Closing the gap between statistical and scientific workflows for improved forecasts in ecology

Victor Van der Meersch, J. Regetz, T. J. Davies & EM Wolkovich

March 19, 2025

Deadline: 1 May 2025 (was 1 April 2025)

For: Scientific and Statistical Workflow theme issue for *Phil Trans A* as an *Opinion*

Abstract

Increasing biodiversity loss and climate change have led to greater demands for useful ecological models and forecasts. Relevant datasets to meet these demands have also increased in size and complexity, including in their geographical, temporal and phylogenetic scales. While new research often suggests that accounting for these complexities variously increases, removes or otherwise alters major trends, I argue that the fundamental approach to model fitting in ecology makes it impossible to evaluate and compare models. These problems stem in part from continuing gaps between statistical workflows – where the data processing and model development are often addressed separately from the ecological question and aim – and scientific workflows, where all steps are integrated. Yet, as ecologists become increasingly computational, and new tools make it easier to share data, the opportunity to close this gap has never been greater. I outline how increased data simulation at multiple steps in the scientific workflow could revolutionize our understanding of ecological systems, yielding new insights. Combining these changes with more open model and data sharing – and developing new efforts to race the same data – could be transformative for ecological forecasting.

1. Problem

    (a) Climate change, biodiversity crisis etc. has made it critical to understand trends to date and be able to forecast future trends (for policy)

    (b) Current workflows in ecology are not up to the task

        i. For trends: lots of different stuff reported for seemingly same data/question, makes people outside ecology wonder if we can even well document them, let alone understand them enough to suggest policy

        ii. For forecasting

    (c) Here we outline the problem and provide a solution!

2. What are current workflows and where are they limiting us?

    (a) For trends ...

     i. easy to find different trends through small model tweaks to analyses and/or different data

    ii. For example, right now many different papers report different biodiversity trends (LPI example?)

   iii. New workflow would make ecologists understand uncertainty in their model data/combo (and perhaps not see/publish results as so divergent?)

  (b) For forecasting ... (somehow jump to our focus on process-based models PBMs here? Something like, forecasting is big and there are diverse methods! Near-term iterative, correlative niche models, but PBMs are often considered the gold standard ... mention machine learning?)

3. Better workflows to the rescue!

  (a) General overview of new workflows

  (b) New vision of each workflow

4. Conclusion: world is better