

Response to Review Round 2

Associate Editor

One of the previous reviewers and a third reviewer have evaluated your resubmission. They both seem very pleased with this version and see this as an important first step to the standardization and “cyber-implementation” of iterative near-term forecasting. Both reviewers and authors argue well that, while many challenges remain to be solved, this is a great first version and discussion template. As such, it can be a very valuable contribution to the field. To best fulfill this purpose, I invite the authors to read the last reviewer’s list of issues and challenges that remain to be solved. Enumerating them in the discussion would constitute a valuable roadmap of future developments and likely make this article highly influential.

Thanks! We agree that the last reviewer has made a number of valuable suggestions for helping lay out a roadmap for the future and we have added discussion of many of them. See details in the response to this reviewer below.

Reviewer 2

I think the authors have addressed what they needed for this to be accepted for publication.

Thanks.

But I think that they misinterpreted my comments about audience. If the point is to get more people using this workflow (for themselves and for the good of science), thinking more about where people are coming and adding sentences to usher them along is key.

We certainly agree that this is the goal and have done our best to accomplish it both here and in our paper on using continuous analysis for data management.

Reviewer 3

In this paper the authors introduce a prototype automated workflow for producing iterative ecological forecasts for the Portal project’s long-term rodent monitoring. At a high level I think this is a really important paper. The interest in iterative ecological forecasting is growing rapidly and new projects keep coming online. However, setting up the cyberinfrastructure (CI) surrounding such forecasts can be nontrivial and, to my knowledge, there has been no previously published discussion outlining the challenges, potential solutions, and emerging best practices for how to do this. While this manuscript is not going to be the last word on how to set up a forecast architecture, it is a very useful starting point as it highlights and makes explicit many of the issues faced and identified approaches for dealing with these issues that leverage modern CI advances. These emerging CI tools make the ecological forecasting process more robust than hacking a workflow together, while at the same time being more nimble and easier to use for smaller forecasts than the enormous mainframe solutions employed by numerical weather forecast centers. As someone who has thought about these issues a lot, I think the authors have done a great job of covering a lot of issues and in bringing a robust system into operations. In my detailed comments below I do have some suggestions for their system, but I want to make clear that I don’t think these are problems they need to solve (and code implemented) for this paper to be published. More often they are things that I think would be good to bring up in the Discussion – I think there’s a bit more to be said about the limitations current approaches and the future directions for this area of Methods research.

We thank the reviewer both for their insightful comments (our responses to which are detailed below) and for being explicit that these are issues to be discussed and worked toward in future development rather than things that need to be implemented immediately.

Detailed comments:

L51: Might be useful to add a few more examples of ecological forecasts occurring on short time scales (daily to quarterly)

L76: You might consider a slight restructuring of the introduction. I feel like some of the ideas being ‘introduced’ in this paragraph were already being referenced earlier in the introduction.

The goal of this paragraph is distinct from the prior paragraphs in that to it introduces the specific recommendations made by Dietze et al. There is some overlap with the general concepts that motivate these recommendations, but since this paper builds directly on Dietze et al. we consider it important to clearly link what we are doing directly to their ideas. So, the first two paragraphs introduce the general space the this paragraph introduces the important intellectual contribution of Dietze et al. We have tried to make this clear in the topic sentence.

L107: On the topic of model averaging, I’d recommend checking out the new Ecological Monograph by Dormann et al <https://doi.org/10.1002/ecm.1309>

We agree that the Dormann et al paper is a valuable contribution in this space and have added a citation to it to the paper.

L148: I agree that the Portal project is a great example for prototyping an ecological forecast. Perhaps say a bit more about how these forecasts are being used (or how the Portal team plans to use them) to inform their basic science. In general I think forecasting experiments is a great idea, but I also wonder if there are any conservation/management implications of this specific forecast.

We have added a sentence on the applied importance of forecasting rodent population dynamics, related to their importance in several major zoonotic diseases, as well as a sentence describing current and future research that we are conducting using this system.

L151 and elsewhere: When forecasting iteratively, this manuscript tends to talk about “re-building” or “refitting” the models. It’s worth noting that one of the most common approaches to iterative forecasting, data assimilation, does not refit/rebuild the model at all, but instead just updates the state of the system (and sometimes the parameters) given the previous forecast and the new data. You might think about a slight tweak to your word choice to be more inclusive of these approaches.

This is an excellent point. We have change the language to “updating” throughout the manuscript to capture the range of possible approaches.

L180 and elsewhere: I find your use of Travis CI to facilitate Continuous Analysis to be clever, but I also think it’s worth acknowledging to the reader, either here or in the Discussion, that this is a bit of a hack and not a system that’s designed for Continuous Analysis. Yes, it can get the job done but it sees like an abuse of the system, it places non-trivial computational constraints on what you can do, and it fundamentally isn’t designed to return quantitative outputs. The authors don’t really explain how they get their output files out of Travis CI and into their database, and looking at the code it appears that this occurs by having Travis CI itself trigger a push into various Github repositories. By contrast, there are other containerized systems like Clipper (UC Berkeley) and OpenWisk (IBM) that are designed for these types of workflows that seem worth mentioning in the Discussion as possible future directions.

We disagree with the reviewer that this is a “hack”. Travis is used and discussed in the main continuous analysis paper (Beaulieu-Jones & Greene 2017). It is also used in the software development community for non-trivial computational tasks and for returning and deploying outputs. In addition, we did fully explain our method for deploying the outputs (see lines 349-359 of the previous submission).

That said, as previously noted (lines 422-423 in the previous submission), we agree that Travis places meaningful computational limitations on the analysis that can be conducted. We have added a new paragraph on future directions to the Discussion that addresses this and mentions alternatives and routes forward for making this kind of Continuous Analysis more scalable.

L188: This paragraph feels like it rambles a bit and would benefit from being tightened up and shortened.

This paragraph has been tightened and shortened substantially in the process.

L245: You mention how you grab previous, observed weather data but don't mention what you're using as the weather forecast data to drive the forecasts. In the repo portalPredictions/data you appear to provide models with both weather data and NDVI to use as part of model calibration, but I don't see the inclusion of forecasts of these quantities, which makes me wonder if they're being used in the forecast at all

We do use these data in the forecasts. The pevGARCH model uses temperature, precipitation, and NDVI as predictors. Forecasts for temperature and precipitation are obtained from the North American Multi Model Ensemble (NMME; Kirtman et al. 2014) downscaled to the Portal site. Forecasts for NDVI are not currently available and so we use a seasonal auto ARIMA model to forecast future values. However, we have avoided discussion of the specific models in this paper since exploration of the infrastructure is already a sufficiently lengthy topic for discussion.

L290: The idea of "plugin infrastructure" is important, but a bit vague. Seems like a good place to point readers to documentation on exactly what you mean here and how to plug new things in.

We have added a page to the GitHub repository expanding on the previous description from the README about how to do this and added a link to that page from this paragraph.

Figure 2: With regards to the predefined structure of the forecasts, for many forecast methods and scoring metrics the mean and interval are insufficient to capture the full, probabilistic nature of the forecasts or to score forecasts against future observations. For example, with ensemble and Monte Carlo methods, should you be storing the full ensemble of predictions? In a multivariate forecast, the current format has no way of recording covariances between state variables. It is also interesting that none of the proposed approaches seem to take the previous forecasts as inputs to the next prediction, and it's not clear whether the current architecture would support that. Again, there are not specific requests where I'm asking the authors to change their system – this is a great first start – but things they and the community should think about when building upon this prototype or developing alternatives. Perhaps worth mentioning briefly in the Discussion?

This is an excellent point and one we've been discussing extensively to try to figure out how to improve going forward. We have now added a section on future directions to the Discussion that addresses this and other forward looking suggestions by the review.

L315: I think the idea of developing standards for archiving ecological forecasts is really important. We're at a point where this community is still relatively small and we should tackle these issues now, rather than face painful interoperability issues in the future that we could have avoided. There's an extensive literature on data standards, meta-data, and ontologies that could be briefly mentioned here.

We have now added a section on future directions to the Discussion that addresses this.

L316: Is this the same rule being used in your calibration? (probably if assuming Gaussian error). If not, then this score is not "proper", in the sense that the best fitting model might not give the highest score. Also what about more probabilistic scores? RMSE doesn't penalize a model for being overconfident.

We have expanded the second sentence of this paragraph to address this point by highlighting deviance as one of the next scores to be added to our assessment.

L335: I really like the idea of thinking of ecological forecasts as a form of pre-registration

Thanks!

L373: I think all of this is great. In looking at the webpage one suggestion I'd make would be to add actual model-data comparisons under the evaluation, not just summary statistics (RMSE, coverage) – the patterns to how forecasts fail are very informative. That said, if I were to add one more metric it would be bias. Almost all calibration approaches force this to be zero, but the forecasts could easily over- or under-predict consistently.

Thanks for the suggestions. We do feature direct model-data comparisons for each previous forecast on the Rodent Report page, but it's a good idea to show these over time. We've added it to the issue queue.

L424: Add discussion of alternatives?

Yes, this is definitely an important oversight. We did initially have some discussion of this and aren't actually sure why we cut it. We have added a new paragraph on future directions that addresses this point and integrates the reviewers suggestions above related to possible alternatives.

L427: If current options are imperfect, could you state very explicitly what the key features are that you think we need to see in new/updated community tools?

We have added a new paragraph on future directions that addresses this point.