Ten quick tips to get you started with Bayesian statistics

Olivier Gimenez, Andy Royle, Marc Kéry, Chloé Nater

Ten quick tips to get you started with Bayesian statistics

Olivier Gimenez
CEFE, Univ Montpellier, CNRS, EPHE, IRD, 1919 Route de Mende, 34090 Montpellier

Andy Royle
U.S. Geological Survey, Eastern Ecological Science Center, 12100 Beech Forest Rd, Laurel MD 20708

Marc Kéry
Swiss Ornithological Institute, Seerose 1, 6204 Sempach, Switzerland

Chloé Nater
Norwegian Institute for Nature Research, Postbox 5685 Torgarden, 7485 Trondheim, Norway

Introduction

Bayesian statistics is a framework in which our knowledge about unknown quantities of interest (especially parameters) is updated with the information in observed data, though it can also be viewed as simply another method to fit a statistical model. It has become popular in many branches of biology [1–3]. For context, five of the ten most cited papers in Web of Science with keywords 'Bayesian statistics' are related to biology (as of August 19, 2024). Bayesian statistics is particularly valuable for biology because it allows researchers to incorporate prior knowledge, handle complex systems, and work effectively with limited or messy data. However, most biologists are trained in frequentist techniques, and the learning curve to become fluent in Bayesian statistics may be perceived as too time-consuming to undertake, or the prospect of adopting an unfamiliar statistical framework can simply appear too daunting.

We provide a list of 10 tips, summarized in Fig. 1, to help you get started with Bayesian statistics. You can also refer to the Glossary for definitions of the technical terms. This paper isn't just for newcomers; even those with some experience in Bayesian methods may find it a useful roadmap to design, conduct, and publish Bayesian analyses. We've drawn mainly on our experience teaching and working with ecologists, but we hope these tips will be relevant to a broader audience of biologists. For those seeking to deepen their understanding, we point to more comprehensive resources that offer in-depth exploration of Bayesian statistics.

The purpose of our paper is not to persuade you to abandon frequentist methods in favor of Bayesian methods. Instead, we advocate for a dual approach where you master both methods as part of your analytical toolkit.

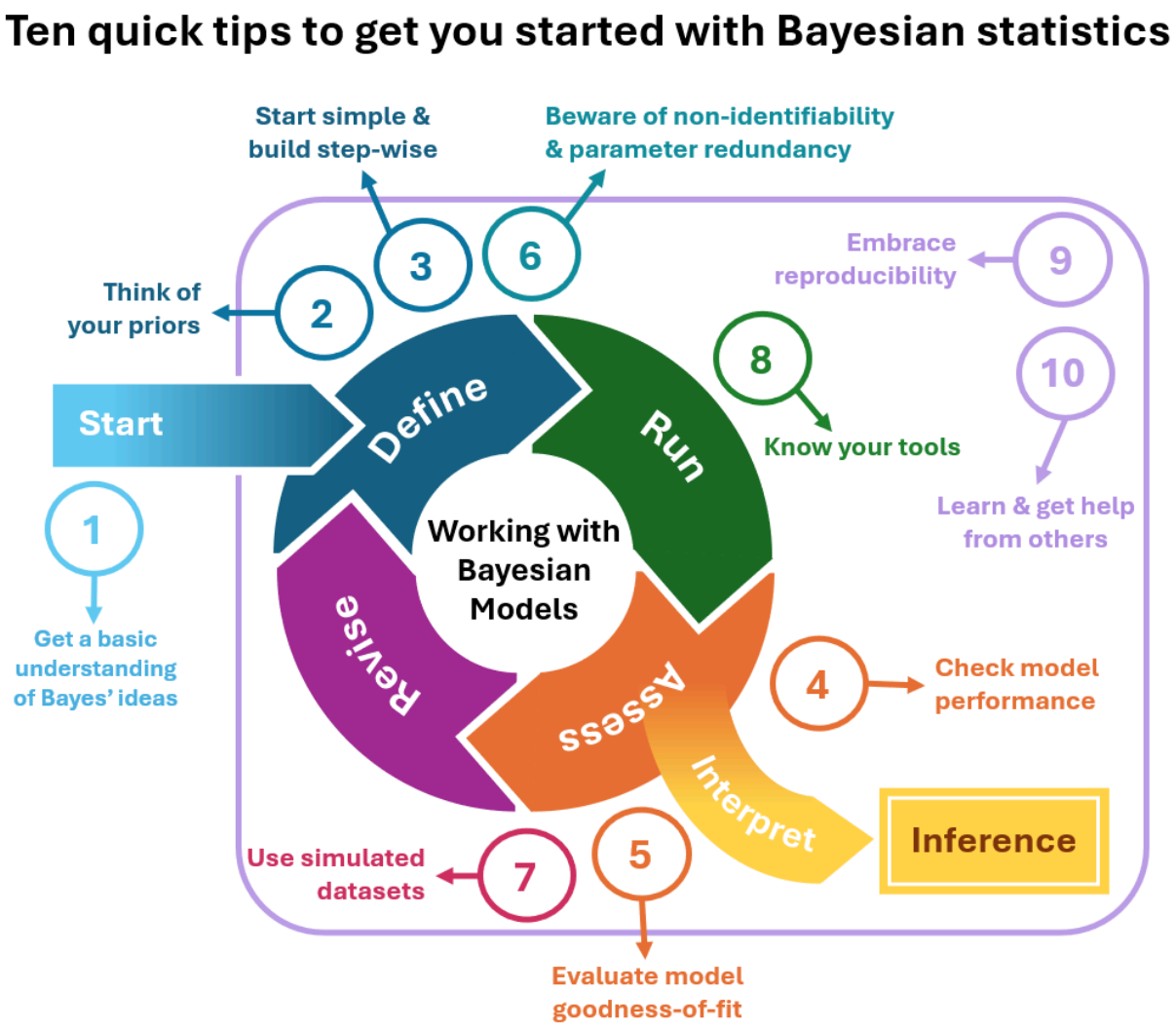

Figure 1: A graphical summary of the 10 quick tips to get you started with Bayesian statistics and how they fit into a larger view of an analytical workflow with Bayesian models.

1. Get a basic understanding of the ideas of Bayes

Bayesian statistics infers unknowns, such as model parameters, using conditional probability. Probability in classical statistics describes the variability of observable data, treated as random variables with distributions. In Bayesian statistics, however, probability quantifies our knowledge of unobservable parameters. For example, we might say, "I am 99% certain it will rain." This approach allows for Bayes' rule to update prior knowledge (from independent data or assumptions) to a new state of knowledge represented by a posterior distribution.

The likelihood function is defined by the statistical model; in a way, the likelihood function *is* the model fitted to the data – here the information in the data enters the estimation. It underpins Bayesian and frequentist methods alike. While frequentists rely solely on likelihood, Bayesians also incorporate priors (Tip #2), allowing external information to influence estimation. A Bayesian model thus combines

a likelihood with priors, and the posterior distribution reflects our knowledge of a parameter as a probability distribution. This can be summarized using metrics like the mean, median, or mode, and uncertainty via standard deviation or percentiles.

Most Bayesian models require approximations to evaluate Bayes' rule, commonly through simulation methods such as Markov chain Monte Carlo (MCMC). MCMC generates a sequence of random values, converging on a stationary distribution that aligns with the desired posterior distribution. The magic of MCMC simplifies the otherwise complex integrals involved in Bayes' rule by sampling values proportional to their posterior density.

2. Think of your priors

In Bayesian statistics, the prior distribution is a fundamental part of the model. It delimits the parameter space and can describe our expectations regarding the value of a parameter. This raises the possibility of formally incorporating "prior information" in the analysis (informative prior), but most analysts specify priors that express a lack of prior information (“vague” or “non-informative”) – or, as is often said, to "let the data speak for themselves" – even if we do have actual prior information from previous studies or expert opinion. There are some situations in which using informative priors is reasonable [4,6]. For example if a similar study was done previously, then using the posterior of the parameter estimate as the prior for the new study is an uncontroversial way to incorporate prior information [5, chapter 20]. A second case is when a complex model has a known non-identifiable or weakly identifiable parameter (Tip #6), then use of an informative prior might make this parameter identifiable.

So exactly *which* prior distribution do we choose for a parameter? For vague priors, you want a distribution that is i) "locally uniform" in the vicinity of the true value of the parameter and ii) covers the range of permissible values of the parameter. For example, for regression parameters (intercept, slopes), we could use a normal distribution with mean 0 and large variance, while for probabilities, we would opt for a uniform distribution between 0 and 1. Others will recommend alternative default priors [6]. The effect of the prior on inferences can be checked with a prior sensitivity analysis – when you are concerned whether a specific prior is sufficiently vague compared to some other choice, test them against each other by rerunning the analysis!

3. Start simple, increase complexity little by little

Arguably, this is the most important general modeling tip. Sometimes people get into Bayesian analysis because they are confronted with a complex modeling problem that cannot be accommodated with a canned R, Python, or Julia package. Upon learning how to write models in one of the languages compatible with Bayesian software, it is tempting to jump straight in and try to implement that complex model in the Bayesian framework of your choice. Do not do that. Chances are high that you will wait for half an eternity, only to be confronted with all kinds of arcane errors that you cannot make sense of. You can avoid this by following one simple rule: always start from the simplest possible version of your model! That is, you should take a modular, stepwise approach to model fitting – this allows you to test whether "things are in order": your data are formatted properly, arrays, matrices and vectors are all of the correct dimension, there are no strange data issues such as data values out of range, or bad prior distributions that suggest inconsistent data. Doing the first analysis on

the simplest model also gives you a sense of MCMC run-time. MCMC run-time scales with the size of your data set *and* model complexity, so starting with a simple model allows you to benchmark the run-time and to check that it's viable within the scope of your work plan. After you get this first, simple model running well and the results are sensible, you should incrementally build your model up to be more and more complex until arriving at your desired model.

4. Check model performance

Working with Bayesian statistics is an iterative process containing multiple rounds of building, assessing, and revising models (Tip #3). Model assessment typically has several steps too, and most focus on either MCMC performance or model goodness-of-fit (Tip #5). Obtaining reliable inferences from a Bayesian model requires that we run multiple chains and confirm that they converge and mix appropriately and consequently provide us with valid posterior samples.

"Convergence" implies the Markov chains have stabilized and samples are being generated from the desired posterior distribution. This is a necessary condition for making inferences from the random numbers produced by an MCMC algorithm. "Mixing" refers to the degree to which different MCMC chains sample the same parameter space. Visualization is one of the best tools for Bayesian model assessments in general [7] and visualizations of the trajectories of the MCMC chains (trace plots) allow us to check convergence and mixing and subsequently adapt our model accordingly (Fig. 2).

Non-convergence can often be addressed by running the MCMC for more iterations. Poor mixing can sometimes be improved by using different MCMC algorithms but can also suggest that there are underlying issues with our model that need addressing, e.g. non-identifiable parameters (Tip #6) or violated assumptions. There are also quantitative metrics for checking MCMC convergence and chain mixing that you can use alongside visualizations; the most frequently used one is the potential scale reduction statistic [8], which measures divergence in the behavior of different chains.

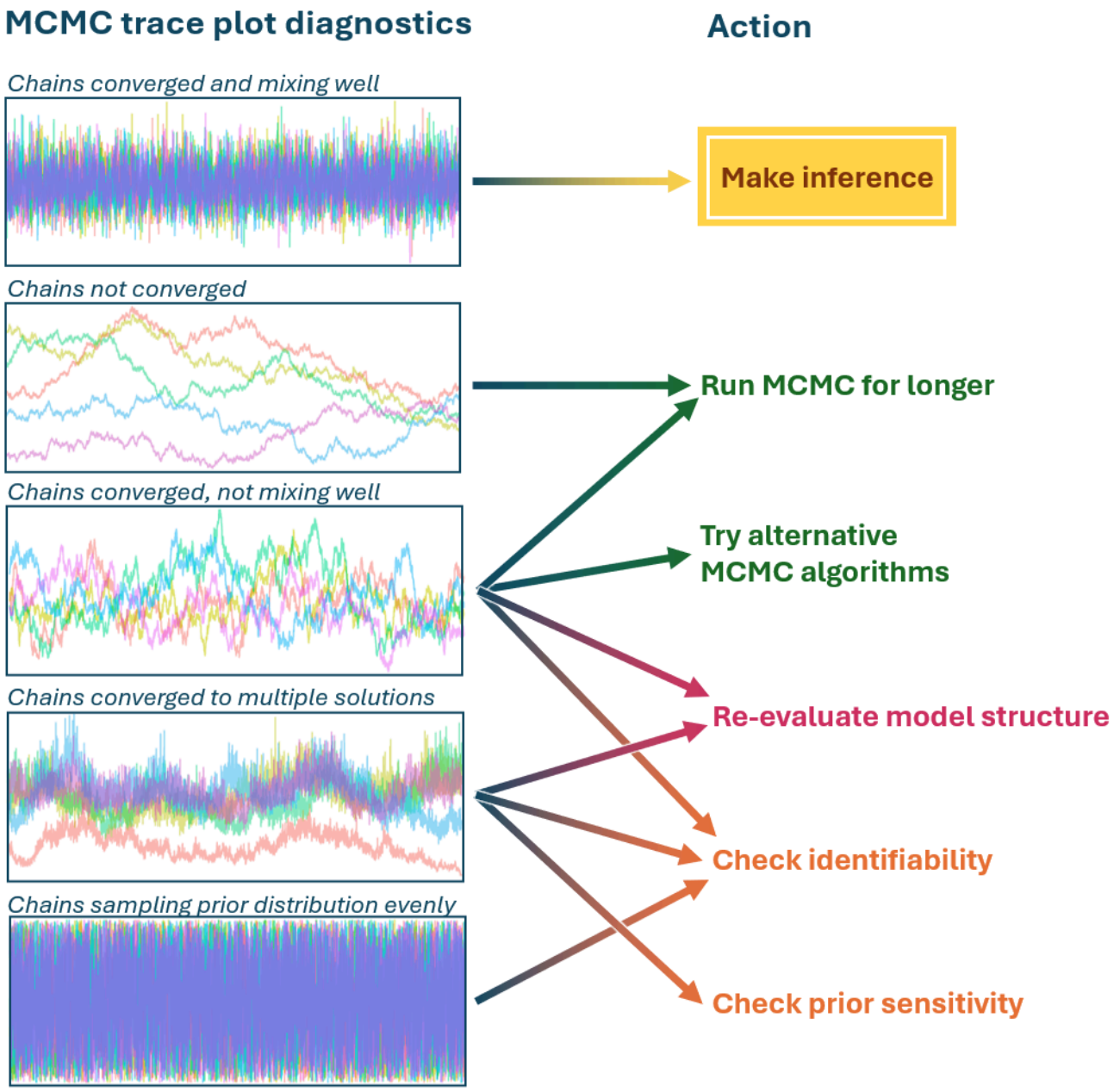

Figure 2: Overview of commonly observed patterns in MCMC traceplots and how to interpret them for diagnostics. Converged chains look like they are stable and fluctuating around the same average value. Well mixing chains have a high degree of overlap with each other. Suboptimal mixing, chains converging to different solutions, and chains that - after reaching convergence - still sample the entire parameter space defined by the prior distribution can indicate problems with the model that warrant further investigation.

5. Evaluate model goodness-of-fit

How well a model fits the data is crucial for trusting the parameter estimates it produces. Goodness-of-fit (GOF) tests are well-established in frequentist statistics [9], and many can be applied to simple Bayesian models (e.g., residual analysis). However, Bayesian methods are often used for more complex models, which lack simple off-the-shelf GOF tests (see review in [9]).

Posterior predictive checks (PPCs), which simulate new datasets from the model's posterior and compare them to the original, are commonly used. The more similar the simulated datasets are to the original, the more likely the model fits well; this can be assessed visually and using a 'Bayesian p-value' [7,10]. Prior predictive checks, based on prior distributions alone (without using data from the analysis), assess adequacy using biological knowledge [11]. Models predicting impossible data (e.g. negative body weight, fewer animals alive than killed per year, etc.) are likely inadequate.

Expert knowledge is essential for model checking, starting with asking: "Do the estimates make biological sense?" This is also key in model selection. Complex Bayesian models are not suited to comparing hundreds of slight variations. Tools like WAIC and cross-validation (review in [12]) can assist you with Bayesian model selection, but we should focus on a few biologically meaningful models likely to solve the inference problem.

6. Beware of model non-identifiability

The Bayesian approach and MCMC methods enable the construction of complex models, but this risks over-parameterization and non-identifiable parameters—those that are confounded and not independently estimable [13]. Models with non-identifiable parameters often show poor MCMC mixing and may converge slowly. However, detecting non-identifiability can be challenging, as parameters may seem estimable from available data even when they are not [13].

To diagnose non-identifiability, simulations can be useful (Tip #7). If true parameter values cannot be recovered from simulated data, it suggests potential non-identifiability. Comparing posteriors from different models can reveal discrepancies, indicating non-identifiability. Examining the overlap between prior and posterior distributions is also useful; non-identifiable parameters typically show substantial overlap and limited 'Bayesian learning' [14]. However, large overlap may also result from an informative prior aligning with the posterior, so it alone does not confirm non-identifiability. Frequentist methods can also help. For instance, profiling the likelihood [15] involves maximizing the likelihood with respect to all parameters except one, held constant at various values. A flat profile likelihood indicates non-identifiability for that parameter. Other methods using symbolic algebra can diagnose non-identifiability but require advanced expertise and may become impractical with many parameters [13].

To address non-identifiability, first ask yourself whether your data is adequate in both quantity and quality, and simplify your model (Tip #3) until the issue is resolved. If the problem persists, consider using informative priors to provide additional information for non-identifiable parameters [16–19]. Second, consider reparameterizing your model or employing other techniques to improve MCMC chain mixing [16,20].

7. Use simulated data sets

Data analysis and data simulation are almost the same thing, but use a model in different directions: analysis takes data and models and estimates parameters, while simulation assumes parameter values and a model and generates potential data sets.

The first goal of data simulation is to enforce an understanding of a model: if you are unable to simulate data under a model, you probably haven't fully understood it. Similarly, computer code for data simulation under a model is arguably one of the best ways to explain or define a statistical model for non-statisticians.

Additional advantages of data simulation include: (1) Truth is known, hence, you can validate your model or your code to fit it. (2) It may help you understand complicated statistical concepts. For instance, if you don't understand what a standard error is, you can repeatedly simulate a data set, estimate some parameter and realize that the standard error is simply the standard deviation of these estimates over replicates. (3) You can evaluate bias and precision of your estimators. (4) Power analysis evaluates the probability with which you can detect a non-zero parameter in a significance test, and data simulation is the most general manner of evaluating power. (5) Study design (e.g., what's the minimum required sample size) is also best approached with simulated data. (6) To check the robustness of a model to assumption violations, we can simulate data under a more general model and then fit a simpler model that lacks crucial assumptions.

8. Know your tools

Just as you wouldn’t typically write iterative optimization code from scratch in the frequentist framework, we recommend using established software for reliable, well-tested, and optimized MCMC algorithms. We focus on free options, and there are numerous choices for Bayesian statistics available in R, Python and Julia [21].

Selecting software based solely on speed and efficiency is challenging [22], as it involves more than just measuring raw computation time, which is complex in itself. You also need to factor in your experience level and (un)familiarity with a modeling language. The ideal choice should balance your specific needs, programming experience, user-friendliness, efficiency, available documentation, and the size and helpfulness of the user community.

Bayesian software generally falls into two categories: general-purpose model-fitting engines requiring coding such as JAGS [23], Stan [24], or Nimble [25] and those with built-in models such as JASP [26], PyMC3 [27] or brms [28]. Coding offers theoretically unlimited flexibility for custom models, but non-coding options can handle many tasks with minimal effort. Prioritize getting your code to work before optimizing for performance.

For a beginner, it may be wise to pick a software and stick with it, as familiarity helps with debugging, useful tricks, and implementing advanced techniques. However, as you gain confidence, consider diversifying your toolkit to tackle tasks that may be challenging with your default software.

9. Embrace reproducibility

Awareness of reproducibility's importance is growing in the biological sciences [29,30], with publishers and funders emphasizing the publication of well-documented data and code [31]. Considering reproducibility from the start—not just as an afterthought—will become a superpower for your work with Bayesian models. Modern research workflows begin with data collection and end with presenting results, and ensuring reproducibility increases the quality and credibility of your work, enhances efficiency in debugging and re-running analyses, and enables broader use of your model (Tip #10).

Four pillars support reproducibility in your workflow: (1) Control of randomness: Specifying seeds for random number generators when simulating data, setting initial values and running MCMC ensures consistent results. This not only enhances reproducibility but also simplifies debugging, especially with a complete set of initial values. (2) Clean coding [32]: Adopting best practices for writing code [33] helps others and yourself understand and re-run analyses later. Clean coding is also the first step in automating your workflow and facilitates running code on high-performance computing setups. (3) Good documentation means thoroughly annotating your code and documenting workflow steps, which benefits others and increases potential for broader impact. (4) Version control involves using git repositories (GitHub/GitLab), which facilitates code development and management and can transform research collaborations [34].

10. Learn and get help from others

Learning Bayesian statistics can be daunting. Fortunately, you are not alone in this journey. We recommend you engage with the Bayesian community within your field of research by leveraging available resources, such as online forums, workshops, and textbooks.

A good starting point is to identify an introductory textbook relevant to your field of biology that covers Bayesian statistics. To build confidence in your skills, consider conducting side-by-side analyses using both frequentist and Bayesian approaches [5,35,36]. This comparative practice can help you appreciate that the results often align. Begin with a simple project of your own or replicate a past analysis from your work or published studies.

If you have questions about Bayesian statistics, don't hesitate to seek help from forums such as Cross Validated or Stack Overflow or more specialized, subject-matter lists. For software-specific inquiries, consider reaching out through mailing lists, and best include a small, reproducible example to illustrate your problem. The feedback you receive can be invaluable and often benefits other beginners as well.

Attending workshops can be an excellent way to learn under the guidance of experienced teachers. These events also offer excellent opportunities to network with others. As you become more advanced, workshops provide avenues to master new methods and further enhance your skills.

Remember, modern science is a collaborative effort, and Bayesian statistics is no exception. If a task feels beyond your current expertise, seek collaboration to learn from others and achieve your research goals.

Conclusions

Learning a new skill can be challenging and time-consuming. While the material we discussed is not entirely new and has been explored by others [3,37,38], our tips aim to offer a quick and practical guide to help you get started. These insights are designed to support your journey through Bayesian analysis in biological research, regardless of your specific field.

Finally, remember that general statistical principles still apply [39,40] [appendix B in 41]: always clarify your modeling goals (whether to describe, understand, explain, or predict), show/visualize your data and more [7], check assumptions, plan your study design carefully [42], report effect size, etc.

References

1. Clark JS. Why environmental scientists are becoming Bayesians. Ecology Letters. 2005;8: 2–14. doi:10.1111/j.1461-0248.2004.00702.x
2. van de Schoot R, Kaplan D, Denissen J, Asendorpf JB, Neyer FJ, van Aken MAG. A Gentle Introduction to Bayesian Analysis: Applications to Developmental Research. Child Development. 2014;85: 842–860. doi:10.1111/cdev.12169
3. van de Schoot R, Depaoli S, King R, Kramer B, Märtens K, Tadesse MG, et al. Bayesian statistics and modelling. Nat Rev Methods Primers. 2021;1: 1–26. doi:10.1038/s43586-020-00001-2
4. McCarthy MA, Masters P. Profiting from prior information in Bayesian analyses of ecological data. Journal of Applied Ecology. 2005;42: 1012–1019. doi:10.1111/j.1365-2664.2005.01101.x
5. Kéry M, Kellner K. Applied Statistical Modelling for Ecologists: A Practical Guide to Bayesian and Likelihood Inference Using R, JAGS, NIMBLE, Stan and TMB. Elsevier, Academic Press; 2024.
6. Lemoine NP. Moving beyond noninformative priors: why and how to choose weakly informative priors in Bayesian analyses. Oikos. 2019;128: 912–928. doi:10.1111/oik.05985
7. Gabry J, Simpson D, Vehtari A, Betancourt M, Gelman A. Visualization in Bayesian workflow. J R Stat Soc A. 2019;182: 389–402. doi:10.1111/rssa.12378
8. Gelman A, Rubin DB. Inference from Iterative Simulation Using Multiple Sequences. Statistical Science. 1992;7: 457–472. doi:10.1214/ss/1177011136
9. Conn PB, Johnson DS, Williams PJ, Melin SR, Hooten MB. A guide to Bayesian model checking for ecologists. Ecological Monographs. 2018;88: 526–542. doi:10.1002/ecm.1314
10. Gelman A, Meng X-L, Stern H. Posterior Predictive Assessment of Model Fitness Via Realized Discrepancies. Statistica Sinica. 1996;6: 733–760.
11. Box GEP. Sampling and Bayes' Inference in Scientific Modelling and Robustness. Journal of the Royal Statistical Society Series A (General). 1980;143: 383–430. doi:10.2307/2982063
12. Hooten MB, Hobbs NT. A guide to Bayesian model selection for ecologists. Ecological Monographs. 2015;85: 3–28. doi:10.1890/14-0661.1
13. Cole D. Parameter redundancy and identifiability. Boca Raton: CRC Press; 2020.
14. Garrett ES, Zeger SL. Latent class model diagnosis. Biometrics. 2000;56: 1055-1067-.
15. Raue A, Kreutz C, Maiwald T, Bachmann J, Schilling M, Klingmüller U, et al. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. Bioinformatics. 2009;25: 1923–1929. doi:10.1093/bioinformatics/btp358
16. Gustafson P. On Model Expansion, Model Contraction, Identifiability and Prior Information: Two Illustrative Scenarios Involving Mismeasured Variables. Statistical Science. 2005;20: 111–140. doi:10.1214/088342305000000098
17. Gelfand AE, Sahu SK. Identifiability, Improper Priors, and Gibbs Sampling for Generalized Linear Models. Journal of the American Statistical Association. 1999;94: 247–253. doi:10.1080/01621459.1999.10473840
18. Swartz T, Haitovsky Y, Vexler A, Yang T. Bayesian Identifiability and Misclassification in Multinomial Data. The Canadian Journal of Statistics / La Revue Canadienne de Statistique. 2004;32: 285–302. doi:10.2307/3315930
19. Eberly LE, Carlin BP. Identifiability and convergence issues for Markov chain Monte Carlo fitting of spatial models. Stat Med. 2000;19: 2279–2294. doi:10.1002/1097-0258(20000915/30)19:17/18<2279::aid-sim569>3.0.co;2-r
20. Gelman A, van Dyk DA, Huang Z, Boscardin JW. Using Redundant Parameterizations to Fit Hierarchical Models. Journal of Computational and Graphical Statistics. 2008;17: 95–122. doi:10.1198/106186008X287337
21. Štrumbelj E, Bouchard-Côté A, Corander J, Gelman A, Rue H, Murray L, et al. Past, Present and Future of Software for Bayesian Inference. Statistical Science. 2024;39: 46–61. doi:10.1214/23-STS907
22. Bolker BM, Gardner B, Maunder M, Berg CW, Brooks M, Comita L, et al. Strategies for fitting

nonlinear ecological models in R, AD Model Builder, and BUGS. Ramula S, editor. Methods Ecol Evol. 2013;4: 501–512. doi:10.1111/2041-210X.12044
23. Plummer M. JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling. Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003), Vienna, 20-22 March 2003. 2003; 1–10.
24. Stan Development Team. Stan Modeling Language Users Guide and Reference Manual, version 2.26.1. 2024. Available: https://mc-stan.org
25. de Valpine P, Turek D, Paciorek CJ, Anderson-Bergman C, Lang DT, Bodik R. Programming With Models: Writing Statistical Algorithms for General Model Structures With NIMBLE. Journal of Computational and Graphical Statistics. 2017;26: 403–413. doi:10.1080/10618600.2016.1172487
26. Love J, Selker R, Marsman M, Jamil T, Dropmann D, Verhagen J, et al. JASP: Graphical Statistical Software for Common Statistical Designs. Journal of Statistical Software. 2019;88: 1–17. doi:10.18637/jss.v088.i02
27. Salvatier J, Wiecki TV, Fonnesbeck C. Probabilistic programming in Python using PyMC3. PeerJ Comput Sci. 2016;2: e55. doi:10.7717/peerj-cs.55
28. Bürkner P-C. brms: An R Package for Bayesian Multilevel Models Using Stan. Journal of Statistical Software. 2017;80: 1–28. doi:10.18637/jss.v080.i01
29. Cohen-Boulakia S, Belhajjame K, Collin O, Chopard J, Froidevaux C, Gaignard A, et al. Scientific workflows for computational reproducibility in the life sciences: Status, challenges and opportunities. Future Generation Computer Systems. 2017;75: 284–298. doi:10.1016/j.future.2017.01.012
30. Powers SM, Hampton SE. Open science, reproducibility, and transparency in ecology. Ecological Applications. 2019;29: e01822. doi:10.1002/eap.1822
31. Jenkins GB, Beckerman AP, Bellard C, Benítez-López A, Ellison AM, Foote CG, et al. Reproducibility in ecology and evolution: Minimum standards for data and code. Ecology and Evolution. 2023;13: e9961. doi:10.1002/ece3.9961
32. Filazzola A, Lortie C. A call for clean code to effectively communicate science. Methods in Ecology and Evolution. 2022;13: 2119–2128. doi:10.1111/2041-210X.13961
33. Cooper N, Hsing P-Y. A Guide to Reproducible Code in Ecology and Evolution. London: British Ecological Society; 2017. Available: https://nhm.openrepository.com/handle/10141/622618
34. Braga PHP, Hébert K, Hudgins EJ, Scott ER, Edwards BPM, Sánchez Reyes LL, et al. Not just for programmers: How GitHub can accelerate collaborative and reproducible research in ecology and evolution. Methods in Ecology and Evolution. 2023;14: 1364–1380. doi:10.1111/2041-210X.14108
35. Gelman A, Hill J. Data Analysis Using Regression and Multilevel/Hierarchical Models. 1st edition. Cambridge: Cambridge University Press; 2006.
36. Inchausti P, Inchausti P. Statistical Modeling With R: a dual frequentist and Bayesian approach for life scientists. Oxford, New York: Oxford University Press; 2022.
37. Gelman A, Vehtari A, Simpson D, Margossian CC, Carpenter B, Yao Y, et al. Bayesian Workflow. arXiv; 2020. doi:10.48550/arXiv.2011.01808
38. Wolkovich EM, Davies TJ, Pearse WD, Betancourt M. A four-step Bayesian workflow for improving ecological science. arXiv; 2024. Available: http://arxiv.org/abs/2408.02603
39. Kass RE, Caffo BS, Davidian M, Meng X-L, Yu B, Reid N. Ten Simple Rules for Effective Statistical Practice. PLOS Computational Biology. 2016;12: e1004961. doi:10.1371/journal.pcbi.1004961
40. Zitomer RA, Karr J, Kerstens M, Perry L, Ruth K, Adrean L, et al. Ten simple rules for getting started with statistics in graduate school. PLOS Computational Biology. 2022;18: e1010033. doi:10.1371/journal.pcbi.1010033
41. Gelman A, Hill J, Vehtari A. Regression and Other Stories. 1st edition. Cambridge New York, NY Port Melbourne, VIC New Delhi Singapore: Cambridge University Press; 2020.
42. Bon JJ, Bretherton A, Buchhorn K, Cramb S, Drovandi C, Hassan C, et al. Being Bayesian in the 2020s: opportunities and challenges in the practice of modern applied Bayesian statistics. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering

Sciences. 2023;381: 20220156. doi:10.1098/rsta.2022.0156

Glossary

Random variable (RV)	Usually defined as a real-valued function defined on the outcome of a random experiment. In practice, something unknown that we want to estimate or the probability of which we want to evaluate in a statistical analysis. Continuous random variables (RVs) include measurements of durations, lengths, weights and (in Bayes) the values of most parameters, while discrete RVs include counts or labels such as dead/alive, red/blond/brown, and (in Bayes) the values of discrete latent parameters such as abundance or presence/absence.
Probability density function (PDF)	Probability distribution for a continuous RV. It gives the probability density for any possible value *x* of the continuous RV, which corresponds to the area under the curve of a rectangle with basis (x-d, x+d) as d goes towards zero. Typical examples: normal (or Gaussian), exponential.
Probability mass function (PMF)	Probability distribution for a discrete RV. It gives the probability of every possible value that the RV can take. Typical examples: Poisson, Binomial.
Probability distribution	Mathematical function that assigns a probability (for the PMF) or a probability density (for the PDF) to all possible values of a RV. Sums to 1 for a PMF and integrates to 1 for a PDF.
Random experiment	Taking a measurement or observation that has some stochasticity associated. E.g., tossing a coin, capturing an animal and measuring its mass.
Outcome	A possible value of the result of a random experiment, e.g., the numbers 1–6 when tossing a die.

Probability function	Function that assigns a probability, or a value between 0 and 1, to any outcome of a random experiment. Values 0 and 1 denote respectively an impossible and a certain outcome.
Conditional probability	The probability of one RV given ('conditioned on') the known value of another RV, e.g. P(mass = 700 | sex = male), which gives the probability density of the mass of an animal of 700 g, given that it is a male. In general, written as P(A|B) for 'probability of A given B'. This is defined as P(A and B) / P(B), i.e., as the joint probability of A and B, divided by the marginal probability of B.
Joint probability	The probability of a combination of two (or more) RVs.
Marginal probability	The probability of a random variable averaging over (or integrating over all possible values of) another random variable.
Joint density of a data set under a model	The joint density of obtaining the observed values of all data (and possibly random effects in the case of random-effects models) under a statistical model. Expressed in terms of the PDFs or PMFs of the model.
(Parametric) statistical model	A set of PDFs/PMFs for all observed random variables (i.e., data) or unobserved random variables (i.e., random effects, latent variables).

Bayes' rule	A derivation of the definition of the conditional probability, which in general can be written as P(A|B) = P(B|A)P(B) / P(B) and which can be used for non-Bayesian calculations when applied to observable quantities, e.g. in clinical testing. In Bayesian statistics, Bayes' rule is used for inferences about unknown, unobservable quantities, and especially for parameters of a statistical model. Denoting parameters theta and a data set as y, Bayes' rule can then be written as P(theta|y) = P(y|theta)P(theta) / P(y). Ignoring the constant in Bayes' rule, we can also write this as P(theta|y) prop. P(y|theta)P(theta), i.e., the posterior is proportional to the product of the likelihood and the prior. Evaluation of P(y) often requires intractable integrals over all parameters and didn't allow application of Bayes' rule for most practical applications for centuries. The discovery of simulation algorithms (MCMC; see below) circumvented this impasse and led to the great advance of Bayesian statistics observed during the last 30 years.
Prior, also prior distribution	A statement of how likely different values are for a parameter in a Bayesian model before any information in a data set to be analyzed is incorporated, P(theta). This is a probability distribution, thus it integrates/sums to 1 over the entire parameter space.
Likelihood function	Joint density function of all data under a model, when viewed as a function of the parameters. Represents the formal connection between data and parameters or loosely also the statistical model fitted to a data set.
Maximum likelihood	A principle that says that the best possible 'guess' for a parameter is that value which maximizes the likelihood function when evaluated for the observed data set.

Posterior, also posterior distribution	A statement of how likely different values are for a parameter in a Bayesian model when we incorporate the information in our data set, i.e. P(theta | y). This is another proper probability distribution that integrates to 1 over the parameter space.
Prior predictive distribution	This is P(y) in Bayes' rule and is the probability distribution when averaged over all possible values of the priors. When evaluated for a given data set, also called the 'normalizing constant' since it ensures that the integral of P(theta | y) becomes equal to 1. The value of the normalizing constant is obtained by integration over all parameters of the model, which in practice can hardly ever be done. MCMC algorithms circumvent this.
Bayesian model	A statistical model for a data set in which we combine a likelihood with priors that we chose for each parameter. Note that a Bayesian model is not fundamentally different from its corresponding non-Bayesian analog except for the priors. Thus a linear regression model (in terms of its likelihood) is the same whether we fit it using least-squares, maximum likelihood or Bayesian posterior inference.
Bayesian statistics, also Bayesian inference or Bayesian posterior inference	The use of conditional probability, via Bayes' rule, to update one state of knowledge using the information contained in some data set and embodied by the likelihood function and to arrive at a new state of knowledge, usually with reduced uncertainty about parameter and other estimated quantities.
Markov chain Monte Carlo (MCMC) algorithm	A vast family of iterative algorithms that are typically used to fit Bayesian models. In essence, they function like random number generators (RNGs) for the posterior distributions that arise from Bayes' rule when combining the likelihood of the data under a model and the priors chosen for the model's parameters. These distributions can be approximated to an arbitrary degree of accuracy by drawing increasing samples of all parameters.