

**Date:** December 19, 2016

**Descriptive Title:** Synthesizing time series of plant and animal populations to understand the limits to ecological forecasts

**Short Title:** Limits to Ecological Forecasting

**PI Contact Information:**

**Andrew Tredennick** (atredenn@gmail.com; 970-443-1599)

Utah State University

5230 Old Main Hill, Logan, UT 84332

**Mevin Hooten** (Mevin.Hooten@colostate.edu)

United States Geological Survey

Colorado State University

address

**Peter Adler** (peter.adler@usu.edu)

Utah State University

5230 Old Main Hill, Logan, UT 84332

**Project Summary:** Forecasting the impacts of global climate change is a major challenge facing ecologists and land managers in the 21<sup>st</sup> century. For plant populations, the challenge lies in adequately representing internal population dynamics and external environmental forcing. Traditionally, plant population modelers have relied on demographic data that are difficult to collect, but time series of plant percent cover are an untapped resource for population modeling and forecasting. We propose to convene a diverse group of ecologists with expertise in plant population monitoring and modeling to build dynamic forecasting models based on percent cover time series from across the United States. Using the ecological forecast horizon as a unifying comparative framework, we will discover where and when plant populations are predictable, and when they are not. Furthermore, for every data set that we use, we will make yearly forecasts of plant population trajectories up to the year 2100 under projected climate change. The data we use will be compiled into a publicly available database, population forecasts will be archived for future validation, and all computer code will be made publicly available.

**Proposed Start and End Dates:** October 2017 to September 2019, with two 4-day meetings at the Powell Center

**Proposed Data Release Date:** September 2019

**Total Requested Budget:** \$XX0,000

**Is this a resubmission?** No

**Conflicts of Interest with Reviewers:** None

**Keywords:** climate and land use change; ecosystems

# 1 Problem Statement

A fundamental challenge facing society is to predict the impacts of global environmental changes such as nitrogen deposition, climate change, and habitat fragmentation on ecosystems. Each of these global change drivers have now exceeded their historical ranges of variability (Steffen et al. 2015), meaning we are entering a no-analog world in which we can no longer look to the past to predict the future. We can, however, look to the past to parameterize models that allow us to *forecast* the future states of ecological systems (Clark et al. 2001). Ecologists are in an excellent position to meet this forecasting challenge because we have spent decades gaining understanding of the processes that regulate populations, communities, and ecosystems. But, we currently lack a systematic understanding of the limits to ecological forecasts and whether those limits are surmountable.

To understand the “forecasting problem” in ecology, Dietze (2017) proposes a first-principles approach to partitioning forecast uncertainty. Consider a dynamical model designed to predict some state  $y$  in the future ( $y_{t+1}$ ) based on the current state ( $y_t$ ), an external covariate ( $x$ ), parameters ( $\theta$ ), and process error ( $\epsilon$ ). We can then write a general form of the model as:

$$y_{t+1} = f(y_t, x_t | \theta) + \epsilon, \quad (1)$$

which states that  $y$  at time  $t + 1$  is a function of  $y$  and  $x$  at time  $t$  conditional on the model parameters ( $\theta$ ) plus process error ( $\epsilon$ ). Using a Taylor expansion “delta method”, Dietze (2017) shows that forecast variance ( $Var[y_{t+1}]$ ) is:

$$Var[y_{t+1}] = \underbrace{\left(\frac{\delta f}{\delta y}\right)^2}_{\text{stability}} \underbrace{Var[y_t]}_{\text{IC uncert.}} + \underbrace{\left(\frac{\delta f}{\delta x}\right)^2}_{\text{driver sens.}} \underbrace{Var[x_t]}_{\text{driver uncert.}} + \underbrace{\left(\frac{\delta f}{\delta \theta}\right)^2}_{\text{param sens.}} \underbrace{Var[\theta]}_{\text{param. uncert.}} + \underbrace{Var[\epsilon]}_{\text{process error}}, \quad (2)$$

where each additive term follows a pattern of *sensitivity* times *variance*. Thus, the variance attributable to any particular factor is a function of how sensitive the model is to the factor and the variance of that factor. For example, large sensitivity to the covariate  $x$  can be compensated for if the uncertainty of the covariate is low.

## 2 Proposed Activities

PIs Tredennick, Hooten, and Adler have led several efforts to forecast the response of plant populations to climate change (Tredennick et al. 2016a, 2016b). Our failures to produce forecasts with reasonable levels of uncertainty, and our inability to attribute that uncertainty to specific causes, has motivated this proposal. We seek to (1) assemble a database of all publicly available time series of plant and animal abundance, (2) use those data to fit forecasting models, and (3) partition the forecast uncertainty from those models to better understand the limits to ecological forecasting. Along the way, as mentioned above, we will test fundamental hypotheses in ecology about chaos

and density-dependence under a novel framework. We have learned that collating large datasets and rigorously fitting statistical population models requires dedicated effort. We therefore request funding for a Powell Fellow (PI Tredennick) to lead all aspects of our proposed work.

## 2.1 Data Synthesis

The first goal of this project is to assemble a database of plant and animal abundance time series from around the globe. We have identified several potential data sets (Table 1) and we anticipate our working group members (Table 2) will be able to identify more based on their professional networks. The assembled database will be the first of its kind and will be extremely useful to the research community. We also anticipate the database to provide examples that can be used to teach advanced statistical methods.

### 2.1.1 Time Series Selection

Our forecasting approach requires data sets of sufficient length to fit dynamic statistical models that include environmental covariates. Therefore, one of our first tasks as a working group will be to develop a set of criteria for selecting data. These criteria will include, but are not limited to, cut-offs for length of time series (probably ~10 years), availability of suitable climate data, and the number of replicate observations within a time period. Our goal is to assemble a database equally represented by plant and animal species.

### 2.1.2 Data Sources and Formatting

We have already identified six publicly-available time series of plant and animal abundance, some of which contain data on multiple species (Table 1). The working group will identify the best data sets from this preliminary list and we will add more as feasible (depending on work load and quality of the data sets). All abundance data sets will be combined into a single database that can be queried using standard SQL software (e.g., SQLite). The abundance data will be linked to a second database of climate or climate-related (e.g., snow depth) covariates. The working group will identify which studies have associated climate data, and for the remainder we will extract climate data from gridded data products (e.g., PRISM). We will work with study area experts, some of whom will be in the working group, to decide on appropriate climate covariates for each population time series. This is a preliminary data processing workflow, and we anticipate working closely with Powell Center experts to synthesize our data sets into a single data base.

*Table 1: Publicly-available abundance time series for plants and animals. Animal abundance is the number of individuals in a given area. Plant abundance is percent cover or biomass.*

Taxa	Species (common name)	Length (yrs.)	Citation/Website
Animal	<i>Bison bison</i> (Bison)	35	Hobbs et al. (2015)
Animal	<i>Grus canadensis</i> (Sandhill crane)	42	Gerber et al. (2015)
Animal	Several (Breeding Bird Survey)	47	<a href="https://www.pwrc.usgs.gov/bbs/">https://www.pwrc.usgs.gov/bbs/</a>

Taxa	Species (common name)	Length (yrs.)	Citation/Website
Animal	<i>Dipodomys</i> spp. (Kangaroo rats)	39	Ernest et al. (2015)
Plant	<i>Artemisia tripartita</i> (Threetip sagebrush)	22	Zachmann et al. (2010)
Plant	<i>abundant plant</i> (name)	30	<a href="http://jornada.nmsu.edu/lter/data/plant">http://jornada.nmsu.edu/lter/data/plant</a>

## 2.2 Analysis

### 2.2.1 Dynamic Forecasting Models

With our assembled time series, we will be in the position to fit dynamic statistical models for each represented species. The particular details of each model may vary, but our general approach will be to fit Bayesian hierarchical state-space models. State-space models are well-suited for ecological forecasting because they inherently incorporate and propagate the sources of uncertainty we seek to partition: process error, observation error, and parameter error. State-space models also explicitly acknowledge that our observations are imperfect, meaning that the state we want to estimate (e.g., abundance of elk in Rocky Mountain National Park) is unknowable, or *latent*. A typical state-space model takes the form:

$$y_t \sim \mathcal{N}(z_t, \sigma_{obs.}) \quad (3)$$

$$z_t \sim \mathcal{N}(\mu_t, \sigma_{proc.}) \quad (4)$$

$$\mu_t = \alpha + \beta_1 z_{t-1} + \beta_2 x_{t-1}, \quad (5)$$

where  $y_t$  is the observed state at time  $t$ ,  $z_t$  is the latent state at time  $t$ ,  $\mu_t$  is the deterministic prediction of the state at time  $t$  given the regression parameters  $\alpha$ ,  $\beta_1$ , and  $\beta_2$ , and  $x_{t-1}$  is a climate covariate at time  $t-1$ . The two error terms represent observation error ( $\sigma_{obs.}$ ) and process error ( $\sigma_{proc.}$ ). The model is *dynamic* because the future state ( $z_t$ ) depends on the previous state ( $z_{t-1}$ ). For clarity we show the data and process models (Eqs. 3 and 4, respectively) with normal likelihoods ( $\mathcal{N}$ ), but these distributions will vary depending on the particular data generating process (Hobbs and Hooten 2015).

#### 2.2.2 An Example: Yellowstone Bison

Yellowstone National Park uses an adaptive management program to regulate the size of its bison (*Bison bison*) population, which requires a model that uses historical data to forecast probable outcomes from management decisions. Long-range forecasts of bison abundance, those greater than five years, suffer from large uncertainty; to the point where the forecasts cover almost every possible outcome (Hobbs et al. 2015). Forecast uncertainty will always increase with time, but the rate of increase can be reduced if we can target specific sources of uncertainty. Do we need to better estimate initial conditions? Or, is effort better spent developing a better model of population growth? We can answer these questions by partitioning forecast uncertainty.

We used annual counts of the Yellowstone bison from 1975 to 2010 to fit a population growth model using a state space approach. Proper distributions were used for the data and likelihood models to account for the data being greater than 0 and integers. After fitting the model, we made annual forecasts for 10 years into the future with all uncertainty propagated (Fig. 1A). We made those same forecasts with (i) just initial condition uncertainty, (ii) initial condition and parameter uncertainty, and (iii) initial condition, parameter, and process uncertainty. This set of forecasts allows us to partition the relative variance of different sources, taking a numerical approach that is analogous to analytical approach shown in Equation 2.

For our forecasts of Yellowstone bison abundance, process error dominates forecast uncertainty at all horizons (Fig. 1B). Parameter error is relatively constant throughout the forecast horizons. Initial condition uncertainty decays very quickly, indicating strong internal population regulation and the lack of chaotic dynamics. In this respect, forecasting bison abundance is a very different problem than forecasting weather. Weather forecasts are limited by certainty in initial conditions because the weather system is chaotic. Therefore, resources are allocated to constantly updating weather observations. On the contrary, bison population dynamics appear to not be chaotic so updating population estimates more than yearly would not help forecast certainty. Rather, our analysis suggests that the forecast is limited by the process model. We could improve the process model by including external climate covariates, which we plan to do as part of this proposal. Doing so would also allow us to determine the relative contribution of driver error to forecast uncertainty through time.

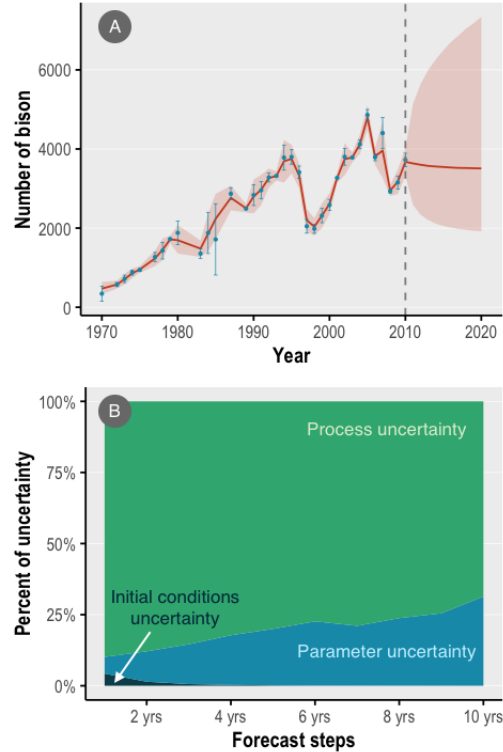


Figure 1: An example state space model fit and forecast (A) and example of forecast error partitioning (B). In (A), points are observed counts and errorbars show the observed standard deviation, solid line is the median of the posterior distribution of the number of bison in Yellowstone and shaded area shows the 0.025 and 0.975 quantiles from the posterior distribution. Within sample estimates are suggested that the forecast is process model by including external climate covariates. Doing so would also allow us to determine the relative contribution of each source of forecast error over time.

### 3 Participants

We have assembled a diverse working group, with gender and career stage diversity. We will also advertise two travel awards for participation in our working group: one for a postdoc to participate and one for a graduate student to participate. The travel awards will be advertised on ESA's SEEDS<sup>1</sup> and Inclusive Ecology<sup>2</sup> webpages to ensure we receive applications from underrepresented minorities.

<sup>1</sup><http://esa.org/seeds/>

<sup>2</sup><http://www.esa.org/inclusive-ecology/>

## 4 Time Table of Activities

**October to December 2017** Monthly PI Skype meetings to prepare for first meeting. Begin email exchanges to confirm participants. Write R scripts to download, clean, and aggregate data in Table 1. Refine YNP bison example for first meeting.

**January 2018** First meeting for four days at Powell Center: Identify other appropriate data sets. Decide on model structures for each data set and identify climate data availability. Form two-person analysis teams for each data set. Go through YNP bison analysis as an exemplar. Define manuscripts.

**February to December 2018** Bi-monthly PI Skype meetings. Continue and finalize data set acquisition (including climate data) and aggregate into single database for analysis teams. Analysis teams, with the PC Fellow, identify appropriate climate drivers for their system and begin fitting state-space models. PC Fellow writes generalizable R functions for analysis teams to use for forecast partitioning. PC Fellow will have at least one Skype meeting with each analysis team during this period. Write first drafts of manuscripts.

**January 2019** Second meeting for four days at Powell Center: Analysis teams present results for feedback from the working group. Finalize data base and identify any outstanding issues. Evaluate manuscript drafts and form writing teams. Create a GITHUB repository for individual forecasts.

**February to June 2019** Finalize data base and forecast repository. Finalize all forecasts and forecast partitions. Continue writing manuscripts. PC Fellow will have at least one Skype meeting with each analysis team during this period.

**July to September 2019** Complete manuscripts and submit for publication. Release database and forecast repository to Powell Center for permanent archiving. Write a blog post for Dynamic Ecology<sup>3</sup> on the limits to ecological forecasting, introduce to database and forecast repository, and pose a forecasting challenge.

## 5 Anticipated Results and Benefits

## 6 References

Clark, J. S., S. R. Carpenter, M. Barber, S. Collins, A. Dobson, J. A. Foley, D. M. Lodge, M. Pascual, R. Pielke, W. Pizer, C. Pringle, W. V. Reid, K. A. Rose, O. Sala, W. H. Schlesinger, D. H. Wall, and D. Wear. 2001. Ecological forecasts: an emerging imperative. *Science* (New York, N.Y.) 293:657–660.

Dietze, M. 2017. *Ecological Forecasting*. Princeton University Press, Princeton.

Hobbs, N. T., and M. B. Hooten. 2015. *Bayesian Models: A Statistical Primer for Ecologists*. Princeton University Press, Princeton.

---

<sup>3</sup><http://dynamicecology.wordpress.com/>

- Hobbs, N. T., C. Geremia, J. Treanor, R. Wallen, P. J. White, M. B. Hooten, and J. C. Rhyan. 2015. State-space modeling to support management of brucellosis in the Yellowstone bison population. *Ecological Monographs* 85:525–556.
- Steffen, W., K. Richardson, J. Rockström, S. Cornell, I. Fetzer, E. Bennett, R. Biggs, and S. Carpenter. 2015. Planetary boundaries: Guiding human development on a changing planet. *Science (New York, N.Y.)* 348:1217.
- Tredennick, A. T., M. B. Hooten, C. L. Aldridge, C. G. Homer, A. R. Kleinhesselink, and P. B. Adler. 2016a. Forecasting climate change impacts on plant populations over large spatial extents. *Ecosphere* 7:e01525.
- Tredennick, A. T., M. B. Hooten, and P. B. Adler. 2016b. Do we need demographic data to forecast plant population dynamics? *Methods in Ecology and Evolution*:n/a–n/a.