



May 27, 2024

Dear Dr. Drake:

We would like to propose a *Method* piece for *Ecology Letters* outlining a four-step simulation based workflow for statistical analyses in ecology, and show how it can enhance data collection, forecasting, and statistical training. The workflow is designed to be broadly generalizable and practical. To make sure the steps are clear and accessible we would present full example of the workflow and accompanying code (in R Markdown) to estimate trends over time in plant and animal phenology. Given the rise in bigger datasets and relatedly more complex models we think this piece would be very timely and have broad interest to readers of *Ecology Letters*.

Our manuscript would briefly review the changing landscape of data, models and aims in ecology today (Anderson *et al.*, 2021; Muff *et al.*, 2022), outlining how bigger data generally require more complex models that many researchers—ourselves included—were not trained in. While many ecologists may not be formally trained in the fitting of large, complex models, we suggest that they often have the computational toolkit to approach such models, but lack an organizational framework to develop, test and improve bespoke models. To address this gap, we outline a generalizable workflow (see Fig. 1 below, and see Grinsztajn *et al.*, 2021; van de Schoot *et al.*, 2021), which is built on fundamental scientific principles. This approach moves away from a focus on null hypothesis testing, towards estimating effect sizes, using models calibrated and better understood through simulating data at multiple steps—using a number of skills more often associated with theoretical than empirical ecology. We conclude by highlighting how adopting this workflow has changed our science and how it may improve statistical and mathematical training in ecology.

The manuscript would be authored by an international and interdisciplinary group of ecologists, evolutionary biologists and statisticians. The workflow follows the basics of how authors EM Wolkovich, TJ Davies and WD Pearse approach model building and leverages the insights and skills of computational statistician M Betancourt who has developed fundamental statistical workflows for diverse scientific disciplines.

We hope our proposed *Method* will provide a road-map for the many ecologists now building more complex models and inspire new training approaches. By integrating simulation more fully in model building and testing this workflow can fit models that are more robust and well-suited to provide new ecological insights—allowing us to refine where to put resources for better estimates, better models, and better forecasts.

We have a draft manuscript that could be submitted upon request. We hope you will consider it as potential submission for *Ecology Letters*.

Sincerely,



Elizabeth M Wolkovich (on behalf of my co-authors also)  
Associate Professor of Forest & Conservation Sciences  
University of British Columbia

## References

- Anderson, S.C., Elsen, P.R., Hughes, B.B., Tonietto, R.K., Bletz, M.C., Gill, D.A., Holgerson, M.A., Kuebbing, S.E., McDonough MacKenzie, C., Meek, M.H. *et al.* (2021) Trends in ecology and conservation over eight decades. *Frontiers in Ecology and the Environment* **19**, 274–282.
- Grinsztajn, L., Semenova, E., Margossian, C.C. & Riou, J. (2021) Bayesian workflow for disease transmission modeling in stan. *Statistics in Medicine* **40**, 6209–6234.
- Muff, S., Nilsen, E.B., O’Hara, R.B. & Nater, C.R. (2022) Rewriting results sections in the language of evidence. *Trends in ecology & evolution* **37**, 203–210.
- van de Schoot, R., Depaoli, S., King, R., Kramer, B., Maertens, K., Tadesse, M.C., Vannucci, M., Gelman, A., Veen, D., Willemsen, J. & Yau, C. (2021) Bayesian statistics and modelling. *Nature Reviews Methods Primers* **1**.

## Example figures

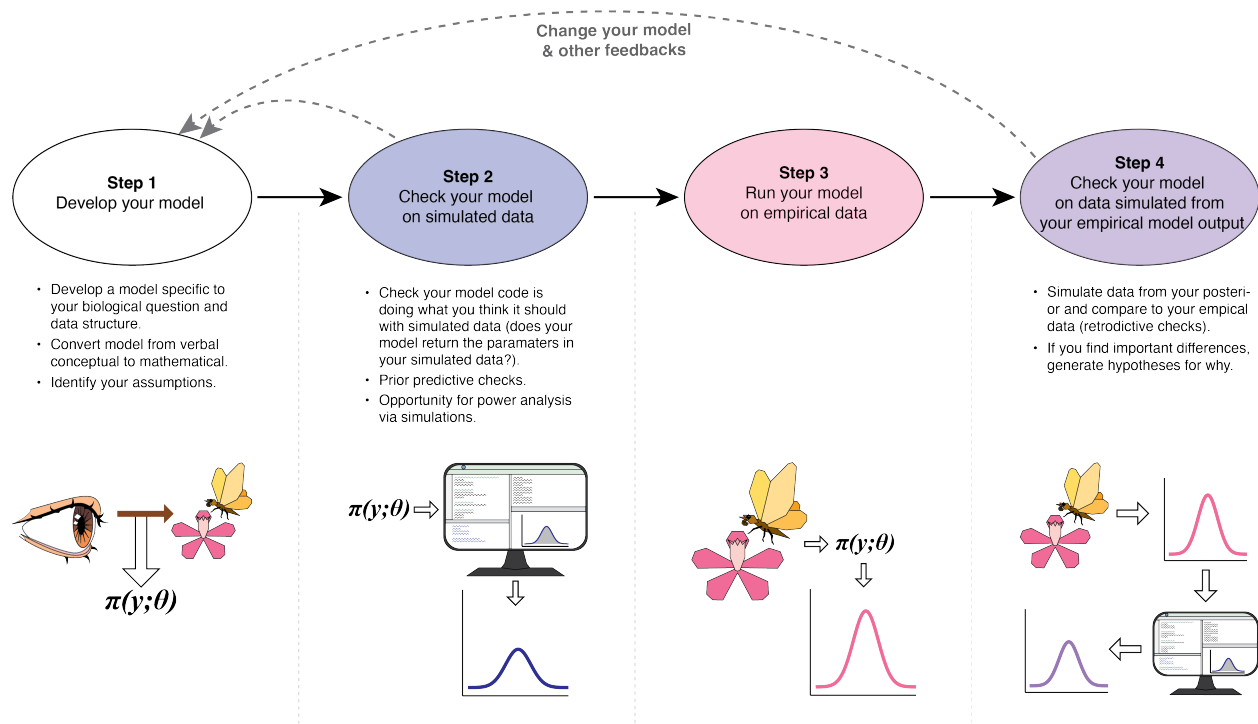


Figure 1: The four-step iterative workflow we outline can help design models for specific ecological questions, data and aims—which makes this a statistical workflow that can naturally become a scientific workflow. It makes the step that many of us focus on—running your model on your empirical data (Step 3)—far more straightforward and insightful by using simulations both before (Step 2) and after (Step 4) it to better understand the model and data together.

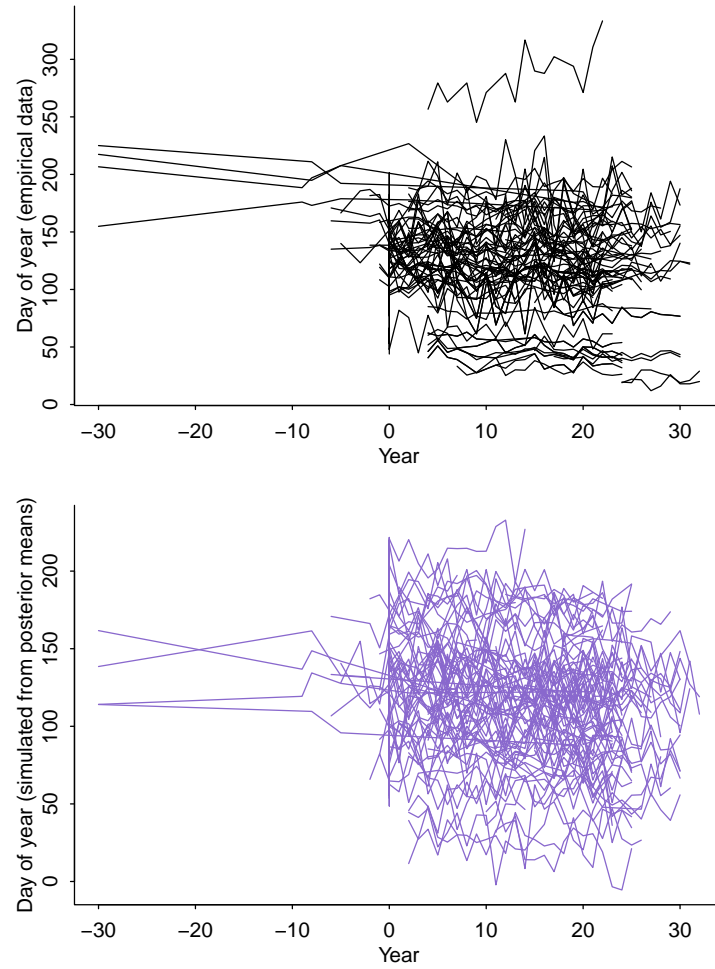


Figure 2: Example of a single retrodictive check from time-series data of phenological events over time. The raw data (top, black) looks similar to one simulated dataset (bottom, purple), based on existing species number, their respective  $x$  data, and simulating from the parameters for each species. See ‘An example workflow’ in the Supplement for more details.