

Reviewer comments (We provide below the full context of the three reviewers' comments) are in *italics*, while our responses are in regular text.

**Editor – comments:**

*The main point coming from both referees is that, because this is not an ecology journal, we can expect most of the readers of the paper to not be ecologists. Thus it would be helpful to add a bit to the abstract and introduction, right at the beginning, explaining why non-ecologists should be interested in this paper. Such an explanation should not be difficult, because these workflow ideas are generalizable. You just want to do something here to reach this audience, as not everyone is going to make it all the way to page 8 to see the section, “How this workflow extends to other fields.”*

**Reviewer 1 – comments:**

*Thank you for the opportunity to review your manuscript “A four-step simulation-based workflow for ecology.” I found this paper to be a useful summary of the Bayesian workflow using ecological examples, and I’m wholeheartedly on board with the value of this workflow in ecological data analysis, which often features observational studies and/or highly complex systems with complex dynamics and potential confounding.*

Yay, thanks.

*As Philosophical Transactions A is primarily mathematical rather than ecological, my biggest critique of the manuscript is that it seems to be intended for an ecological audience. That is, the manuscript digests and distills concepts that are well known in the statistical literature (Gelman et al’s Bayesian Workflow paper, Betancourt’s Bayesian Workflow case study, etc) for ecologists. But in my opinion the manuscript doesn’t successfully use ecological examples to shed light on the Bayesian workflow in a way that statisticians would find particularly novel or illuminating—probably because the manuscript doesn’t *\*intend\** to do so. However, given the likely audience for this journal, I believe the paper would have more impact for its readership if it could leverage ecological examples to drive statistical insight. Something like a dispatch from the field: “here are some of the opportunities, successes, and complications that arise when putting these ideas into practice in the ecological domain.”*

*Below, I provide some minor, specific comments.*

*Title: I think it would be clarifying to ecologists (who are, I believe, the intended audience of this manuscript) to call this a workflow for “ecological data analysis”, rather than for “ecology”. True, this workflow extends beyond data analysis and across to elements of study design, but to a limited degree. For example, this is not a workflow for the development of ecological theory or the conduct of ecological fieldwork.*

*Lines 10-11: Arguably, these lines oversell the novelty of this manuscript. To be “Here we describe” in a paper abstract suggests that the thing being described is new. “Building on new advances” suggests that the present manuscript is doing the building. I don’t think that’s entirely the case here.*

*Lines 109-121: The introduction, as well as lines 98-100, explicitly set this paper up as being about something bigger than mechanistic Bayesian modeling. Does the claim here at line 110 hold up in general, including when the model under consideration is a black-box technique from a machine learning toolkit?*

*Line 128 “i.e., you fix values for your model parameters, then test how well your model re-*

*covers them”: This is the case only for parametric models, but again I think this paper’s aim is to be more broadly applicable.*

*Line 157-159: This is another example where the workflow being presented is uniquely relevant to MCMC based approaches to Bayesian model fitting. I’ll cease calling these out, but it’s worth a think about how important it is that this manuscript engage ecologists whose statistical practice is heavily skewed towards non-bayesian techniques, be they frequentist, black-box machine learning, or something else.*

*Lines 181-195: I would suggest placing more emphasis here on the judicious choice of quantities (or other data summary, e.g. graphical outputs) for PPCs. Too often in ecology and elsewhere I see people using dumb quantities for their PPCs that are nearly guaranteed to fail to detect misspecification and/or to fail to provide useful insight into what model assumptions are being violated in the event that misspecification is detected. For a great example of judiciously chosen quantities that will be accessible to ecologists, see Minmo et al 2015 <https://www.pnas.org/doi/full/10.1073/pnas.1412301112>*

*Line 210: suggest changing to “non-existent calendar days beyond the 365th”. I suppose it probably \*is\* pretty uncommon to see predictions for ordinal day 1000 in particular.*

*Line 217: “We never noticed [non-identifiability or degeneracy] before using this workflow.” What do you mean by this? Surely you got some convergence failures from lmer, or some absurd MCMC posteriors with no convergence, probably with greater frequency prior to adopting this workflow than after. Do you mean that before adopting the workflow, you didn’t associate these issues with the concept of degeneracy?*

*Line 223: I think it would be highly relevant and useful (and eye-opening for the readership) to provide an example of a non-identifiable model that gets caught by this workflow but that common software packages are happy to pretend to give an answer about.*

*Line 234: What do you mean by “their underlying theorems”? The theorems in statistics that underpin the techniques for fitting statistical models? I’m not sure I grasp what this sentence is getting at.*

*Line 237-238: “solve [an ecological model] analytically” to me suggests solving for the equilibrium in an ODE model or something, but here it seems to be applied to a statistical model, of the sort that one can simulate data from. What is meant by “solve analytically” here? Analytically finding the Bayesian posterior?*

*Short Case Study 1: This entire case study, up until the last paragraph, doesn’t really make clear how the workflow proposed here provides benefits beyond standard power analysis.*

*Short Case study 2: “Researchers commonly used this to estimate the effect of daylength in models of leafout”. Really? I appreciate not wanting to call anyone out for this, and I don’t doubt that this has been done before, but commonly?*

*Short Case Study 2 (again): “but that would show us that many variables now matter, including a number of ones, such as total precipitation, that we did not use to simulate our data either.” This is a great example of how the workflow can guide people to better theory-motivated data and model structures, and as such I’d love to see it fleshed out a bit more in the case study—I think this is the interesting part. I also think it’s worth clarifying the*

following. As written, the idea that the lasso is failing because it thinks precipitation matters is confusing, because at first glance it looks like the lasso procedure is broken. We put GDD—the only thing we used to simulate the data—into the model, and it came back selecting a bunch of other stuff as important, like precip. I think it would be useful to spell out for the reader that the thing that matters causally in the simulation—GDD on the date of leaf-out—was not put in the model. Instead, the model contains variably good or poor proxies for roughly when we might expect the GDD threshold to have been reached, and it turns out that precip is a relevant proxy, alongside mean spring temp and others. The “aha” moment, then, is realizing that by throwing mean spring temp into the model, we haven’t captured the key generative structure around GDDs, and indeed precip might well be just as useful a predictor as mean spring temperature.

## **Reviewer 2 – comments:**

*The paper A four-step simulation-based workflow for ecology presents a workflow for statistical modeling of ecological data. They argue that traditional statistical methods and a focus on null hypothesis testing does not lend itself to meet the modeling challenges that ecology finds itself in.*

*It’s easy enough for me as someone who uses the workflow to agree with the authors throughout the paper. However, it’s not always clear to me from the text how the need for folks to learn and adopt more complex models for complex data requires the workflow. For instance, in the introduction: “Such data generally require more complex models, such as those that accommodate both the underlying biological processes and how the measurements were made.” From this, could folks who feel confident in using complex models bypass this workflow and still do good science? To me the simulation-based approach is really more like a different way of building models in general. My biggest challenge has been trying to convince folks to adopt the workflow as a way to do good science irrespective of the complexity of the data or of the model.*

*- Page 3, line 50 – Why do the authors refer to it as Bayesian methods and not Bayesian inference? I find ecologists tend to use these sorts of terms as a way to say that a ‘Bayesian model/method’ is better without really understanding that the data generating mechanism is not fundamentally different in a frequentist/Bayesian context. I think a clearer explanation would help clarify what the authors mean by Bayesian methods.*

*- Perhaps the authors can discuss that larger, messier data requiring more complex models may be more easily fit in a Bayesian framework and thus lends itself to new modeling opportunities.*

*Overall I enjoyed the paper and hope that a larger audience will start moving beyond p-values and adopt more simulation-based approaches.*