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

Victor Van der Meersch[1*], James Regetz[2], T. Jonathan Davies[1,3] & EM Wolkovich[1]

July 10, 2025

[1] Forest and Conservation Sciences, University of British Columbia, Vancouver, BC V6T 1Z4, Canada

[2] National Center for Ecological Analysis and Synthesis, 1021 Anacapa St, Santa Barbara, CA 93101, United States

[3] Botany, University of British Columbia, Vancouver, BC V6T 1Z4, Canada

[*] `mailto:victor.vandermeersch@ubc.ca`

For: Workflow for Applied Data Analysis theme issue for *Phil Trans A* as an *Opinion*

Abstract

Concerns about increasing biodiversity loss and climate change have led to greater demands for useful ecological models. Datasets relevant for developing these models have also increased in size and complexity, including in their geographical, temporal and phylogenetic dimensions. New research often suggests that models accounting for these complexities can yield more accurate trends and predictions. We argue, however, that the usual workflows for model fitting in ecology make it difficult to evaluate and compare current models for several reasons. First, the research community is split between two disconnected paradigms: using data to fit simple, trend-like models with few parameters, or developing forecasting models that include complex, mechanistic submodels only indirectly informed by data. Second, in both cases, models tend to be developed not only in isolation of one another, but also without a coherent framework for linking scientific questions and understanding to statistical and modeling decisions, throughout the entire modeling process. We propose a principled workflow for end-to-end empirical model development that bridges the gap between process-based and statistical approaches, integrates sound statistical and scientific practices, and especially relies on data simulation to inform decisions at multiple steps in the process. We argue that this approach, coupled with a shift toward universal training, more open model sharing, and alignment on common datasets, could transform ecological modeling.

1 Introduction

Anthropogenic drivers are reshaping natural systems (Díaz *et al.*, 2019). Impacts are projected to increase in coming decades, as climate change accelerates biodiversity loss, altering ecosystem services and human well-being (IPBES, 2019). Implementing sustainable policies to mitigate these impacts is thus a global priority, but designing the best policies requires estimating and understanding biodiversity and ecosystem trends to date, alongside the skill to forecast future dynamics.

Meeting these policy needs has led often to two separate paths: one focused on estimating trends from new global datasets, and another focused on forecasting from generally distinct datasets or

mechanistic models based on less data. Newly available large-scale, long-term datasets have provided our first 'global' estimates of biodiversity trends (e.g. Loh *et al.*, 2005; Dornelas *et al.*, 2018), but these data—gathered opportunistically from multiple sources—are unbalanced and suffer from large geographic, temporal and taxonomic biases. Models to date have failed to fully address these challenges and, perhaps because of these limitations, are rarely if ever used for forecasting. Instead, forecasting—under different plausible scenarios—has generally relied on entirely different datasets combined with either correlative or process-based models (IPBES, 2019), with process-based models often promoted as the most realistic approach (Urban *et al.*, 2016; Pilowsky *et al.*, 2022) because they focus on mechanistic representations of ecosystem functioning. These approaches have failed to yield clear agreement on current species trends, leading to ongoing debates about the magnitude and even direction (Dornelas *et al.*, 2014; Leung *et al.*, 2020; Buschke *et al.*, 2021; Johnson *et al.*, 2024), and producing forecasts that diverge due to high model uncertainty at the ecological level (Cheaib *et al.*, 2012; Thuiller *et al.*, 2019).

The inferential goals of these two modeling paradigms—explanation vs. prediction—seem disconnected at first glance, but remains in reality much more blurred. Trend estimation, often with simpler models, tend to focus on an explanatory purpose, but often lacks a clear link with theories that could provide causal interpretation. For forecasting, the use of more complex process-based models for a predictive purpose is often justified by the assumption that "we must understand to predict", contributing to the confusion between explanation and prediction (Shmueli, 2010).

We argue that current debates, confusions and diverging forecasts are driven in large part by the incoherent and disconnected workflows used today in ecology (Loreau *et al.*, 2022; Talis & Lynch, 2023; Johnson *et al.*, 2024). Research estimating biodiversity trends has become focused on methodological aspects; the current workflow fails to examine the gap between ideal and available data, and rarely tests for predictive accuracy that could scale up to allow forecasting. At the same time, process-based models developed for forecasting often evolve through the addition of new separate layers or components. These new parts are often disconnected from the original research aim, its data stream, and the previous scientific insights, because current approaches rarely examine the model as a functioning whole and thus ignore major problems (e.g., non-identifiability, discussed below).

Workflows that fully integrate all the steps required to build a model from an ecological question, with evaluation of limitations and potential problems before estimating its parameters and making projections, could reduce many of these problems. In particular, we argue that workflows that incorporate data simulation at multiple steps can quickly identify flaws in model structure and constraints in data, and allow us to understand when, where, and why different models diverge (McElreath, 2018; Betancourt, 2020; Gelman *et al.*, 2020; Schad *et al.*, 2020; Grinsztajn *et al.*, 2021; van de Schoot *et al.*, 2021; Wolkovich *et al.*, 2024). Towards this aim, we outline the steps of a universal workflow that could harmonize both trend estimation and forecasting.

2   Scientific method and workflows

Quantitative science relies on a model-based framework to confront hypothesis with data (Chamberlin, 1965). In an idealized scientific method, we would formulate a research question and hypotheses, design an experiment accordingly, build a model, collect data, and use this data to inform our model and differentiate between hypotheses. This method underlies much of the recent pre-registration movement, where hypotheses and methodology are defined prior to data collection (Nosek *et al.*, 2018). But this idealized method often does not apply to the reality of ecological research. Many

important questions cannot be addressed through controlled experiments and replications. In such cases, we must rely on existing, heterogeneous datasets alongside uncertain and incomplete theory to provide a large-scale and long-term perspective (Hilborn & Mangel, 1997). Indeed, most macroecological insights have emerged from exploring patterns in these datasets (exploratory data analysis).
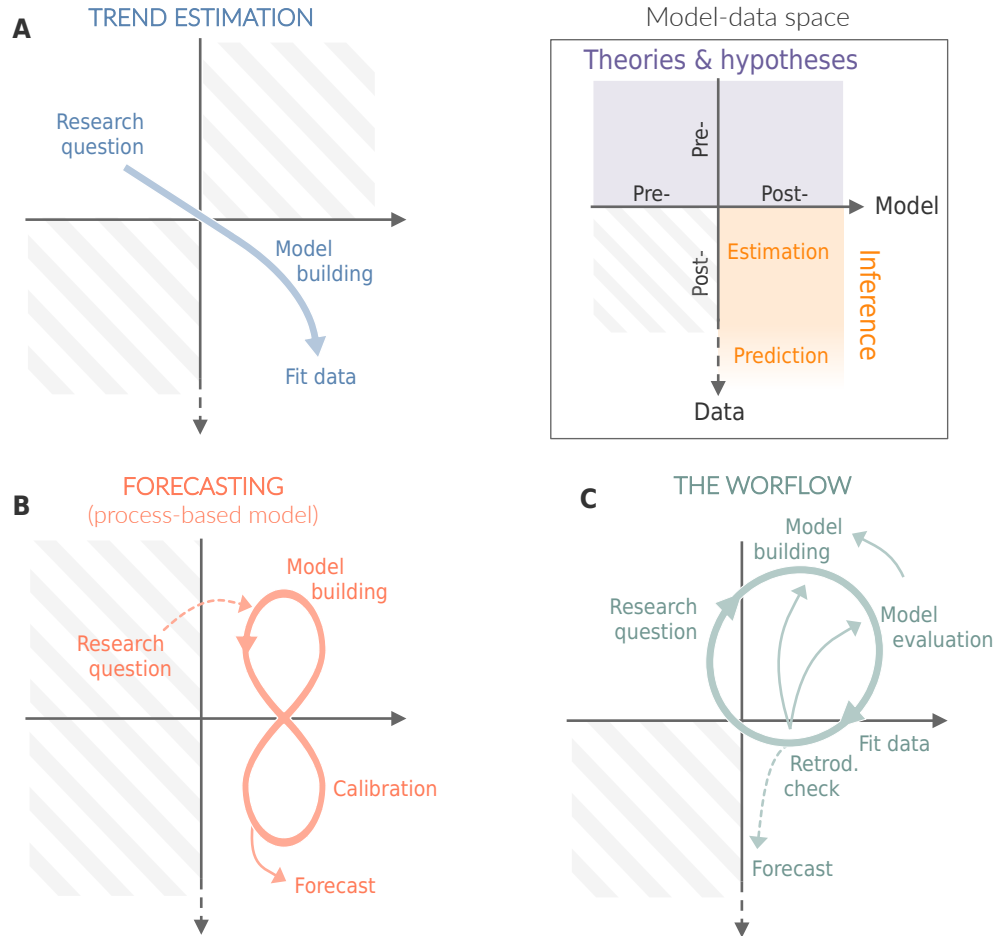
Figure 1: Different trajectories through model-data space during model development. Ideally, this begins in the pre-model/pre-data quadrant, where the research question is defined. The next step moves to the post-model/pre-data quadrant, where a model is built—independently of the data. This is where we should spend a good deal of time evaluating model behavior using simulated data. Once we are satisfied with our model, we can move to the post-model/post-data quadrant, where the model is fitted to real data. Then, we perform retrodictive checks to compare model predictions to observations, which likely give us some feedbacks to refine our model.

This reality should drive researchers to use more robust and coherent methods. But the current workflows combined with the challenges ecologists are facing—both in terms of data complexity and societal needs—instead may lead to persistent problems. Trend estimation has focused mostly on fitting a model to empirical data in a uni-directional way—without the checks and likely feedbacks that often highlight uncertainty and related limitations in the model and/or data (figure 1a). For forecasting, researchers have focused on making predictions with increasingly complex mechanistic models, frequently obscuring the steps underlying model building and parameterization (figure 1b). Researchers often calibrate the different parts of these models separately, and fix some parameter

values based on experiments and expert knowledge, to avoid problems when trying to fit the model as a whole. Addressing these problems while accounting for the realities of working with ecological data requires a more comprehensive workflow.

We argue a workflow that moves along the data-model space in a coherent sequence of steps (figure 1c) could reduce many of these problems and thus improve ecological science. The first step of this workflow is to define an explicit research question and formulate hypotheses (step 1, figure 2). This involves making clear assumptions about the most influential drivers, within the specific context of our study. This should guide the construction of a narrative model of how we believe the system works, focusing on the mechanisms that could generate the data we observe, including the observational error. This step of carefully formulating hypotheses has gained increasing attention in the ecological literature (Grace & Irvine, 2020). From this narrative, we can then develop a mathematical model—an ensemble of equations that encapsulates our knowledge and is designed to answer our research question (step 2, figure 2). The general idea is to start with a relatively simple model that we could refine later (see example workflow we provide). At this stage, prioritizing biologically meaningful parameters is crucial, as it allows us to have a sense of plausible parameter values. This means choosing a model formulation where each parameter corresponds to an interpretable behavior (which sometimes requires considering alternative parameterizations).

With a model in place, the next step focuses on testing and understanding it via data simulation (step 3, figure 2). 'Fake' or 'test' data are generated directly from the model by fixing parameters to some reasonable values (which is straightforward if the parameters are interpretable) and from fake predictor data. We then fit our model to this simulated dataset and evaluate its ability to recover the prescribed values. At this stage, the focus is on understanding the model, so we may need to spend time thinking about whether each parameter is realistic or not. This is also a step for making sure the model is working as expected (see example worklow). Ideally, this data simulation step should be repeated several times with different parameter values, sampled within a reasonable range. In a Bayesian framework, simulated parameters can be sampled from the prior distribution (allowing us to check the implications of our priors, Gelman *et al.*, 2020).

Once we are confident about our model structure, we can introduce real data as part of an initial model fitting step (step 4, figure 2). This way, we obtain parameter estimates constrained by observations. These parameter estimates lead to the second data simulation step, this time using our fitted model parameters to generate predictions (step 5, figure 2). This—which we call a retrodictive check (Betancourt, 2020, also called a posterior predictive check in a Bayesian workflow, Gelman *et al.*, 2020)—allows model output to be compared to observations. There is no general way to do this: the goal is to tailor checks that assess key features of the data in regards to our research question (Gelman *et al.*, 2020). It's only once all steps have been completed that we can interpret parameter values with respect to our research question. The workflow encourages a focus on the full model, where any parameter (such as a trend estimate) must be carefully interpreted alongside others, as all are fundamental components that shape both inference and forecasting.

Within such a workflow, forecasting emerges as a natural outcome: rather than being a final goal, it only involves jointly modeling new circumstances along with the original data. The adoption of the workflow in macroecological studies—where model building is often informed by patterns in the data—would also make the exploratory analysis more transparent (as an explicit preliminary step for entering the workflow) and would compel researchers to more clearly develop a research question before extensive model fitting.

A key feature of this workflow is the central role of data simulation, which introduces two feedback loops—and should be applied not only in a Bayesian framework but across broader modeling

approaches. The first feedback arises when we evaluate the model on simulated data. The failure of the model to recover known parameter values and handle the complexity of the simulated data should prompt reconsidering the model, or even reformulating the research question. Further, this step might reveal that some parameters are highly non-identifiable (meaning the parameter(s) cannot be uniquely estimated), flagging the need to change the model structure—before incorporating empirical observations. The second feedback loop comes from the retrodictive check. Discrepancies here may indicate a missing key driver, and suggest the current model is too simplistic. We can refine the model to integrate the missing process(es) (if we can identify them) and return to the start of the workflow. Insights from the retrodictive check can also lead us to introduce additional complexity when simulating fake data, such as phylogenetic structure or observational biases (e.g. unbalanced data). For example, if a researcher realizes their empirical data is geographically biased, this bias should be built into the model and thus into this data simulation step. This iterative evaluation of the model moves beyond a simple reliance on goodness-of-fit metrics. At each iteration, we are able to evaluate the model behavior, both with simulated and real data, taking into account our expert knowledge of the ecological processes.

3 The workflow in practice

Across the different fields of ecology—for both parameter estimation and forecasting—a systematic application of a coherent workflow could highlight the best opportunities to accurately quantify uncertainties and yield new scientific insights. This will help refocus the debate on designing new hypotheses, formulating new questions—and guiding efforts to collect new data. Here, we illustrate how such a workflow could lead to significant improvements in two case studies: (i) estimating global biodiversity trends and (ii) forecasting future species and ecosystem dynamics using process-based models.

3.1 Trends

Ecologists today have amassed data on populations and species across the globe; they have also engaged in an increasing number of debates on regional and global trends over time, with arguments over the magnitude and even direction of population and biodiversity metrics (Dornelas *et al.*, 2014; Leung *et al.*, 2020; Terry *et al.*, 2022; Müller *et al.*, 2024). While shifting estimates are part of the process of science—refining our approaches and thus estimates over time—we believe these debates would be fewer and they would be more rapidly resolved through use of an improved workflow.

An improved workflow that required data simulation and retrodictive checks would lead to larger model advances and a greater recognition of uncertainty—thus highlighting likely consistency in estimates across models—that could better aid policy. Using the workflow would make it more obvious that what now appear as major discrepancies are shifts in point estimates that are generally all in the same uncertainty space (Johnson *et al.*, 2024)—and it would challenge modelers to show major predictive advances, which is not currently part of the process. Explanatory power in most models of observational data is usually very low (Low-Décarie *et al.*, 2014; Møller & Jennions, 2002) and thus tests of models' predictions rarely expected. But the workflow highlights that predictions from the model—what we call retrodictive checks—are part of the process of science, and critical to testing for what may be missing in a model. We expect retrodictive checks on most published trend analyses would highlight major missing components in these models, and drive changes both in the models themselves and in the simulated data to check the models. This step builds somewhat

on the skills needed for null models (Gotelli & Ulrich, 2012), but with a shift in focus towards the specific bespoke model at hand. Ecologists have started to use simulated data more to understand potential limitations of their models and data combined, but this is still extremely rare, and efforts to date often treat simulations as separate from the statistical model (Buschke *et al.*, 2021; Dove *et al.*, 2023), short-circuiting their full utility if used in an iterative workflow.

Applying the workflow to current trend estimates could importantly highlight the best way to improve data collection for more reliable estimates. Returning to the example of a global estimate of trends in vertebrate populations of species over time (see Box) and applying our proposed workflow would mean more efforts to define the goal and question—is it a simple global estimate? Or a need to also find which species are declining most, including those that may have poor or no data? From there a generative model using simulated data for testing could incorporate many aspects of the populations, and data, that are often only included in 'null' or 'synthetic data generation' currently (Buschke *et al.*, 2021; McRae *et al.*, 2025) but could be built into the models fit to the empirical data. Eventually fitting the empirical data and performing retrodictive checks would likely highlight major missing components of the generative model and, ultimately, this would help inform our global estimates of mean trends (see workflow example). For example, certain populations are recovering for very specific reasons (e.g., elephants in regions where the ivory trade drove declines in the past) that perhaps should be modeled. From this model, what data are most critically needed to address the updated aims would become clearer and could drive new data collection (Toszogyova *et al.*, 2024).

3.2 Forecasting from process-based models

Ecological forecasting is a broad field with a diverse range of methods. Our second case-study focuses on process-based modeling, which is often considered the gold standard for forecasting in ecology (Urban *et al.*, 2016; Pilowsky *et al.*, 2022) and beyond. Newer models, however, generally incorporate greater complexity (and an ever-growing number of parameters), which can make it difficult to increase scientific understanding (Franklin *et al.*, 2020), and suggest or model potential policies.

Increasing model complexity can be beneficial, especially when it helps better quantify the uncertainty; however, this is not always the outcome. In process-based modeling for climate forecasts, the uncertainty range on the effect of increasing CO_2 concentration on temperature have remained largely unchanged (Zelinka *et al.*, 2020). This has driven calls for more rigorous and transparent calibration processes (Balaji *et al.*, 2022). Similar concerns arise in ecology, where strong disagreements exist about the effect of climate change on future species distributions (Cheaib *et al.*, 2012) and ecosystem dynamics (Lovenduski & Bonan, 2017). These uncertainties have large implications beyond ecology, as they influence simulations of biosphere-atmosphere interactions and, ultimately, future climate projections (Bonan & Doney, 2018; Simpson *et al.*, 2025). Some researchers now advocate for simplifying models, to avoid over-parametrization when the data provide little information to constrain some parameters (Wang *et al.*, 2017; Harrison *et al.*, 2021). If a model becomes too complex, it may become a black box, and understanding the sources of uncertainty and how they propagate through the model may become nearly impossible (see Box). Each additional process and parameter can increase overall uncertainty to the point where model projections lose their usefulness for decision makers (Saltelli *et al.*, 2020).

Process-based models used to project species or ecosystem distributions highlight some of these problems. Most of the focus is on the model projections—often represented as maps without

uncertainties. Model evaluation generally focuses on how well these projections match observed distributions, sometimes under different climatic conditions to challenge the models (Van der Meersch *et al.*, 2025)—most of the time evaluating only one of the many different output variables of the model. The complexity of these models can make it difficult to assess how well the potential parameter space was searched, and whether there was any potential non-identifiability—such work is generally done as a separate effort, exposing potential model problems later on (Van der Meersch & Chuine, 2025), and not required as a preliminary step.
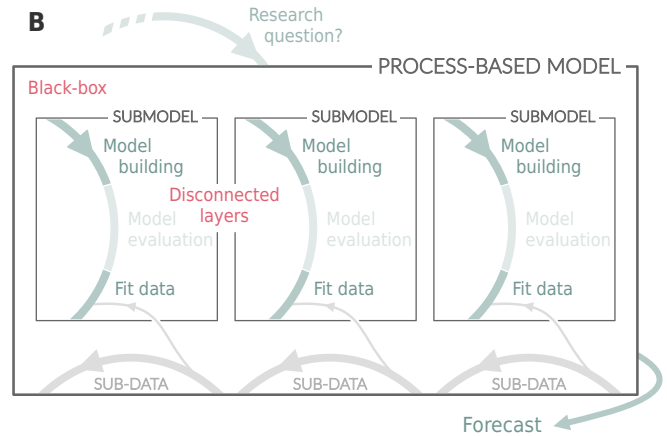
Applying the workflow to process-based models would help open the black box. Each successive step of model development in the workflow may highlight current problems and a path to solutions. Incorporating data simulation would introduce a crucial step between model building and data fitting (also called calibration), ensuring a clear delineation between the two, and help expose potential parameter identifiability issues in the model design. Uncovering identifiability issues would likely force researchers to begin with a simpler version of the model, which they could build on iteratively, testing for support—or lack thereof—when adding model complexity. The workflow may also highlight model problems by requiring more explicit model calibration (i.e. data fitting), which is currently hidden within opaque 'model building', making it easy to hide non-identifiability in the model. Through the workflow non-identifiable parts of models could be addressed by reformulating the mathematical structure of certain processes, or finding ways to apply additional constraints (e.g., narrower ranges for certain parameters or developing new hypotheses that target the appropriate level of complexity). The resulting process-based models would likely be simpler and thus more tractable for quantifying parameter uncertainty and propagating it through projections.

Trend estimation

Mechanistic forecasting
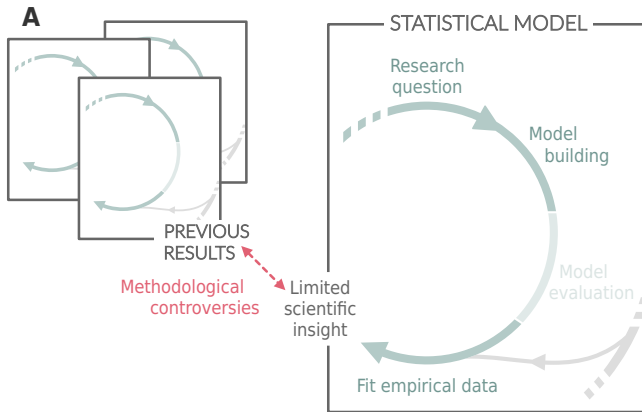
In the current workflow for estimating trends over time, a new model with a new estimate often leads to a paper (see figure A above) because ecologists spend less time interrogating their models with simulated data, or their model performance fit to empirical data. The Living Planet Index (LPI, `www.livingplanetindex.org`), which aims to include long-term data on vertebrate populations of species across the globe, is emblematic of these conflicting results. With updated data released semi-annually alongside new datasets and new estimates of decline, a growing number of high-profile papers have challenged how strong the evidence is for population decline (Dornelas *et al.*, 2014; Gonzalez *et al.*, 2016; Wagner *et al.*, 2021; Müller *et al.*, 2024), with each paper taking a slightly different analytical approach. For examples using LPI data, Leung *et al.* (2020) published a mixture model that suggested most populations were not significantly declining, followed by other alternative modeling approaches (Buschke *et al.*, 2021; Puurtinen *et al.*, 2022) including a recent one suggesting a basic analysis of the dataset should always include three sources of autocorrelation, finding trends that encompassed many previous results (Johnson *et al.*, 2024).

Model development is the central step of the process-based workflow, typically requiring several years, yet it often remains opaque for anyone who has not worked the model itself. The step of designing the model—translating knowledge and hypotheses into mathematical equations and parameters—is often blurred with the step of model calibration (or tuning), where parameter values are inferred. Models are often treated as an accumulation of multiple submodels, each governing one or several ecological processes (see figure B above). Rather than being fitted as a whole, submodels are calibrated separately against specific subsets of data, and some parameters are simply prescribed (i.e., fixed to a value found in the literature) or tuned to reproduce certain observations or theory. The way models are currently calibrated is likely not a coincidence, but rather a workaround to avoid confronting the full complexity of the model. Calibrating submodels separately allows to avoid the fact that, if the model were fitted as a whole, many parameters would compensate for one another—revealing structural degeneracies and making the model far more difficult to use.

A common workflow to bridge trend estimation and forecasting
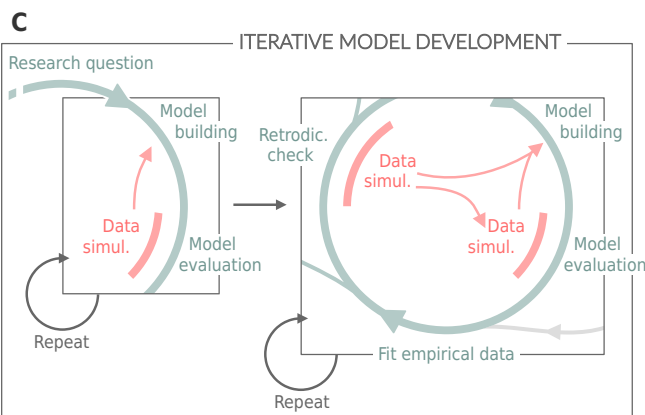
A universal workflow offers an opportunity to bridge statistical and process-based frameworks, integrating mechanistic knowledge and leveraging robust statistical approaches (e.g. Rounce *et al.*, 2020). Process-based models would no longer be perceived as deterministic black boxes by other researchers but rather as robust statistical frameworks encapsulating both data structure and mechanistic knowledge—and where the full model is fitted jointly. It would be an opportunity to spread the incorporation of mechanistic assumptions beyond the process-based modeling community, potentially improving trend estimates. For both trend estimation and forecasting, the workflow would refocus attention on the research question, highlighting the ecological hypotheses that justify the use and design of the model.
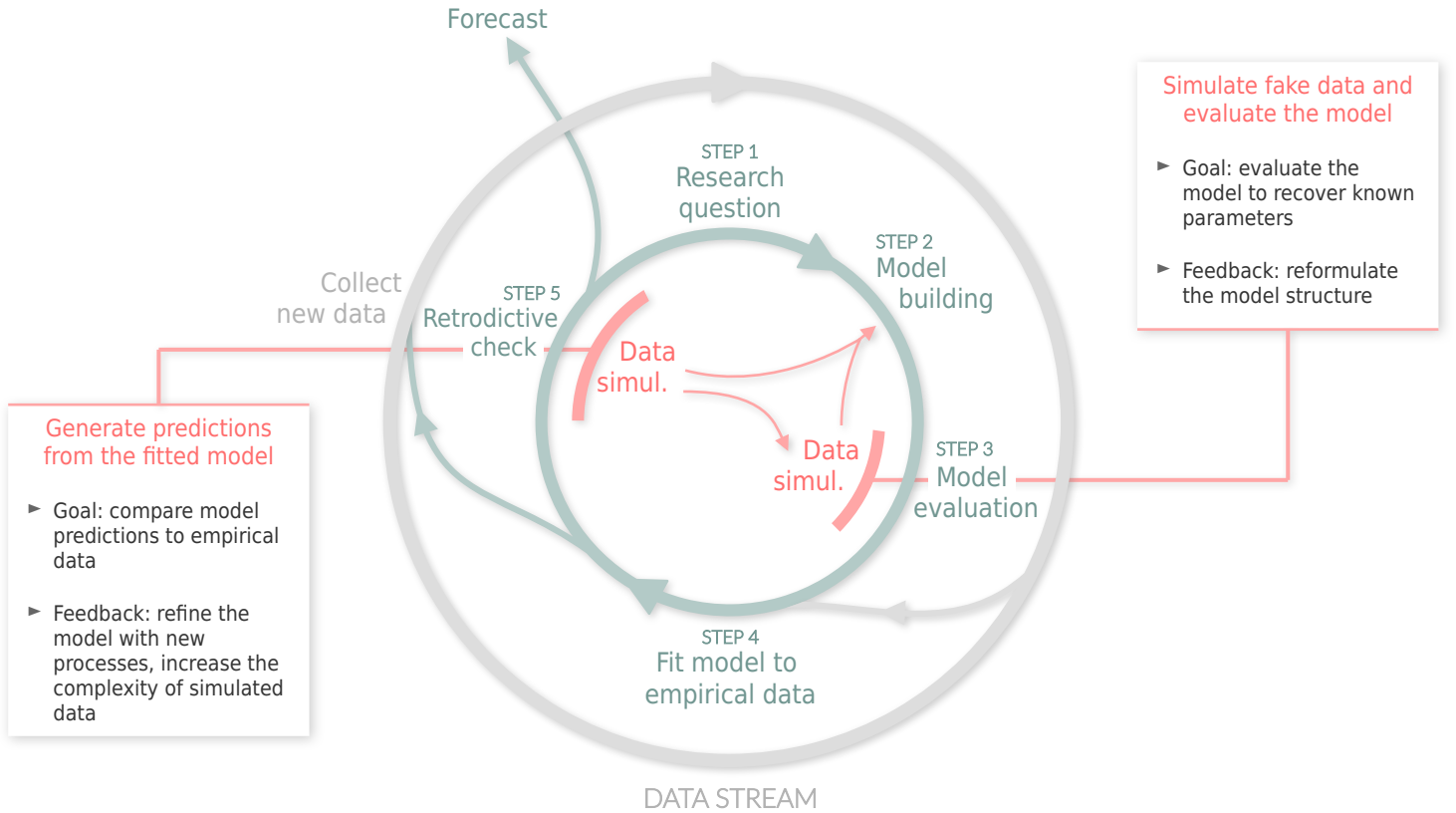
Figure 2: The workflow we propose here (which builds from recent advances in workflows, Betancourt, 2020; Gelman *et al.*, 2020; Schad *et al.*, 2020; Grinsztajn *et al.*, 2021; van de Schoot *et al.*, 2021; Wolkovich *et al.*, 2024) focuses on iterative feedbacks between the research question, model and data. Most research should start at Step 1, with the research question, followed by extensive model building and evaluation using simulated data (Steps 2 and 3) before proceeding to fitting the model on empirical data and examining it through retrodictive checks (Steps 4 and 5). Within this workflow, forecasting is a natural output of the process and not a separate process or one available for only certain modeling approaches. The integration of observations (the data stream) occurs at Step 4—only after the model has been thoroughly evaluated—and can also highlight opportunities to collect new data and enrich the data stream.

4 Barriers and opportunities

We believe our workflow could help advance ecological science and its applications, but widespread use of it requires overcoming major hurdles that pervade science. The first is pressure to publish quickly, which can be at odds with the reality that good model fitting is inherently iterative and takes time. The second is reluctance to embrace adoption of open science practices that ensure modeling efforts are fully transparent and reproducible. We believe adopting a workflow such as the one we propose may change this reluctance, and aligns with ongoing efforts to place increasing value on research that is carefully developed, openly collaborative (including both data and code), and transparent about areas of uncertainty.

4.1 Adopting the workflow

Advancing ecology to where most researchers use models built more flexibly from ecological theory and insights applied to their ecological systems will not happen rapidly without a major shift in training. Much of ecology still divides the world into training for those who gather data and learn a limited set of pre-built models versus those who develop more complex models. In ecological training today, researchers who conduct field and lab studies often learn a limited set of particular statistical tests matched to particular experiment designs and simple information on their variable types (e.g., categorical x and y leads to using a chi-squared test).

When ecologists trained in a limited set of tests require more complex models, they are expected to collaborate with others trained more in model development (though often for highly specific applications, such as wildlife population estimates, where generative models are often predefined). These two groups further differ from process-based modelers, who often train in physical and eco-physiological processes and how to abstract them into mainly deterministic models. Neither group has fully integrated data simulation into their statistical or scientific workflows. Simulation is generally reserved as a form of training needed mainly by those specializing in theoretical ecology, who often solve analytical equations but rarely link to empirical data. While specialization is valuable, we argue the fundamental separation in ecology has overly-siloed these groups and prevented more rapid progress.

We suggest unified training in our proposed workflow, or a similar one (Betancourt, 2020; Gelman *et al.*, 2020; Grinsztajn *et al.*, 2021; van de Schoot *et al.*, 2021), would focus on learning to generate questions and then models, and then how to simulate data from them. Through this and the use retrodictive checks, most ecologists would be better equipped to think through what parameters are most critical to their question and/or aim (e.g., management), and also gain a much stronger connection to the level of uncertainty in many of ecological estimates. Empiricists would be more likely to recognize critical gaps in current models fit by those specializing in ecological modeling and help advance those models. Process-based modelers may start a new generation of simpler models that are more tractable to theoretical ecologists, who may see new bridges from their work to empirical data and forecasting.

4.2 A tractable alternative to machine learning

With rapid advances in machine learning, improving current methods to gain greater scientific insights to drive better forecasting seems increasingly important. Machine learning will likely surpass process-based models for forecasting accuracy if the latter lack a robust estimation of their parameters and fall in a complexity trap, at the cost of their interpretability. Similarly, estimates

of trends from empirical data using models without clear mechanistic drivers may soon offer fewer advantages over machine learning.

Beyond improving model building and evaluation, our proposed workflow also has the potential to shift how process-based models are perceived, particularly by those unfamiliar with them. Process-based models could once again be a way to answer a research question—whereas today, model simulations have increasingly become a subject of study on their own. Ideally, applying the workflow would help to move away from the traditional process-based model paradigm, where parameters are typically assigned fixed values without properly accounting for their uncertainty. Instead, it would guide a step-by-step model fitting, parameter estimation, and uncertainty quantification—preventing modelers from making biased inferences and unfounded assumptions beyond what the data can support. It would thus define a clear and limited context in which the model should apply, and limit discussion of adding increasing complexity.

Conclusions: Today model development in ecology is rarely transparent, which limits how easily the research community can understand models, and thus identify potential issues. Instead of broad inclusive conversations about how to improve models to advance our ecological understanding, a significant portion of scientific debate has become mired in methodological considerations. However, we believe our workflow provides a tractable step to fixing this. By focusing on model development more tightly tied to ecological expertise, we argue this workflow should broaden the community that contributes to model development. It may also help resolve apparent conflicts by identifying where divergences in model predictions emerge, whether from differences in assumptions, model structures, or other aspects of the modeling process. As ecologists are increasingly expanding their computational toolkits, many field, and lab, and other forms of 'empirical' ecologists have the basic tools to follow this workflow to build models that better represent their ecological domains of interest, and—most importantly—to interrogate those models.

References

Balaji, V., Couvreux, F., Deshayes, J., Gautrais, J., Hourdin, F. & Rio, C. (2022) Are general circulation models obsolete? *Proceedings of the National Academy of Sciences* **119**, e2202075119.

Betancourt, M. (2020) Towards A Principled Bayesian Workflow. `https://betanalpha.github.io/assets/case_studies/principled_bayesian_workflow.html`.

Bonan, G.B. & Doney, S.C. (2018) Climate, ecosystems, and planetary futures: The challenge to predict life in earth system models. *Science* **359**.

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.

Chamberlin, T.C. (1965) Method of multiple working hypotheses. *Science* **148**, 754.

Cheaib, A., Badeau, V., Boe, J., Chuine, I., Delire, C., Dufrêne, E., François, C., Gritti, E.S., Legay, M., Pagé, C., Thuiller, W., Viovy, N. & Leadley, P. (2012) Climate change impacts on tree ranges: model intercomparison facilitates understanding and quantification of uncertainty. *Ecology Letters* **15**, 533–544.

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**.

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.

Dove, S., Böhm, M., Freeman, R., McRae, L. & Murrell, D.J. (2023) Quantifying reliability and data deficiency in global vertebrate population trends using the living planet index. *Global Change Biology* **29**, 4966–4982.

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.

Gelman, A., Vehtari, A., Simpson, D., Margossian, C.C., Carpenter, B., Yao, Y., Kennedy, L., Gabry, J., Bürkner, P.C. & Modrák, M. (2020) Bayesian workflow.

Gonzalez, A., Cardinale, B.J., Allington, G.R., Byrnes, J., Arthur Endsley, K., Brown, D.G., Hooper, D.U., Isbell, F., O'Connor, M.I. & Loreau, M. (2016) Estimating local biodiversity change: a critique of papers claiming no net loss of local diversity. *Ecology* **97**, 1949–1960.

Gotelli, N.J. & Ulrich, W. (2012) Statistical challenges in null model analysis. *Oikos* **121**, 171–180, gotelli, Nicholas J. Ulrich, Werner.

Grace, J.B. & Irvine, K.M. (2020) Scientist's guide to developing explanatory statistical models using causal analysis principles. *Ecology* **101**.

Grinsztajn, L., Semenova, E., Margossian, C.C. & Riou, J. (2021) Bayesian workflow for disease transmission modeling in stan. *Statistics in Medicine* **40**, 6209–6234.

Harrison, S.P., Cramer, W., Franklin, O., Prentice, I.C., Wang, H., Brännström, A., de Boer, H., Dieckmann, U., Joshi, J., Keenan, T.F., Lavergne, A., Manzoni, S., Mengoli, G., Morfopoulos, C., Peñuelas, J., Pietsch, S., Rebel, K.T., Ryu, Y., Smith, N.G., Stocker, B.D. & Wright, I.J. (2021) Eco-evolutionary optimality as a means to improve vegetation and land-surface models. *New Phytologist* **231**, 2125–2141.

Hilborn, R. & Mangel, M. (1997) *The Ecological Detective: Confronting Models with Data.* Princeton University Press.

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.

Loh, J., Green, R.E., Ricketts, T., Lamoreux, J., Jenkins, M., Kapos, V. & Randers, J. (2005) The living planet index: using species population time series to track trends in biodiversity. *Philosophical Transactions of the Royal Society B: Biological Sciences* **360**, 289–295.

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.

Lovenduski, N.S. & Bonan, G.B. (2017) Reducing uncertainty in projections of terrestrial carbon uptake. *Environmental Research Letters* **12**, 044020.

Low-Décarie, E., Chivers, C. & Granados, M. (2014) Rising complexity and falling explanatory power in ecology. *Frontiers in Ecology and the Environment* **12**, 412–418.

McElreath, R. (2018) *Statistical Rethinking: A Bayesian Course with Examples in R and Stan.* Chapman and Hall/CRC.

McRae, L., Cornford, R., Marconi, V., Puleston, H., Ledger, S.E., Deinet, S., Oppenheimer, P., Hoffmann, M. & Freeman, R. (2025) The utility of the living planet index as a policy tool and for measuring nature recovery. *Philosophical Transactions B* **380**, 20230207.

Møller, A. & Jennions, M.D. (2002) How much variance can be explained by ecologists and evolutionary biologists? *Oecologia* **132**, 492–500.

Müller, J., Hothorn, T., Yuan, Y., Seibold, S., Mitesser, O., Rothacher, J., Freund, J., Wild, C., Wolz, M. & Menzel, A. (2024) Weather explains the decline and rise of insect biomass over 34 years. *Nature* **628**, 349–354.

Nosek, B.A., Ebersole, C.R., DeHaven, A.C. & Mellor, D.T. (2018) The preregistration revolution. *Proceedings of the National Academy of Sciences* **115**, 2600–2606.

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.

Puurtinen, M., Elo, M. & Kotiaho, J.S. (2022) The living planet index does not measure abundance. *Nature* **601**, E14–E15.

Rounce, D.R., Khurana, T., Short, M.B., Hock, R., Shean, D.E. & Brinkerhoff, D.J. (2020) Quantifying parameter uncertainty in a large-scale glacier evolution model using bayesian inference: application to high mountain asia. *Journal of Glaciology* **66**, 175–187.

Saltelli, A., Bammer, G., Bruno, I., Charters, E., Di Fiore, M., Didier, E., Nelson Espeland, W., Kay, J., Lo Piano, S., Mayo, D., Pielke Jr, R., Portaluri, T., Porter, T.M., Puy, A., Rafols, I., Ravetz, J.R., Reinert, E., Sarewitz, D., Stark, P.B., Stirling, A., van der Sluijs, J. & Vineis, P. (2020) Five ways to ensure that models serve society: a manifesto. *Nature* **582**, 482–484.

Schad, D.J., Betancourt, M. & Vasishth, S. (2020) Toward a principled bayesian workflow in cognitive science.

Shmueli, G. (2010) To explain or to predict? *Statistical Science* **25**.

Simpson, I.R., Shaw, T.A., Ceppi, P., Clement, A.C., Fischer, E., Grise, K.M., Pendergrass, A.G., Screen, J.A., Wills, R.C., Woollings, T. *et al.* (2025) Confronting earth system model trends with observations. *Science advances* **11**, eadt8035.

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.

Terry, J.C.D., O'Sullivan, J.D. & Rossberg, A.G. (2022) No pervasive relationship between species size and local abundance trends. *Nature Ecology & Evolution* **6**, 140–144.

Thuiller, W., Guéguen, M., Renaud, J., Karger, D.N. & Zimmermann, N.E. (2019) Uncertainty in ensembles of global biodiversity scenarios. *Nature Communications* **10**.

Toszogyova, A., Smyčka, J. & Storch, D. (2024) Mathematical biases in the calculation of the living planet index lead to overestimation of vertebrate population decline. *Nature Communications* **15**, 5295.

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.

van de Schoot, R., Depaoli, S., King, R., Kramer, B., Maertens, K., Tadesse, M.C., Vannucci, M., Gelman, A., Veen, D., Willemsen, J. & Yau, C. (2021) Bayesian statistics and modelling. *Nature Reviews Methods Primers* **1**.

Van der Meersch, V., Armstrong, E., Mouillot, F., Duputié, A., Davi, H., Saltré, F. & Chuine, I. (2025) Paleorecords reveal biological mechanisms crucial for reliable species range shift projections amid rapid climate change. *Ecology Letters* **28**.

Van der Meersch, V. & Chuine, I. (2025) Can inverse calibration help improving process-explicit species distribution models? *Ecological Modelling* **506**, 111132.

Wagner, D.L., Grames, E.M., Forister, M.L., Berenbaum, M.R. & Stopak, D. (2021) Insect decline in the anthropocene: Death by a thousand cuts. *Proceedings of the National Academy of Sciences* **118**, e2023989118.

Wang, H., Prentice, I.C., Keenan, T.F., Davis, T.W., Wright, I.J., Cornwell, W.K., Evans, B.J. & Peng, C. (2017) Towards a universal model for carbon dioxide uptake by plants. *Nature Plants* **3**, 734–741.

Wolkovich, E., Davies, T.J., Pearse, W.D. & Betancourt, M. (2024) A four-step bayesian workflow for improving ecological science.

Zelinka, M.D., Myers, T.A., McCoy, D.T., Po-Chedley, S., Caldwell, P.M., Ceppi, P., Klein, S.A. & Taylor, K.E. (2020) Causes of higher climate sensitivity in CMIP6 models. *Geophysical Research Letters* **47**.