Trends in Ecology & Evolution

CellPress
REVIEWS

Opinion

Ecological Dynamics: Integrating Empirical, Statistical, and Analytical Methods

Amanda N. Laubmeier,[1,*] Bernard Cazelles,[2] Kim Cuddington,[3] Kelley D. Erickson,[4] Marie-Josée Fortin,[5] Kiona Ogle,[6] Christopher K. Wikle,[7] Kai Zhu,[8] and Elise F. Zipkin[9]

Understanding ecological processes and predicting long-term dynamics are ongoing challenges in ecology. To address these challenges, we suggest an approach combining mathematical analyses and Bayesian hierarchical statistical modeling with diverse data sources. Novel mathematical analysis of ecological dynamics permits a process-based understanding of conditions under which systems approach equilibrium, experience large oscillations, or persist in transient states. This understanding is improved by combining ecological models with empirical observations from a variety of sources. Bayesian hierarchical models explicitly couple process-based models and data, yielding probabilistic quantification of model parameters, system characteristics, and associated uncertainties. We outline relevant tools from dynamical analysis and hierarchical modeling and argue for their integration, demonstrating the value of this synthetic approach through a simple predator–prey example.

Opportunities Using Modern Ecological Data

A proliferation of large-scale observation networks, automated sensors, and citizen science initiatives (e.g., NEON, FLUXNET, eBird [1–3]) has resulted in vast quantities of data on environmental factors, ecosystem properties, population and community dynamics, and species distributions, among others. While not without their challenges [4], these data sources allow ecologists to address research questions that were previously unapproachable, such as how ecological processes interact across spatial and temporal scales. A major focus of this burgeoning research has been on identifying relationships between patterns and processes, using either machine learning techniques [5–7] or Bayesian hierarchical modeling (BHM) [8], to inform simple models of ecological dynamics with empirical data [9,10]. Recent pushes towards using empirical data to more accurately predict and forecast ecological responses have the potential to greatly enhance understanding in a changing world [11–13]. Process-based models are also important tools for understanding and prediction and they have a long history in ecology [14]. Here, we propose to leverage mathematical tools and modern statistical methods alongside emerging sources of novel and big data to maximize understanding and improve forecasts of ecological dynamics.

Novel Insights from Dynamical Analysis of Ecological Systems

Process-based models are commonly employed to understand the behavior of ecological systems under a variety of scenarios [15,16]. For example, to determine whether a population will increase or decrease, we might construct a model describing the influence of growth and mortality processes on changes in the population. Mathematical analysis of this model can yield assessments of the influence of different parameters describing key biological processes underlying population growth. Such **dynamical analysis** (see Glossary) presents an advantage over *post hoc* approaches that examine variations in realized population growth rates. Compared

Highlights

Increasing availability of data sets from diverse sources over a range of spatial and temporal scales presents an opportunity to address important research questions, including how to efficiently use such data to understand and predict the dynamical behavior of ecological systems.

Recent mathematical advances, stemming from traditional dynamical analysis, describe long-term system behavior in relation to governing ecological properties, allowing for the exploration of short-term (transient) system behaviors.

Bayesian hierarchical models (BHMs) facilitate the integration of multiple data sources with theoretical models, providing great potential to improve understanding and predictions of ecological dynamics when combined with mathematical dynamical analysis.

Previous work with integral projection models and integrated population models also suggests that pairing BHMs with process-based models can yield novel ecological insights.

[1]Department of Mathematics & Statistics, Texas Tech University, Lubbock, TX, USA
[2]Eco-Evolutionary Mathematics, CNRS UMR 8197, Ecole Normale Supérieure, Paris, France
[3]Department of Biology, University of Waterloo, Waterloo, Ontario, Canada
[4]Center for Conservation and Sustainable Development, Missouri Botanical Garden, St. Louis, MO, USA
[5]Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, Ontario, Canada

https://doi.org/10.1016/j.tree.2020.08.006

with phenomenological methods, dynamical analysis leads to a deeper understanding of the properties driving system behavior, by explicitly quantifying change in terms of underlying mechanisms (e.g., nutrient quality, birth rates, predation rates). Such techniques can also be employed to predict qualitative characteristics of a system (e.g., whether it tends toward steady-state equilibrium, oscillations, or chaotic fluctuations). Characterizing these outcomes is critically important in conservation biology and can reveal conditions under which perturbations (e.g., from environmental stochasticity or chronic anthropogenic disturbance) are likely to permanently disrupt ecosystems.

[6]School of Informatics, Computing, and Cyber Systems, Northern Arizona University, Flagstaff, AZ, USA
[7]Department of Statistics, University of Missouri, Columbia, MO, USA
[8]Department of Environmental Studies, University of California, Santa Cruz, CA, USA
[9]Department of Integrative Biology, Program in Ecology, Evolution, and Behavior, Michigan State University, East Lansing, MI, USA

*Correspondence:
amanda.laubmeier@ttu.edu
(A.N. Laubmeier).

Many dynamical analyses focus on long-term outcomes, such as a stable equilibrium, yet most ecological systems do not complete their response to a perturbation before the next one occurs. Moreover, a system's 'short-term' behavior can be long lived [17], appearing stable before an unexpected shift to a completely different state. Identifying this type of behavior is critical to management because different intervention strategies are required for a system experiencing transient behavior compared with a stable system [17]. However, the change in behavior might not be apparent from time-series data prior to a state shift and, importantly, can occur in the absence of any concurrent changes in driving variables (e.g., nutrient loads or temperature). This dynamical behavior differs from the classical 'tipping point' scenario where a regime shift suddenly occurs as a result of a change in driving variables or system parameters [18], and identification of such behavior can be complicated by the fact that models are incapable of including all variables that affect a system. In addition to describing these transient shifts, the dynamic properties of a system determine the prevalence of certain types of stochastically driven behavior (e.g., cohort resonance [19]) or switching between stable states [20]. Quantifying these different types of dynamic behavior yields estimates of the probability that a population or ecosystem is less stable than it appears. Predicting probable behaviors is crucial for understanding ecological processes and ecosystem responses to environmental change.

Mathematical Tools for Dynamical Analysis

A variety of quantities can be derived from model parameters and evaluated to determine the potential dynamics of ecological systems. For example, the parameters of a matrix population model may be of interest (e.g., the probability of transitioning between different age classes; [21]), but derived quantities may be just as useful. The eigenvalues and eigenvectors of the matrix, which are derived (calculated) from model parameters, give information about a population's long-term expected growth rate and stable age structure. Eigenvalues and eigenvectors are often assessed in **local stability analysis**, a common method to identify parameter combinations and initial conditions determining qualitative changes in model behaviors. System behavior is often determined from the behavior of linear approximations of the system in the vicinity of an equilibrium point (e.g., abundance of predators and prey, Box 1). Importantly, this means analysis is limited to exploring the behavior of systems with only small perturbations around an equilibrium point (where the linear approximation is appropriate). Furthermore, inferences could be misleading for highly nonlinear responses to perturbation [22]. Even if local stability analysis indicates strong **resilience** and a stable equilibrium, a system might temporarily move away from its equilibrium following a perturbation.

Persistent transient states are common and may differ substantially from long-term behaviors. These differences can emerge from frequently used models, such as the Rosenzweig-MacArthur predator–prey model (e.g., Box 1), without a change in parameter values. The **reactivity** of this predator–prey system indicates its potential to experience increasing deviations away from its equilibrium following a perturbation (Box 1). In systems with high maximum amplification, perturbations often produce large proportional deviations from equilibrium [23]. Reactivity and maximum amplification are derived quantities that change with model parameters and changes in these quantities do not directly align with changes in stability or resilience. Transient dynamics might also occur after a disturbance marked by a change in driving variables. Although local stability analysis does not predict system behavior immediately following a change, knowledge of how stability depends on model parameters can guide our expectations regarding transient behavior. For example, following a bifurcation where a stable equilibrium disappears, the shape of the **quasi-potential surface** [20] indicates if the system will linger in the vicinity of the former equilibrium [17].

Informing Dynamical Analysis with Ecological Data

Although dynamical analysis has predictive power, its accuracy is intrinsically linked to the quality of the underlying model and associated parameter values. Environmental stochasticity or variation in parameter values might lead to amplification of disturbances [24] or differences in expected dynamics. For example, varying parameter values in a differential equation model can determine whether a monotonic or oscillating approach to a stable equilibrium is expected (Box 1). Therefore, uncertainty in parameter values will lead to uncertainty about which dynamic behaviors are most likely. This sensitivity is particularly important when considering nonlinear models or processes beyond their stable equilibrium states. In theoretical studies, predictions might incorporate parameter values estimated from system observations (e.g., via statistical techniques such as regression), measurements of underlying processes (e.g., empirical growth rates), or rough biological reasoning (e.g., relative scales of parameters). Such methods can introduce unquantified uncertainty in parameter estimates and likely do not take advantage of increasing streams of ecological data. Conversely, approaches that effectively leverage ecological data and statistical models to quantify uncertainty often do not take advantage of unique insights provided by dynamical analysis. This shortcoming motivates this opinion paper.

In Figure 1, we illustrate how dynamical analysis and statistical methods can be combined to obtain analytical descriptions of ecological processes with associated uncertainty, informed by diverse data streams. To summarize, we propose analyzing process-based models (e.g., a differential equation model for predator–prey dynamics, Box 1) to derive biologically meaningful quantities of interest from model parameters (e.g., conditions for stability). The dependencies and outcomes of these models are integrated into statistical models conditioned on environmental processes, model parameters, and observed data. Such statistical models can accommodate various data sources [25], such as recent examples coupling diverse data with ecological models,

Glossary

Amplification envelope: the combination of the largest possible amplification of a perturbation and time when it occurs.

Dynamical analysis: mathematical analysis of a time-varying model to draw conclusions about the behavior of the underlying ecosystem.

Latent variable: processes, often time-varying, that are not directly observable and thus not explicitly known, but that describe important (ecological) system behavior.

Local stability analysis: determining long-term trajectories of a modeled system, dependent on model parameters and initial states of the ecosystem.

Parameter nonidentifiability: when a parameter cannot be reliably estimated, sometimes due to a lack of available information or indistinguishable model features, or where different combinations of parameter values can lead to the same posterior probability.

Quasi-potential surfaces: potential and quasi-potential surfaces are functions describing the amount of 'work' required to move from one state to another.

Reactivity: the maximum amplification rate of a disturbance, overall initial perturbations, immediately following the perturbation.

Resilience: an asymptotic approximation of the decay rate of perturbations to the linear system. The larger the resilience, the faster perturbations eventually decay.

Box 1. Asymptotic Analysis of a Predator–Prey Model

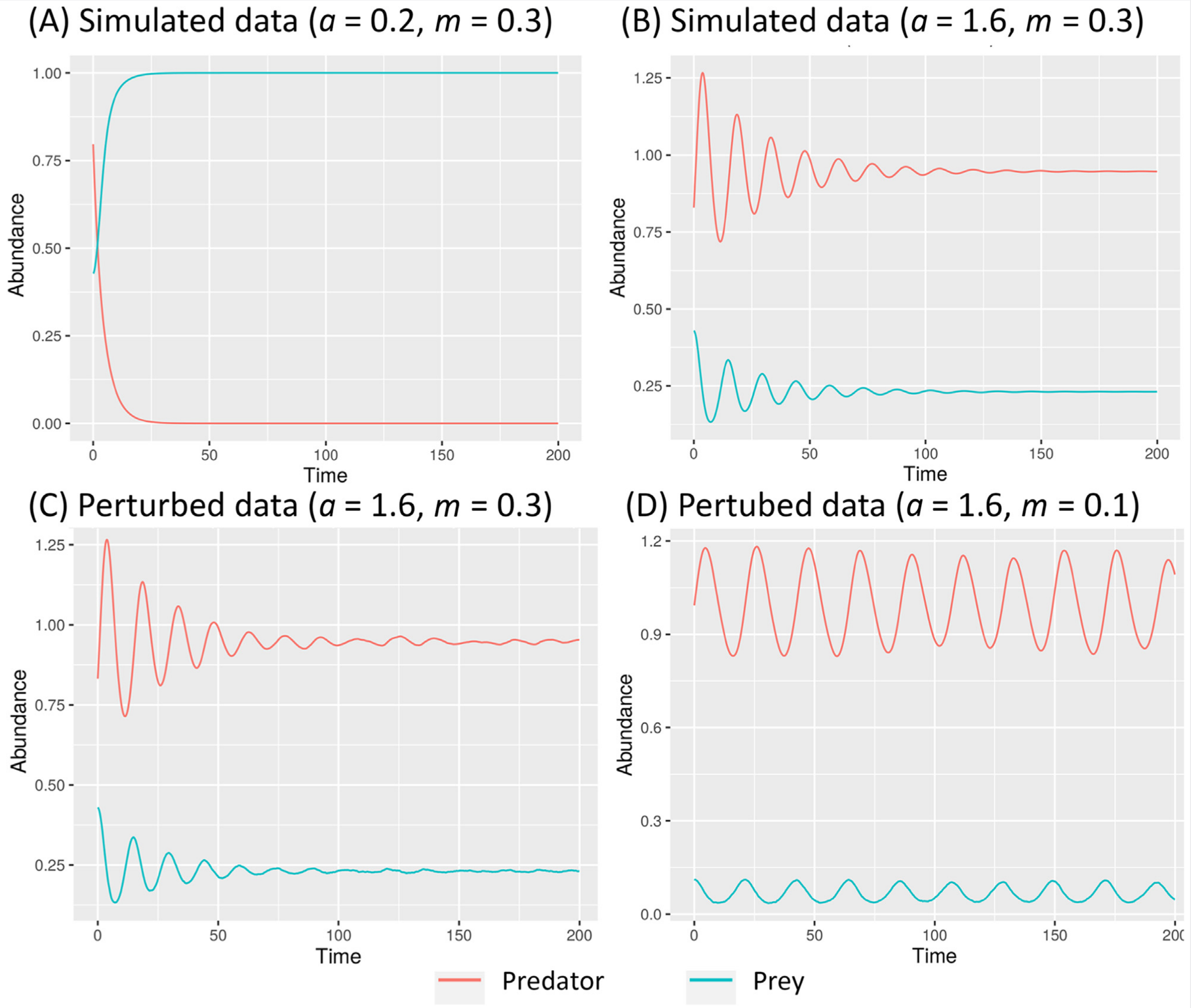
The Rosenzweig-MacArthur model describes dynamics between prey with density-dependent population growth and a specialist predator with a saturating consumption rate. A nondimensionalized version of this model for the numbers of prey n and predator p is given as:

$$\frac{dn}{dt} = n\left(1 - \frac{n}{k}\right) - \frac{np}{n+1} \qquad \text{[I]}$$

$$\frac{dp}{dt} = ap\left(\frac{n}{n+1} - m\right) \qquad \text{[II]}$$

where k is an aggregate quantity related to carrying capacity, a is an aggregate quantity representing the predation gains scaled by the prey growth rate, and m is the density-independent death rate of the predator scaled by predation benefits. There is a unique positive equilibrium at $n^* = \frac{m}{1-m}, p^* = (1+n^*)(1-\frac{n^*}{k})$. Analysis of eigenvalues for the linearized system near this equilibrium shows that the system is locally stable when $\frac{k-1}{2} < n^* < k$. When these eigenvalues are real, the population monotonically approaches the stable equilibrium (Figure IA), but when the eigenvalues have imaginary parts, the population oscillates around the stable equilibrium (Figure IB). These outcomes are determined by the relationship between a and m, and uncertainty in these parameters is important when determining dynamics. Knowing the probability of different parameter values can determine possible dynamics in the system.

Neubert and Caswell [23] demonstrate that even when an equilibrium is stable, the reactivity of this system can change with model parameters. When the reactivity is low, perturbations away from equilibrium quickly return to the stable value (Figure IC). When the reactivity is high, perturbations can result in long-lived oscillations with high amplitude (Figure ID). Correctly identifying the reactivity of the system, which is calculated from model parameters, is therefore important to understanding system behavior.



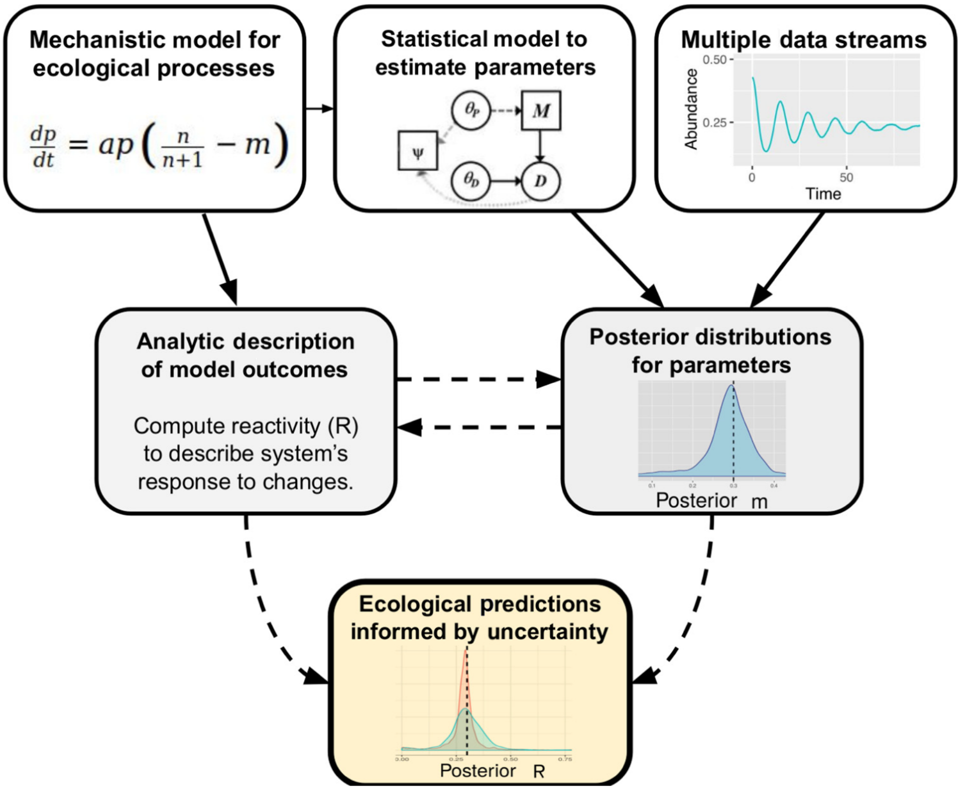
Trends in Ecology & Evolution

Figure I. Predator and Prey Abundances for the Rosenzweig-MacArthur Model. For different values of *a* and *m* (where *a* is an aggregate quantity representing the predation gains scaled by the prey growth rate and *m* is the density-independent death rate of the predator scaled by predation benefits), populations can exhibit a monotonic approach to equilibrium (A), oscillations towards equilibrium (B, C), or long-lived oscillations in the presence of noise (D).

including integral projection models [26] and integrated population models (e.g., [27]). We elaborate on this Bayesian approach to combining models and data in the next section, but note that this type of inference provides a basis for understanding limitations in data (e.g., from measurement error, sampling error, or missing data). Such limitations can inform design strategies for future data collection to facilitate and improve inference in the presence of uncertainty.

Statistical Tools for Quantifying Uncertainty

When performing ecological prediction or inference using process-based models, it is crucial to account for multiple sources of uncertainty. Ecological data may have significant measurement uncertainty (e.g., detection probability in occupancy models) and might only be partially



Trends in Ecology & Evolution

Figure 1. Schematic of Proposed Integration of Mathematical and Statistical Models to Gain Insight into Ecological Dynamics. Process-based models, statistical models, and empirical data (top row) yield analytic characterizations of the processes driving ecological change and statistical descriptions of uncertainty in model parameters (middle row). These valuable results can inform one another, leading to improved ecological understanding and predictive power (bottom row). For example, in Box 1 we consider a mechanistic model for a predator population p consuming some prey n according to parameters a (aggregate quantity representing the predation gains scaled by the prey growth rate) and m (density-independent death rate of the predator scaled by predation benefits). The statistical model (see Box 2) describes how different variables affect observations of the population. In Box 3, we demonstrate the results that emerge from integrating these descriptions.

observed. Additionally, many ecological phenomena are informed by data from multiple sources (e.g., observations of population counts, survival, mortality, fecundity, or mark-recapture, etc.). Incorporating multiple data sources can introduce substantial complexity into traditional likelihood-based analyses, especially when combined with nonlinear dynamics [28]. Such process-based modeling may stem from competing scientific theories, all of which are a simplification of reality. This introduces an additional source of uncertainty: process uncertainty (e.g., influenced by demographic and environmental stochasticity). Furthermore, some parameters associated with both the data and process models may also be the focus of inference. In many cases, these parameters can be stochastic or influenced by covariates, which can introduce parameter uncertainty (e.g., spatially varying, habitat-dependent carrying capacities in a population growth model).

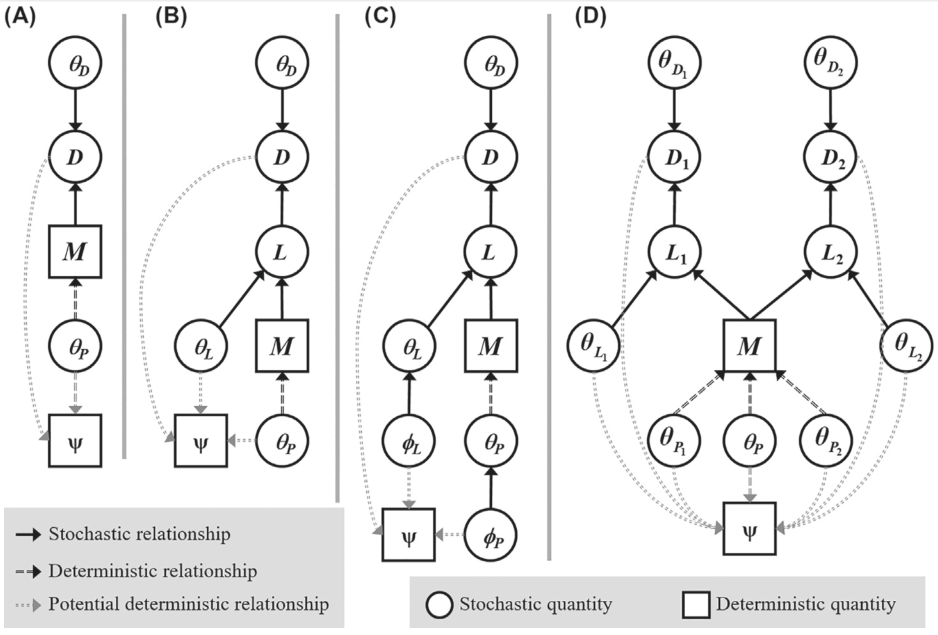
BHM is now widely used to effectively account for these various sources of uncertainty (Box 2; see also [8,29,30]). In the context of dynamical analysis of ecological systems, a BHM approach can be used to fit process-based models to available data for inferences of key model parameters. The BHM framework draws upon probability rules to model conditional data components

Box 2. Overview of the Bayesian Hierarchical Model (BHM) Approach Applied to Dynamic Ecological Systems

A BHM provides a statistical framework for linking empirical data to process-based models. This is illustrated by graphical models in Figure I, which give rise to a probabilistic expression of the BHM (see the supplemental information online), including the joint posterior distribution of all unknowns (parameters, derived quantities, latent variables), enabling inference about system behavior while accounting for noise/error in the data.

Let $\boldsymbol{D}$ and $\boldsymbol{\theta}$ denote observed data and unknown parameter(s), respectively. In the nonhierarchical Bayesian model (Figure IA), the data (e.g., population counts) are treated as stochastic, arising from some sampling distribution conditional on data-related parameters ($\boldsymbol{\theta_D}$; e.g., measurement error variances) and some estimated value. The estimated values (e.g., predicted population counts) are given by the process model ($\boldsymbol{M}$), conditional on process parameters ($\boldsymbol{\theta_P}$; e.g., growth rate, carrying capacity). We can also obtain posteriors for derived quantities of interest ($\boldsymbol{\psi}$; e.g., steady-state population size) that may be deterministic functions of stochastic quantities (e.g., $\boldsymbol{\theta_P}$, $\boldsymbol{D}$).

In hierarchical extensions (Figure IB–D), the latent process ($\boldsymbol{L}$; e.g., true population size) is observed with error (via $\boldsymbol{D}$) and imperfectly described by the process model ($\boldsymbol{M}$). Here, $\boldsymbol{D}$ is conditional on the truth ($\boldsymbol{L}$) and measurement error (quantified by $\boldsymbol{\theta_D}$). $\boldsymbol{L}$ varies stochastically around the modeled values, with process error (quantified by $\boldsymbol{\theta_L}$) describing variability not captured by $\boldsymbol{M}$. A further extension (Figure IC) specifies hierarchical distributions for process parameters ($\boldsymbol{\theta_P}$; e.g., a and m in Box 1), which may vary by levels of different factors (e.g., genotype, site, population) modeled hierarchically around global-level (e.g., species, region) parameters ($\boldsymbol{\varphi_P}$). When applied to multiple data sources (Figure ID), the data ($\boldsymbol{D_1}$ and $\boldsymbol{D_2}$; e.g., counts of a prey and predator, Box 1) are linked by a process model ($\boldsymbol{M}$; e.g., pair of differential equations for n and p, Box 1); the process parameters ($\boldsymbol{\theta_P}$) may be 'grouped' according to their relationship to different processes (e.g., parameters solely related to prey or predator dynamics, or those that influence both).



Trends in Ecology & Evolution

Figure I. Graphical Representations for Relationships in Bayesian Models. The directional edges (arrows) indicate conditional relationships among quantities (i.e., circle and square nodes) and whether these relationships are stochastic or deterministic functions.

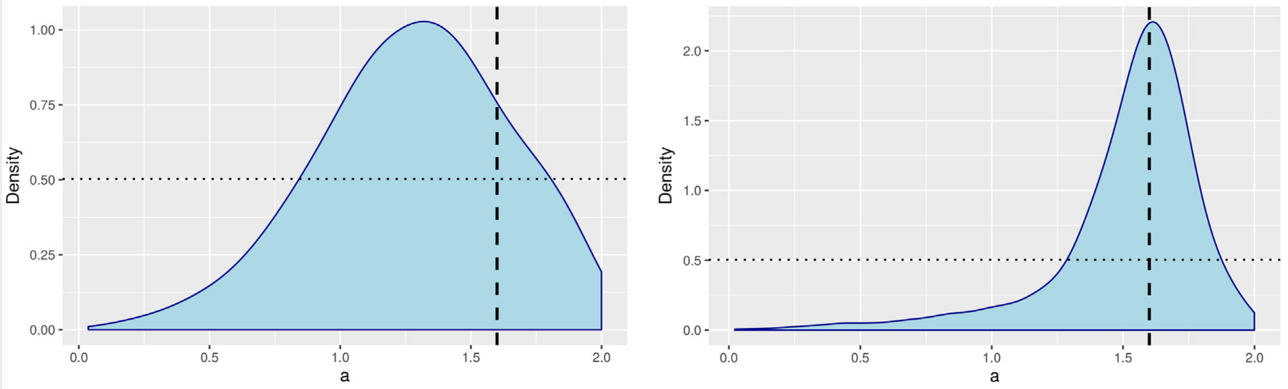
and **latent variables** linked to the process-based models (see the supplemental information online). Hierarchical models are particularly valuable in ecological modeling, since many processes and parameters of interest are only partially observed or latent, and hierarchical models can be used to explicitly separate these from the observation processes. The BHM framework produces a full, joint posterior distribution of all unknown quantities, which can be evaluated to obtain point estimates of parameters along with descriptions of uncertainty. It is then straightforward to obtain the posterior distribution of analytically derived quantities that are explicit (or implicit) functions of model parameters (e.g., eigenvalues, eigenvectors, and bifurcation points; Box 3). Additionally,

Box 3. Fitting and Transient Analysis of a Predator–Prey Model

For $k = 1$, we generated synthetic data for the Rosenzweig-MacArthur model (Box 1) and estimated a and m from the data (see Figure IC,D in Box 1). We use uniform prior distributions and a simple Metropolis-Hastings algorithm to update the posterior distributions and Gelman-Rubin diagnostics indicate that the algorithm converges. Please see the supplemental material online for R code to simulate the model and estimate parameters.

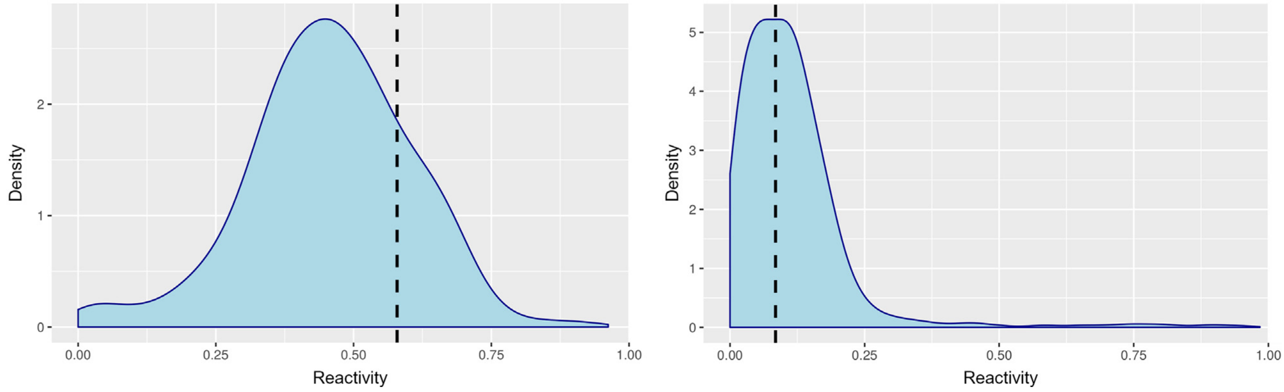
When $a = 1.6$ and $m = 0.1$, the system is expected to undergo damped oscillations in its approach to the equilibrium. However, in this part of parameter space the system is also highly reactive and, when perturbed, it undergoes lengthy transient oscillations (Box 1). As a result, posteriors for a and the reactivity are wide and shifted away from the true values (Figure IA,B). Different reactivity values correspond to drastically different predictions about the qualitative behavior of the system's transient dynamics. Importantly, these possible outcomes would not be apparent from a point estimate of reactivity.

In contrast, when $m = 0.3$, the system still converges to the equilibrium via damped oscillations, although the oscillations are less pronounced. Due to a lower reactivity, these oscillations do not grow or persist following a perturbation (Box 1). This leads to more precise posteriors for a and the reactivity (Figure IC,D). In this situation, increasing m (dependent on predator mortality) reduces reactivity and permits a better understanding of a (dependent on predator–prey interactions). An understanding of how the system's reactivity responds to model parameters could guide an intentional management decision to improve estimates. Alternatively, an understanding of the relationship between a and m in reactive regions could provide an informative prior to improve estimated posteriors (see Figure S1 in the supplemental information online).



(A) Posterior a ($a = 1.6$, $m = 0.1$)

(B) Posterior a ($a = 1.6$, $m = 0.3$)



(C) Posterior reactivity ($a = 1.6$, $m = 0.1$)

(D) Posterior reactivity ($a = 1.6$, $m = 0.3$)

Trends in Ecology & Evolution

Figure I. Posterior Distributions for a (A,B) and System Reactivity (C,D) in the Rosenzweig-MacArthur Model, Using Data from Box 1. Simulated data use the same value of a (a is an aggregate quantity representing the predation gains scaled by the prey growth rate) and two different values of m (m is the density-independent death rate of the predator scaled by predation benefits). The prior distribution for a is plotted as a horizontal dashed line and there is no explicit prior distribution for reactivity, which is derived from estimated parameters.

distributional inference on these quantities provides an understanding of the appropriateness of the underlying model, allowing for a scientific approach to model validation and selection.

Computational Challenges and Data Needs for Meaningful Analysis

The advantage of the BHM approach is that complexities in the data, ecological model, and parameters are all accounted for in a tractable fashion. The immediate challenge in applying BHM techniques to process-based models and diverse data streams is computational. These challenges may be exacerbated by complex, nonlinear models, but they can typically be overcome with efficient numerical sampling procedures. However, for large models with many state variables and parameters, the estimation of the likelihood may be nontractable (curse of dimensionality). In this case, so-called likelihood-free methods (e.g., approximate Bayesian computation [29,31]) can improve results.

A less straightforward challenge in implementing a BHM approach arises from model complexity and available data sources. Realistic models might incorporate many biological details at the expense of tractability. The complexity of these details might not be conducive to mathematical analysis or parameter estimation and **parameter nonidentifiability** can arise due to limited data availability, uncertainty in observations, model insensitivity to certain parameters, or model structure more generally. Furthermore, model parameterization can influence the ability to estimate parameters and achieve convergence with BHMs. In some cases, it may be necessary to reparameterize models to improve convergence and obtain identifiable parameter estimates. A robust statistical framework and underlying process-based model will help identify sources of prediction uncertainty and information gaps (e.g., uncertain estimates in a reactive region; Box 3), which can be used to prioritize future data collection efforts (e.g., modifying population death rates to move to a less reactive region; Box 3). Mismatches between theoretical predictions and ongoing data collection, combined with an improved understanding of underlying processes, might help pinpoint limitations of existing data sources and biological assumptions.

Richer Parameter Models

An important consideration in applying an approach that marries process-based models, dynamical analysis, and BHM statistical methods is whether meaningful analysis is possible from estimated posterior distributions. Wide posteriors may be induced by multiple sources of uncertainty, which can propagate through dynamical analysis and lead to imprecise estimates. Flat or wide posteriors may result from parameter nonidentifiability (see earlier) and inferences based on wide or flat posteriors will be noninformative. To remedy this, one must limit the number of inferred parameters, consider informative priors for parameters where such information is available, or reformulate the process-based model to reduce nonidentifiability [32,33].

Even in cases where the posterior distribution is precise, ecological systems can still be greatly influenced by varying external factors such as climatic or anthropogenic forcing. If these factors are not incorporated into the model, then resulting predictions may be misleading or inaccurate. Data on external factors can be directly accommodated within the BHM framework, often via multilevel models that specify submodels for parameters in the process-based model (e.g., [34]). Modeling such parameters extends typical approaches to dynamical analysis, which often neglect temporally varying parameters. One can also infer time-varying parameters to indirectly account for the influence of external forcing factors on these parameters [35]. Importantly, long-term monitoring of ecological systems and additional data sources will improve the viability of the proposed, integrative approach and lead to more meaningful analysis.

Concluding Remarks

We have outlined fundamental mathematical and statistical tools for studying ecological systems, which are commonly employed independently of one another. For a simple example, we demonstrated how dynamical analysis informed by robust statistical models, and vice versa, can improve estimates of model parameters and characterizations of long-term behavior through evaluation of derived quantities. Further integrating mathematical and statistical approaches to ecological research can yield unprecedented understanding of the dynamical properties of ecological systems and processes driving ecological change. However, to fully utilize these tools together, researchers must continue to address the computational problems that arise when working with complex models and integrating increasingly diverse and large datasets. Careful evaluation of the conditions under which dynamical analysis is likely to produce fruitful results can help researchers understand when to use such approaches with BHMs (see Outstanding Questions). Exploring insight gained from the integration of dynamical analysis, BHM statistical methods, and varied empirical datasets can lead to improved predictions of ecological dynamics, particularly in the presence of uncertainty and unprecedented ecological change.

Outstanding Questions

Incorporating biological realism into ecological models often increases complexity (e.g., a greater number of model parameters and/or state variables), leading to a tradeoff between representing sufficient biological detail and permitting tractable, informative mathematical analysis. How can researchers identify the necessary details to incorporate into such models? What indicators can be used to verify that relevant details have not been omitted from the mathematical model?

Nonidentifiable parameters may arise from complex models. How can researchers preserve the biological meaning of key parameters contributing to nonidentifiability? Common solutions include incorporating different data sources, using informative priors, or replacing mechanistic processes with time-varying, phenomenological parameters. In doing so, how can researchers evaluate and control the influence of different data sets and assumptions on the resulting analysis?

A priori analysis of model behavior over the parameter space might indicate that variation in some parameters will have a limited effect on system state, while others are important in determining long-term behavior. How can researchers leverage the integration of mathematical and statistical models to identify informative data sources or experiments and to understand *a priori* the range of realistic parameter values and biologically realistic covariation among parameters? How can such information be used to prioritize data collection in ecological systems?

Acknowledgments

We are grateful to Bo Zhang and Orou Gaoue for their comments on this paper, which began at a workshop hosted by the National Institute for Mathematical and Biological Synthesis (NIMBioS). Author participation in the workshop was supported by the National Science Foundation through NSF Award #DBI-1300426, with additional support from The University of Tennessee, Knoxville.

Supplemental Information

Supplemental information associated with this article can be found online at https://doi.org/10.1016/j.tree.2020.08.006

References

1. Franklin, J. *et al.* (2017) Big data for forecasting the impacts of global change on plant communities. *Glob. Ecol. Biogeogr.* 26, 6–17
2. Hampton, S.E. *et al.* (2013) Big data and the future of ecology. *Front. Ecol. Environ.* 11, 156–162
3. Theobald, E.J. *et al.* (2015) Global change and local solutions: tapping the unrealized potential of citizen science for biodiversity research. *Biol. Conserv.* 181, 236–244
4. LaSorte, F.A. *et al.* (2018) Opportunities and challenges for big data ornithology. *Condor* 120, 414–426
5. Christin, S. *et al.* (2019) Applications for deep learning in ecology. *Methods Ecol. Evol.* 10, 1632–1644
6. Hastie, T. *et al.* (2005) *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, Springer
7. Peters, D.P. *et al.* (2014) Harnessing the power of big data: infusing the scientific method with machine learning to transform ecology. *Ecosphere* 5, 1–15
8. Cressie, N. *et al.* (2009) Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling. *Ecol. Appl.* 19, 553–570
9. Evans, M.E. *et al.* (2016) Towards process-based range modeling of many species. *Trends Ecol. Evol.* 31, 860–871
10. Talluto, M.V. *et al.* (2016) Cross-scale integration of knowledge for predicting species ranges: a metamodeling framework. *Glob. Ecol. Biogeogr.* 25, 238–249
11. Clark, J.S. *et al.* (2001) Ecological forecasts: an emerging imperative. *Science* 293, 657–660
12. Dietze, M.C. *et al.* (2018) Iterative near-term ecological forecasting: needs, opportunities, and challenges. *Proc. Natl. Acad. Sci. U. S. A.* 115, 1424–1432
13. LaDeau, S.L. *et al.* (2017) The next decade of big data in ecosystem science. *Ecosystems* 20, 274–283
14. Grimm, V. (1994) Mathematical models and understanding in ecology. *Ecol. Model.* 75, 641–651
15. Kot, M. (2001) *Elements of Mathematical Ecology*, Cambridge University Press
16. May, R.M. (2001) *Stability and Complexity in Model Ecosystems* (Vol. 6), Princeton University Press.
17. Hastings, A. *et al.* (2018) Transient phenomena in ecology. *Science* 361, eaat6412
18. Scheffer, M. *et al.* (2001) Catastrophic shifts in ecosystems. *Nature* 413, 591–596
19. Bjørnstad, O.N. *et al.* (2004) Trends and cohort resonant effects in age-structured populations. *J. Anim. Ecol.* 73, 1157–1167
20. Nolting, B.C. and Abbott, K.C. (2016) Balls, cups, and quasi-potentials: quantifying stability in stochastic systems. *Ecology* 97, 850–864
21. Caswell, H. (2001) *Matrix Population Models: Construction, Analysis, and Interpretation* (2nd edn), Sinauer Associates
22. Tremblay, R.L. *et al.* (2015) When stable-stage equilibrium is unlikely: integrating transient population dynamics improves asymptotic methods. *Ann. Bot.* 116, 381–390
23. Neubert, M.G. and Caswell, H. (1997) Alternatives to resilience for measuring the responses of ecological systems to perturbations. *Ecology* 78, 653–665
24. Freilich, M.A. *et al.* (2020) Reconstructing ecological networks with noisy dynamics. *Proc. R. Soc. A* 476, 20190739
25. Li, L.M. *et al.* (2017) Quantifying transmission heterogeneity using both pathogen phylogenies and incidence time series. *Mol. Biol. Evol.* 34, 2982–2995
26. Elderd, B.D. and Miller, T.E. (2016) Quantifying demographic uncertainty: Bayesian methods for integral projection models. *Ecol. Monogr.* 86, 125–144
27. Zipkin, E.F. and Saunders, S.P. (2018) Synthesizing multiple data types for biological conservation using integrated population models. *Biol. Conserv.* 217, 240–250
28. Pontarp, M. *et al.* (2019) Inferring community assembly processes from macroscopic patterns using dynamic eco-evolutionary models and approximate Bayesian computation (ABC). *Methods Ecol. Evol.* 10, 450–460
29. Clark, J.S. (2005) Why environmental scientists are becoming Bayesians. *Ecol. Lett.* 8, 2–14

30. Wikle, C.K. (2003) Hierarchical Bayesian models for predicting the spread of ecological processes. *Ecology* 84, 1382–1394
31. Beaumont, M.A. *et al.* (2002) Approximate Bayesian computation in population genetics. *Genetics* 162, 2025–2035
32. Haneuse, S.J.P. and Wakefield, J.C. (2007) Hierarchical models for combining ecological and case–control data. *Biometrics* 63, 128–136
33. Luo, Y. *et al.* (2009) Parameter identifiability, constraint, and equifinality in data assimilation with ecosystem models. *Ecol. Appl.* 19, 571–574
34. Guo, J.S. *et al.* (2020) Temporal shifts in iso/anisohydry revealed from daily observations of plant water potential in a dominant desert shrub. *New Phytol.* 225, 713–726
35. Cazelles, B. *et al.* (2018) Accounting for non-stationarity in epidemiology by embedding time-varying parameters in stochastic models. *PLoS Comput. Biol.* 14, e1006211