



August 5, 2024

Dear Dr. Stephens:

We would like to propose an *Opinion* for *Trends in Ecology & Evolution* entitled ‘A four-step Bayesian workflow for improving ecological science.’ Given the increasing aims of forecasting and prediction today, ecologists are using more complex models to leverage larger datasets (Anderson *et al.*, 2021; Muff *et al.*, 2022), but many researchers—ourselves included—were not trained in best statistical practice for these approaches. This can lead to poor models and incorrect predictions and decisions. While the field of statistics has worked to develop robust workflows for analysing complex data, these advances have largely failed to percolate into ecological and evolutionary research.

- To bridge this gap, we outline a generalizable workflow for ecological and evolutionary research (see Fig. 1 below), which is built on fundamental scientific principles and new insights from statistics and data science (Grinsztajn *et al.*, 2021; van de Schoot *et al.*, 2021). This approach moves away from a focus on null hypothesis testing, towards estimating effect sizes, with 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 may improve statistical and mathematical training in ecology, including how we implement machine learning models and other new approaches.
- There have been many in-depth articles on various aspects of model fitting and validation in recent years (e.g. Conn *et al.*, 2018; Gabry *et al.*, 2019; Tredennick *et al.*, 2021; DiRenzo *et al.*, 2023), but there have been no attempts—to our knowledge—to generate a statistical workflow aimed directly at evolutionary biologists and ecologists. While we build on recent advances in scientific and statistical workflows in other disciplines (Gelman *et al.*, 2020; Grinsztajn *et al.*, 2021; Schad *et al.*, 2021), we provide a unique perspective by explaining the workflow and why it is both tractable and critical now for ecological and evolutionary science.
- We believe this article would be ideal for *TREE* given our aims for it to be concise and extremely approachable. The workflow is designed to be generalizable across disciplines of ecology and evolution. To make sure the steps are clear and accessible we would present a full example of the workflow and accompanying code (in R Markdown) to estimate trends over time in plant and animal phenology.

The manuscript is 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 realize our author list is one person beyond the desired size, but we found multiple perspectives on how to implement this workflow have been especially useful.

We hope that you will find our proposed submission, which provides a road-map for the many ecologists now building more complex models, suitable for publication in *Trends in Ecology & Evolution*. By integrating simulation more fully in model building and testing this workflow can help 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.

Sincerely,

A handwritten signature in black ink, appearing to read 'Elizabeth M. Wolkovich', with a stylized flourish at the end.

Elizabeth M Wolkovich
Associate Professor of Forest & Conservation Sciences
University of British Columbia

Relevant recent articles:

As we mention above, our article leverages advances in statistical workflows, which were developed especially by more Bayesian-focused statisticians, but designed to be applicable widely (Betancourt, 2020; Gelman *et al.*, 2020). These new approaches have been adapted to epidemiology and cognitive science (Grinsztajn *et al.*, 2021; Schad *et al.*, 2021); our article would be the first to introduce them specifically to evolutionary and ecological research. We believe an approachable, concise treatment of this workflow may be especially powerful in ecology where many researchers now have the computational toolkit to fit larger more complicated models, but lack an organizational framework on how best to develop, test and improve their models. By adopting a standardised workflow, such as the one we propose, we hope more ecologists will be confident to fit their data to bespoke mathematical models of biological systems, which in turn would yield better inference, predictions and forecasts.

Example figure

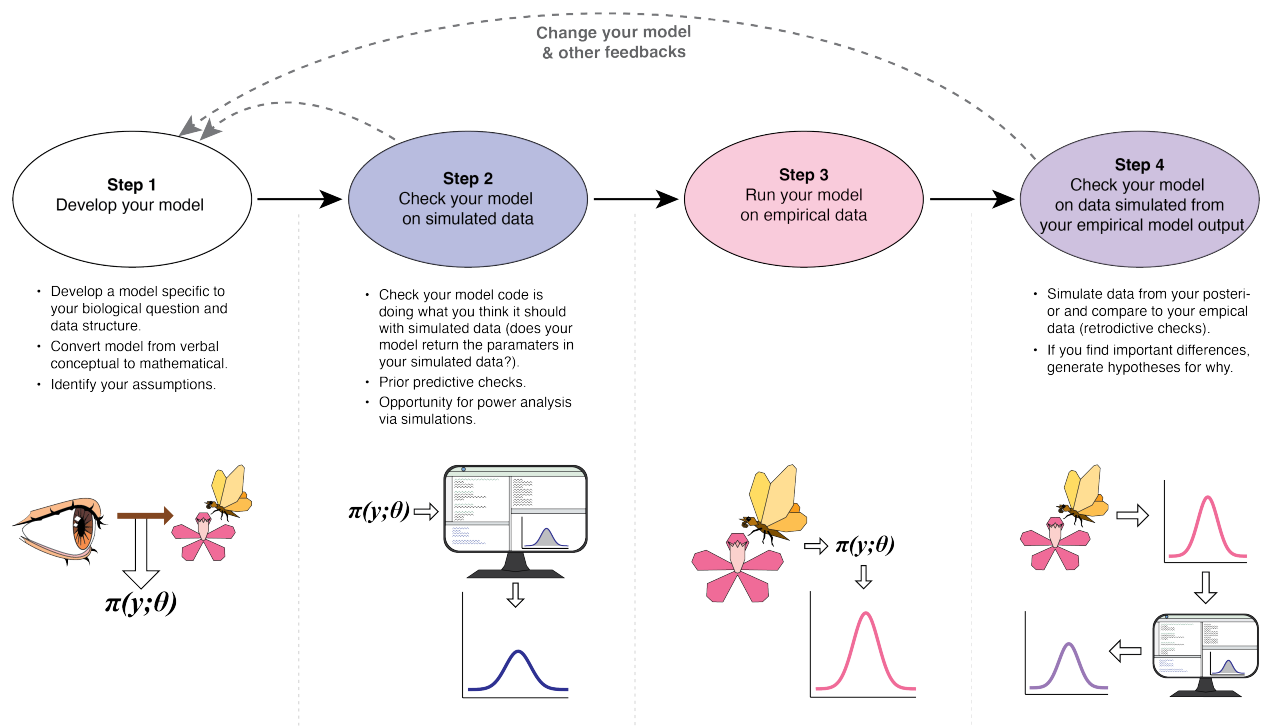


Figure 1: The four-step iterative workflow we outline can help design models for specific ecological questions, data and aims—which makes this a statistical workflow that can naturally become a scientific workflow. It makes the step that many of us focus on—running your model on your empirical data (Step 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.

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.
- Betancourt, M. (2020) Towards A Principled Bayesian Workflow. https://betanalpha.github.io/assets/case_studies/principled_bayesian_workflow.html.
- Conn, P.B., Johnson, D.S., Williams, P.J., Melin, S.R. & Hooten, M.B. (2018) A guide to Bayesian model checking for ecologists. *Ecological Monographs* **88**, 526–542.
- DiRenzo, G.V., Hanks, E. & Miller, D.A. (2023) A practical guide to understanding and validating complex models using data simulations. *Methods in Ecology and Evolution* **14**, 203–217.
- Gabry, J., Simpson, D., Vehtari, A., Betancourt, M. & Gelman, A. (2019) Visualization in bayesian workflow. *Journal of the Royal Statistical Society Series a-Statistics in Society* **182**, 389–402.
- Gelman, A., Vehtari, A., Simpson, D., Margossian, C.C., Carpenter, B., Yao, Y., Kennedy, L., Gabry, J., Bürkner, P.C. & Modrák, M. (2020) Bayesian workflow. arXiv.
- 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.
- Schad, D.J., Betancourt, M. & Vasisht, S. (2021) Toward a principled Bayesian workflow in cognitive science. *Psychological Methods* **26**, 103–126.
- Tredennick, A.T., Hooker, G., Ellner, S.P. & Adler, P.B. (2021) A practical guide to selecting models for exploration, inference, and prediction in ecology. *Ecology* **102**, e03336.
- 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**.