Supplement: Simulation as a best practice in Bayesian workflows and beyond

EM Wolkovich, TJ Davies, WD Pearse & M Betancourt

September 26, 2023

1 Which workflow?

Formally, all a 'workflow' does is organize various steps together in a systematic fashion, but there are many different workflows depending on what the aim is, which will determine which steps a workflow should include. For example a workflow aimed at calibration could look like an expanded version of our Step 1, where all the steps focus on investigating the assumptions encoded in a given model using simulated data. Or a workflow aimed at inference could expand Step 3, to focus on constructing a posterior, then investigating its model adequacy via several criteria. An inferential workflow can also be extended into a model development workflow. If the model adequacy criteria inform not only that something is inadequate about the current model assumptions but what is inadequate (ideally this happens some in Step 4) then one can use those hints to iterative improve the modeling assumptions. We present in the main text a very simplified model development workflow that combines calibration, inference and some model development, but it is not necessarily appropriate for everyone, depending on their aims.

2 What's a model?

We cavalierly use the terms model, mechanistic model, process model, data generating process, and statistical model in this paper, which follows their current use in ecology. This reality comes naturally from divergent fields using them in different ways, but it's important to recognize that what is 'mechanistic' or 'statistical' is not usually a clear distinction in ecology. Further, applying specific terms to certain modelling approaches, or to specific parts of a model should be done with care, especially if it impacts how you interpret the model. For example, consider:

$y \sim \text{normal}(f(x), \sigma)$.

Within this $f(x)$ is sometimes called the 'mechanistic' or 'process' model, and $\text{normal}(, \sigma)$, the 'noise', 'error', or even 'measurement error', however, these terms are only accurate in certain (in our experience: rare) cases. Given that $f(x)$ will never contain the true underlying process, $\text{normal}(, \sigma)$ is functionally capturing everything not in $f(x)$.

3 References