Date: January 9, 2017

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
Fort Collins, CO 80523

Peter Adler (peter.adler@usu.edu)
Utah State University
5230 Old Main Hill, Logan, UT 84332

Project Summary: Forecasting the impacts of global environmental change is a major challenge facing ecologists and land managers in the 21st century. More here...

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

"...ecologists do not predict because when we do it reveals how little we understand about the natural world." –Houlahan et al. 2016

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.

Making poor forecasts is inevitable as ecology matures as a more predictive science. The key is to learn from our failures so that forecasts become more accurate over time. The success story of weather forecasting tells us that basic research on the contributions to forecast uncertainty is essential (Bauer et al. 2015). In ecology, a powerful approach is to combine our detailed knowledge of organisms and ecosystem processes with emerging knowledge on forecast uncertainty (Petchey et al. 2015).

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+1}), an external covariate (x), parameters (θ), and process error (ε). We can then write a general form of the model as:

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

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

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

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.

The goal of this project is to synthesize forecasts of plant and animal populations to advance our fundamental knowledge on the limits to ecological forecasts. We will convene a diverse group of population ecologists to partition forecast uncertainty for several species and systematically link sources of uncertainty to characteritics of the focal populations (e.g., life history traits). Along the way, we will develop the first ever repository of ecological forecasts and a publicly-available data base of abundance time series. The data base will become a go-to resource for teaching and research, and will be used for a forecasting challenge.

2 Hypotheses

All forecasts are uncertain, and reducing forecast error hinges upon knowing the source of uncertainty. The premise of this proposal is that ecological characteristics of organisms (e.g., life history) are related to sources of forecast uncertainty. Specifically, we pose three hypotheses:

H1. The contribution of initial conditions error to forecast uncertainty is greater for fast growing populations than for slow growing populations. Fast growing populations are more likely to exhibit chaotic or near-chaotic dynamics, which inflates forecast error due to initial conditions uncertainty. To test this hypothesis we will partition forecasts and regress the proportion of uncertainty attributable to initial conditions against estimated population growth rates.

H2. The contribution of driver error to forecast uncertainty is greater for species that "predict" than for species that "bet-hedge" against environmental conditions. In variable environments, plants have developed two main strategies for maintaining fitness despite changing conditions: (i) bet-hedging, where germination is low but constant, and (ii) predictive germination, where germination is cued by the environment. Bet-hedgers should be insensitive to environmental drivers, so their forecast uncertainties should not be driven by driver error. We will compare forecast partitions for winter annual plants that span a spectrum from bet-hedgers to predictors to test this hypothesis.

H3. The contribution of process error to forecast uncertainty increases with the number of "specialist" trophic links. We are using population models, which ignore trophic interactions. Therefore, if trophic interactions are important and/or plentiful, our models will poorly represent the process. We will compare forecast partitions for all data sets to test this hypothesis.

Our hypotheses are motivated by recent reviews of ecological forecasting (Petchey et al. 2015, Houlahan et al. 2016) that suggest research should be organized around understanding the effects of ecological processes and variables on forecast uncertainty. By testing our hypotheses, we will gain a systematic understanding of what sources of uncertainty dominate ecological forecasts and how those sources relate to ecological characteristics. We note that these three hypotheses are likely just a starting point, as we anticipate the working group will generate additional hypotheses.

3 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.

3.1 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 be provide examples that can be used to teach advanced statistical methods.

3.1.1 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.

3.1.2 Data Sources and Formatting

We have already identified nine 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, many 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	*Bison bison* (American Bison)	35	Hobbs et al. (2015)
Animal	*Grus canadensis* (Sandhill Crane)	42	Gerber et al. (2015)
Animal	Several (Breeding Bird Survey)*	47	http://www.pwrc.usgs.gov/bbs/
Animal	*Dipodomys* spp. (Kangaroo rats)*	39	Ernest et al. (2015)
Plant	*Artemisia tripartita* (Threetip Sagebrush)	22	Zachmann et al. (2010)
Plant	*Eriastrum diffusum* (Miniature Woollystar)*	14	Ernest et al. (2015)
Plant	*Artemisia* spp. (Sagebrush spp.)	27	Homer et al. (2013)
Plant	*Artemisia scopulorum* (Alpine Sagewort)*	11	http://niwot.colorado.edu/
Plant	Winter annuals (Desert Lab LTREB)	30	Gremer and Venable (2014)

*These datasets contain observations for many species, not just the one listed as an example.

3.2 Analysis

3.2.1 Dynamic Forecasting Models

With our assmebled 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 propogate 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:

$$y_t \sim \text{Normal}(z_t, \sigma_{obs.}) \quad (3)$$

$$z_t \sim \text{Normal}(\mu_t, \sigma_{proc.}) \quad (4)$$

$$\mu_t = f(\theta_p, z_{t-1}, \mathbf{x}), \quad (5)$$

where y_t is the observed state at time t, z_t is the latent state at time t, μ_t is the determinstic prediction of the state at time t given estimated parameters (θ_p), the latent state z at time t-1, and $\mathbf{x}$ is a vector climate covariates. The two error terms represent observation error ($\sigma_{obs.}$) and process error ($\sigma_{obs.}$). 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, but these distributions will vary depending on the particular data generating process (Hobbs and Hooten 2015).

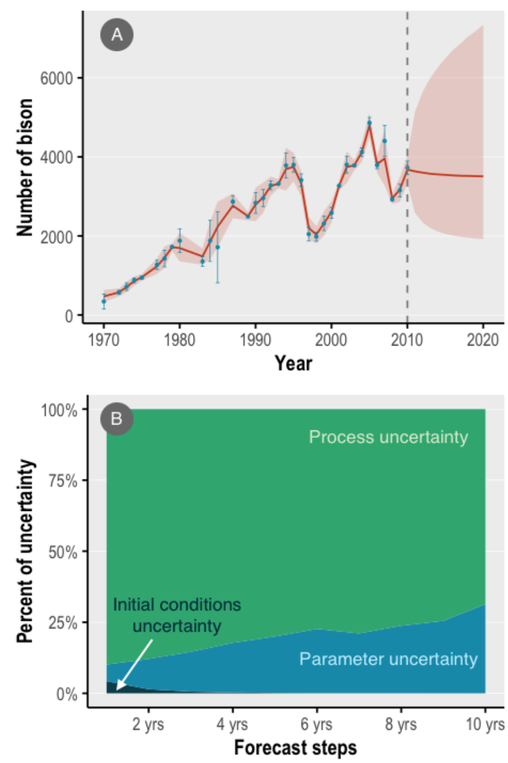

Figure 1: An example state space model fit and forecast (A) and example of forecast error partitioning (B). In (A), points are observed counts and errorbars show the observed standard deviation; solid line is the median of the posterior distribution of the number of bison in Yellowstone and shaded area shows the 0.025 and 0.975 quantiles from the posterior distribution. Within-sample estimates are constrained by data, whereas the forecasts (past 2010) are not. Panel (B) shows the relative contribution of each source of forecast error over time.

3.2.2 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.

We used annual counts of the Yellowston bison from 1975 to 2010 to fit a population growth model using a state-space approach (Equations 3-5). 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 analagous 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, which is consistent with our **Hypothesis H1**. The next step is to compare Fig. 1B to similar results from a fast-growing population, as planned for this proposal.

4 Participants

We have assembled a diverse working group, with gender and career stage diversity (Table 2). Of the 10 listed participants, four are women and three are early career scientists. The proposed Powell Center Fellow is Andrew Tredennick.

Table 2: List of participants.

Name	Affiliation	Expertise	Associated Dataset
Andrew Tredennick*[1,2,3]	Utah State University	Data management/synthesis, population forecasting	n/a
Mevin Hooten*	U.S. Geological Survey	Bayesian modeling, statistical forecasting	Sandhill crane
Peter Adler*	Utah State University	population ecology/modeling, data synthesis	Idaho sagebrush
Lauren Buckley	University of Washington	ecological forecasting climate change	*Anolis* spp.
Michael Dietze*	Boston University	ecological forecasting, partitioning uncertainty	n/a
Emily Farrer*	Tulane University	population modeling, Bayesian analysis	Niwot plants
Jennifer Gremer*	University of California, Davis	plant population modeling data management	Winter annuals
Janneke HillRisLambers*	University of Washington	plant population modeling, climate change	CA annual plants
N. Thompson Hobbs*	Colorado State University	population ecology, state space models	Yellowstone bison
Ethan White*	University of Florida	ecological forecasting, data synthesis	BBS & Portal Data

*Confirmed participant
[1]Powell Center Fellow
[2]Technical liaison to Powell Center computing staff
[3]Party responsible for adherence to Powell Center Data and Information Policy

5 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 aquisition (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[1] on the limits to ecological forecasting, introduce to database and forecast repository, and pose a forecasting challenge.

6 Anticipated Results and Benefits

The greatest benefit of this project will be the conceptual advance of linking ecological theory to sources of forecast uncertainty. We anticipate our research associated with this project will catalyze a new area of research on ecological forecasting. The assembled data base of plant and animal abundance time series will be an invaluable resource for teaching and research, and the repository of population forecasts will provide new opportunities to validate, and improve, forecasts. We expect to produce at least three publications:

1. A synthesis paper testing our hypotheses using all the data sets aimed at *PNAS* or *Ecology Letters*.
2. A data paper for *Ecology* or *Scientific Data* describing the assembled data base and forecast repository.
3. Several in-depth papers on forecasts of specific populations aimed at top-tier applied journals such as *Ecological Applications* and *Journal of Applied Ecology*.

7 References

Bauer, P., A. Thorpe, and G. Brunet. 2015. The quiet revolution of numerical weather prediction. Nature 525:47–55.

[1] http://dynamicecology.wordpress.com/

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.

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.

Houlahan, J. E., S. T. Mckinney, T. M. Anderson, and B. J. Mcgill. 2016. The priority of prediction in ecological understanding.

Petchey, O. L., M. Pontarp, T. M. Massie, S. Kéfi, A. Ozgul, M. Weilenmann, G. M. Palamara, F. Altermatt, B. Matthews, J. M. Levine, D. Z. Childs, B. J. McGill, M. E. Schaepman, B. Schmid, P. Spaak, A. P. Beckerman, F. Pennekamp, and I. S. Pearse. 2015. The ecological forecast horizon, and examples of its uses and determinants. Ecology Letters 18:597–611.

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.