

# A simulation-based approach for quantifying and partitioning uncertainty to improve ecological forecasts

*Andrew T. Tredennick<sup>1</sup>, Peter B. Adler<sup>1</sup>, and others<sup>2</sup>*

<sup>1</sup> Department of Wildland Resources and the Ecology Center, Utah State University, Logan, UT,  
United States

<sup>2</sup> Department of somewhere

## **Abstract**

Making informed ecosystem management 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. Here we develop a simulation-based approach for 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 partitioned among its constituent parts, namely initial conditions uncertainty, parameter uncertainty, driver uncertainty, process error, and their interactions. We demonstrate the approach with simulated data and with an

empirical example using data from the Yellowstone bison population. 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: forecast, Markov chain Monte Carlo, prediction, population model, uncertainty*

## Introduction

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), ushering in a no-analog era in which the past cannot 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. 2018). 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 an essential component of this learning process (Bauer et al. 2015).

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 ( $\theta$ ), and process error ( $\epsilon$ ). We can then write a general form of the model as:

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

which states that  $y$  at time  $t + 1$  is a function of  $y$  and  $x$  at time  $t$  conditional on the model parameters ( $\theta$ ) plus process error ( $\varepsilon$ ). 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:

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

where each additive term follows a pattern of *sensitivity* times *variance* and “IC uncert.” refers to “Initial Conditions 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 meteorological variables – meteorologists 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

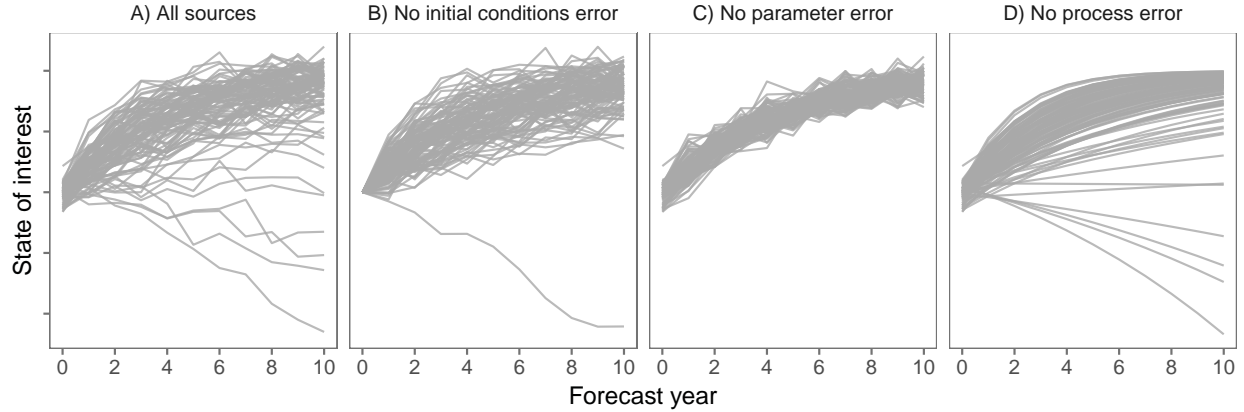


Figure 1: Example of forecast uncertainty with different sources of error set to zero. Each line represents one realization, out of 200, from an order-one autoregressive model (AR1 process). Contrary to the analytical expression (Eq. 2), initial conditions uncertainty and parameter uncertainty clearly interact. The spread of lines in (A) is not wholly because of initial conditions uncertainty (B) or parameter uncertainty (C), it is their combined influence that causes the spread of realizations in (A). At least in this example, process error (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 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). Therefore, progress in quantifying and partitioning forecast uncertainty requires a more flexible approach than that provided by Eq. 2.

## A simulation-based approach for partitioning 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 test the general idea

75 using simulated data and extend the approach to consider interactions among sources of uncertainty.

76 As a starting point, consider a generic Bayesian state-space model

$$\textbf{Data model: } y_t \sim [y_t | z_t, \sigma_o^2], \quad t = 1, \dots, T, \quad (3)$$

$$\textbf{Process models: } z_t \sim [z_t | \mu_t, \sigma_p^2], \quad (4)$$

$$\mu_t = g(z_{t-1}, \mathbf{x}'_t, \boldsymbol{\theta}), \quad t = 2, \dots, T, \quad (5)$$

$$\textbf{Parameter model: } \boldsymbol{\theta} \sim [\boldsymbol{\theta}], \quad (6)$$

77 where  $y_t$  is the observed state at time  $t$ ,  $z_t$  is the latent state at time  $t$ ,  $\mu_t$  is the deterministic  
 78 prediction of  $z$  at time  $t$  from the process model  $g$ , which is a function of  $z$  at time  $t-1$ , a vector of  
 79 covariates ( $\mathbf{x}$ ) at time  $t$ , and a set of unknown parameters,  $\boldsymbol{\theta}$ .  $\sigma_o^2$  is observation error and  $\sigma_p^2$  is  
 80 process error. The notation  $[a | b, c]$  reads, “the probability of  $a$  given  $b$  and  $c$ ,” and  $[\boldsymbol{\theta}]$  refers to the  
 81 prior probability distributions for all parameters in the parameter vector,  $\boldsymbol{\theta}$ .

82 For our purposes, we are interested in the probability distributions of the true state  $\mathbf{z}$  at future points  
 83 in time, conditional on previous observations ( $\mathbf{y}$ ). This is referred to as the forecast distribution or  
 84 the predictive process distribution (Hobbs and Hooten 2015), which, for one time step ahead of the  
 85 final observation ( $T + 1$ ), is defined as

$$[z_{T+1} | y_1, \dots, y_T] = \int \dots \int [z_{T+1} | z_T, \boldsymbol{\theta}] \quad (7)$$

$$[z_1, \dots, z_T, \boldsymbol{\theta} | y_1, \dots, y_T] d\boldsymbol{\theta} dz_1 \dots dz_T.$$

86 This type of model is easily fit using Bayesian methods and Markov chain Monte carlo (MCMC)  
 87 algorithms. The posterior distribution of all unknowns, including states and parameters, is estimated  
 88 over  $K$  MCMC iterations.

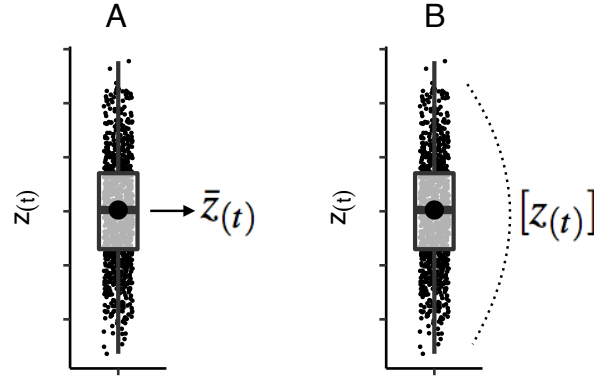


Figure 2: Forecasts can be made using (A) a point estimate of the median of the latent state  $z$ ,  $\bar{z}(t)$ , as a starting value or (B) using the full distribution of  $z$ ,  $[z(t)]$ . In both panels, the small points are the estimates of  $z$  at time  $t$  from each of 1000 MCMC iterations, the boxplots show the distribution, and the large point shows the median. The scenario in A represents the case where initial conditions uncertainty is set to zero. Comparing the variance of forecasts made under scenarios A and B allows us to quantify the amount of uncertainty attributable to initial conditions.

## References

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

- 103 National Academy of Sciences 115:1424–1432.
- 104 Hobbs, N. T., and M. B. Hooten. 2015. Bayesian Models: A Statistical Primer for Ecologists.  
105 Princeton University Press, Princeton.
- 106 Sobol', I. 1993. Sensitivity Estimates for Nonlinear Mathematical Models.
- 107 Steffen, W., K. Richardson, J. Rockström, S. Cornell, I. Fetzer, E. Bennett, R. Biggs, and S.  
108 Carpenter. 2015. Planetary boundaries: Guiding human development on a changing planet. Science  
109 348:1217.