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 28, 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 (Díaz *et al.*, 2019). This ongoing biodiversity crisis is expected to increase in the next decades because of climate change, and will continue to alter ecosystem services and human well-being (IPBES, 2019). 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 (e.g. Dornelas *et al.*, 2018). These data, gathered opportunistically and from multiple sources, are often unbalanced and have geographic, temporal and taxonomic biases. Addressing these biases requires the use of appropriate statistical inference. Forecasting future changes—under different plausible scenarios—generally relies on either correlative models or process-based models (IPBES, 2019). The latter, which focus on a mechanistic representation of ecosystem functioning, are often promoted as the most realistic approach (Urban *et al.*, 2016; Pilowsky *et al.*, 2022).

The urgent need to answer policy-relevant questions has favored the proliferation of diverse methods developed by different researchers, lacking an overall coherence. Though there is no doubt nature is declining globally, significant uncertainty remains. There is no consensus on current species trends, with ongoing debates driven by widely varying reports that sometimes show conflicting trend directions (Dornelas *et al.*, 2014; Leung *et al.*, 2020; Buschke *et al.*, 2021; Johnson *et al.*, 2024). Future projections also diverge considerably, due to a high model uncertainty at the ecological level. Predictive modeling is increasingly relying on overly complex models (with a huge number of parameters), making it less adequate to generate new scientific insights (Franklin *et al.*, 2020).

Current controversies and focus on methodological aspects stem from the lack of coherence between current workflows in ecology (Loreau *et al.*, 2022; Talis & Lynch, 2023; Johnson *et al.*, 2024). Each new model development is added as a separate layer, disconnected from the original research aim, the data stream and the previous scientific insights. Workflows should fully integrate all the steps required to build a model from an ecological question, evaluate its limitations and degeneracies, before estimating its parameters and making projections. Here, we introduce an universal workflow that proposes to iteratively build upon all these steps, harmonizing both trend estimation and forecasting, with the aim of refocusing the debate on ecological questions and increasing the speed of scientific progress.

Scientific method and workflows

Quantitative science rely on a model-based framework to confront hypothesis and data, making some approximations (). The general scientific method stresses out that the research question should guide both the design of the experiment and the corresponding model building. The experiment should then be conducted—according to this specific design. Finally, the resulting experimental data should be used to inform our model and differentiate between hypotheses—hopefully answering our initial question. However, this is often an idealized view and does not reflect the complexity of ecology ().

Divergences from this scientific method are common. One explanation is the lack of rigor, leading to questionable practices, such as retrospectively crafting hypothesis to explain the results of a model rather than testing a clear pre-defined question (data-driven analysis, Fig. 1). But the reality of ecological research is also a major driver of the divergences from the ideal scientific method. Many important questions often cannot be addressed by conducting experiments and replications, and we must often rely on existing datasets to have a large-scale and long-term perspective. However, this does not explain the persistent flaws and lack of overall coherence in ecological modeling. Trend estimation still often relies on a one-way 'inference', and most of the time is spent fitting the model to empirical data (Fig. 1). For forecasting, researchers focus on making predictions with complex models, but the steps of model building and parameterization are not very transparent and not clearly delineated (Fig. 1). The different parts of the model are often calibrated separately rather than as a whole, and some parameter values are just fixed based on experiments and expert knowledge.

[...]

[...]

The workflow in pratice

[...]

[...]

Forecasting is a broad field with a diverse range of methods. Here, we focus on process-based modeling, which is often considered the gold standard in ecology and beyond. Process-based models are built on explicit mathematical equations to describe (supposedly causal) relationships between environmental drivers and ecological responses. They also often incorporate empirical relationships, particularly when knowledge is incomplete or when some processes are intentionally omitted. Processes are often represented at different nested spatiotemporal scales, depending on the underlying assumptions. Model development is the central step, typically requiring several years, yet it often remains opaque from an external perspective. The step of designing the model—translating knowledge and hypotheses into mathematical equations and parameters—is often blurred with the step of model calibration, where parameter values are inferred. Models are often treated as an accumulation of multiple submodels, each governing one or several ecological processes. Rather than being fitted as a whole, submodels are calibrated separately against specific subsets of data, and some parameters are simply prescribed (i.e. fix to a value found in the literature).

Applying the workflow to process-based models is a key for opening the black box. It would serve as a guide through the successive steps of model development. In particular, incorporating data simulation would introduce a crucial step between model building and data fitting, ensuring a clear delineation between the two and exposing strong degeneracies in the model design. This approach would force researchers to begin with a simpler version of the model, providing a clear pathway to support—or reject—the additional complexity and new parameters along the iterative development of the model. Data fitting would no longer be just a hidden aspect of model building but a step as crucial as forecasting to gain new ecological insights. The workflow could also refocus attention on the research question, defining a clear and limited context in which the model should apply. Process-based model would once again be a way to answer a research question—whereas today, model simulations have increasingly become a subject of study on their own. Finally, an universal workflow provides an opportunity to merge statistical and process-based approaches, integrating mechanistic knowledge and leveraging robust statistical approaches.

Across the different fields of ecology—for both parameter estimation and forecasting—a systematic application of a coherent workflow holds the promise to highlight the opportunities to best reduce uncertainties through new scientific insights, toward the most critical steps. This will help refocus the debate on designing new hypothesis, formulating new questions—and guiding efforts to collect new data. In a world where machine learning is rapidly advancing, there is no point of sticking to traditional methods if no changes are made. Machine learning may surpass process-based models if the latter lack a robust estimation of their parameters and fall in a complexity trap, at the cost of their interpretability. Similarly, trend analysis, when the focus is on methodological controversies (due to the lack of an iterative workflow) rather than on a robust mechanistic foundation, offers no clear advantage over machine learning.

Wrap up: how to make it happen?

References

Buschke, F.T., Hagan, J.G., Santini, L. & Coetzee, B.W.T. (2021) Random population fluctuations bias the living planet index. *Nature Ecology & Evolution* **5**, 1145–1152.

Dornelas, M., Antão, L.H., Moyes, F., Bates, A.E., Magurran, A.E., Adam, D., Akhmetzhanova, A.A., Appeltans, W., Arcos, J.M., Arnold, H., Ayyappan, N., Badihi, G., Baird, A.H., Barbosa, M., Barreto, T.E., Bässler, C., Bellgrove, A., Belmaker, J., Benedetti-Cecchi, L., Bett, B.J., Bjorkman, A.D., Błażewicz, M., Blowes, S.A., Bloch, C.P., Bonebrake, T.C., Boyd, S., Bradford, M., Brooks, A.J., Brown, J.H., Bruelheide, H., Budy, P., Carvalho, F., Castañeda-Moya, E., Chen, C.A., Chamblee, J.F., Chase, T.J., Siegwart Collier, L., Collinge, S.K., Condit, R., Cooper, E.J., Cornelissen, J.H.C., Cotano, U., Kyle Crow, S., Damasceno, G., Davies, C.H., Davis, R.A., Day, F.P., Degraer, S., Doherty, T.S., Dunn, T.E., Durigan, G., Duffy, J.E., Edelist, D., Edgar, G.J., Elahi, R., Elmendorf, S.C., Enemar, A., Ernest, S.K.M., Escribano, R., Estiarte, M., Evans, B.S., Fan, T., Turini Farah, F., Loureiro Fernandes, L., Farneda, F.Z., Fidelis, A., Fitt, R., Fosaa, A.M., Daher Correa Franco, G.A., Frank, G.E., Fraser, W.R., García, H., Cazzolla Gatti, R., Givan, O., Gorgone-Barbosa, E., Gould, W.A., Gries, C., Grossman, G.D., Gutierréz, J.R., Hale, S., Harmon, M.E., Harte, J., Haskins, G., Henshaw, D.L., Hermanutz, L., Hidalgo, P., Higuchi, P., Hoey, A., Van Hoey, G., Hofgaard, A., Holeck, K., Hollister, R.D., Holmes, R., Hoogenboom, M., Hsieh, C., Hubbell, S.P., Huettmann, F., Huffard, C.L., Hurlbert, A.H., Macedo Ivanauskas, N., Janík, D., Jandt, U., Jażdżewska, A., Johannessen, T., Johnstone, J., Jones, J., Jones, F.A.M., Kang, J., Kartawijaya, T., Keeley, E.C., Kelt, D.A., Kinnear, R., Klanderud, K., Knutsen, H., Koenig, C.C., Kortz, A.R., Král, K., Kuhnz, L.A., Kuo, C., Kushner, D.J., Laguionie-Marchais, C., Lancaster, L.T., Min Lee, C., Lefcheck, J.S., Lévesque, E., Lightfoot, D., Lloret, F., Lloyd, J.D., López-Baucells, A., Louzao, M., Madin, J.S., Magnússon, B., Malamud, S., Matthews, I., McFarland, K.P., McGill, B., McKnight, D., McLarney, W.O., Meador, J., Meserve, P.L., Metcalfe, D.J., Meyer, C.F.J., Michelsen, A., Milchakova, N., Moens, T., Moland, E., Moore, J., Mathias Moreira, C., Müller, J., Murphy, G., Myers-Smith, I.H., Myster, R.W., Naumov, A., Neat, F., Nelson, J.A., Paul Nelson, M., Newton, S.F., Norden, N., Oliver, J.C., Olsen, E.M., Onipchenko, V.G., Pabis, K., Pabst, R.J., Paquette, A., Pardede, S., Paterson, D.M., Pélissier, R., Peñuelas, J., Pérez-Matus, A., Pizarro, O., Pomati, F., Post, E., Prins, H.H.T., Priscu, J.C., Provoost, P., Prudic, K.L., Pulliainen, E., Ramesh, B.R., Mendivil Ramos, O., Rassweiler, A., Rebelo, J.E., Reed, D.C., Reich, P.B., Remillard, S.M., Richardson, A.J., Richardson, J.P., van Rijn, I., Rocha, R., Rivera-Monroy, V.H., Rixen, C., Robinson, K.P., Ribeiro Rodrigues, R., de Cerqueira Rossa-Feres, D., Rudstam, L., Ruhl, H., Ruz, C.S., Sampaio, E.M., Rybicki, N., Rypel, A., Sal, S., Salgado, B., Santos, F.A.M., Savassi-Coutinho, A.P., Scanga, S., Schmidt, J., Schooley, R., Setiawan, F., Shao, K., Shaver, G.R., Sherman, S., Sherry, T.W., Siciński, J., Sievers, C., da Silva, A.C., Rodrigues da Silva, F., Silveira, F.L., Slingsby, J., Smart, T., Snell, S.J., Soudzilovskaia, N.A., Souza, G.B.G., Maluf Souza, F., Castro Souza, V., Stallings, C.D., Stanforth, R., Stanley, E.H., Mauro Sterza, J., Stevens, M., Stuart-Smith, R., Rondon Suarez, Y., Supp, S., Yoshio Tamashiro, J., Tarigan, S., Thiede, G.P., Thorn, S., Tolvanen, A., Teresa Zugliani Toniato, M., Totland, Ø., Twilley, R.R., Vaitkus, G., Valdivia, N., Vallejo, M.I., Valone, T.J., Van Colen, C., Vanaverbeke, J., Venturoli, F., Verheye, H.M., Vianna, M., Vieira, R.P., Vrška, T., Quang Vu, C., Van Vu, L., Waide, R.B., Waldock, C., Watts, D., Webb, S., Wesołowski, T., White, E.P., Widdicombe, C.E., Wilgers, D., Williams, R., Williams, S.B., Williamson, M., Willig, M.R., Willis, T.J., Wipf, S., Woods, K.D., Woehler, E.J., Zawada, K. & Zettler, M.L. (2018) Biotime: A database of biodiversity time series for the anthropocene. *Global Ecology and Biogeography* **27**, 760–786.

Dornelas, M., Gotelli, N.J., McGill, B., Shimadzu, H., Moyes, F., Sievers, C. & Magurran, A.E. (2014) Assemblage time series reveal biodiversity change but not systematic loss. *Science* **344**, 296–299.

Díaz, S., Settele, J., Brondízio, E.S., Ngo, H.T., Agard, J., Arneth, A., Balvanera, P., Brauman, K.A., Butchart, S.H.M., Chan, K.M.A., Garibaldi, L.A., Ichii, K., Liu, J., Subramanian, S.M., Midgley, G.F., Miloslavich, P., Molnár, Z., Obura, D., Pfaff, A., Polasky, S., Purvis, A., Razzaque, J., Reyers, B., Chowdhury, R.R., Shin, Y.J., Visseren-Hamakers, I., Willis, K.J. & Zayas, C.N. (2019) Pervasive human-driven decline of life on earth points to the need for transformative change. *Science* **366**.

Franklin, O., Harrison, S.P., Dewar, R., Farrior, C.E., Brännström, A., Dieckmann, U., Pietsch, S., Falster, D., Cramer, W., Loreau, M., Wang, H., Mäkelä, A., Rebel, K.T., Meron, E., Schymanski, S.J., Rovenskaya, E., Stocker, B.D., Zaehle, S., Manzoni, S., van Oijen, M., Wright, I.J., Ciais, P., van Bodegom, P.M., Peñuelas, J., Hofhansl, F., Terrer, C., Soudzilovskaia, N.A., Midgley, G. & Prentice, I.C. (2020) Organizing principles for vegetation dynamics. *Nature Plants* **6**, 444–453.

IPBES (2019) Global assessment report on biodiversity and ecosystem services of the intergovernmental science-policy platform on biodiversity and ecosystem services. Tech. rep.

Johnson, T.F., Beckerman, A.P., Childs, D.Z., Webb, T.J., Evans, K.L., Griffiths, C.A., Capdevila, P., Clements, C.F., Besson, M., Gregory, R.D., Thomas, G.H., Delmas, E. & Freckleton, R.P. (2024) Revealing uncertainty in the status of biodiversity change. *Nature* **628**, 788–794.

Leung, B., Hargreaves, A.L., Greenberg, D.A., McGill, B., Dornelas, M. & Freeman, R. (2020) Clustered versus catastrophic global vertebrate declines. *Nature* **588**, 267–271.

Loreau, M., Cardinale, B.J., Isbell, F., Newbold, T., O'Connor, M.I. & de Mazancourt, C. (2022) Do not downplay biodiversity loss. *Nature* **601**, E27–E28.

Pilowsky, J.A., Colwell, R.K., Rahbek, C. & Fordham, D.A. (2022) Process-explicit models reveal the structure and dynamics of biodiversity patterns. *Science Advances* **8**, eabj2271.

Talis, E.J. & Lynch, H.J. (2023) Capturing stochasticity properly is key to understanding the nuances of the living planet index. *Nature Ecology & Evolution* **7**, 1194–1195.

Urban, M.C., Bocedi, G., Hendry, A.P., Mihoub, J.B., Pe'er, G., Singer, A., Bridle, J.R., Crozier, L.G., De Meester, L., Godsoe, W., Gonzalez, A., Hellmann, J.J., Holt, R.D., Huth, A., Johst, K., Krug, C.B., Leadley, P.W., Palmer, S.C.F., Pantel, J.H., Schmitz, A., Zollner, P.A. & Travis, J.M.J. (2016) Improving the forecast for biodiversity under climate change. *Science* **353**, aad8466.