

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 25, 2025

* Says he is happy to help and give friendly review, but not sure he will reach level of co-author.

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.

Goal: Increase awareness of how we can merge statistical and scientific workflows in ecology (especially forecasting) and what we would get out of it.

Introduction

Nature is increasingly threatened by multiple drivers of change, with a largely dominant influence of human activities. This ongoing biodiversity crisis is expected to increase in the next decades, and will continue to alter ecosystem services and human well-being. To support implementation of sustainable policies among the socioeconomic and environmental dimensions, it is critical to understand trends to date and be able to forecast future dynamics.

Estimation of global biodiversity indicators and current trends depends on large-scale and long-term datasets, across terrestrial, freshwater, and marine ecosystems. These data, gathered opportunistically and from multiple sources, are often unbalanced and have geographic, temporal and taxonomic biases. Forecasting future changes—under different plausible scenarios—generally relies on either correlative models or process-based models. The latter, which focus on a mechanistic representation of ecosystem functioning, are often promoted as the most realistic approach.

The urgent need to answer policy-relevant questions has favored the proliferation of diverse methods developed by different researchers, and lacking an overall coherence. Though there is no doubt nature is declining globally, significant uncertainty thus remains. There is no consensus on current species trends, with ongoing debates driven by widely varying reports that sometimes show conflicting trend directions. Future projections also diverge considerably. Predictive modeling is increasingly relying on overly complex models (with a huge number of parameters), making it less adequate to generate new scientific insights.