How to fit Bayesian models and influence people *or* How to do Bayesian model fitting in ecology *or* The best way to be a Bayesian in ecology today

EM Wolkovich and WD Pearse and TJ Davies and M Betancourt?

August 26, 2023

Abstract

Improvements in algorithms and computational speed have heralded a new era of Bayesian model fitting. Models are easier to fit, faster to run, and more flexible, making them ideal for many of the complexities of ecological sampling designs. This has opened up Bayesian approaches to a new world of users, but often ecologists are not aware of some of the advantages now offered. Ecologists frequently use the 'same-old' tricks, approaches, and work-arounds that served them well for frequentist methods, and are not taking advantage of some of the new possibilities on offer. Here we present a Bayesian workflow—centering on model and data simulation —that represents a ground shift of not just how to do fit Bayesian models, but how we should approach statistical model fitting to advance our science. This workflow integrates mechanistic and statistical models with a computational toolkit, which we argue would accelerate ecological science. Steps of this workflow will often highlight current limits of our inference and suggest paths forward, as they provide a clear method to test how uncertain models are given small sample sizes, or how often a poor model may appear 'best' given certain diagnostics. By interrogating our methods through generative models and data from such models, we can refine where to put resources for better estimates, better models, and better forecasts. While we outline these methods from the happy surf of our Bayesian beach, the ocean we describe works for anyone fitting models to ecological data.

1 Main text

Recent years have seen the explosion of Bayesian models across fields (van de Schoot *et al.*, 2021; Schad *et al.*, 2021; Grinsztajn *et al.*, 2021), including ecology. This change comes in part from increased computational power, but more from new algorithms (e.g. Hamiltonian Monte Carlo, Hoffman & Gelman, 2014; Betancourt, 2019) that have made Bayesian models faster and arguably easier to fit (Carpenter *et al.*, 2017). Capitalizing on these advances R packages, such as brms, have seen much wider use than the previous generation of packages that streamlined fitting a set of pre-determined models using Bayesian approaches (e.g. MCMCglmm). Bayesian approaches, long heralded as more powerful, flexible and especially adept at capturing the multiple levels of variance in ecological data, seem poised to help ecology as a discipline advance in leaps and bounds, and to avoid contributing to the replication crisis.

Ecology itself also seems poised to apply a century of theory and empirical insights to solve the global challenges society is placing upon it. Policy-makers demand predictive models of how communities and ecosystems will shift with climate change, habitat degradation, and other anthropogenic forces to balance conflicting societal needs and desires. Ecology, now three decades on after the start of the synthesis movement, has greatly advanced its statistical and computational approaches to test fundamental theory across systems (Hampton *et al.*, 2013). Early concerns have given way to a widespread appreciation of the value of these new approaches (alongside a resounding and bizarre fervor over 'natural history'). At the same time, growing anthropogenic challenges have increased the need for predictive models and forecasts from ecologists. Policy-makers are faced with so-called 'wicked' problems that cannot be simply solved, and so they need forecasts that communicate uncertainty so that they can communicate the inherent risks and trade-offs in decision-making. [JD adds: models that embrace uncertainty, something we still find hard to communicate well, and variability vs mean responses...]

The result is that the average ecologist today has a diversified skillset compared to that of decades ago. Ecology has long been a field with deep statistical training, but the modern ecologist is now also expected to be computational: to handle large datasets, produce repeatable workflows, and translate models into forecasts for policy and planning. Many ecologists now bridge field, lab and computational methods. Bayesian methods can help facilitate this new job description.

Bayesian approaches are not new to ecology (indeed they are over twice as old as the field of ecology itself), and they have long been used in certain subdisciplines related to estimating population sizes of things people want to eat or manage. For example, wildlife biologists frequently use mark-recapture data and associated models to estimate population sizes; such methods almost always use Hidden Markov models, which can rarely be fit without Bayesian methods (Muthukumarana *et al.*, 2008; Zheng *et al.*, 2007). Similarly, fisheries biology has tended towards more complex models—such as state-space models—to estimate stocks, which are similarly easiest to fit with Bayesian approaches (Trijoulet *et al.*, 2018; Millar & Meyer, 2000).

However, away from these fields, ecologists have largely focussed on frequentist statistics and null hypothesis testing. It's easy to see why: they are analytically straightforward (high-schoolers can carry out t-tests and ANOVAs armed only with pen, paper, and patience) and centuries of effort have led to standardised ways to many questions. However, such one-size-fits-all approaches can lead underlying assumptions on the assumed distribution of data to be ignored, spatial and phylogenetic data introduce non-independencies, and in large ecological datasets we often fit multi-way interaction terms that make main effects hard to interpret, or random effects to correct for group level factors. Most concerningly, as datasets become larger null hypotheses tend to be rejected even when the underlying effect sizes are weak. The modern frequentist statistician is now well-versed in the dangers and mitigations for such issues, but arguments for simplicity start to become weak.

Today, as large-scale ecological data becomes available for more diverse systems and for questions addressing other aims, ecologists are using Bayesian models in new ways. Yet, there is naturally a lag in training in statistics, and in Bayesian modeling in particular. Bayesian approaches provide a pathway to powerful models that can transform how we understand our systems, but they can also lead to pitfalls most ecologists are not trained to notice or deal with. Many of these pitfalls can be avoided by approaching Bayesian analyses through specific workflows (Betancourt, 2020; van de Schoot *et al.*, 2021), which themselves are built on a process of how to do not just stats, but how to do science.

Here we describe a broadly generalizable workflow for Bayesian analysis and show how it can revolutionize training in ecology by integrating more model building and model understanding. Once you start doing this workflow, your scientific life will never be the same.

1.1 A bluffer's guide to what Bayesian statistics is

Many of us have some intuition as to what a probability is, and for many of us it's "the long-term frequency with which something happens". We would expect, if we tossed a coin 100 times, that we would get roughly 50 heads: we would say that the probability of tossing a coin and getting a 'head' is $\frac{50}{100}$, which is 50% or $\frac{1}{2}$. We wouldn't be very surprised if we got 49 or 55 heads, but we would be surprised if we got 99.

This definition of probability—which is the *frequentist* definition—is useful in many situations, but it has a few disadvantages. the first is frequentist definitions aren't very useful when dealing with totally unexpected situations. Frequentist probabilities are grounded in *data*, and as such if we don't have a lot of data we struggle to generate even estimates of our uncertainty. We don't, for example, have an idea what the probability of life on Mars is because we've only looked at one planet, and so we don't have a long-term frequency with which, when we're exploring planets, we find life. Similarly, graduate students are often sent out into places where no one has worked before, which can make it difficult to figure out whether what they see in their first year of data collection is unusual, or to make predictions for their second year. Thus it would be nice to have a statistical way that can *propagate our (un)certainties about we do(n't) know* into our statistical modelling.

Luckily there is an alternative definition of probability that allows us to incorporate so-called *belief* into our definition of probability: Bayesian probability. The Bayesian definition of probability is:

$$probability = \frac{likelihood \cdot prior}{normaliser} \tag{1}$$

Where *probability* is the Bayesian probability, *likelihood* is exactly the same as a frequentist likelihood , `prior` is your best-guess of the probability (your 'belief prior' to collecting data), and `normaliser` is a mathematical constant that makes sure our probability cannot go above 100% or below 0% (statisticians are lazy, and will not 'give 110%'). This mathematical constant (technically it is the probability of our data; $p(data)$) is a nuisance term that is extremely challenging to estimate (sometimes it is impossible!) and held back the practical application of Bayesian statistics for many years. Nowadays, with increases in computer power, we can use numerical methods such as 'Markov Chain Monte Carlo' (MCMC) methods to avoid having to estimate its value precisely. Such methods are simulations that involve chance at each step (that the process proceeds in iterations or steps, and only the last iteration affects the next, is why this is a 'Markov Chain'), tweaking and changing coefficients within your model until a distribution of plausible coefficients are found. This distribution is called the 'posterior' distribution to distinguish how it is our view of probability *after* the data and the prior have been considered. There is no guarantee that the posterior distribution will, itself, be found: the user must check for so-called 'convergence' onto the true posterior distribution as the process involves randomness. It is for this randomness that 'Monte Carlo' is added into the name, after a city famous for gambling.

It turns out there is a further advantage that such an approach gives us. While frequentist probabilities are all about the probabilities of observing data ($p(data|model)$: what are the chances of getting 99 heads in 100 coin-tosses?), Bayesian probabilities are all about the probabilities of observing models ($p(model|data)$: what are the chances this coin is biased towards heads?). This makes it remarkably straightforward to report the results of Bayesian analyses in terms of estimates of uncertainty and precision of coefficients (contrast "I'm 80% certain that coin comes up heads at least two-thirds of the time" with "I'm less than 5% certain those last twenty coin tosses came from a fair coin").

1.2 A brief overview of the benefits—and pitfalls—of Bayesian models

Bayesian models have many benefits, but an often-mentioned one is that 'you can fit any model you want.' While this is not entirely true (Gelman *et al.*, 2014; Beaumont, 2019), compared to the models ecologists can fit in popular modeling packages (e.g. lme4) Bayesian modeling options can feel limitless. As long as you can write out the likelihood of your desired model and assign priors to all parameters, you can generally 'fit' any model. This includes non-linear ones, non-Gaussian families (e.g. Poisson, beta or combinations thereof, such as hurdle models), hierarchical designs and any combination of these, as well as 'joint' models where parameters estimated in one equation appear in another, thus carrying through estimated uncertainty. Such flexibility is incredibly powerful in ecology where data are often influenced by complex spatial or temporal patterns, non-linear processes are widespread and common data types, are non-Gaussian (e.g. counts, percent cover etc.).

Fitting a bespoke model to your data allows us to get the numbers we often really want but don't have access to. Thus, instead of reporting a treatment's p-value and accompanying F statistic and degrees of freedom, models can be designed to estimate and report effects per °C of warming or at what level non-linearities due to high temperature begin—always with estimated uncertainty. While replication crises in other fields, driven in part by a overly zealous focus on p-values (Halsey *et al.*, 2015; Loken & Gelman, 2017), and the rise of meta-analyses in ecology (Hampton *et al.*, 2013) have led to a somewhat greater focus on 'effect sizes' in ecology (often used to refer to very specific unitless statistics, such as Cohen's d), bespoke models take this to a new level. Researchers can easily estimate comparable effect sizes from z-scored data (in units of standard deviation), alongside estimates in meaningful natural units, such as per °C of warming or per hectare of habitat lost.

Further, ecologists can report estimates of probabilities about coefficients (*e.g.*, "I'm 80% certain the flowers open earlier") rather than null models (*e.g.*, "I'm less than 5% certain these data came from a model where flowers aren't responding"). Importantly, frequentist tests are based around a particular α_{crit}, which is an accepted level of Type I error rates. α is not the same as statistical power (β), which is the probability that we would detect a true effect. Setting α_{crit} to 5% does not make β 95%: if we define significance as a less than 5% chance that some data came from a null hypothesis, that does not mean we have a 95% chance of detecting a true effect. Bayesian statistics, by allowing us to compare the probabilities that *models* are correct, and not the probabilities that *data* are taken from a give model, allows us to test what we have always cared about: the ecological processes themselves.

But this valuable flexibility is also one of the greatest pitfalls of Bayesian models. You can fit almost whatever you want, but critical parts of your model might be almost entirely unimpacted by your data. And in ecological model fitting, we're currently most often interested in parameter estimates strongly informed by our data.

This 'pitfall' of Bayesian is not new, nor unique to ecology—though the complexity of ecological data and processes may make it especially pernicious in ecology—and decades of statistical research has aimed to develop best practices when using Bayesian models to avoid this. These best practices generally center around a specific workflow; a variety of which can be found in exquisite detail elsewhere (Betancourt, 2020; Grinsztajn *et al.*, 2021; van de Schoot *et al.*, 2021). We do not aim to repeat those here, but instead to provide a highly simplified—but powerful—workflow that, when applied to Bayesian modeling in ecology, could greatly accelerate progress.

1.3 A very basic Bayesian workflow

We outline a Bayesian workflow below that includes what we consider the major steps for Bayesian model fitting (Fig. 1). Many steps should be familiar to statistical ecologists, but are often overlooked, whereas other steps may appear particular to Bayesian (e.g. prior predictive checks), but are actually useful for anyone—using Bayesian models or not—to challenge their models of how the world works.

As we focus on a simplified list of major steps, many of the smaller but still critical steps are omitted. For example, visualization is required at every step—especially 2 and 4; while we do not discuss this explicitly we refer to relevant publications. We also jump in, and somewhat over, one of the biggest steps.

We present the workflow assuming users have a model already in hand, which—if you have collected data—is true. Your model may be only verbal or conceptual; however, for this workflow, you'll need to convert such models into mathematical versions. This can be challenging at first, as ecologists often learn only a simplified version of this step (often focused on identifying which distribution—normal, binomial, zero-inflated Poisson, etc.—their response variable most closely resembles), and approach it after data collection. In reality, skilled users of this workflow will develop their model before they collect and refine it—or develop possible additions to it—as they collect data; models often fit best when they include both the underlying biological model being studied and a model of the data generating process (e.g., gaps between sampling dates, biases in sampling etc.).

Getting to the point where the Bayesian workflow is part of data design and collection, however, requires starting somewhere—with some model in hand. A suite of resources on generative modeling can help (MacElreath, 2016; Betancourt, 2021b), along with two points. First, you must start somewhere, so know that you can and will improve on this skill. Second, as you start, ask lots of questions—and push yourself on your answers—what do you expect and what's reasonable biologically from your model? For example, instead of simply identifying to which distribution your response variable looks most similar, ask yourself what generates that distribution and what you think its mean, minimum and maximum is. Do you expect data below zero? Up to infinity? If not, why not? You'll be generating your model—including its priors—as you do this.

Our workflow is explained mostly program-agnostically. Though at times we assume a user of Stan, a relatively new probabalistic programming language, that interfaces with R, Python, Julia (and more) to write bespoke Bayesian models and underpins the R packages brms and rstanarm, which fit a suite of specific (pre-defined) models (Carpenter *et al.*, 2017). We focus on Stan as its MCMC algorithm (a variant of Hamiltonian Monte Carlo, HMC) is fast and produces specific output to warn of model fit issues (i.e., divergent transitions) in a way other MCMC algorithms do not (e.g. Metropolis-Hastings or GIBBS), but the basic workflow should apply to diverse implementations of Bayesian modeling.

Step 1: Check your model code.
To start the workflow, you need to write up your model and check it. As with all code, just

because it runs, does not mean it does what you think it does. Whether writing it out in Stan, where you need to be able to write out the full likelihood and set all your own priors, or using a package that writes much of the model for you (e.g. rstanarm), you need the code and a way to verify the code is correct—test data.

Test data (aka 'simulated data', or 'fake data,' etc.), and the skills required to build it, are central to this workflow. The idea of 'test data' is that you simulate data from your model in such a way that you can use the resulting data to test your model code is correct. This means that to build test data you need to understand your model well enough to generate data from known parameters (which you select); you then run that data through your model and confirm it returns the parameters.

This very basic model checking step is uncommon for many ecologists, but critical in our view. If you can simulate data from your model, then you can powerfully—and easily—answer questions related to statistical power, what effect sizes are reasonable, and—most likely—have new insights into how your model suggests the world works. 'All models are wrong; some models are useful,' becomes much clearer when you have the power to generate data from your model under any parameter set you want. Conversely, if cannot complete this step, you'll struggle to understand if the model fit well, and struggle further to meaningfully interpret the model output, making this apparently simple programmatic step actually encapsulate a far deeper understanding of your model. Once your Stan output returns the parameters you expect from your test data, you can move onto interrogating your priors.

Step 2: Check your priors.
Priors are the source of much discussion and focus by Bayesian and non-Bayesian researchers alike. Often treated as the big bad wolf of Bayesian, or the unseen hand producing the model fits you get, according to some. More accurately they are half of the equation that gives you your model posterior; and they are *designed to be that way.* Bayesian philosophically is built around the idea that you have prior knowledge that you trust and want to compare to new data (with your new data showing up through your likelihood).

How much priors influence your model fit is in many ways up to, and your data. Depending on your model and the data, the likelihood (influenced by your data) can easily overwhelm your priors. On the other hand, as you can fit a Bayesian model with 100 parameters and only 5 datapoints, there can be cases where your priors matter a lot. Most dangerously, failing to think about them, check them and understand your model enough, can mean they may matter more than you know.

Luckily you can use the code that you wrote to build test data (Step 1) and can use it for prior predictive checks. In these, you explore a distribution of potentially reasonable values for your priors, then see how they influence your resulting output. How exactly to do this depends on your question, model and aims, but many guides can help you think through this (Betancourt, 2021a; Wesner & Pomeranz, 2021; Winter & Depaoli, 2023).

In our experience, prior predictive checks can quickly disabuse newbies to Bayesian skeptics of their prejudices. For example, you may not think you have a prior on how sunlight affects plant growth, until you realize your 'agnostic prior' actually allows plant growth of hundred of meters per day for one individual plant. Prior predictive checks in this way serve both as a check on the priors you're using, and to further explore the model of the world to which you're planning to fit your data.

Step 3: Run your model on your empirical data.

The next step is to run the model—you've now validated, test-run and have ready to go—on your exciting new empirical data. Check diagnostics so you know it's running well and adjust until it is; this includes a suite of convergence and efficiency metrics ($\hat{R}$, ESS, lack of divergent transitions etc.) that are well-discussed elsewhere (Betancourt, 2020; Gelman *et al.*, 2020; van de Schoot *et al.*, 2021; Gabry *et al.*, 2019, and not our focus here)

This is the step many ecologists skip straight to, ourselves included. It's easy to see the appeal—fitting our new data to the model can feel like the moment when we'll learn something new. But, at least in our experience, this is not always the case. When we rush to this step, that first model we fit is often followed by another, and another—perhaps because one does not converge, or the results of another do not make immediate sense. After a while of this process, it can easily feel like we're not sure what we learned, if anything. And we can get distracted from what we are actually most interested in—the biological effects. In contrast, by approaching the model through Steps 1-2, it's often much easier to quickly see through the results of the model fit. And also easier to plan next steps.

Step 4: Retrodictive predictive checks
Once you have your posterior based on your model and new empirical data, it's time to interrogate it. Steps 1-2 have set you up well for this, as you have a sense of what different parameter estimates do to the model.

Now that you have the parameter estimates from your posterior you can simulate new data from them and see how that new world looks to you—called retrodictive predictive checks (or posterior predictive checks, Fig. 2-3). Exactly how to do these are—again—dependent on your question, model and aims (but there is lots written on this, Held *et al.*, 2010; Gelman *et al.*, 2000; Conn *et al.*, 2018). In contrast to prior predictive checks, however, this step is built into some software. If you use rstanarm, then the package shinystan will automatically give you a set of retrodictive predictive checks, including comparisons of the mean and variance of simulated datasets with those of from your empirical data. Often here you may find big differences from your empirical data, and can start to generate hypotheses for why.

With those hypotheses in hand, you may very well want to tweak your model—this is all part of the workflow. At that point you return to Step 1, tweak your code, and repeat the process. In this way, fitting multiple models is encouraged, but in a far more structured and careful way than traditional ecological model fitting in our experience. And there are big benefits to it.

This process more fully integrates mathematical modeling into statistical modeling. To complete Step 1, you have to understand the underlying math of your model enough to simulate data from it. This can be challenging at first (e.g. in our courses many students cannot recall how to simulate y data for a simple linear regression), but is immensely beneficial to understanding your model. Indeed, we have found the greatest insights come not from the step we all know best—fitting the model with empirical data—but from every other step in this workflow. Developing simulated data to test the model, running prior and retrodictive checks all dive you deep into understanding your statistical model, which suddenly you may find yourself thinking through much more mechanistically. In our experience this process has quickly translated into insights for our biological systems, and changed how we approach statistical models.

1.4 How this workflow changed our science

As we have used this workflow, how we approach our statistical models has changed. These changes have generally been similar for each of us. We suspect they are not unique to us, our study systems or questions. Instead, we think they represent common approaches to statistical modeling in ecology that could help the field progress, much as we believe they have helped our science advance.

Understanding nonidentifiability

Identifiability and—more common in our experience: nonidentifiability—refers to when all parameters in a model can be uniquely identified. Models can be nonidentifiable in several ways, including when mathematically some parameters cannot be uniquely defined. Statistically, nonidentifiability often is an outcome of the empirical data combined with the model. In this definition nonidentifiability occurs when the data do not contain "sufficient information for unique estimation of a given parameter or set of parameters in a particular model" (Gelman & Hill, 2009).

Nonidentifiability can come up in many ways in ecology—and be hard to see, especially if you rush through model fitting. But if you have to write out your model and simulated data, you may suddenly realize lots of places for nonidentifiability to live. For example, when species do not occur across most sites, a model including separate parameters for site and species is often nonidentifiable, but there's no warning in packages to tell you this; you have to learn it from the data. We have become far better experts at nonidentifiability based on this workflow—and we have adjusted how we collect data and interpret results because of it.

Simplify

Once we noticed how pervasive nonidentifiability and weak identifiability (where parameter estimates are highly uncertain because the model and data together are nearly nonidentifiable) were we started simplifying our models. Whereas previously when we had data that qualitatively appeared complexly nested, crossed, split or twisted, we would have tried to fit all of these intricacies (on the intercept) we are now more slow to add these to our models. By both understanding these terms better (including the many different ways each can be modeled), and understanding them better—depending on the data and model in each unique context—we work more carefully through what to include.

Before using this workflow, some of us would start with complex models, then simplify models until they converged. Often these were hierarchical models with many levels, built based more on a dogma of non-independence (and, somewhat relatedly, correct degrees of freedom) than an understanding of the model. We also often fit a suite of interactions: multiple two-way interactions and the occasional three-way interaction were common fare. But in simulating data, and fitting models to real, messy, imbalanced data using the workflow we came to see how much we were asking of our data and models together. Fitting a two-way interaction with half the effect size of a main effect takes a 16X sample size, compared to fitting the main effects alone (which then average over the interactions Gelman & Hill, 2020). This is sobering. It's more sobering when you see it played out again and again through this workflow.

We now simplify based on a more careful reckoning. Usually this starting model is not simple, it often includes grouping factors that may be difficult to fit, but that we see as absolutely critical to the question, model and data at hand. We build up from there, often grouping posteriors by some factor (e.g. site or biome) before we add it to our model. However, the ending model is often not as complex as we may have fit if we did not follow all the steps. And—through them, especially Step 4—we can feel confident our model is capturing important variation.

Understanding this variation includes re-approaching how we plot our model and data. While many of us are good at plotting raw data, and plotting our main model parameters (and plotting both these pieces together), the world in between is rarely taught. And it's in this in-between world that we have found a deep understanding of our models, what they're doing and—relatedly—what our results show. Many packages offer model predictions—based on a full or near-full model, but being able to decompose model predictions into varying components can give powerful insights. For example, model predictions can be plotted with and without major grouping factors such as 'block,' 'plot', or 'species,' and suddenly how predictable a model is (or not) based on other factors like 'treatment' or 'time' becomes clear.

Looking at parameters, not p-values
Before this workflow, not all of us commonly discussed the values that parameters in our model took—things like the slope and intercept (two common model parameters) were sometimes reported, but we did not know them as well as we knew whether the p-value for the slope was < 0.05. This changes quickly when you need to build simulated data, for example, for a phenological event (a day within the calendar year so, ideally between 1 and 366) and suddenly find it's >1000 based on the quick parameters you put into your code.

This focus on the value of parameters scales up through the workflow and across projects. Having a better sense of parameter values across different biological contexts, model parameterizations, and time periods gives a better sense of how the biological world works, including what's reasonable, possible or wildly unrealistic.

1.5 How this workflow intersects with ecological training

This four-step workflow is a simplified version of the current best practices for Bayesian model fitting (Betancourt, 2020; van de Schoot *et al.*, 2021), but many of the skills required are not part of traditional ecological training. Writing out the math behind most statistical models enough to complete Step 1 bleeds into the skillset usually reserved for those working on theory, where coding and simulating from a model are common tasks. In contrast field, lab and otherwise empirical-data based ecologists often fit models they could not simulate data from. This dichotomy seems short-sighted in our current era, of bigger, messier data and greater computational methods poised to handle such messy data—if scientists are trained in how to build the right models for their questions and data. A reasonably competent coder could easily simulate data under a complex model that they might not have the mathematical expertise to solve.

This workflow also requires a shift in how we approach Bayesian methods—and training in Bayesian models—in ecology. Already, uptake of new Bayesian R packages highlight that Bayesian methods are no longer the purview of only wildlife and fisheries biologists, and these changes come alongside advances in Bayesian workflows, algorithms, and visualiziations (e.g. Betancourt, 2020; van de Schoot *et al.*, 2021; Gabry *et al.*, 2019), that ecology must adapt its training for. While this is an active area, we highlight three major changes.

(1) Prior 'beliefs' are changing. Best practices for determining priors is an active area of statistical research (Gelman *et al.*, 2014; Gelman & Hill, 2020), and training should reflect current best practices. These include that 'non-informative priors' are a misnomer, as they are often informative (Lemoine, 2019), and priors can easily be 'weakly informative.' Thus a strong focus on the dangers of priors in training can be overkill (see supp or box where WDP includes part of his rant?).

Current training often includes a very strong focus on priors, including conjugate priors, because of the importance of conjugate priors in closed form solutions for particular posteriors. Modern algorithms, such as HMC, do not require conjugate priors, which are now antiquated (and thus sometimes referred to as 'the crystal deodorant of Bayesian statistics'). Prior predictive checks provide a far more powerful way to understand how priors work within a particular model, and are far more useful than rules about which priors should be fit in certain cases or memorizing which priors are conjugate.

(2) ‘Random effects’ are not just random. Hierarchical models contain grouping factors, sometimes referred to as ‘random effects,’ such as species or individual. This term, however, is misleading, imprecise and thus no longer recommended (Gelman & Hill, 2009). In ecology, it also carries with a heavy weight of older ‘rules’ of what is ‘random’ versus ‘fixed,’ including that ‘random effects are things you don’t care about (for example the ‘block’ effect from a randomized block design). After retrodictive posterior checks (Step 4), you might feel differently, as hierarchical effects are (by definition) drawn from an underlying distribution—meaning you can predict outside of the specific set you sampled (for example, you can predict for a new species or individual), whereas you cannot do so for most ‘fixed effects.’

(3) P-values are easily misleading (‘I am arbitrary but my story is often told ... ’), and there’s no easy fix for that. The replication crisis, rampant in other fields, is based in part on an overly hopeful belief that p-values will separate the signal from the noise, with one easy number. In reality, small sample sizes, lack of routine reporting of interpretable effect sizes, fitting of many models without adequate explanation, poor data and code reporting habits all increase the chance of finding ‘significance’ at a level of ≤ 0.05 (Halsey *et al.*, 2015; Loken & Gelman, 2017). This reality means a similar crisis is likely lurking in ecology, especially given small sample sizes alongside a tendency to fit complicated models with multiple interactions. The answer to this, however, is not to make p-values smaller (Halsey *et al.*, 2015; Colquhoun, 2017), nor is it Bayesian approaches , which (as we touch on above), bring their own ways to sift exciting results from a pile of chafe.

The answer is a workflow for careful model building, model fitting and model interrogation—as we outline here. This workflow depends strongly on simulating data—for testing your model (Step 1), testing your priors (Step 2) and understanding your model (Step 4)—an area we actively under-train in ecology. This workflow makes clear how relevant and important simulation is, but the relevance of simulation extends well beyond Bayesian model fitting. As larger datasets and machine learning increase their utility, and prevalence, methods to test our understanding, interrogate our models and develop new models, will depend strongly on simulation, potentially transforming ecology, as it is transforming science in general (Flynn *et al.*, 2022; Kuntz & Wilson, 2022; Oren *et al.*, 2017).

Take home messages (maybe)

1. You should not fit a model you cannot simulate
2. Fit simpler models
3. Know your nonidentifiability

2 References

Beaumont, M.A. (2019) Approximate bayesian computation. *ANNUAL REVIEW OF STATISTICS AND ITS APPLICATION, VOL 6* **6**, 379–403.

Betancourt, M. (2019) The convergence of markov chain monte carlo methods: From the metropolis method to hamiltonian monte carlo. *ANNALEN DER PHYSIK* **531**.

Betancourt, M. (2020) Towards a principled bayesian workflow. `https://betanalpha.github.io/assets/case_studies/principled_bayesian_workflow.html`.

Betancourt, M. (2021a) Prior modeling. https://betanalpha.github.io/assets/case$_{s}tudies/prior_{m}odeling.htmll$.

Betancourt, M. (2021b) (what's the probabilistic story) modeling glory? `https://betanalpha.github.io/assets/case_studies/generative_modeling.html`.

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M.A., Guo, J., Li, P. & Allen, R. (2017) Stan: A probabilistic programming language. *Journal of Statistical Software* **76**, 10.18637/jss.v076.i01.

Colquhoun, D. (2017) The reproducibility of research and the misinterpretation of p-values. *Royal Society Open Science* **4**.

Conn, P.B., Johnson, D.S., Williams, P.J., Melin, S.R. & Hooten, M.B. (2018) A guide to bayesian model checking for ecologists. *Ecological Monographs* **88**, 526–542.

Flynn, K.J., Torres, R., Irigoien, X. & Blackford, J.C. (2022) Plankton digital twins-a new research tool. *JOURNAL OF PLANKTON RESEARCH* **44**, 805–813.

Gabry, J., Simpson, D., Vehtari, A., Betancourt, M. & Gelman, A. (2019) Visualization in bayesian workflow. *Journal of the Royal Statistical Society Series a-Statistics in Society* **182**, 389–402.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A. & Rubin, D.B. (2014) *Bayesian Data Analysis*. CRC Press, New York, 3rd edn.

Gelman, A., Goegebeur, Y., Tuerlinckx, F. & Van Mechelen, I. (2000) Diagnostic checks for discrete data regression models using posterior predictive simulations. *JOURNAL OF THE ROYAL STATISTICAL SOCIETY SERIES C-APPLIED STATISTICS* **49**, 247–268.

Gelman, A. & Hill, J. (2009) *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge, New York.

Gelman, A. & Hill, J. (2020) *Regression and Other Stories*. Cambridge University Press.

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.

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

Halsey, L.G., Curran-Everett, D., Vowler, S.L. & Drummond, G.B. (2015) The fickle p value generates irreproducible results. *Nature Methods* **12**, 179–185.

Hampton, S.E., Strasser, C.A., Tewksbury, J.J., Gram, W.K., Budden, A.E., Batcheller, A.L., Duke, C.S. & Porter, J.H. (2013) Big data and the future of ecology. *Frontiers in Ecology and the Environment* **11**, 156–162.

Held, L., Schroedle, B. & Rue, H. (2010) Posterior and cross-validatory predictive checks: A comparison of mcmc and inla. *STATISTICAL MODELLING AND REGRESSION STRUCTURES: FESTSCHRIFT IN HONOUR OF LUDWIG FAHRMEIR* (eds. T. Kneib & G. Tutz), pp. 91–110.

Hoffman, M.D. & Gelman, A. (2014) The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. *JOURNAL OF MACHINE LEARNING RESEARCH* **15**, 1593–1623.

Kuntz, D. & Wilson, A.K. (2022) Machine learning, artificial intelligence, and chemistry: how smart algorithms are reshaping simulation and the laboratory. *PURE AND APPLIED CHEMISTRY* **94**, 1019–1054.

Lemoine, N.P. (2019) Moving beyond noninformative priors: why and how to choose weakly informative priors in bayesian analyses. *Oikos* **128**, 912–928.

Loken, E. & Gelman, A. (2017) Measurement error and the replication crisis. *Science* **355**, 584–585.

MacElreath, R. (2016) *Statistical Rethinking*, vol. 469 pp. CRC Press, New York.

Millar, R. & Meyer, R. (2000) Non-linear state space modelling of fisheries biomass dynamics by using metropolis-hastings within-gibbs sampling. *JOURNAL OF THE ROYAL STATISTICAL SOCIETY SERIES C-APPLIED STATISTICS* **49**, 327–342.

Muthukumarana, S., Schwarz, C.J. & Swartz, T.B. (2008) Bayesian analysis of mark-recapture data with travel time-dependent survival probabilities. *CANADIAN JOURNAL OF STATISTICS-REVUE CANADIENNE DE STATISTIQUE* **36**, 5–21.

Oren, T., Turnitsa, C., Mittal, S. & Diallo, S.Y. (2017) Simulation-based learning and education. *GUIDE TO SIMULATION-BASED DISCIPLINES: ADVANCING OUR COMPUTATIONAL FUTURE* (eds. S. Mittal, U. Durak & T. Oren), Simulation Foundations Methods and Applications, pp. 293–314.

Schad, D.J., Betancourt, M. & Vasishth, S. (2021) Toward a principled bayesian workflow in cognitive science. *PSYCHOLOGICAL METHODS* **26**, 103–126.

Trijoulet, V., Holmes, S.J. & Cook, R.M. (2018) Grey seal predation mortality on three depleted stocks in the west of scotland: What are the implications for stock assessments? *Canadian Journal of Fisheries and Aquatic Sciences* **75**, 723–732.

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**.

Wesner, J.S. & Pomeranz, J.P.F. (2021) Choosing priors in bayesian ecological models by simulating from the prior predictive distribution. *ECOSPHERE* **12**.

Winter, S.D.D. & Depaoli, S. (2023) Illustrating the value of prior predictive checking for bayesian structural equation modeling. *STRUCTURAL EQUATION MODELING-A MULTIDISCIPLINARY JOURNAL* .

Zheng, C., Ovaskainen, O., Saastamoinen, M. & Hanski, I. (2007) Age-dependent survival analyzed with bayesian models of mark-recapture data. *ECOLOGY* **88**, 1970–1976.

3 Figures

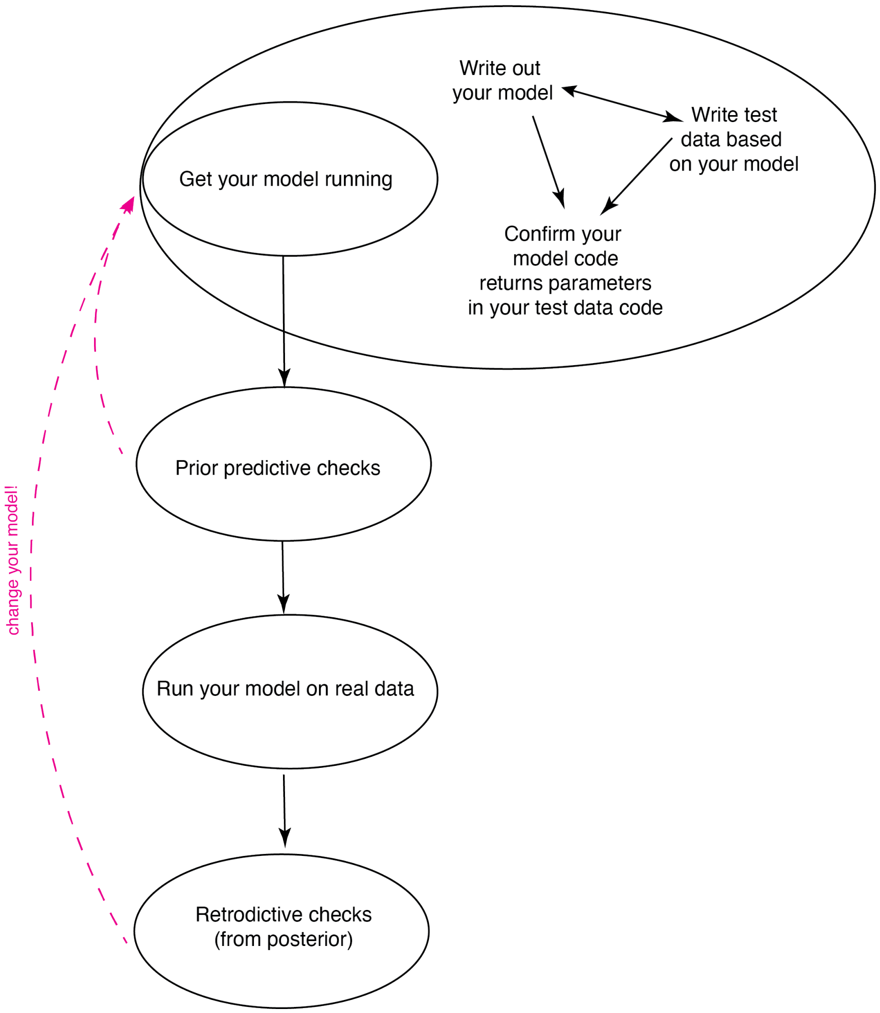

Figure 1: A very basic workflow for Bayesian model fitting include four major steps with potential feedbacks (pink dashed arrows) and begins with testing your model through simulated (test) data.

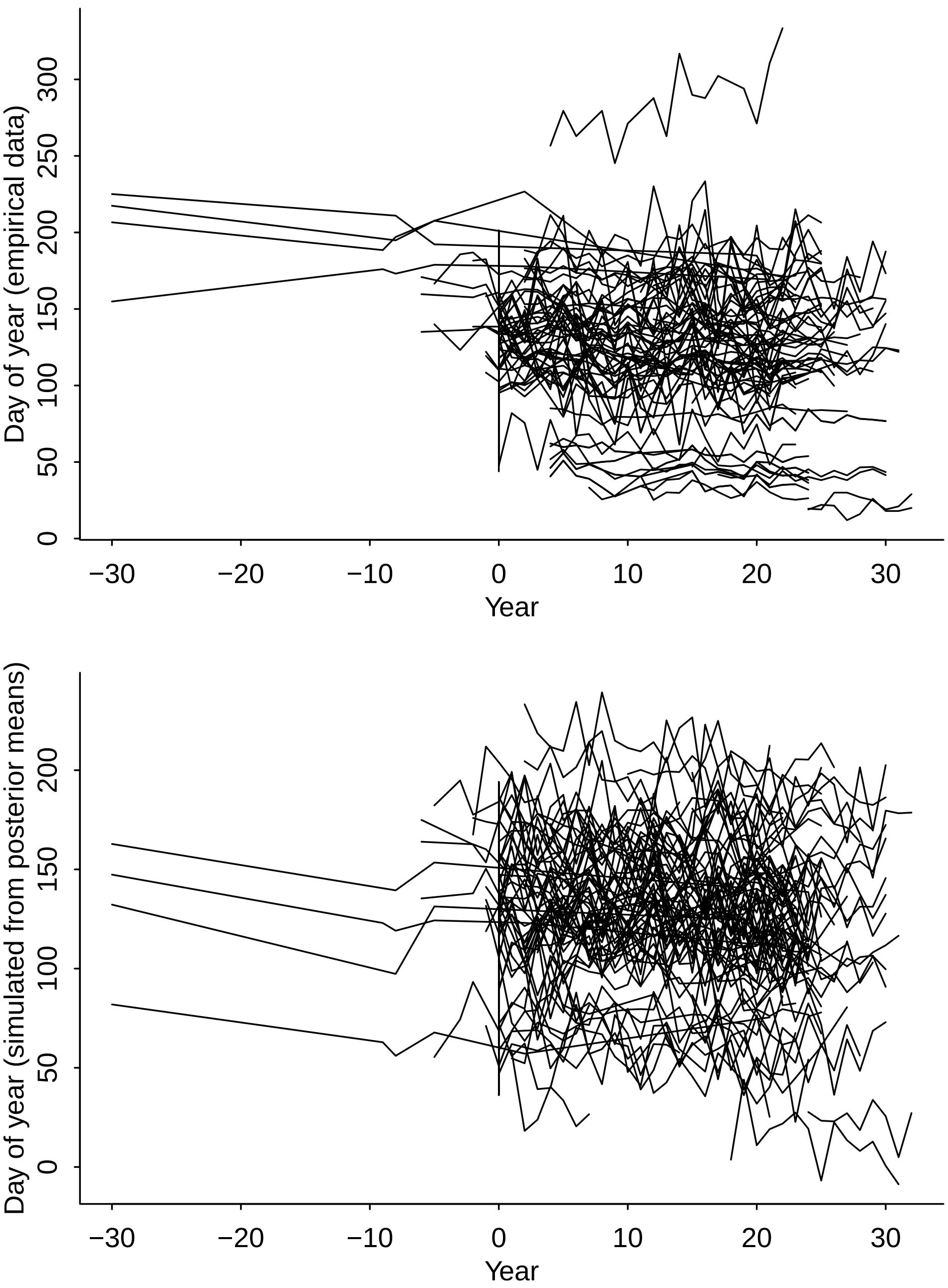

Figure 2: Example of a retrodictive check from time-series data of phenological events over time (raw data on top; one simulated dataset on the bottom, based on existing species number and their respective x data).

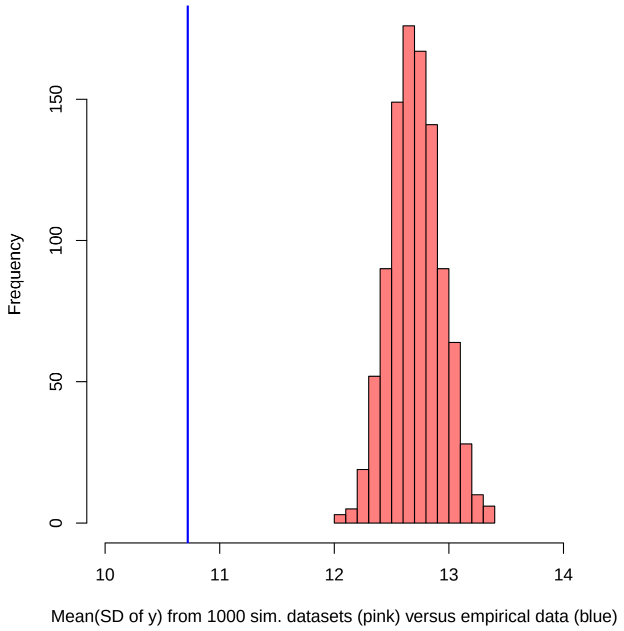

Figure 3: Example of a retrodictive check from time-series data of phenological events over time, averaging across 1000 simulations.