gradient.

Species A occurs?

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

Figure 3 taxonomy of false negatives

2.2. Methods

Split the process into parts.

- 1) Building an interaction prediction model. 2) Make it spatial based on distributions. 3) Forecast distributions based on CMIP6.

2.3. Preliminary Results

- 1) we got a tree

Transition to next chapter by discussing uncertainty in interaction prediction across space.

3

Chapter Two: Optimizing spatial sampling of species interactions

There are false-negatives in interation data. Co-occurrence is not the same thing as interaction ([cite?](#)), but often is used as a proxy.

This chapter unravels the relationship between a given species relative abundance and the sampling effort needed to adequately understand this species distribution and interactions.

There is more than one way to observe a false-negative.

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.

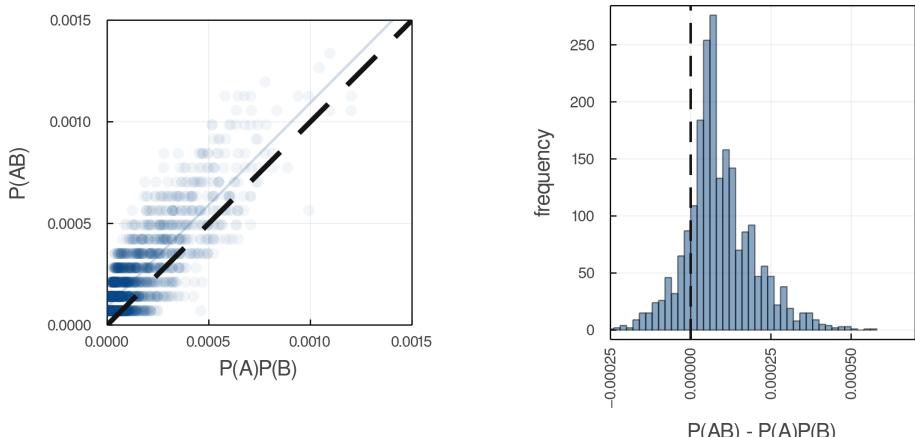
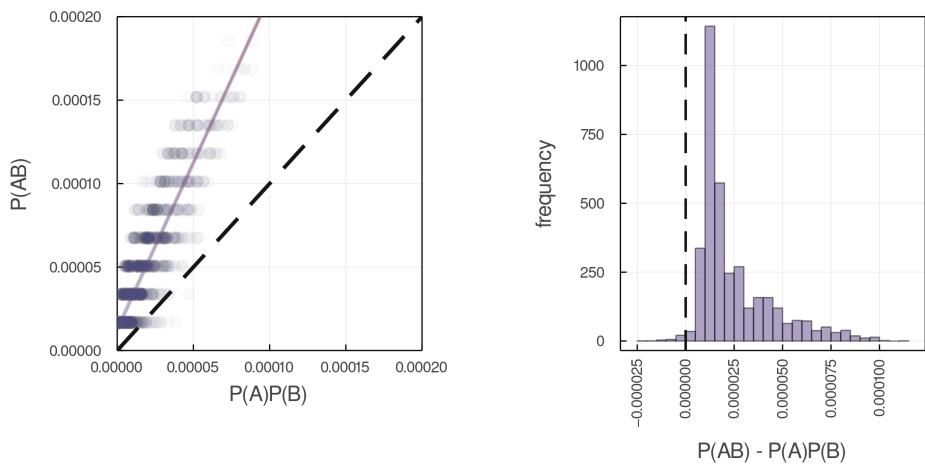


Figure 4 f



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 (Thompson & Townsend 2000; Hadfield *et al.* 2014)—further I'm planning to add the field data from chapter one into this analysis once complete.

new addition: - simulate species distribution and efficacy of detection given a set of observation points where the dist from observation site decays. optimize set of repeated sampling locations L for a *known* distribution D. address SDM not being the territory

3.1. Results

- nonrandom association figure sampling effort under neutral model

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

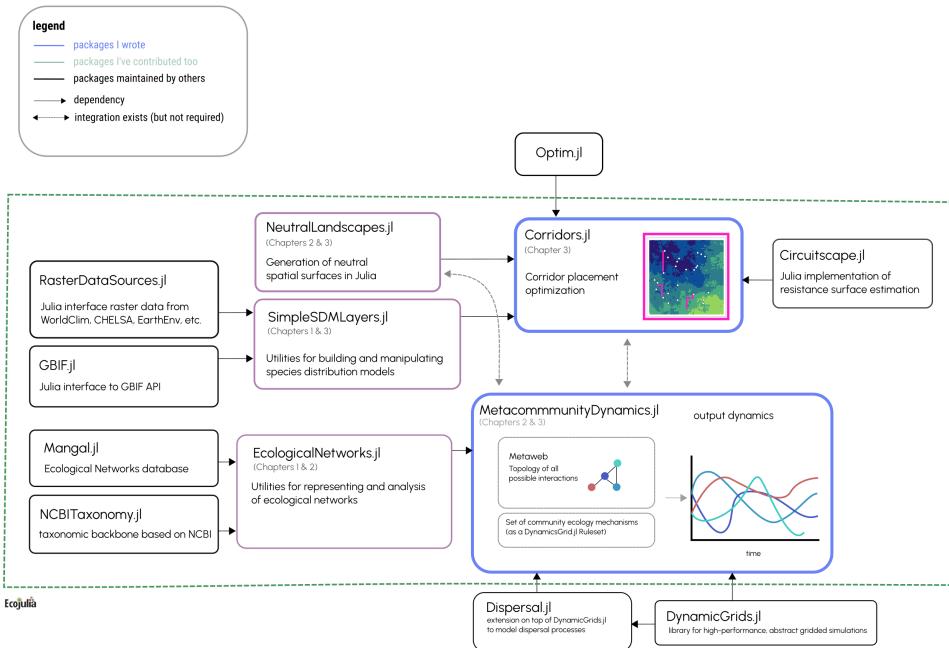


Figure 5 todo

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 (Hanski & Ovaskainen 2000; Ovaskainen *et al.* 2002) to estimate a distribution of time until extinction for each landscape modification (fig. 6, right).

4.1. Methods

- land cover -> resistance -> extinction time simulated annealing to
- optimize landscape optimization

5

CH4 a software note on the resulting packages.

(MetacommunityDynamics.jl: a virtual laboratory for community ecology): a collection of modules in the Julia language for different aspects of metacommunity ecology, including most of the code used for the preceding chapters.

6

Conclusion

7

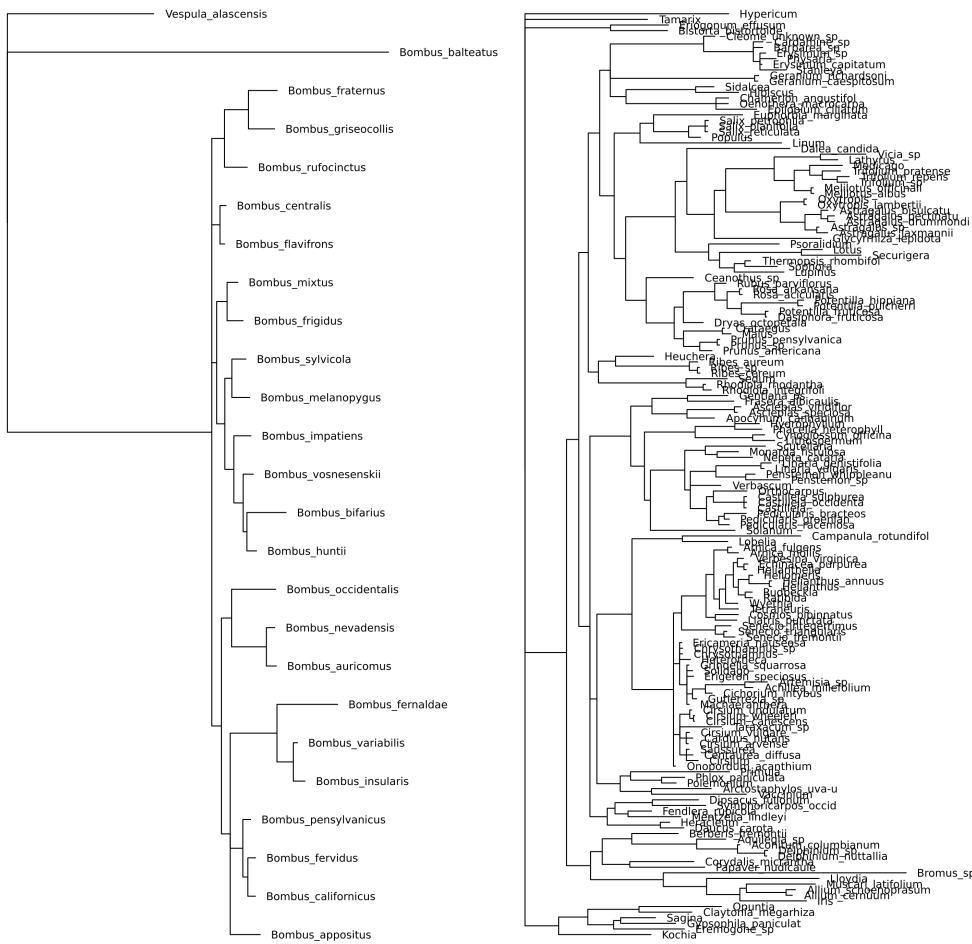


Figure 6 trees

Appendix

References

- Bauer, P., Thorpe, A. & Brunet, G. (2015). The quiet revolution of numerical weather prediction. *Nature*, 525, 47–56.
- Beckage, B., Gross, L.J. & Kauffman, S. (2011). The limits to prediction in ecological systems. *Ecosphere*, 2, art125.
- Chen, Y., Angulo, M.T. & Liu, Y.-Y. (2019). Revealing Complex Ecological Dynamics via Symbolic Regression. *BioEssays*, 41, 1900069.
- Dietze, M.C. (2017). Prediction in ecology: A first-principles framework. *Ecological Applications*, 27, 2048–2060.
- Hadfield, J.D., Krasnov, B.R., Poulin, R. & Nakagawa, S. (2014). A Tale of Two Phylogenies: Comparative Analyses of Ecological Interactions. *The American Naturalist*, 183, 174–187.
- Hanski, I. & Ovaskainen, O. (2000). The metapopulation capacity of a fragmented landscape. *Nature*, 404, 755–758.
- Hill, C., DeLuca, C., Balaji, Suarez, M. & Da Silva, A. (2004). The architecture of the Earth System Modeling Framework. *Computing in Science Engineering*, 6, 18–28.
- Hubbell, S.P. (2001). *The unified neutral theory of biodiversity and biogeography*. Monographs in population biology. Princeton University Press, Princeton.
- Levin, S.A. (1992). The Problem of Pattern and Scale in Ecology: The Robert H. MacArthur Award Lecture. *Ecology*, 73, 1943–1967.
- Makiola, A., Compson, Z.G., Baird, D.J., Barnes, M.A., Boerlijst, S.P., Bouchez, A., et al. (2020). Key Questions for Next-Generation Biomonitoring. *Frontiers in Environmental Science*, 7.
- Ovaskainen, O., Sato, K., Bascompte, J. & Hanski, I. (2002). Metapopulation Models for Extinction Threshold in Spatially Correlated Landscapes. *Journal of Theoretical Biology*, 215, 95–108.
- Ovaskainen, O., Sato, K., Bascompte, J. & Hanski, I. (2002). Metapopulation Models for Extinction Threshold in Spatially Correlated Landscapes. *Journal of Theoretical Biology*, 215, 95–108.
- Pennekamp, F., Iles, A.C., Garland, J., Brennan, G., Brose, U., Gaedke, U., et al. (2019). The intrinsic predictability of ecological time series and its potential to guide forecasting. *Ecological Monographs*, 89, e01359.
- Petchey, O.L., Pontarp, M., Massie, T.M., Kéfi, S., Ozgul, A., Weilenmann, M., et al. (2015). The ecological forecast horizon, and examples of its uses and determinants. *Ecology Letters*, 18, 597–611.
- Peterman, W.E. (2018). ResistanceGA: An R package for the optimization of resistance surfaces using genetic algorithms. *Methods in Ecology and Evolution*, 9, 1638–1647.
- Strydom, T., Catchen, M.D., Banville, F., Caron, D., Dansereau, G., Desjardins-Proulx, P., et al. (2021). A Roadmap Toward Predicting Species Interaction Networks (Across Space and Time) (Preprint). Eco-EvoRxiv.
- Thompson, R.M. & Townsend, C.R. (2000). Is resolution the solution?: The effect of taxonomic resolution on the calculated properties of three stream food webs. *Freshwater Biology*, 44, 413–422.
- Urban, M.C., Travis, J.M.J., Zurell, D., Thompson, P.L., Synes, N.W., Scarpa, A., et al. (2021). Coding for Life: Designing a Platform for Projecting and Protecting Global Biodiversity. *BioScience*.