Date: January 10, 2018

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: January 2019 to December 2020, with two 4-day meetings at the Powell Center

Proposed Data Release Date: January 2021

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

Is this a resubmission? Yes

Conflicts of Interest with Reviewers: None

Keywords: climate and land use change; ecosystems

Problem Statement

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, Dietze et al. n.d.). 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, the Earth's weather is a chaotic system, meaning its dynamics are internally unstable and sensitive to initial conditions uncertainty. This is why billions of dollars are spent each year to measure meterological variables – meterologists learned that the key to reducing forecast error ($Var[y_{t+1}]$) was to reduce the uncertainty of initial conditions ($Var[y_t]$).

In contrast, ecologists are attempting to make actionable forecasts with little knowledge of which term in Eq. 2 dominates forecast error. Knowing which term dominates forecast error in different ecological settings will advance our fundamental understanding of the natural world and immediately impact practical efforts to monitor, model, and predict ecological dynamics. While it is impossible to partition forecast uncertainty for every species, links among the dominate source of forecast error and species' ecologies might serve as prior information that can guide initial monitoring and modeling schemes.

Thus, 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). Doing so will (1) provide actionable information on new ways to improve ongoing ecological forecasts, (2) bridge ecological theory and ecological forecasts, and (3) quantify the current limits to ecological forecasts.

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.

Testing our hypotheses requires applying a standardized analytical approach to a synthetic data set representative of many species – an activity that can only be done at the Powell Center as a working group. These three hypotheses are just a starting point, as the working group will generate additional hypotheses.

Proposed Activities

PIs Tredennick, Hooten, and Adler have led several efforts to forecast the response of plant populations to climate change (Tredennick et al. 2016, 2017). 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 (to be recruited) to lead all aspects of our proposed work.

Data Synthesis

Time Series Selection

We have identified 12 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. 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, 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.

All abundance data sets will be combined into a single database that can be queried using standard SQL software. The abundance data will be linked to a second database of environmental covariates. We will extract climate data as needed from gridded data products (e.g., PRISM). We will work with study area experts (Table 2) to decide on appropriate climate covariates for each time series. This is a preliminary data processing workflow; we will work with Powell Center experts to synthesize our data sets.

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 (weekly census)	http://ghopclimate.colorado.edu
Animal	Antarctic penguin spp.*	38	http://www.penguinmap.com/
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.

Analysis

Dynamic Forecasting Models

Using the assembled time series, we will fit Bayesian hierarchical state-space models, which 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 bison in Yellowston) is unknowable, or *latent*. A typical state-space model takes the form:

$$y_t \sim \text{Normal}(z_t, \sigma_\text{o}^2) \tag{3}$$

$$z_t \sim \text{Normal}(\mu_t, \sigma_\text{p}^2) \tag{4}$$

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

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 **x** is a vector of environmental covariates. The two error terms represent observation error (σ_o^2) and process error (σ_p^2). 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 the distribution of the data model will vary depending on the particular data generating process (Hobbs and Hooten 2015). Using different distributions for the data and process models also helps make variance parameters identifiable when replicate observations are not available (Hobbs and Hooten 2015), thus avoiding a potential pitfall of state-space models (Auger-Méthé et al. 2016)

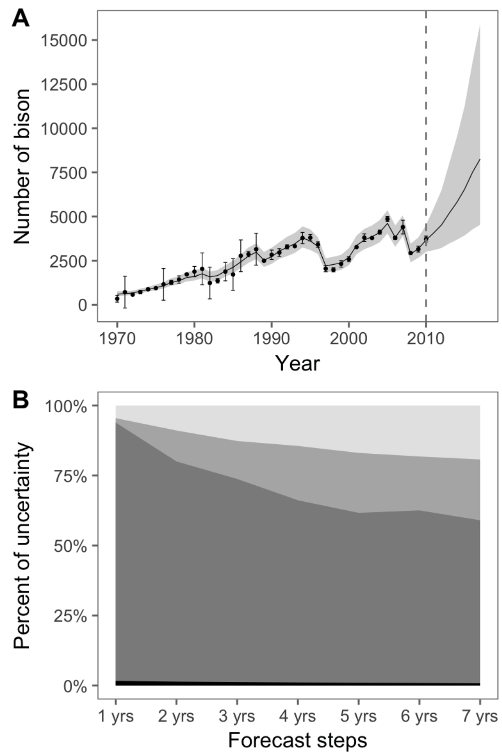

Figure 1: An example state space model fit and forecast (A) and example of forecast uncertainty partitioning (B). In (A), points are observed counts and errorbars show the observed standard deviation; solid line is the median of the posterior distribution of the number of bison in Yellowstone and shaded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

An Example: Yellowstone Bison

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 7 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; (*iii*) initial condition, parameter, and driver uncertainty; and (*iv*) initial condition, parameter, driver, 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 driver and 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.

Participants

We have assembled a diverse working group, with gender and career stage diversity (Table 2). Of the 14 listed participants, five are women and six are early career scientists.

Table 2: List of participants.

Name	Affiliation	Expertise	Associated Data set
Andrew Tredennick*[1,2]	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
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
Heather Lynch*	Stony Brook University	population modeling, population forecasting	Antarctic penguins
Ethan White*	University of Florida	ecological forecasting, data synthesis	BBS & Portal Data
Perry Williams*†	Colorado State University	spatiotemporal modeling, population forecasts	Sea otter
Postdoctoral Fellow	TBD	population ecology, population forecasts	n/a

*Confirmed participant
†Local (Ft. Collins) participant
[1]Technical liaison to Powell Center computing staff
[2]Party responsible for adherence to Powell Center Data and Information Policy

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.

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. Define manuscripts.

February to December 2018 Bi-monthly PI Skype meetings. Continue and finalize data set aquisition 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. PC Fellow Skypes with each analysis team as needed to complete analyses. Write first drafts of manuscripts.

January 2019 *Second meeting for four days at Powell Center*: Analysis teams present results for feedback. 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.

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*.

Literature Cited

Auger-Méthé, M., C. Field, C. M. Albertsen, A. E. Derocher, M. A. Lewis, I. D. Jonsen, and J. M. Flemming. 2016. State-space models' dirty little secrets: Even simple linear Gaussian models can have estimation problems. Scientific Reports 6.

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

Botero, C. A., F. J. Weissing, J. Wright, and D. R. Rubenstein. 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., 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 293:657–660.

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

Dietze, M. C. 2017. Prediction in ecology: A first-principles framework. Ecological Applications 27:2048–2060.

Dietze, M., A. Fox, L. Beck-Johnson, J. Betancourt, M. Hooten, C. Jarnevitch, T. Kiett, M. Kenney, C. Laney, L. Larsen, L. Loescher, C. Lunch, P. Pijanowski, J. Randerson, E. Reid, A. T. Tredennick, R. Vargas, K. Weathers, and E. P. White. (n.d.). Iterative near-term ecological forecasting: Needs, opportunities, and challenges. Proceedings of the National Academy of Sciences.

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

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

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

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 348:1217.

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

Tredennick, A. T., M. B. Hooten, and P. B. Adler. 2017. Do we need demographic data to forecast plant population dynamics? Methods in Ecology and Evolution 8:541–551.

Williams, P. J., M. B. Hooten, J. N. Womble, G. G. Esslinger, M. R. Bower, and T. J. Hefley. 2017. An integrated data model to estimate spatiotemporal occupancy, abundance, and colonization dynamics. Ecology 98:328–336.