

Quantifying and partitioning uncertainty to improve ecological forecasts

Andrew T. Tredennick¹, Peter B. Adler¹, and others²

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

² 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 herd. We also provide general functions written in the statistical programming language R, which should allow others using Bayesian

21 state-space models to employ our approach in their own research.

22 *Keywords: forecast, Markov chain Monte Carlo, prediction, population model, uncertainty*

23 **Introduction**

24 A fundamental challenge facing society is to predict the ecological impacts of global environmental
25 changes such as nitrogen deposition, climate change, and habitat fragmentation. Each of these
26 global change drivers have now exceeded their historical ranges of variability (Steffen et al. 2015),
27 ushering in a no-analog era in which the past cannot predict the future. We can, however, look to
28 the past to parameterize models that allow us to forecast the future states of ecological systems
29 (Clark et al. 2001, Dietze et al. 2018). Ecologists are in an excellent position to meet this
30 forecasting challenge because we have spent decades gaining understanding of the processes that
31 regulate populations, communities, and ecosystems. However, we lack a systematic understanding
32 of the current limits to ecological forecasts. As a result, we do not know how to allocate research
33 effort to improve our forecasts.

34 Making poor forecasts is inevitable as ecology matures into a more predictive science. The key is to
35 learn from our failures so that forecasts become more accurate over time. The success of
36 meteorological forecasting tells us that basic research on the causes of forecast uncertainty is an
37 essential component of this learning process (Bauer et al. 2015).

38 Various approaches have been used to characterize and partition forecast uncertainty (Sobol' 1993,
39 Cariboni et al. 2007). For example, consider a dynamic model designed to predict some state y in
40 the future (y_{t+1}) based on the current state (y_t), an environmental driver(s) (x), parameters (θ), and
41 process error (ϵ). We can then write a general form of the model as:

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

which states that y at time $t + 1$ is a function of y and x at time t conditional on the model parameters (θ) plus process error (ε) . Ignoring covariance among factors, 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.

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

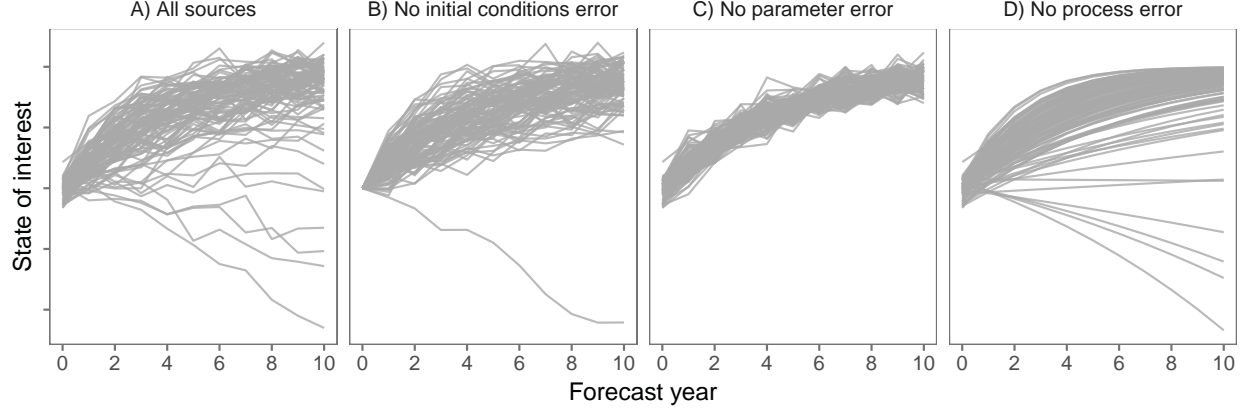


Figure 1: Example of forecast uncertainty with different sources of error set to zero. Each line represents one realization, out of 200, from an order-one autoregressive model (AR1 process). Contrary to the analytical expression (Eq. 2), initial conditions uncertainty and parameter uncertainty clearly 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`.

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 using simulated data and extend the approach to consider interactions among sources of uncertainty. As a starting point, consider a generic Bayesian state-space model

$$y_{(t)} \sim [y_{(t)} | z_{(t)}, \sigma_o^2], \quad (3)$$

$$z_{(t)} \sim [z_{(t)} | \mu_{(t)}, \sigma_p^2], \quad (4)$$

$$\mu_{(t)} = g(z_{(t-1)}, \mathbf{x}'_{(t)}, \boldsymbol{\theta}), \quad (5)$$

where $y_{(t)}$ is the observed state at time t , $z_{(t)}$ is the latent state at time t , $\mu_{(t)}$ is the deterministic 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 at time t , and a set of unknown parameters, $\boldsymbol{\theta}$. σ_o^2 is observation error and σ_p^2 is process error. The notation $[a | b, c]$ reads, “the probability of a given b and c .”

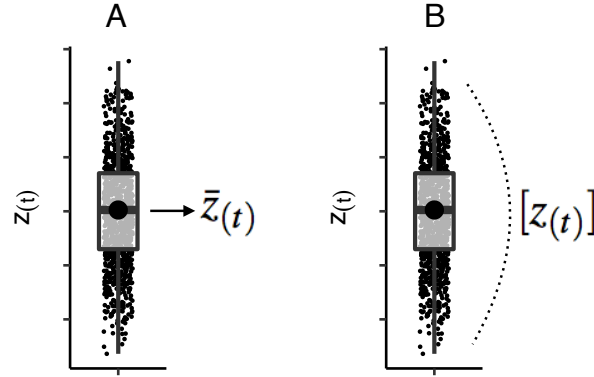


Figure 2: Forecasts can be made using (A) a point estimate of the median of the latent state z , $\bar{z}(t)$, as a starting value or (B) using the full distribution of z , $[z(t)]$. In both panels, the small points are the estimates of z at time t from each of 1000 MCMC iterations, the boxplots show 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.

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

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*

82 27:2048–2060.

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

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

89 Steffen, W., K. Richardson, J. Rockström, S. Cornell, I. Fetzer, E. Bennett, R. Biggs, and S.
90 Carpenter. 2015. Planetary boundaries: Guiding human development on a changing planet. *Science*
91 348:1217.