

# Thesis proposal

Michael D. Catchen<sup>1,2</sup>

---

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

---

1 \_\_\_\_\_

## Introduction

---

P1

---

Within the last several hundred years, human activity have rapidly changed Earth's atmosphere, oceans, and surface. Greenhouse gas emissions have caused an increase the temperature of both Earth's surface and oceans (resulting in acidification), and both agricultural and urban development has rapidly

---

## P2

However, robust forecasting of ecological processes will change in the future is, to say the least, quite difficult. This difficulty is compounded by a few factors, the first being that sampling ecosystems is not easy. Ecological data is often biased, and noisy, spatially and temporally sparse. As a result *ecosystem monitoring* (Makiola *et al.* 2020) has emerged as an imperative. Developing a system for ecological observation, which is able to coordinate across locations. (**AndyUrbanBiomonitoring?** paper).

---

The second major challenge in forecasting ecosystems is that the underlying dynamics of most ecological processes are fundamentally unknown (and unknowable) and instead must be inferred.

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 as over time. It turns out to be much more effective to, rather than attempt to directly model  $\vec{x}(t)$  itself, to instead describe how  $\vec{x}$  changes from one timestep to the next, 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 form of this functional response in real systems is effectively unknown, and some forms are inherently more “forecastable” than others (Chen *et al.* 2019).

---

## P3

However,  
we  
run  
into  
many  
prob-  
lems  
when  
aim-  
ing  
to  
ap-  
ply  
this  
type  
of  
model  
to  
em-  
piri-  
cal  
data  
in  
ecology.

---

The initial success of ODE models can be traced back to the larger program of ontological reductionism, which became the de facto approach model physical sciences after its early success in physics, which, and by the time ecology was becoming a quantitative science (some time in the

---

But  
ecosys-  
tems  
are  
per-  
haps  
the  
quintessen-  
tial  
ex-  
am-  
ple  
of  
sys-  
tem  
that  
can-  
not  
be  
un-  
der-  
stood  
sim-  
ply  
by  
iter-  
ative  
re-  
duc-  
tion  
of its  
com-  
po-  
nents.  
Emer-  
gent  
phe-  
nom-  
ena,  
mech-  
a-  
nisms  
at  
dif-  
fer-  
ent  
scales,  
etc.

---

Some  
have  
been  
ex-  
plored  
in  
the  
eco-  
logi-  
cal  
liter-  
a-  
ture:  
(1)  
Some  
ap-  
pli-  
ca-  
tions  
of  
dy-  
namic  
mod-  
els  
in  
ecol-  
ogy  
as-  
sume  
long-  
run  
equi-  
lib-  
rium.  
(2)  
Stochasticity

---

(3)  
Eco-  
logi-  
cal  
pro-  
cesses  
vary  
across  
more  
vari-  
ables  
than  
the  
tools  
of  
ana-  
lytic  
mod-  
els  
are  
suited  
for.  
As  
the  
num-  
ber  
of  
vari-  
ables  
in an  
ana-  
lytic  
model  
in-  
creases,  
so  
does  
the  
abil-  
ity of  
the  
sci-  
en-  
tist  
to  
de-  
cern  
clear  
rela-  
tion-  
ships  
be-  
tween  
them,  
and  
so  
does  
over-  
fit-  
ting  
po-  
ten-  
tial.  
Curse  
of  
dimensionality—

---

P4

The term *ecological forecasting* implicitly creates an analogy between predicting how ecosystems will change in the future by using the term “forecasting”—the most immediate analog being the success story of weather forecasting via numerical weather prediction (NWP).

Although it is become almost hack to complain about the dang weather forecast being wrong, over the least 50 years the (**Bauer2015QuiRev?**).

The success of NWP, and the Earth observations that support it 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.

Much as one would not aim to forecast the weather in Quebec by applying Navier-Stokes. NWP has worked because it incorporates information about data and meteorological processes collected at difference scales into models that. Use of computational methods in NWP.

Transition to simulation as the solution: shift toward approach of building models that *generate* data.

(resolving the semantic ambuity of what differentiates “mechanistic” vs “phenomological” models is out of scope for now).

More broadly a reflection reflect ecology lagging behind the statistical methods used in sciences that face similar challenges (many dimensions, many mechanisms at different scales, each with stochasticity). Chaotic dynamics emerge from simple analytic models, and . Whether ecosystems actually exhibit chaotic behavior is a different question.

---

P5

But  
fore-  
casting  
isn’t  
the  
only  
diffi-  
cult  
prob-  
lem  
here.



---

Transition  
to  
theme  
of op-  
timiza-  
tion  
given  
un-  
known  
infor-  
ma-  
tion. A  
fore-  
cast  
gives  
us a  
range  
of  
future  
values  
with  
uncer-  
tainty  
around  
them.  
Fur-  
ther a  
conve-  
nient  
prop-  
erty  
that a  
fore-  
casting  
model's  
uncer-  
tainty  
goes  
up  
over  
time  
(if we  
as-  
sume  
the  
under-  
lying  
pro-  
cess is  
Markov-  
this is  
a  
strong  
as-  
sump-  
tion  
but oft  
true of  
the  
mod-  
els we  
fit to  
tempo-  
ral  
data)

---

In face of uncertainty, decision making is an optimization problem. 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.

---

Indeed  
Marx's  
most  
well  
known  
quote  
that  
“philoso-  
phers  
have  
hith-  
erto  
only  
inter-  
preted  
the  
world  
in var-  
ious  
ways;  
the  
point  
is to  
change  
it.”  
and a  
neces-  
sary  
step  
toward  
estab-  
lishing  
a just  
and  
sus-  
tain-  
able  
world.

---

#### **P6 – final intro para**

Three major components here: 1) Ecosystem monitoring, 2) Forecasting using the products of that monitoring, and 3) Choosing the best possible mitigation strategy.

This flow is outlined in the left panel of fig. [1](#)

## **2**

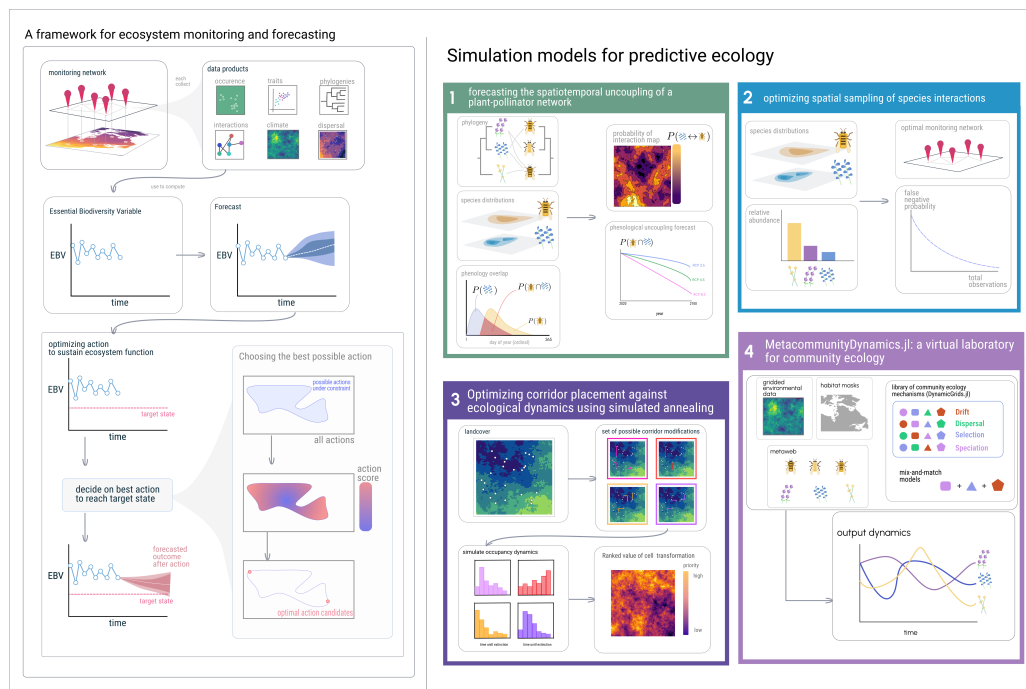
---

### **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” (**Strydom2021?**)



**Figure 1** thesis concept

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

## CH1 concept figure

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

**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

## CH2 optimizing sampling of interactions

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

For a given sample of interaction data, and proposes a method for optimizing spatial sampling of a possible interaction between species as a function of the estimated distribution of both species.

### 3.1. Methods

- the missing link paper, turn this into optimizing with two different SDMs
- relative abundance and its effect on false negative
- non-independent associations in samples
- 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

### 3.2. Results

#### 3.2.1 *In-progress results*

4

---

## CH3 optimizing corridor placement

This chapter proposes an algorithm for optimizing restoration across space

(corridorplacement/restoration effort) given a raster where each cell indicates land-cover. The optimization method uses the result of a simulated process (specifically occupancy dynamics in the landscape) and uses simulated annealing to estimate the global optimum of the targetstate (specifically mean-time-to-extinction for the occupancy dynamics example).

### 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.

- TK conceptual figure with interfaces between what I'm writing / have contributed to and linked with other libraries
- Observatories.jl, Corridors.jl, MCD.jl

6

---

## concl

// this is a discussion para An oft applied definition of the origin of is “the application of the scientific method to natural history.” Since its origin ecology has been a descriptive science. This is a natural by-product of the immense variability of Earth’s biosphere. emerged to explain particular phenomena at particular scales. In recent years, there has been an interest in an epistemological shift in ecology. To shift ecology into a predictive science. The justification for this shift is twofold: (1) bogged down philosophy of science, by further rooting our understanding of ecosystem function and dynamics in an ability to predict their structure (Dietze 2017). and (2) the practical need for models for *ecological forecasting*.

---

## References

- 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.
- 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.