Date: January 30, 2017

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

Short Title: 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. Unfortunately, our knowledge of the limits to ecological forecasts lags behind our deep understanding of ecological systems. Fortunately, emerging theory on partitioning forecast uncertainty provides an opportunity to better understand the limits to ecological forecasts. We propose to convene a diverse working group of quantitative ecologists, each with detailed knowledge of a particular study system, to develop forecasting models of plant and animal populations. As a team, we will partition the forecast uncertainty for each focal population to test fundamental hypotheses about the relationship between particular sources of uncertainty and ecological characteristics. We will develop a forecast repository and a synthetic database of population time series coupled with environmental covariates that will be ideal for teaching and research. We anticipate our work will catalyze a new research area in ecology focused on identifying the boundaries of ecological forecasts based on ecological theory.

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: $130,801 (Year 1 $23,565; Year 2 $107,236)

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. 2017

A fundamental challenge facing society is to predict the ecological impacts of global environmental changes such as nitrogen deposition, climate change, and habitat fragmentation. 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. However, we lack a systematic understanding of the current limits to ecological forecasts. As a result, we do not know how to allocate research effort to improve our forecasts.

Making poor forecasts is inevitable as ecology matures into a more predictive science. The key is to learn from our failures so that forecasts become more accurate over time. The success of meteorological forecasting tells us that basic research on the causes of 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) proposed 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 environmental driver(s) (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) showed 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* ("IC uncert." refers to "*I*nitial *C*onditions uncertainty"). 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 offset if the uncertainty of the covariate is low. Knowing the relative size of these uncertainties is important to guide future research on ecological processes, data collection schemes, and the application of forecasting models to applied questions. For example, knowing that a large proportion of forecast uncertainty stems from uncertainty in projections of the covariate x means that excluding the covariate may be the best choice when making forecasts. Understanding how, and if, sources of forecast uncertainty systematically covary with ecological characteristics will allow us to identify the potential limits to forecasts based on our deep understanding of ecological systems.

The goal of this project is to synthesize forecasts of plant and animal populations to advance our fundamental knowledge of 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 characteristics of the focal populations (e.g., life history traits). Along the way, we will develop a repository of ecological forecasts and a publicly-available database of abundance time series with environmental covariates. The database will become a go-to resource for teaching and research on ecological forecasting.

2 Hypotheses

All forecasts are uncertain, and reducing forecast uncertainty 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 motivated by recent reviews of ecological forecasting (Petchey et al. 2015; Houlahan et al. 2017).

H1. The contribution of initial conditions uncertainty 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 inflate forecast uncertainty 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 intrinsic population growth rates.

H2. The contribution of driver uncertainty to forecast uncertainty is greater for species that "predict" than for species that "bet-hedge" against environmental conditions. In variable environments, annual 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 rates, cued by the environment, vary substantially from year to year. Bet-hedgers should be insensitive to environmental drivers, so their forecast uncertainties should not be driven by driver uncertainty. We will compare forecast partitions for winter annual plants that span a spectrum from bet-hedgers to predictors (Gremer et al. 2014) to test this hypothesis. We also anticipate applying this approach to our other focal data sets by quantifying variation of annual vital rates to determine where species lie on the bet-hedging spectrum (e.g., Botero et al. 2015).

H3. The contribution of process uncertainty to forecast uncertainty increases with the number of trophic links. We are using single-species population models in which trophic interactions are implicit, not explicit. Therefore, if trophic interactions are important our models will poorly represent the process. We will compare forecast partitions for all data sets to test this hypothesis. Our working group will consist of experts for each data set (Tables 1 and 2), meaning we will be able to develop food webs for each focal species based on expert knowledge. We recognize that constructing subjective food webs is a coarse approach, but in practice our knowledge of a species' trophic connections is typically coarse. Thus, we contend that our proposed approach is actually well-suited to addressing the practical challenge of linking forecast uncertainty to trophic connections.

By testing our hypotheses, we will gain a systematic understanding of what sources of uncertainty dominate ecological forecasts for different types of populations. These three hypotheses are just a starting point, as 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 frustration with the high levels of uncertainty in our forecasts, and our inability to attribute that uncertainty to specific causes, has motivated this proposal. We seek to (1) assemble a database of publicly available time series of plant and animal abundances which contain the necessary data, (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. We have learned that collating large data sets 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

3.1.1 Time Series Selection

The first goal of this project is to assemble a database of plant and animal abundance time series. 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. 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 and other environmental data, and the number of replicate observations within a time period. Our goal is to assemble a database with equal representation of plant and animal species.

The assembled database will contain yearly abundance estimates coupled with environmental covariates, setting it apart from other population time series that do not include potential drivers (e.g., The Global Population Dynamics Database and BioTIME). We will also contribute our assembled database to other synthetic databases where appropriate.

3.1.2 Data Sources and Formatting

We have already identified eleven 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 database.

Table 1: Publicly-available abundance time series for plants and animals.

Taxa	Species (common name)	Length (yrs.)	Citation/Website
Animal	*Bison bison* (American Bison)	35	Hobbs et al. (2015)
Animal	*Enhydra lutris* (Sea otter)	20	Williams et al. (2017)
Animal	Breeding birds*	47	http://www.pwrc.usgs.gov/bbs/
Animal	*Dipodomys* spp. (Kangaroo rats)*	39	Ernest et al. (2015)
Animal	Grasshopper spp.*	10 (with weekly census)	http://ghopclimate.colorado.edu
Plant	Sagebrush steppe perennial plants*	22	Zachmann et al. (2010)
Plant	AZ desert annuals*	14	Ernest et al. (2015)
Plant	*Artemisia* spp. (Sagebrush spp.)	27	Homer et al. (2013)
Plant	Alpine tundra plants*	11	http://niwot.colorado.edu/
Plant	Winter annuals (Desert Lab LTREB)*	30	Gremer and Venable (2014)
Plant	Mt. St. Helens plants*	30	del Moral (2010)

*These data sets contain observations for many species.

3.2 Analysis

3.2.1 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 propogate the sources of uncertainty we seek to partition: process uncertainty, observation uncertainty, and parameter uncertainty. 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:

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

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

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

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) for the specified model function [*f()*], the latent state z at time t-1, and $\mathbf{x}$ is a vector of environmental covariates. 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 simplicity 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 & Hooten 2015).

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 of management decisions. Long-range forecasts of bison abundance, those greater than five years, are so uncertain that 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 the analytical approach shown in Equation 2.

The key result is that process uncertainty dominates forecast uncertainty at all horizons (Fig. 1B). Parameter uncertainty 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 in testing **H1** is to compare Fig. 1B to results from fast-growing populations.

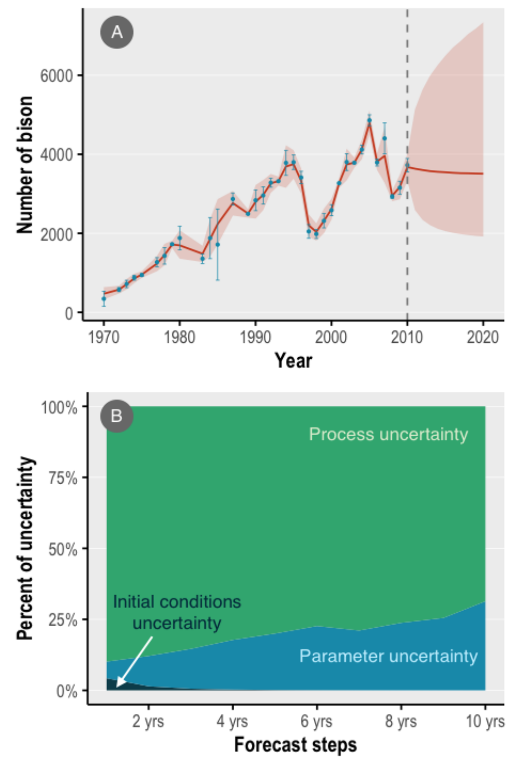

Figure 1: An example state space model fit and forecast (A) and example of forecast uncertainty partitioning (B). In (A), points are observed counts and errorbars show the observed standard deviation; solid line is the median of the posterior distribution of the number of bison in Yellowstone and shaded 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 uncertainty over time.

4 Participants

We have assembled a diverse working group, with gender and career stage diversity (Table 2). Of the 12 listed participants, four are women and five are early career scientists. The proposed Powell Center Fellow is Andrew Tredennick. PI Tredennick has expertise in Bayesian hierarchical modeling, data management, and forecasting.

Table 2: List of participants.

Name	Affiliation	Expertise	Associated Data set
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	Sea otter
Peter Adler*	Utah State University	population ecology/modeling, data synthesis	Idaho sagebrush
Lauren Buckley*	University of Washington	ecological forecasting climate change	Grasshopper spp.
Michael Dietze*	Boston University	ecological forecasting, partitioning uncertainty	n/a

Name	Affiliation	Expertise	Associated Data set
George Esslinger*	U.S. Geological Survey Alaska Science Center	population monitoring, GIS analysis	Sea otter
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	Mt. St. Helens 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
Perry Williams*†	Colorado State University	spatiotemporal modeling, population forecasts	Sea otter

*Confirmed participant
†Local (Ft. Collins) 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 example. 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 Skype meetings with each analysis team (as many as possible to complete analyses). 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 database and identify any outstanding issues. Evaluate manuscript drafts and form writing teams. Create a GitHub repository for forecasts.

February to June 2019 Finalize database 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. In addition to peer-reviewed papers, write a blog post for Dynamic Ecology[1] on the limits to ecological forecasting to reach an even broader audience. The blog post will introduce the database and forecast repository, and pose a forecasting challenge for non-working group members to apply different methods in an effort to "beat" our forecast skill.

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 that research associated with this project will catalyze a new area of research on ecological forecasting. The working group participants will gain experience using cutting-edge quantitative techniques. The assembled database of plant and animal abundance time series, coupled with environmental covariates, will be a valuable resource for teaching and research, and the repository of population forecasts will provide new opportunities to validate, and improve, forecasts. No repository of ecological forecasts currently exists, but one of our participants (M. Dietze) plans to develop one as part of the NSF-funded *Near-term Ecological Forecasting Initiative*. Our working group and his project will work synergistically to create a repository. In addition, we expect to produce several publications, including:

1. A synthesis paper testing our hypotheses using all the data sets aimed at *PNAS* or *Ecology Letters*.
2. A forum-style paper on the limits to ecological forecasting, how to overcome them, and the usefulness of even uncertain forecasts for ecosystem management for *Frontiers in Ecology and the Environment*.
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*.

References

Bauer, P., Thorpe, A. & Brunet, G. (2015). The quiet revolution of numerical weather prediction. *Nature*, 525, 47–55.

Botero, C.A., Weissing, F.J., Wright, J. & Rubenstein, D.R. (2015). Evolutionary tipping points in the capacity to adapt to environmental change. *Proceedings of the National Academy of Sciences*, 112, 184–189.

Clark, J.S., Carpenter, S.R., Barber, M., Collins, S., Dobson, A., Foley, J.A., et al. (2001). Ecological forecasts: an emerging imperative. *Science*, 293, 657–660.

del Moral, R. (2010). Thirty years of permanent vegetation plots, Mount St. Helens, Washington, USA. *Ecology*, 91, 2185.

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

Gremer, J.R. & Venable, D.L. (2014). Bet hedging in desert winter annual plants: Optimal germination strategies in a variable environment. *Ecology Letters*, 17, 380–387.

Hobbs, N.T. & Hooten, M.B. (2015). *Bayesian Models: A Statistical Primer for Ecologists*.

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

Princeton University Press, Princeton.

Hobbs, N.T., Geremia, C., Treanor, J., Wallen, R., White, P.J., Hooten, M.B., et al. (2015). State-space modeling to support management of brucellosis in the Yellowstone bison population. *Ecological Monographs*, 85, 525–556.

Houlahan, J.E., Mckinney, S.T., Anderson, T.M. & Mcgill, B.J. (2017). The priority of prediction in ecological understanding. *Oikos*, 126, 1–7.

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.

Steffen, W., Richardson, K., Rockström, J., Cornell, S., Fetzer, I., Bennett, E., et al. (2015). Planetary boundaries: Guiding human development on a changing planet. *Science*, 348, 1217.

Tredennick, A.T., Hooten, M.B. & Adler, P.B. (2016a). Do we need demographic data to forecast plant population dynamics? *Methods in Ecology and Evolution*, n/a–n/a.

Tredennick, A.T., Hooten, M.B., Aldridge, C.L., Homer, C.G., Kleinhesselink, A.R. & Adler, P.B. (2016b). Forecasting climate change impacts on plant populations over large spatial extents. *Ecosphere*, 7, e01525.

Williams, P.J., Hooten, M.B., Womble, J.N., Esslinger, G.G., Bower, M.R. & Hefley, T.J. (2017). An integrated data model to estimate spatio-temporal occupancy, abundance, and colonization dynamics. *Ecology*, 0, 1–9.