A simulation-based approach for quantifying and partitioning uncertainty to improve forecasts of dynamic processes

Andrew T. Tredennick[1,2*]

[1] Odum School of Ecology, University of Georgia, Athens, GA, United States

[2] Center for the Ecology of Infectious Diseases, University of Georgia, Athens, GA, United States

Abstract

Making informed decisions in the face of rapid environmental change requires forecasts from models of ecological processes. However, forecasts from ecological models are often associated with high degrees of uncertainty, making it difficult for such forecasts to inform decision-making processes. To make progress toward the goal of reliable and informative ecological forecasts, we need to know from where forecast uncertainty arises. Such knowledge can guide investment in future research that will most improve forecast skill. There is a rich history of analytical expressions that partition the variance of future dynamics, but these expressions suffer from necessary assumptions such as linear dynamics and small-variance approximations that exclude interactions. Similarly, the earth systems modeling community has developed methods for paritioning uncertainty of model projections, but these operate at a different modeling grain than most of ecology. Building on these approaches, we develop a simulation-based approach for

*Corresponding author, Email: atredenn@gmail.com

quantifying and partitioning forecast uncertainty from Bayesian state-space models that overcomes the limitations of previous analytical approaches. Our approach is similar to an Analysis of Variance, where the total variance of a forecast is paritioned among its constituent parts, namely initial conditions uncertainty, parameter uncertainty, driver uncertainty, process error, and their interactions. We apply the approach to near-term forecasts of the Yellowstone bison population and measles in Niger, demonstrating the broad utility of our approach. We also provide functions written in the statistical programming language R, which will allow others using Bayesian state-space models to employ our approach in their own research.

Keywords: Bayesian state-space model, forecast, Markov chain Monte Carlo, measles, prediction, population model, uncertainty, Yellowstone bison

Introduction

Ecology is entering an era of prediction. This new era is made possible by continuous data streams (e.g., the National Ecological Observatory Network, citizen science data, remote sensing), increased statistical literacy among ecologists, and the need to provide actionable information for decision makers (Dietze et al. 2018). Thus, we are on the verge of answering Clark et al.'s (2001) challenge to make forecasting a core goal of ecology. 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 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 an essential component of this learning process (Bauer et al. 2015). Therefore, we need a systematic

and robust way to quantify and partition forecast uncertainty into its constituent parts (Dietze 2017).

Various approaches have been used to characterize and partition forecast uncertainty (Sobol' 1993, Cariboni et al. 2007). For example, consider a dynamic 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:

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

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). Ignoring covariance among factors and assuming linear dynamics, Dietze (2017), following Sobol' (1993) and Cariboni et al. (2007), suggests that forecast variance ($Var[y_{t+1}]$) is approximately:

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

where each additive term follows a pattern of *sensitivity* times *variance* and "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 atmosphere 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 having an analytical expression such as Eq. 2 is satisfying, arriving at the expression involves strict assumptions. First, Eq. 2 only holds when the underlying dynamics are linear, which may not be the case for many populations and models. Second, Eq. 2 only includes additive effects of each factor because the Taylor series decomposition requires small-variance approximations that eliminate interactions. But, interactions among the factors are probably common. For example, in a simple simulation of an AR(1) process, we show that initial conditions uncertainty and parameter error interact to generate the full spread of forecast variance (Figure 1). Analyzing only the main effects of each source of uncertainty would lead to the false conclusion that initial conditions uncertainty is not important. Progress in quantifying and partioning forecast uncertainty therefore requires a more flexible approach than that provided by Eq. 2.

We have four objectives: First, we introduce a canonical equation for error propagation and review approaches for quantifying sources of forecast uncertainty from other fields. Second, we describe a general, simulation-based method for quantifying and partitioning forecast uncertainty from Bayesian state-space models. Third, we apply our method to near-term forecasts of the Yellowstone bison population and measlse cases in China. These two applications demonstrate the generality of our approach. Fourth, through our applications we introduce functions written in the statistical programming language R ({R Core Team} 2016), which can be used to implement our method for many Bayesian state-space models.

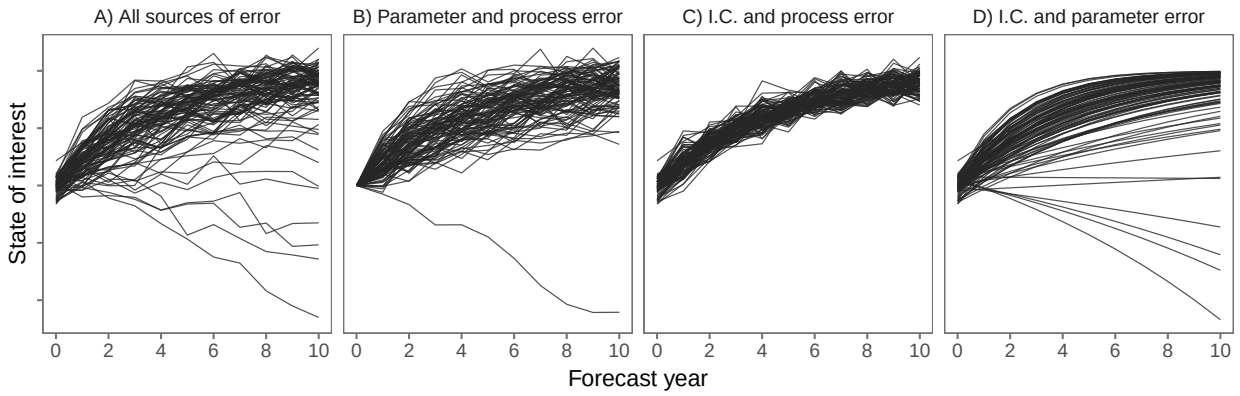

Figure 1: Example of forecast uncertainty with different sources of error set to zero. Each line represents one realization, out of 200, from an order-one autoregressive model (AR(1) process). Contrary to the analytical expression (Eq. 2), initial conditions (I.C.) uncertainty and parameter uncertainty clearly interact. The spread of lines in (A) is not wholly because of initial conditions uncertainty or parameter uncertainty (compare panels B and C). It is their combined influence that causes the spread of realizations in (A and D). At least in this example, process error (set to zero in D) does appear to be independent, but we used a small value of process error to highlight other interactions. Source code: `generate_forecast_fxns.R`.

A Brief History of Quantifying and Partitioning Forecast Uncertainty

Error propagation

Methods for quantifying and partitioning forecast uncertainty all share common roots in statistical error propagation. Error propagation is concerned with translating the effects of variable (or parameter) uncertainty into the uncertainty of a function based those variables. That is, for the output q of some function $q = f(x_1, x_2, \ldots, x_n)$, we are interested in the variance of q given the variance of input variables ($\mathbf{x}$). It is this generic model formulation that leads to the canonical expression of error propagation

$$\sigma_q^2 = \left(\frac{\delta q}{\delta x_1}\sigma_{x_1}\right)^2 + \left(\frac{\delta q}{\delta x_2}\sigma_{x_2}\right)^2 + \cdots + \left(\frac{\delta q}{\delta x_n}\sigma_{x_n}\right)^2 \quad (3)$$

$$= \sum_{i=1}^{n}\left(\frac{\delta q}{\delta x_i}\sigma_{x_i}\right)^2. \quad (4)$$

This expression states that the variance of the output of function q (σ_q^2) is equal to the sum of squares of the input variables weighted by the sensitivity of the output variable to each input variable, quantified as the partial derivative. *Add more here on interactions, which require knowing the correlations among x's.*

Dynamic theoretical models

Lotka-Volterra style models and variance of *t+1* (draws from error propagation and sensitivity analysis).

Weather forecasting

Chaos and initial conditions.

Earth system models

Carbon and uncertainty from models and scenarios.

A Simulation-Based Approach for Partitioning Forecast Uncertainty

Analytical expressions of forecast uncertainty must rely on simplifying assumptions. Two important assumptions are (1) that different sources of uncertainty do not interact and (2) that the system of equations is linear. These analytical expressions are important for guiding our intuition, but these strict assumptions limit our ability to partition forecast uncertainty in practice. Thus, we present a simulation approach that is entirely model-based and requires no assumptions, other than those embedded in the model itself. We are building on the ideas put forth by Dietze (2017), who suggested a simulation approach for quantifying the terms in Eq. 2. Here we generalize the approach to fully partition forecast uncertainty among the main effects of different sources and their interactions.

As a starting point, consider the Bayesian state-space model

$$
\begin{aligned}
\textbf{Data Model:} \quad & y_t \sim \left[y_t \mid z_t, \sigma_{\mathrm{o}}^2\right], && t = 1, \ldots, T, \\
\textbf{Process Models:} \quad & z_t \sim \left[z_t \mid \mu_t, \sigma_{\mathrm{p}}^2\right], && \\
& \mu_t = g\left(z_{t-1}, \mathbf{x}_t', \theta\right), && t = 2, \ldots, T, \\
\textbf{Parameter Models:} \quad & \boldsymbol{\phi} \sim \left[\theta, \sigma_{\mathrm{p}}^2, \sigma_{\mathrm{o}}^2, z_{t=1}\right], && (5)
\end{aligned}
$$

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 z at time t from the process model g, which is a function of z at time *t-1*, a vector of covariates ($\mathbf{x}$) at time t, and a set of unknown parameters, θ (Berliner 1996). σ_{o}^2 is observation error and σ_{p}^2 is process error. The notation $[a \mid b, c]$ reads, "the probability of a given b and c" (Gelfand and Smith 1990), and $\boldsymbol{\phi}$ refers to the prior probability distributions for all unknown parameters and the initial conditions for the latent state, $z_{t=1}$.

For our purposes, we are interested in the probability distributions of the true state $\mathbf{z}$ at

future points in time, conditional on previous observations ($\mathbf{y}$). This is referred to as the forecast distribution or the predictive process distribution (Hobbs and Hooten 2015 pp. 199–200), which, for one time step ahead of the final observation ($T+1$), is defined as

$$\begin{aligned}[z_{T+1}|y_1,\ldots,y_T] = \int\int\cdots\int \left[z_{T+1}|z_T,\mathbf{x}_T,\theta,\sigma_{\mathrm{p}}^2\right] \\ \times \left[z_1,\ldots,z_{T+1},\theta,\sigma_{\mathrm{p}}^2|y_1,\ldots,y_T\right] d\theta d\sigma_{\mathrm{p}}^2 dz_1\ldots dz_T.\end{aligned} \tag{6}$$

The model in Eq. 5 can be fit using Markov chain Monte Carlo (MCMC) algorithms, which makes calculating the forecast distribution a relatively simple task. To obtain $[z_{T+1}|y_1,\ldots,y_T]$, we can change the indexing of t in Eq. 5 to $t = 2,\ldots,T+1$ and then sample $z_{T+1}^{(k)}$ from $\left[z_{T+1}|g(z_T^{(k)},\mathbf{x}_{T+1}^{(j(k))},\theta^{(k)}),\sigma_{\mathrm{p}}^{2(k)}\right]$ given the current values for $\theta^{(k)}$ and $z_T^{(k)}$ on each $k = 1,\ldots,K$ iteration of the MCMC algorithm (Hobbs and Hooten 2015, Williams et al. 2018). Note that we index the external covariate vector $\mathbf{x}$ using *j* and *k*, where *j(k)* is realization *j* of the covariate vector $\mathbf{x}$ associated with MCMC sample *k*. In some cases, the external covariate will have only one value, in which case all *K* MCMC samples will share the same $\mathbf{x}$. In other cases, their may be a distribution of external covariate values associated with uncertainty from the forecast of $\mathbf{x}$, resulting in *n* values for each x_{T+1}. When $n < K$, which we anticipate it often will be, then $\mathbf{x}$ can be sampled with replacement and a value can be assigned to each MCMC sample. Making forecasts further into the future than $T+1$ requires extending $T+1$ to $T+2,\ldots,T+q$ and iteratively sampling $\left[z_{T+q}|g(z_{T+q-1}^{(k)},\mathbf{x}_{T+q}^{(j(k))},\theta^{(k)}),\sigma_{\mathrm{p}}^{2(k)}\right]$ (Hobbs and Hooten 2015).

The forecast distribution (Eq. 6) has all of the quantities that contribute to forecast uncertainty by incorporating their uncertainty explicitly across the *K* MCMC iterations. For example, initial conditions uncertainty is incorporated because $z_{T+1}^{(k)}$ is a function of $z_T^{(k)}$, resulting in a total of *K* point forecasts for *z* that comprise the posterior distribution of *z* at each time *t*. If we wish to ignore initial conditions uncertainty (I.C. uncertainty), we can make all *K* point forecasts starting from the mean of z_T

Table 1: Sampling equations for generating forecast distributions at times $T+q$ (where T is the time of the last observation) across $k = 1, \ldots, K$ MCMC samples with only certain sources of uncertainty present.

Source of Uncertainty	**Notation**	**Sampling Equation**	
I.C. Uncertainty	$V^{(I)} = V^{(I,\overline{PA},\overline{D},\overline{PS})}$	$\left[z_{T+q} \mid g(z_{T+q-1}^{(k)}, \mathbf{x}_T^{(*)}, \theta^{(*)}), 0\right]$	
Parameter Uncertainty	$V^{(PA)} = V^{(\overline{I},PA,\overline{D},\overline{PS})}$	$\left[z_{T+q} \mid g(z_{T+q-1}^{(k)}, \mathbf{x}_T^{(*)}, \theta^{(k)}), 0\right],$	$q > 1$
		$\left[z_{T+q} \mid g(z_T^{(*)}, \mathbf{x}_T^{(*)}, \theta^{(k)}), 0\right],$	$q = 1$
Driver Uncertainty	$V^{(D)} = V^{(\overline{I},\overline{PA},D,\overline{PS})}$	$\left[z_{T+q} \mid g(z_{T+q-1}^{(k)}, \mathbf{x}_T^{(j(k))}, \theta^{(*)}), 0\right],$	$q > 1$
		$\left[z_{T+q} \mid g(z_T^{(*)}, \mathbf{x}_T^{(j(k))}, \theta^{(*)}), 0\right],$	$q = 1$
Process Uncertainty	$V^{(PS)} = V^{(\overline{I},\overline{PA},\overline{D},PS)}$	$\left[z_{T+q} \mid g(z_{T+q-1}^{(k)}, \mathbf{x}_T^{(*)}, \theta^{(*)}), \sigma_{\mathrm{p}}^{2(k)}\right],$	$q > 1$
		$\left[z_{T+q} \mid g(z_T^{(*)}, \mathbf{x}_T^{(*)}, \theta^{(*)}), \sigma_{\mathrm{p}}^{2(k)}\right],$	$q = 1$

Note: The notation $V^{(A,\overline{B},\overline{C})}$ means that uncertainty from A enters the forecast from its posterior distribution (as approximated from MCMC samples), while sources B and C are set to their means (as calculated across the MCMC samples). In the main text we leave out the averaged terms (those with overbars) to reduce clutter. Thus, $V^{(A)} = V^{(A,\overline{B},\overline{C})}$.

$$z_T^{(*)} = E(z_T | y_1, \ldots, y_T) \approx \frac{\sum_{k=1}^{K} z_T^{(k)}}{K}, \tag{7}$$

which we call $z_T^{(*)}$ (as opposed to $z_T^{(k)}$), while retaining the uncertainty for all other parameters. Our iterative sampling to obtain the forecast distribution then becomes a conditional statement,

$$z_{T+q} \sim \begin{cases} \left[z_{T+q} | g(z_{T+q-1}^{(k)}, \mathbf{x}_T^{(j(k))}, \theta^{(k)}), \sigma_{\mathrm{p}}^{2(k)}\right], & q > 1 \\ \left[z_{T+q} | g(z_T^{(*)}, \mathbf{x}_T^{(j(k))}, \theta^{(k)}), \sigma_{\mathrm{p}}^{2(k)}\right], & q = 1. \end{cases} \tag{8}$$

We can extend the basic idea of setting subsets of parameters and states to their posterior means (or medians, depending on the distribution) to make partitioned forecasts where only prescribed sources of uncertainty contribute to forecast uncertainty (Table 1).

It is important to note that although our discussion has centered on obtaining forecast distributions within the MCMC algorithm used to fit the model, it is only feasible to do this for the full forecast distribution (Eq. 6). In all other cases, where states or parameters must be averaged over the K MCMC iterations, forecast simulations must be done *post hoc* using saved MCMC samples (Box 1). In other words, estimating $z_1, z_2, \ldots, z_T$ is done within the MCMC fitting algorithm, while estimating $z_{T+1}, z_{T+2}, \ldots, z_{T+q}$ is done after fitting the model, but with the full MCMC output.

With these basics in mind, we now develop our approach for partitioning and quantifying uncertainty, which is similar to an Analysis of Variance (ANOVA). Let the variance of the forecast distribution at time $T+q$ be $V_{T+q}^{(X)}$, where $X = F$ (full forecast distribution), I (initial conditions uncertainty only), PA (parameter uncertainty only), D (driver uncertainty), or PS (process uncertainty only) (Table 1). $V_{T+q}^{(I)}$, $V_{T+q}^{(PA)}$, $V_{T+q}^{(D)}$, and $V_{T+q}^{(PS)}$ are the main effects of each factor on the forecast distribution such that

$$\begin{aligned} V_{T+q}^{(F)} = {} & V_{T+q}^{(I)} + V_{T+q}^{(PA)} + V_{T+q}^{(D)} + V_{T+q}^{(PS)} \\ & + \varepsilon_{T+q}^{(I,PA)} + \varepsilon_{T+q}^{(I,D)} + \varepsilon_{T+q}^{(I,PS)} + \varepsilon_{T+q}^{(PA,PS)} + \varepsilon_{T+q}^{(PA,D)} + \varepsilon_{T+q}^{(D,PS)} \\ & + \varepsilon_{T+q}^{(I,PA,D)} + \varepsilon_{T+q}^{(I,PA,PS)} + \varepsilon_{T+q}^{(I,D,PS)} + \varepsilon_{T+q}^{(PA,D,PS)} \\ & + \varepsilon_{T+q}^{(I,PA,D,PS)}, \end{aligned} \tag{9}$$

where the notation $\varepsilon_{T+q}^{(X,Y)}$ represents the remaining interactive effect of X and Y on $V_{T+q}^{(F)}$ after accounting for their main effects. For example, if the full forecast variance is a function of only initial conditions uncertainty I and parameter uncertainty PA, then

$$V_{T+q}^{(F)} = V_{T+q}^{(I)} + V_{T+q}^{(PA)} + \varepsilon_{T+q}^{(I,PA)}, \tag{10}$$

which rearranges to

$$\varepsilon_{T+q}^{(I,PA)} = V_{T+q}^{(F)} - \left[V_{T+q}^{(I)} + V_{T+q}^{(PA)}\right]. \tag{11}$$

We show an example of applying Eqs. 10-11 in a hypothetical situation where forecast uncertainty is determined by initial conditions uncertainty and parameter uncertainty alone in Figure 2.

The necessary terms for partitioning forecast variance can be obtained by calculating the variance of the partitioned forecast distributions (equations in Table 1 and combinations thereof). To take this one step further, and to reiterate the core idea, let $V(F)$ be a function of initial conditions uncertainty I, parameter uncertainty PA, and driver uncertainty D. We can then write the equation for forecast uncertainty and the derived interaction effects as

There must be a way to write a concise form of these equations vai summations.

$$V_{T+q}^{(F)} = V_{T+q}^{(I)} + V_{T+q}^{(PA)} + V_{T+q}^{(D)} + \varepsilon_{T+q}^{(I,PA)} + \varepsilon_{T+q}^{(I,D)} + \varepsilon_{T+q}^{(D,PA)} + \varepsilon_{T+q}^{(I,PA,D)}, \qquad \text{where} \tag{12}$$

$$\varepsilon_{T+q}^{(I,PA)} = V_{T+q}^{(I+PA)} - \left[V_{T+q}^{(I)} + V_{T+q}^{(PA)}\right] \tag{13}$$

$$\varepsilon_{T+q}^{(I,D)} = V_{T+q}^{(I+D)} - \left[V_{T+q}^{(I)} + V_{T+q}^{(D)}\right] \tag{14}$$

$$\varepsilon_{T+q}^{(PA,D)} = V_{T+q}^{(PA+D)} - \left[V_{T+q}^{(PA)} + V_{T+q}^{(D)}\right] \tag{15}$$

$$\varepsilon_{T+q}^{(I,PA,D)} = V_{T+q}^{(F)} - \left[V_{T+q}^{(I)} + V_{T+q}^{(PA)} + V_{T+q}^{(D)}\right] - \left[\varepsilon_{T+q}^{(I,PA)} + \varepsilon_{T+q}^{(I,D)} + \varepsilon_{T+q}^{(PA,D)}\right], \tag{16}$$

where the notation $V_{T+q}^{(A+B)}$ is the forecast variance under scenario where both A and B are allowed to contribute to forecast uncertainty (i.e., a combination of the equations in Table 1). We present the equations for the full partition among I, PA, D, and PS in Appendix 1. In the next two sections, we apply our approach (Box 1) to quantify and partition the uncertainty of near-term forecasts of the Yellowstone bison population and measles cases in China.

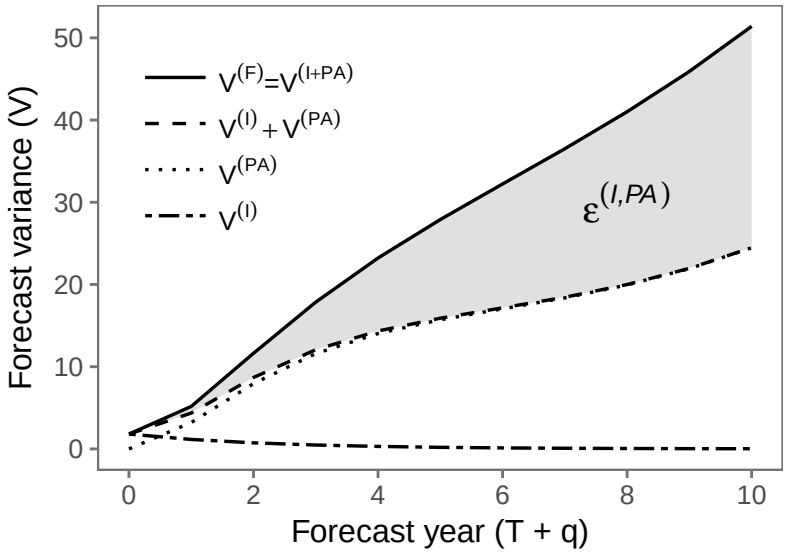

Figure 2: Graphical example of partitioning forecast uncertainty into main effects (V) and interaction effects (ε). The grey shaded area shows the interaction effect ($\varepsilon^{(I,PA)}$) that must be accounted for to fully partition forecast uncertainty between initial conditions uncertainty ($V^{(I)}$) and parameter uncertainty ($V^{(PA)}$). Source code: `generate_forecast_fxns.R`.

Box 1. Pseudocode for quantifying and partitioning forecast uncertainty from a Bayesian state-space model.

1. Fit a Bayesian state-space model (i.e., Eq. 5) with data $y_1, \ldots, y_T$ and save the MCMC samples.
2. Forecast $z_{T+q}^{(k)}$ for all $k = 1, \ldots, K$ MCMC samples to generate the full forecast distribution following Eq. 6 (this can be done within the MCMC algorithm or *post hoc* with saved MCMC samples).
3. Forecast $z_{T+q}^{(k)}$ for all $k = 1, \ldots, K$ MCMC samples to generate the partitioned forecast distributions for each source of uncertainty, averaging quantities over the K MCMC samples as necessary (equations in Table 1 and combinations thereof).
4. For each forecast time q, calculate the variance of each forecast distribution from steps 2-3.
5. Partition forecast variance using Eq. X.

Box 1:

Application: Yellowstone Bison Population

Data

Aerial counts of the Yellowstone bison population from 1970-2017 (Fig. 3A) were used to estimate model parameters and states. Counts typically occurred four times a year. We used summer counts only because the summer aggregation of bison tends to produce more accurate counts (Hobbs et al. 2015). Replicate counts were taken in 41 of the 48 years, and from these replicates we estimated the mean annual total population in the park and the annual standard deviation of counts (i.e., observation error). We also include known harvest counts in our population model (Fig. 3A), as described below.

We used cumulative January precipitation as an environmental covariate in our population model because winter snow limits the ability of bison to access forage (Fig. 3B). We could have used snow water equivalent observations from SNOTEL sites, but future projections of snow pack are much more difficult to generate and obtain the projections of precipitation.

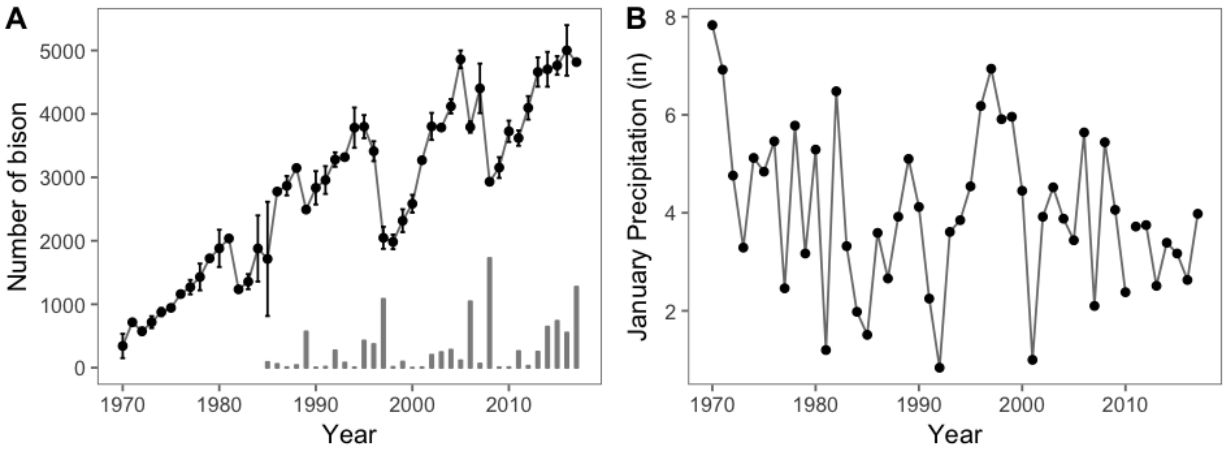

Figure 3: Time series of (A) annual summer counts of Yellowstone bison (mean $\pm$ 1 S.D.) and (B) accumulated annual snow water equivalent from the West Yellowston SNOTEL station. These data were used to estimate parameters and states in our model. Blue points are training data, orange points are validation data for which we made forecasts.

Yellowston bison: state-space model

We used a Bayesian state-space model to obtain posterior distributions of all unknown parameters, latent states, and forecasts of future states. The fully specified model, including a model of the ecological process, a model linking the process to the data, and parameter models, takes the form:

$$\left[\boldsymbol{\theta}_{\mathrm{p}}, \kappa, z_{(t)} | y_{(t)}, x_{(t)}\right] \propto \prod_{t=2}^{48} \underbrace{\left[z_{(t)} | \boldsymbol{\theta}_{\mathrm{p}}, z_{(t-1)}, x_{(t)}\right]}_{\text{process}} \prod_{t=1}^{41} \underbrace{\left[y_{(t)} | \kappa, z_{(t)}\right]}_{\text{data}} \underbrace{\left[\boldsymbol{\theta}_{\mathrm{p}}, \kappa, z_{(t=1)}\right]}_{\text{parameters}}, \tag{17}$$

where $\boldsymbol{\theta}_{\mathrm{p}}$ is a vector of parameters in the process model, $z_{(t)}$ is the latent, or unobservable and true, state of the bison population abundance at time t, $y_{(t)}$ is the observed state of bison population abundance at time t, $x_{(t)}$ is the standardized value of cumulative January precipitation at time t, and κ is a variance term associated with the model likelihood (see below). Note that the product associated with the "process model" applies over seven more years than the product associated with the "data model." The extra seven years are forecasts of the latent state seven years into the future, for which no likelihood can be calculated.

Our process model represents the population dynamics of the Yellowstone bison using the stochastic model

$$\mu_{(t)} = \log(z_{(t-1)}) + e_{(t-1)} + r + b_0 \left(\log(z_{(t-1)}) + e_{(t-1)}\right) + b_1 x_{(t)}, \tag{18}$$

$$\log(z_{(t)}) \sim \text{Normal}\left(\mu_t, \sigma_{\mathrm{p}}^2\right), \tag{19}$$

where the determinstic model for $\mu_{(t)}$ is a Gompertz model of population growth that predicts the mean of $z_{(t)}$ on the log scale as function of the true state of the log population at time $t-1$ ($z_{(t-1)}$), the intrinsic, per capita rate of increase (r), the strength of density-dependence (b_0), the effect (b_1) of January precipitation at time t ($x_{(t)}$), and $e_{(t)}$ is the logarithmic integration of extractions occuring between observations in year $t-1$ and year t. The quantity σ_{p}^2 is the process variance on the log scale, which accounts for all the drivers of the true state that are not found in the determinstic model.

We used a negative binomial likelihood to link the observations ($\mathbf{y}$) to the estimated latent states ($\mathbf{z}$). The likelihood is $y_{(t)} \sim \text{NB}\left(z_{(t)}, \kappa\right)$. We used a negative binomial likelihood rather than a poisson because there was evidence of over-dispersion.

Our state-space model requires prior distributions for all parameters and for the initial condition of z. We used an informative prior for the intrinsic growth rate, r, based on the population growth (λ) reported by Hobbs et al. (2015). The Hobbs et al. (2015) model includes estimates of two population growth rates, one for populations infected with brucellosis and one for populations not infected with brucellosis. Although the population growth rate was depressed in brucellosis-infected populations, the posterior estimate of the difference between the two growth rates overlapped zero, indicating little support for differentiating among the two. We therefore chose to use the estimates for the brucellosis-free population: mean $= 1.11$ and s.d. $= 0.0241$. We converted λ, the population growth rate, to r, the per capita rate of increase, by log-transforming λ: $r = \log(\lambda) = 0.1$. To get the standard deviation on the same scale, we simulated 100,000 numbers from the distribution $\log(\text{Normal}(1.11, 0.0241))$ and calculated the standard deviation of those numbers, which equaled 0.02. Thus, our prior distribution for r was $\text{Normal}(0.1, 0.02)$. We defined the prior distribution of the initial condition $z_{(t=1)}$ as $z_{(t=1)} \sim \text{Normal}\left(y_{(t=1)}, \sigma^2_{\text{o},(t=1)}\right)$.

We chose all other prior distributions to be vague. However, no prior distribution is completely uninformative, so we made sure that our choice of priors did not have large impacts on posteriors by trying several choices of priors and their associated parameters. We then observed their effects on the posteriors (Hobbs and Hooten 2015), which were small. Our final chosen priors were: $1/\sigma^2_{\text{p}} = \tau_{\text{p}} \sim \text{Gamma}(0.01, 0.01)$ and $b_{\in(0,1)} \sim \text{Normal}(0, 1000)$.

We estimated the posterior distributions of parameters and states using Markov chain Monte Carlo (MCMC) methods implemented using JAGS 4.2.0 (Plummer) and the `rjags` package (Plummer2) to connect JAGS to R (R citation). Posterior samples were collected from three chains, each with unique initial values that were variable relative to the posterior distributions. We used 50,000 iterations to adapt the chains for optimal sampling, then accumulated 100,000 samples from

each chain after an initial burn-in period of 50,000 itareations. We checked for convergence and stationarity of the MCMC chains by visually inspecting the trace plots for each parameter and state, and by calculating multivariate scale reduction factors and the Gelman metric (citations) using the `coda` package in R.

Yellowston bison: results

The posterior distributions of model parameters show that density-dependence is evident but weak (β_0 in Fig. 4), accumulated snow water equivalent has a weak negative effect on bison population growth (β_1 in Fig. 4), and that the prior distribution of per capita growth rate entirely informed the posterior distribution (r in Fig. 4). The complete overlap of the prior and posterior distributions for r is not unexpected because we used a strong prior informed by a previous study using this same data (though with a different model) (Hobbs et al. 2015).

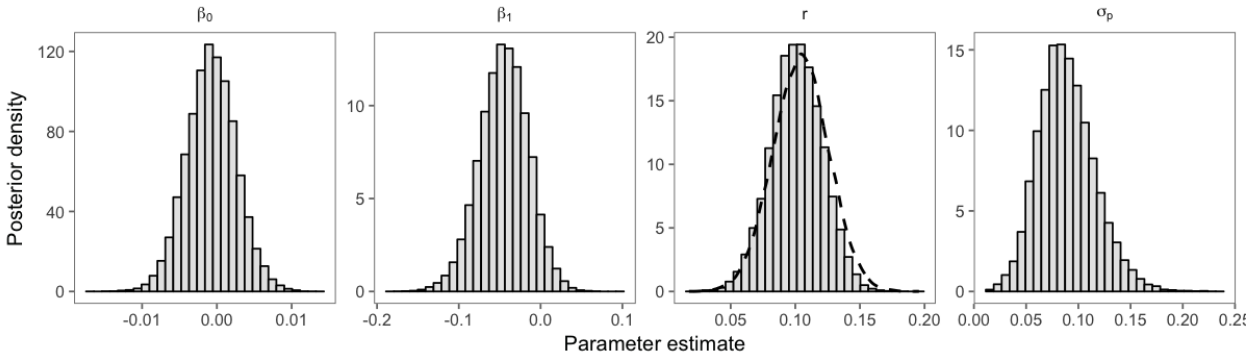

Figure 4: Posterior distributions of select model parameters: β_0 = density-dependence, β_1 = effect of January precipitation, r = per capita growth rate, σ_p = process error. The dashed line in the panel for r is the informed prior distribution. Prior distributions are not shown for other parameters because they are not visible on the scale of the posteriors.

Model forecasts are very uncertain and grew over time (Fig. 5A). Indeed, by 2017 forecasts ranged from ~2,000 bison to over 10,000 bison. Forecast uncertainty mostly comes from driver error, which accounts for about 70% of total forecast variance (Fig. 5B). Initial conditions error is consistently low across all years, as is parameter error. Process error contributes about 15-25% across all years.

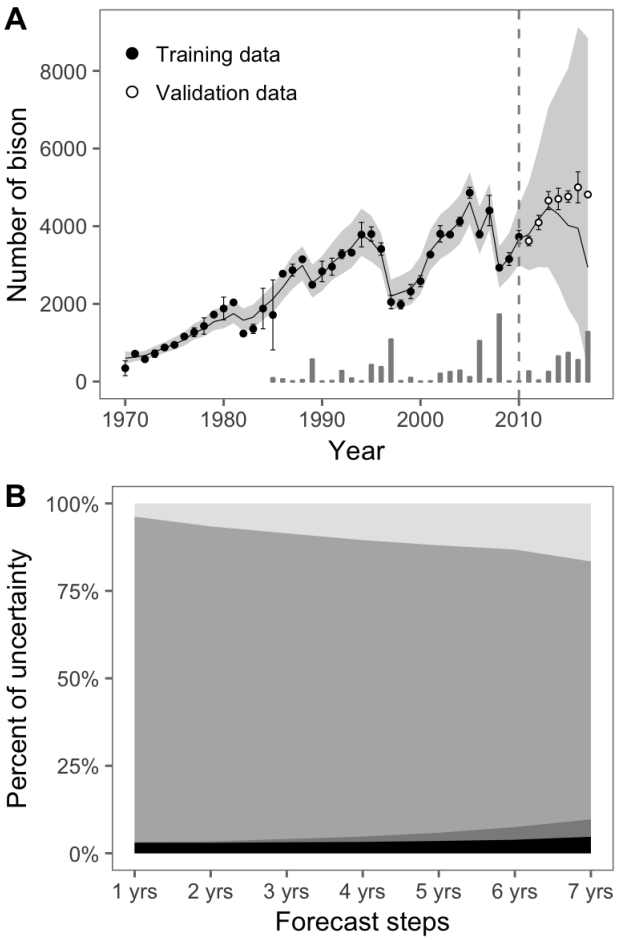

Figure 5: (A) Posterior predictions (before dashed vertical line) and forecasts (after dashed vertical line) of the Yellowstone bison population. Solid line is the median of the posterior predictive distribution and the shaded area is the 95% Bayesian credible interval. Forecasts shown here were made using known values of accumulated snow water equivalent in each year. (B) Partitioned forecast variance using the simulation approach algorithm in Box 1.

Application: Measles in Niger

Discussion

Acknowledgements

This research was funded by National Science Foundation grant DEB-1353078 (awarded to Peter B. Adler) and NIH MIDAS xxxxx (awarded to John M. Drake).

References

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

Berliner, L. M. 1996. Hierarchical Bayesian Time Series Models. Pages 15–22 *in* Maximum entropy and bayesian methods.

Cariboni, J., D. Gatelli, R. Liska, and A. Saltelli. 2007. The role of sensitivity analysis in ecological modelling. Ecological Modelling 203:167–182.

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.

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

Dietze, M. C., A. Fox, L. M. Beck-Johnson, J. L. Betancourt, M. B. Hooten, C. S. Jarnevich, T. H. Keitt, M. A. Kenney, C. M. Laney, L. G. Larsen, H. W. Loescher, C. K. Lunch, B. C. Pijanowski, J. T. Randerson, E. K. Read, A. T. Tredennick, R. Vargas, K. C. Weathers, and E. P. White. 2018. Iterative near-term ecological forecasting: Needs, opportunities, and challenges. Proceedings of the National Academy of Sciences 115:1424–1432.

Gelfand, A. E., and A. F. Smith. 1990. Sampling-based approaches to calculating marginal densities. Journal of the American Statistical Association 85:398–409.

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

Princeton University Press, Princeton.

{R Core Team}. 2016. R: A language and environment for statistical computing.

Sobol', I. 1993. Sensitivity Estimates for Nonlinear Mathematical Models.

Williams, P. J., M. B. Hooten, J. N. Womble, G. G. Esslinger, and M. R. Bower. 2018. Monitoring dynamic spatio-temporal ecological processes optimally. Ecology 99:524–535.