00163 Computer Systems and Models, Use of

Louis J. Gross, Department of Ecology and Evolutionary Biology & Department of Mathematics, University of Tennessee, Knoxville, TN, United States

Brian Beckage, Department of Plant Biology & Department of Computer Science, University of Vermont, Burlington, VT, United States

Abstract

Models in a variety of forms play a critical role in advancing our understanding of natural systems. Models abstract basic principles and derive the implications of such abstractions. This provides a method to analyze alternative hypotheses about natural system responses and the mechanisms that underlie these responses. This article presents an introduction to various modeling approaches, emphasizing those implemented on computers, and discusses how these have been applied to analyze species abundance and distribution.

This is an update of Louis J. Gross, Computer Systems and Models, Use of, Editor(s): Simon A Levin, Encyclopedia of Biodiversity (Second Edition), Academic Press, 2013, Pages 213–220,ISBN 9780123847201, https://doi.org/10.1016/B978-0-12-384719-5.00305-1.

Glossary

Aggregation Combining several potentially separate components of a system to simplify analysis.
Dynamic model Mathematical description of a system with components that vary in time.
Machine learning Methods to train computational algorithms that classify, predict, cluster or discover patterns in a dataset.
Model A construct, physical, mathematical, or computational, that is a simplification of reality developed to meet certain objectives.
Multimodel Single integrated model that links together models utilizing different mathematical or computational approaches.
Parameter Constant in a model that must be estimated from data or assumed to be of a particular value.
System dynamics An approach for analyzing the behavior of complex systems over time.

Key Points

- There are many objectives for models applicable in biodiversity science.
- Models of natural systems can be used for description, understanding mechanism, projection, forecasting, and control.
- Quantitative models can be formulated using dynamical systems, computational, statistical, and data-driven methods.
- Major applications of models in biodiversity include habitat suitability, metapopulations, individual-based approaches, multimodeling and behavioral dynamics.

Introduction

Just as storytellers can take their audience on trips to faraway places and provide a glimpse of life in different cultures, scientists tell their stories about the way the world works by making models. Such models never provide a complete view of how the world works, but do give us glimpses that help us to piece together interactions between different parts of the world and the processes that connect them. These models take many forms, some being mostly verbal, others mostly qualitative and graphical, some phrased in various mathematical forms, and still others set up as collections of rules within a computer program. Modeling is a process to attempt to understand the world around us. Creating a model that can reproduce the observations from or dynamics of systems around us can help us understand how these systems operate. However, there is no guarantee that any particular model is the ‘truth’ because there may be many models that produce similar results.

The Modeling Process

Models provide maps of varying levels of complexity to help us understand the topography of natural systems. There are coarse road maps that provide merely the outline of major arteries for traffic, telling us nothing about buildings or other features of the landscape, but providing an overview of the linkages between key components of a system. More elaborate models show us the buildings and the infrastructure that links these buildings – water and power lines. Even more complex models would indicate the humans in each building, their occupations, their social behavior, and the flow of money or capital goods between them. Similarly, models at many levels of complexity can be useful for addressing different questions in the study of biodiversity. A very coarse model might analyze the effect of land use change worldwide on total species richness. A more complex model could consider particular regions and analyze differentially the changes in land use within them and the associated changes in species richness. A still more complex model could consider the local dynamics of a population of a particular species and its interactions with the local environment and from this elaborate the dynamics of the population′s abundance as well as the abundances of populations of other interacting species. A further level of complexity in model formulation would incorporate the feedbacks of land use on climate change, connect this to socioeconomic models that account for ecosystem services, and project how these forces impact biodiversity from local to global scales.

The modeling process includes several steps. It all starts with data or observations. The next step is specifying a set of questions or objectives for the model to address. The type of model chosen will depend on the model objectives. It is helpful at this early stage of model development to specify criteria for evaluation of the model. Such criteria can then be used to determine whether the model is acceptable for the purposes for which it is being constructed. The model structure is conceptualized based on key components, processes, and interactions to be included. This conceptual model is then implemented in detail by specifying functional forms, mathematical relations, rule sets, and algorithms. Once parameterized with available data, a process called calibration, the model implementation is then used to investigate the behavior of the system being modeled. The model is evaluated based on the specified criteria and potentially applied to address the questions of interest or the model is modified in light of the evaluation criteria.

The Purposes of Modeling

The type of model one constructs depends on the questions being asked and the availability of data. Models are used for a variety of purposes, but include, in no particular order:

1. To suggest experiments or observations to collect.
2. To provide a framework to assemble and organize bodies of observations.
3. To "allow us to imagine and explore a wider range of worlds than ours, giving new perceptions and questions about how our world came to be as it is" (**Jacob, 1982**).
4. To clarify hypotheses and chains of argument.
5. To identify key components in systems.
6. To allow investigation while accounting for societal or ethical constraints.
7. To allow simultaneous consideration of spatial and temporal change.
8. To extrapolate to broad spatial or long temporal scales for which data cannot easily be obtained.
9. To prompt testable hypotheses.
10. To serve as a guide to decision making in circumstances where data are unavailable.
11. To provide an antidote to the helpless feeling that the world is too complex to understand by providing a means to investigate general patterns and trends.
12. To predict how a system will behave under different management strategies or to control the system to meet some objective.

The many specific purposes for constructing models may be grouped into a few general objectives: description, understanding mechanism, projection, forecasting, and control (**Haefner, 2005**). These objectives are not mutually exclusive, so that descriptive and mechanistic approaches may be used to aid forecasting and control.

Description

Sometimes all that might be desired is a simple description of a collection of data such as observations on the state of a system. For example, an arithmetic or geometric average provides a single value to summarize a list of numerical data. The single descriptor used may depend on the type of data – for example, a geometric mean is typically more appropriate as a metric for summarizing varying growth rates than an arithmetic mean. A single descriptor may be sufficient for some purposes that do not require knowledge of how much variation is in the data or in the temporal sequence of observations. To assess variation, a dispersion measure such as variance would be needed. These summary statistics are coarse, ignoring many of the details in the data. Yet they do allow us to easily comprehend major differences between different data sets.

Extensive species lists within certain taxa from two locations may be compared by considering just the total numbers of species in the two locations and the number of species in common between them. Such a summary may be sufficient for a comparison of the two locations, while ignoring details such as the diversity within the taxa included.

Descriptive approaches may be much more complex than simply providing averages and variances. Options for analysis of ecological data have grown tremendously due to the ready availability of computational platforms and open-source tools such as R or Python and associated libraries. Bayesian methods and exploratory statistical approaches are applied to analyze complex multidimensional data (**Whitlock and Schluter, 2008**). Machine learning methods provide an automated means to investigate patterns and create novel descriptions of large, multivariate datasets (**Christin *et al*., 2019**). Spatiotemporal analysis as an outgrowth of traditional time-series and spatial methodologies allows the histories of species richness or abundances to be compared between locations or correlated with the histories of anthropogenic actions in the locations. Geographic Information Systems provide the functionality to analyze biodiversity across multiple regions, accounting for landscape metrics such as distance and connectedness.

Understanding Mechanism

If the objective is to provide an understanding of how a particular system operates, then it is necessary to take account of the processes that govern the system. While all such mechanistic models are descriptive at some level, the point is to represent the basic physical, chemical, and biological processes operating in the system. This requires including those processes that operate at a spatial and temporal extent appropriate for the problem being addressed, and ignoring others. Thus, analysis of how alternative global warming projections would affect worldwide biodiversity might include the geographic variation in the temperature projections at a spatial extent of hundreds of square kilometers, but would no doubt ignore the microclimate variation of every square meter. Even if it were possible to characterize the meter-by-meter temperature differences expected to occur according to alternative warming trends, the lack of available detail on the species present at this detailed spatial resolution limits the utility of including such detail. Despite the lack of data at fine spatial resolution, it is feasible to consider hybrid methods that incorporate mechanistic biophysical models for constraints on species' presence linked to either statistical models based on climate data or down-scaled climate models. (**Buckley *et al.*, 2010**). For a discussion of scaling see the chapter "Concepts and Effects of Scale".

Projection and Forecasting

Forward-looking models are of two general types: those that attempt to project the behavior of the system based on certain explicit assumptions (projection), and those that attempt to forecast the future behavior (forecasting). The difference is between what might be true in the future if certain assumptions hold (projection) and what will be true in the future (forecasting; **Caswell, 2001**). In many biological situations, the forecasting problem is challenging, as it would involve taking account of a wide variety of unpredictable abiotic phenomena (e.g., hurricanes and droughts). It is often feasible to construct a model to forecast over some time period the future dynamics of a system based on current observations and particular reasonable assumptions about the interactions in the system. Ecological forecasting (**Dietz, 2017**) provides formal methods to analyze numerous problems related to biodiversity such as:

(i) Population sizes/demography for which there is a long history in fisheries, wildlife management and epidemiology with direct impact on policies for harvest, quotas, vaccination, etc.
(ii) Environmental processes including: biogeochemistry, temperature and precipitation projection on various time and spatial scales. Models here link ecological states (e.g. landscape-type) to atmospheric and oceanographic physical models including mechanistic and statistical methods and ecosystem models.
(iii) Distribution and abundance in which environmental variables are used to project presence/absence and/or abundance. This typically involves the development of Species Distribution Models (SDM), also called niche models.
(iv) Landscape/functional group/community which uses environmental variables, history and community dynamics models to project what community type and level of biodiversity is present across space.

The majority of population models (discussed in the chapter "Population Dynamics" in these volumes) project the future dynamics of a population based on the biotic forces of demographics, genetics, and social structure within the population. SDMs can project shifts in spatial patterns of species' presence based on assumptions on spatial climatic patterns as well as look at historical patterns of biodiversity dynamics using paleoclimate data (**Carnaval *et al.*, 2009**). Uncertainties associated with unpredictable phenomena can be taken into account by attempting to project just the mean and variance of the variables of interest (e.g., population size), or by developing a stochastic model that incorporates assumptions about the probabilities of alternative future conditions. In the latter situation, projections would typically arise from numerous alternative simulations accounting for the assumed probabilities of different conditions, and the projections are therefore often not single values or single spatial maps, but rather probability distributions for the values of interest.

Control

When a system has one or several components that are under human control, either completely or in part, then a model can be used to help determine how to apply such a control to meet certain objectives. Examples of controls are harvest quotas, fertilization or pesticide application, flows from a dam, constraints on importation of potentially harmful invasive species, mitigation of wildfires, and land use zoning regulations. Examples of objectives are maximizing biodiversity, minimizing population extinction probability, reducing the spread of nonnative species, and maintaining population size above some determined threshold (such as a minimum viable population size). Control models mostly focus on the dynamics of the system, with the simplest form of control being bang/bang, meaning on/off, such as allowing harvest in certain years and not allowing harvest in other years. A related objective is for control models to produce a relative comparison of alternatives in order to rank these alternatives according to some criteria (**DeAngelis *et al.*, 1998**). Still other control models are used to analyze the physiological responses of individual organisms to varying environments and the homeostasis that can arise through these responses. Spatial control involves choosing what action to carry out, where to do it, and when to apply the action. Many problems in spatial management of natural systems can be phrased as optimization problems in which the actions taken at different locations are constrained, often due to some overall budget limitations, with an objective to maximize some characteristic of the entire landscape, such as a diversity metric (**Hof and Bevers, 2002**). Models for environmental organizations with goals of maintaining biodiversity have taken account of economic and effort constraints to provide guidance on optimal spatial and temporal management strategies (**Le Bouille *et al.*, 2022**).

Types of Models

Physical Models

Models can be physical, such as animal models used in drug testing and airplane models used in wind tunnels. In biodiversity contexts, microcosms and mesocosms, which are limited biological systems built in a laboratory setting, play this role. These are meant to mimic the key biotic forces interacting within a natural system, but are constructed at a spatial extent that allows for easy observation and controlled experimentation. They cannot include all of the components of the real system, but do allow for projection of how the real system might respond under particular perturbations. A wide array of model systems have been proposed as appropriate to evaluate responses of organism behavior and species interactions. Physical models are clearly limited, particularly to organisms that are mostly sessile or have very short distance movements.

Mathematical Models

Mathematical models come in a wide variety of forms. Some are simply graphical relations that show the qualitative relationship between certain components of a system, mainly to demonstrate the shape of response and whether one component increases or decreases with another. An example would be the increase and then decrease in species diversity along a gradient from low to high frequency of disturbance (**Fig. 1**). Here, there is no attempt to predict at exactly what disturbance frequency the exact peak in species diversity occurs. Rather, the objective is to illustrate the qualitative behavior of diversity, for example, the intermediate disturbance hypothesis, which posits that high diversity occurs at intermediate disturbance frequencies.

Many mathematical models in ecology deal with the dynamics of populations and communities. Such models consider the basic processes of birth and death, immigration and emigration, and competition and predation to elucidate general theories of population dynamics. Described using differential or difference equations, these models allow for projection of the long-term behavior of populations, and provide methods to project the within-population structure (age, size, genetic, etc.) over time and space. These models can be expanded in numerous ways to incorporate between-species interactions, account for demographics of births and deaths, use discrete spatial units (e.g., islands or patches) as well as continuous spatial spread, and deal with the stochastic aspects arising particularly in small populations (**Murray, 2007**).

Computational Models

Computational models have become a standard tool in ecology. Computers are used to analyze complex data, interpret the implications of mathematical models by carrying out numerical schemes to solve the mathematical models, and to create models that are not readily described through any other means than through computer code. The algorithm is the set of instructions for the computer that represent the set of rules that are hypothesized to govern the behavior of the system. Computer systems designed for ecological monitoring and real-

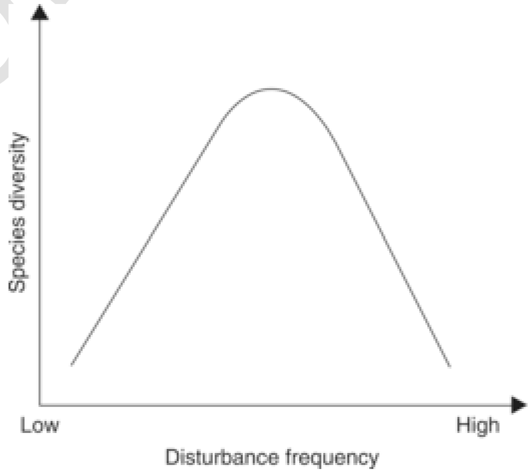

Fig. 1 Illustration of the intermediate disturbance hypothesis.

time analysis, as have been constructed for the National Ecological Observatory Network, also incorporate specially designed workflows to visualize, manage, and analyze a variety of incoming data streams, and provide remote access to diverse users. Cyberinfrastructure projects, such as CyVerse, provide collaborative tools and access to open source and cloud computer systems to link information at various scales of investigation, from the genome to the whole organism, to the ecosystem.

This article emphasizes the use of computers to model biological processes, with the realization that technology affects our capacity to develop relevant models and investigate natural systems. Our ability to mine large data sets arising from field-based and remote sensors, construct computational models from this, and evaluate the utility of these models, continues to expand. The advent of parallel computation, in which computational tasks are split among numerous processors, has greatly expanded the speed of computation, the range of problems addressed, and the methods we can use to address them. The widespread use of Markov Chain Monte Carlo methods to carry out complex Bayesian modeling is one example of the continuing potential for advances in computational power to affect the science of biodiversity. The rise of data science, machine learning, neural networks, and artificial intelligence (e.g., arising from large language models) to investigate problems and provide novel biological insights from various data streams is yet another set of tools applicable to many areas of earth and natural science (**Reichstein *et al.*, 2019**).

Computational models are quite varied in structure. All the standard mathematical models of populations and communities, constructed using differential or difference equations, may be implemented on computers. Indeed, since it may be quite difficult to develop analytical solutions for such models, analysis of their behavior often requires the use of numerical solution methods implemented on a computer. There are many computer models that, although they may have a description that is essentially mathematical in form, are really described by the computer code itself rather than an explicit set of mathematical equations. An example would be cellular automata models, one type of which consists of a two-dimensional lattice, with each point on the lattice having one of a number of states. The simplest situation would be each lattice point being occupied (e.g., in the 1-state) or unoccupied (in the 0-state). The model is then described by a set of rules that determines how the state of a lattice point changes from one time step to another, based on the states of surrounding lattice points. Such a cellular automaton may be used to mimic the spatial dynamics of populations, in which each lattice point represents a possible location of an individual. Alternatively, each lattice point can be interpreted as a local population, and the entire lattice can then follow the collection of such populations, called the metapopulation. It has been argued that the major questions in science are feasible to be investigated using cellular automata (**Wolfram, 2002**).

System Dynamic Models

System dynamic (SD) models are used to represent complex systems characterized by feedback loops among components. SD models are often elaborate computer models that attempt to visually represent most of the major biotic and abiotic factors that affect the system through stocks and flows using systems of differential equations. Many agricultural system models are of this type, and include the crop, its pests, soil nutrients, and weather conditions, among other factors. Some other types of computer models are described in later sections of this chapter and in the chapters "Dynamic Global Vegetation Models" and "Landscape Modeling". In all cases, though the model is in essence specified by the code itself, it is very useful to have some graphical description of the major components of the model. One example is shown in **Fig. 2**. A number of general modeling software packages for system dynamics are designed explicitly to aid construction of computer models through the use of graphical elements. These packages provide users with the capability to rapidly develop and com-

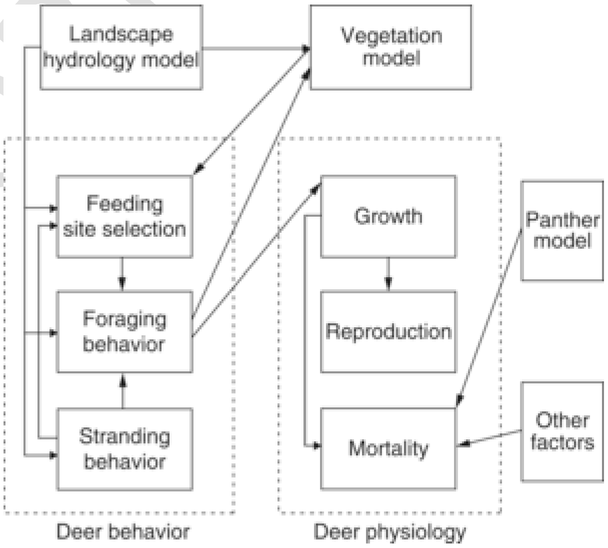

Fig. 2 Graphical depiction of the major components of an individual-based model for deer.

pare models with differing numbers of components, for example, differing species and associated trophic interactions between them, by automatically generating the computer code for the models and allowing users to vary the basic model parameters such as birth rates. Each of these tools has limitations, for example, in the solution methods they employ, but as long as users are aware of these, the tools can greatly reduce the time needed to construct computer models with many interacting components.

Limitations of Models

Trade-offs: Generality, Precision, Realism

No single model can do everything. In the process of deciding what components of a system to include, what processes to consider, and what spatial and temporal extent is appropriate, the model excludes part of reality. Modeling is a process of selective ignorance. We decide what to include and what to exclude. Part of the art of modeling is coming to grips with the issue of which details are important and which ones are not. In most cases the process is iterative, with a sequence of different models being tried until a model is arrived at that includes just the essential details necessary to address the problem of concern. Inherent in the process of modeling is that there are criteria by which the model will be evaluated. These criteria might be best chosen prior to model development, based on the objectives for which the model is being constructed.

One view of the trade-offs in constructing a model is that no single model can be simultaneously general, precise, and realistic (**Levins, 1968**). As **Fig. 3** illustrates, these properties may be viewed as points of a triangle. Generality implies that the model may be useful in many different natural systems. A realistic model is one with components, parameters, and variables that are all possible to estimate from observations. A precise model is one that produces quantitative, accurate descriptions of the natural system. Models for theory development, including most of the classical population and community models, are quite general and somewhat realistic, but lack in precision. Descriptive models designed to mimic the response of particular systems tend to be quite precise, slightly realistic, and not at all general. Much of the approaches in data-driven modeling (regression being a classical one), including models derived from machine learning methods, are of this type. They may provide an accurate portrayal of a particular system, for example, winter wheat growth in Nebraska, but may well not transfer effectively to other situations such as winter wheat growth in eastern Russia. For example, the effectiveness of supervised machine learning methods depends critically upon the training sets used to develop the model. System dynamic and control models take up various positions in the figure, depending on the level of precision desired.

Aggregation and Loss of Detail

A factor that affects where a particular model fits into the scheme shown in **Fig. 3** is the amount of aggregation included. Aggregation is sometimes referred to as coarse graining and trades off detail for predictive power (**Beckage *et al.*, 2011**). Natural systems consist of many components that can be lumped together or disaggregated. Population models, for example, that use a single variable to represent the whole population inherently ignore the within-population structure (e.g., age and size) or equivalently incorporate assumptions about the distribution of this structure within the population (e.g., stable age distribution). Such a model would not be able to discriminate between a population with mostly small individuals and one with mostly large individuals, unless this within-population structure was assumed to affect the population's growth characteristics in some manner. Aggregation implies a loss of detail that allows consideration of quite general models. Often this then requires model parameters that are not at all easily estimated from observations. Examples would be the growth rates in population models and the competition coefficients in community models. The less the aggregation in the model, the more parameters there tend to be and more data are required in order to estimate them.

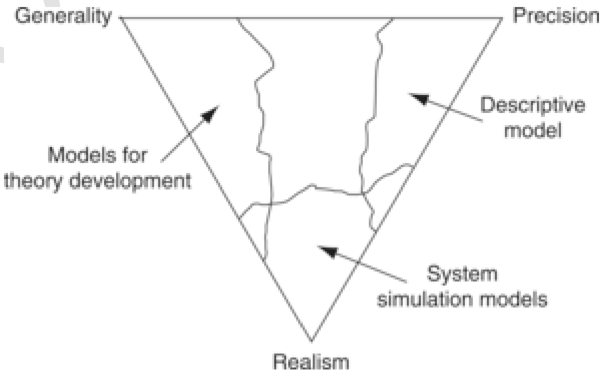

Fig. 3 Trade-offs in modeling.

Uncertainties: Mechanisms, Data for Parameterization, Biotic and Abiotic Forcing Functions

The modeling process is limited by the information available. There may not be basic agreement on the mechanisms that are critical in the system of concern, so that any particular model includes only some of the mechanisms, or just partial information on these mechanisms. Many population models applied to vertebrates ignore social structure, despite evidence that it is often present in such populations. Excluding some mechanisms in a model may occur due to lack of understanding of the effect of such mechanisms at the scale of interest for the model. This can give rise to another set of models designed specifically to investigate these effects. So there might be a hierarchy of models, each of which disaggregates some components in order to investigate the potential impacts of uncertainties in these components on the overall model results of particular interest.

Under situations in which the mechanisms are well understood, it may not be possible to accurately estimate model parameters because adequate data are lacking. A variety of statistical methods are designed specifically to determine the optimal choice of parameters in such situations (**Hilborn and Mangel, 1997**). These methods may take account of parameters estimated either for similar species or for similar locations other than the one being considered. For example, it may be necessary to use observed clutch size distributions for one bird species in application to a similar species about which such information is lacking. Another uncertainty associated with natural system models arises due to the unpredictability of forcing functions such as weather and disturbance. If historical information is available on these functions, this may be used to estimate the stochastic effects of such forcing. Simple ways to incorporate unpredictability of forcing functions would be to determine how the model responds to the maximum and minimum observed values of these forcing functions and compare this to the model results when using the mean value of the forcing function. This is one form of uncertainty analysis in which a protocol is used to evaluate the impact of uncertainties in model inputs and model structure on the results of the model. Computational methods make it very feasible to evaluate model response to more elaborate and realistic assumptions on forcing by making numerous model simulations under different forcing function assumptions. For example, **Fuller *et al*. (2008)** evaluate the robustness of Everglades restoration scenarios under diverse assumptions about future rainfall conditions. Various uncertainties limit the detail at which models may be constructed, and thus limit the types of questions that may be addressed using models.

Some Tools of the Trade

Statistical Approaches

Statistical models usually have a descriptive objective rather than a mechanistic one. The form chosen for the model and parameters within the models are directly estimated by choosing them in a manner that best fits a certain dataset. Thus, any particular statistical model is typically not very general in application to different systems. The structure of such models may be useful in a wide variety of different contexts. Regression models, which assume a particular mathematical relationship between variables and assume that errors in the data take a particular form, are widely applied. Numerous regressions have been estimated for species richness as a function of latitude, altitude, and rainfall (**Huston, 1994**). Discussion of statistical methods applied to estimation of population sizes and densities may be found elsewhere in these volumes.

An alternative view of statistical models is that they are far more than simply descriptive, by providing a method to tease apart potential impacts of complex interactions from limited observations. This is one of the objectives of hierarchical Bayesian methods in which data on uncontrolled variables are used to estimate the interactions at different levels or scales in a system. This might include determining the relative genetic variability within and between populations, developing methods to combine data from quite different experiments, and analyzing how global change impacts plant diversity through its various effects on CO_2, nitrogen deposition, and changes in precipitation. The idea is to consider a hierarchy of processes in which the probability distribution at one level of the process depends on that of components at other levels. Using conditional probability arguments several times allows a Bayesian updating to build up a complex model from simpler conditional relationships. The posterior distribution can then be quite complex and analytically intractable, but can be obtained using Markov Chain Monte Carlo methods in which the Markov chain being simulated has the same stationary distribution as the desired posterior distribution from the hierarchical Bayes model. See **Clark and Gelfand (2006)** for a variety of applications of this methodology to environmental problems.

Data Science and Machine Learning

The broad area of data science deals with the intersection of data collection, curation, integration, and analysis, with the analysis methods including statistical models developed through classical means as well as those developed through recent machine learning methods. Machine learning trains computational algorithms that split, sort, and transform a set of data to maximize the ability to classify, predict, cluster, or discover patterns in a target dataset. Deep learning refers to machine learning algorithms that construct hierarchical architectures of increasing sophistication. Artificial neural networks with many layers are examples of deep learning algorithms. Artificial intelligence is the capacity of an algorithm to assimilate information to perform tasks that are characteristic of human intelligence, such as recognizing objects and sounds, contextualizing language, learning from the environment, and problem solving (**Reichstein *et al*., 2019**). Types of machine learning include:

(i) Supervised learning using algorithms which learn from a training set of “labeled” examples (exemplars which include the outputs, not just the inputs) to generalize to a broader set of possible inputs. Techniques include logistic regression, support vector machines, decision trees, random forests.
(ii) Unsupervised learning uses algorithms which learn from a training set of “unlabeled” examples (e.g. this does not include the outputs) so the goal is to infer the underlying structure of the data according to some statistical, geometric or similarity criterion. Techniques include k–means clustering, kernel density estimation, neural nets.
(iii) Reinforcement learning includes algorithms which learn via reinforcement from criticism that provides information on the quality of a solution, but not on how to improve it. So improved solutions arise from an iterative process to explore the possible solution space. Techniques include dynamic programming for Markov decision processes, Q-learning, policy iteration, and neural nets.

The major classes of problems addressed using machine learning include:

(i) Classification - specifies the class to which data elements belong so this inherently involves “splitting” data into classes and may involve some kind of clustering technique to do so. Classic problems in phylogeny are examples, such as assigning organisms to a grouping based on ancestral relationships, but also classifying images, landscapes, etc.
(ii) Regression – models a target prediction based on a set of inputs and often is used as a technique in supervised learning.
(iii) Anomaly detection – identifying items or events that do not conform with an expected pattern in a broad data set. Examples include extreme weather events and rare behaviors of individuals in a population.
(iv) State prediction or forecasting – projecting outcomes based on properties of data for new experiments or time-dependent conditions. Ecological forecasting is an example in which future trajectories of a natural system are projected based on current conditions and assumptions about future environments.
(v) Dimensionality reduction – identifying a more compact representation of a data set without a great loss of information. Principle components analysis and multidimensional scaling are examples.

Machine learning methods have been applied in many contexts to biodiversity science – an example is the analysis and estimation of conservation risk for thousands of land plant species across the world based on georeferenced data for > 150,000 species (**Pelletier *et al.*, 2018**).

Dynamic Models

Although many of the traditional models for populations and communities are in the form of dynamical systems (e.g., collections of linked differential or difference equations), often the types of analyses performed for these models are based around equilibrium assumptions. The objective is to find long-term asymptotic behavior. This may be a static equilibrium (e.g., population sizes approach a constant value through time) or a dynamic equilibrium (e.g., population sizes follow repeatable patterns through time; **Murray, 2007**). While these situations may arise, many models produce behavior that does not have a long-term equilibrium structure. Another key objective is to determine the stability characteristics of any equilibria that arise, in the mathematical sense of determining whether a model that is perturbed from an equilibrium condition will return to it. The dynamics arising in all these models takes account of the basic demographics of the population, as well as interactions with other populations. Adding abiotic conditions such as temperature and rainfall, or adding spatial components, often requires that the analysis be done using numerical simulations. The availability of rapid solution methods for even quite complex dynamical systems has allowed for extensive investigation of the behavior of these models when parameters are not well known. “Parameter sweeps” are particularly simple to perform in parallel using clusters readily available through cloud computing resources, and provide a means to determine how a system is projected to move from one type of dynamical behavior to another as parameters change (**Beckage *et al.*, 2018**).

Geographical Information Systems (GIS)

Remote-sensing methods have opened new possibilities for following and modeling the responses of the earth′s biota. A key tool that allows the use of such materials is GIS, which enables computers to graphically display the remote-sensing data as two-dimensional maps. Each image may represent one aspect of an underlying landscape, such as landcover or vegetation type. The image value at any particular location (or pixel) in the map is estimated using models that classify the output of the cameras or the multispectral scanners on satellites into types appropriate for the objective. These models require ground-truthing to ensure that the estimated value for a particular location matches what is actually present. GIS methods allow various spatially explicit components of a landscape to be combined by looking at different map layers (different images measuring different aspects of the landscape). A mathematical function is then applied that averages or applies thresholds to these various components. Estimates of regional and worldwide carbon uptake are obtained using such methods applied to vegetation maps, in which different carbon assimilation values are assigned to different vegetation types, linked with weather maps supplying temperature and rainfall patterns. Many specialized tools are available within standard GIS systems, for example, to follow the movements of individual organisms with remote-sensing tools. Temporal GIS methods provide a means to analyze dynamic models across landscapes, and there are 3-D extensions of GIS.

Applications

Habitat Suitability Indices

Habitat Suitability Evaluation Procedures (HEP) are a formalized methodology for impact assessment on wildlife habitat. These are based on Habitat Suitability Index (HSI) models (**Verner *et al.*, 1986**), which attempt to summarize the site characteristics that affect the utilization of particular habitats by a variety of wildlife species. Numerous HSI models have been constructed, typically consisting of very simple regression-type models. The key habitat variables are often some measure of canopy cover in a variety of classes, diameter classes of trees and shrubs, tree stem densities, area of open water, and distance to forest cover, among others. The objective is to combine these variables, based on extensive field observations done in a correlative manner, to provide overall indices of suitability. HSIs are always indexes with values between zero and one, and they are assumed to be proportional to carrying capacity.

HSIs are based only on local habitat variables; they completely ignore any effects due to species interactions, except those due to indirect effects on related habitat variables. The models ignore the spatial interactions of habitat types across a landscape. This leads to difficulty in situations for which the size, shapes, edge effects, and neighborhood relationships have a greater effect on habitat preference than local forest composition and structure variables. The models also do not take account of the issue of presence/absence of a species, and thus ignore any historical influence on potential local abundance. HSIs are inherently static entities, so any dynamics they produce are driven completely by changes in habitat variables and not by the inherent dynamics and demography in the species being considered. Despite these criticisms, HSIs are among the most commonly used set of ecological models, in part because users realize their limitations and view them as a simplified tool to summarize a very complicated situation by a single number. Such simplification must result in a loss of information, but a key issue regarding HSIs is whether they indeed can be used as a useful predictor for abundance.

Metapopulation Models

HEP is an attempt to account for the spatial nature of biodiversity by including explicit maps of basic habitat variables. An intermediate approach between models that include all the spatial detail available from maps and those that ignore all spatial aspects of a system is metapopulation modeling. These models consider a landscape to be split up into a number of localized populations, called subpopulations, with the entire collection of these called a metapopulation. Most of the biotic interactions driving population dynamics occur within the localized subpopulations, but there are exchanges of individuals between these subpopulations. This allows for differing environmental, demographic, genetic, and disease situations to be present in the subpopulations. Depending on the assumptions about transport of individuals between the subpopulations, these can be relatively isolated or closely coupled.

A clear advantage of metapopulation models is the ability to derive analytical results, such as equilibrium and stability behavior, as a function of the within-subpopulation characteristics and the between-subpopulation factors such as movement. The models are particularly appropriate for cases in which a landscape can be reasonably viewed as containing discrete patches of habitat suitable for the population, with the intervening regions not suitable. The level of detail in these models can be quite variable, with the simplest versions just treating subpopulations as either present or absent. More complicated models take account of demographics within each subpopulation or explicit details on the relative spatial locations of each subpopulation that affect movement between them. These can be used for a population viability analysis, in which the probability that the overall population will survive for varying time periods is estimated.

Individual-Based Approaches

Classical ecological modeling approaches include various levels of aggregation in order to simplify the model. An alternative reductionist approach is to take account of differences between individuals within a population, allow the individuals to feed, grow, and interact, and from the aggregated behavior of these individuals build an understanding of population-level responses (**Grimm and Railsback, 2005**). These individual-based approaches (also referred to as agent-based) are increasingly common; thanks to computational capacity, the availability of modeling tools and associated guidance for constructing and describing these models and the increasing availability of data on behavior and physiology of species of particular interest. The advantages of these approaches include the ability to consider the effect of abiotic factors on populations through their direct impact on individual behavior and growth, to take account explicitly of spatial variation in habitat factors, and to deal with small populations in which individual differences within the population can have great impacts on population-level responses. Disadvantages of individual-based approaches include the requirement for a great amount of detailed data to realistically simulate individual behaviors, and the typical necessity of making numerous simulations to evaluate any particular response that may arise because of the stochastic nature of the models. The methodology to describe and interpret these models is now becoming standardized and there are readily accessible general tools for constructing these models, NetLogo being one of the more accessible ones.

Multimodeling and Regional Assessment

Natural systems have many interacting components operating at a variety of temporal and spatial extents and requiring differing levels of detail to describe the interactions between them. One historical approach in ecology to model such a system is to break it into a number of compartments (often for different trophic levels, and sometimes grouping within each trophic level) and consider the dynamics of each compartment with movements of energy, biomass, or nutrients among them. It is quite difficult to make these systems analysis approaches spatially explicit or to link them to GIS. Yet it is now becoming possible to link together a variety of different modeling approaches in order to best utilize the available data, with different resolutions at different trophic levels, and carry this out in a spatially explicit manner. This is part of the general area of hybrid modeling. Such multimodels may use very simple models similar to HSIs for some trophic components, more complex dynamical systems for certain populations with mostly very localized interactions, and individual-based models for organisms that move great distances and average over the spatial heterogeneity. One example of such an approach is the Across Trophic Level System Simulation (ATLSS) Project, a multimodel used to estimate the biotic impacts of alternative water management plans on the Everglades of south Florida (**DeAngelis *et al*., 1998**).

Building multimodels requires extensive landscape data obtainable from remote-sensing and ground efforts. The approach is inherently dynamic, and thus requires methods to estimate the spatial dynamics of key environmental drivers, or else have available a history of this spatially that can be analyzed statistically. In the Everglades, the major driver is water, and both historical data and detailed hydrologic models are available to provide estimates for scenario evaluations. Without such data, assumptions must be made about the dynamics of the landscape. Continued enhancement of remote-sensing methods and sensor data on individual organisms is providing extensive time series of remotely sensed data to both calibrate multimodels and provide the opportunity to iteratively improve their predictive abilities. For problems at regional spatial extents, such multimodels are a rational method to aid planning while taking account of the best scientific data at the variety of resolutions available.

Behavioral Dynamics of Species of Special Concern

Great strides are being made in improving our understanding of the conservation biology of rare, threatened, and endangered species throughout the world. In an effort to better estimate the responses of populations of these species, numerous remote-sensing methods have been employed to track the movements of individual organisms. These include tracking devices implanted within or fixed on sampled individuals, which allow explicit location and physiological data (e.g., body temperature) to be obtained regularly and remotely throughout the individualʹs life. When done for many individuals within a population, it is becoming possible to follow the behavioral dynamics of mixtures of individuals. This includes the ability to ascertain details of mating, territoriality, and aggressive interactions. It is likely in the future that managers concerned with a particular species will be able to observe in real time the movements of many individuals within a population. Then they might apply spatially explicit modeling methods to project the response of the population to particular management alternatives and compare these with optimization models that project the “best” that is possible to do according to some criteria.

Conclusions

Biodiversity science benefits from multiple types of models with a variety of objectives including description, understanding mechanism, projection, forecasting, and control. Quantitative models can be formulated using mathematical, computational, statistical, and machine learning methods. No one model can do everything because of tradeoffs in generality, precision, and realism. Applications of models in biodiversity include habitat suitability, metapopulations, individual-based approaches, multimodeling and behavioral dynamics.

Acknowledgments

LJG appreciates the long-term support of the US Geological Survey through various awards for the ATLSS project, and the support of the National Institute for Mathematical and Biological Synthesis, supported by the National Science Foundation through Award #DBI-1300426, with additional support from The University of Tennessee, Knoxville. Brian Beckage was supported in part by NASA Grant Number 80NSSC20M0122 and by the USDA National Institute of Food and Agriculture Hatch, Project Number 1025208. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of any agency providing funding support.

References

Beckage, B., Gross, L.J., and Kauffman, S. (2011) The limits to prediction in ecological systems. *Ecosphere* 2 (11): 125. https://doi.org/10.1890/ES11-00211.1.
Beckage, B., Gross, L.J., and Lacasse, K., et al. (2018) Linking models of human behavior and climate alters projected climate change. *Nature Climate Change* 8: 79–84. https://doi.org/10.1038/s41558-017-0031-7.
Buckley, L.B., Urban, M.C., and Angilletta, M.J., et al. (2010) Can mechanism inform species distribution models? *Ecology Letters* 13: 1041–1054.
Carnaval, A.C., Hickerson, M.J., Haddad, C.F.B., Rodrigues, M.T., and Moritz, C. (2009) Stability predicts genetic diversity in the Brazilian Atlantic forest hotspot. *Science* 323: 785–789.
Caswell, H. (2001) Matrix population models. second ed. Sinauer Associates, Sunderland, MA.
Christin, S., Hervet, E., and Lecomte, N. (2019) Applications for deep learning in ecology. *Methods in Ecology and Evolution* 10: 1632–1644.
Clark, J.S., Gelfand, A.E. (eds.) (2006) *Hierarchical modelling for the environmental sciences*. Oxford University Press, Oxford.

DeAngelis, D.A., Gross, L.J., and Huston, M.A., et al. (1998) Landscape modeling for Everglades ecosystem restoration. *Ecosystems* 1: 64–75.

Dietz, M. (2017) Ecological forecasting. Princeton University Press, Princeton, NJ.

Fuller, M.M., Gross, L.J., Duke-Sylvester, S.M., and Palmer, M. (2008) Testing the robustness of management decisions to uncertainty: Everglades restoration scenarios. *Ecological Applications* 18: 711–723.

Grimm, V. and Railsback, S.F. (2005) Individual-based modeling and ecology. Princeton University Press, Princeton, NJ.

Haefner, J.W. (2005) Modeling biological systems: Principles and applications. second ed. Springer, Berlin.

Hilborn, R. and Mangel, M. (1997) The ecological detective: Confronting models with data. Princeton University Press, Princeton, NJ.

Hof, J. and Bevers, M. (2002) Spatial optimization in ecological applications. Columbia University Press, New York, NY.

Huston, M.A. (1994) Biological diversity. Cambridge University Press, Cambridge, UK.

Jacob, F. (1982) The possible and the actual. Pantheon Books, New York, NY.

Le Bouille, D., Fargione, J., and Armsworth, P.R. (2022) Spatio-temporal variation in the costs of managing protected areas. *Conservation Science and Practice* 4: e12697.

Levins, R. (1968) Evolution in changing environments. Princeton University Press, Princeton, NJ.

Murray, J.D. (2007) Mathematical biology: I. An introduction. Springer-Verlag, Berlin.

Pelletier, T.A., Carstens, B.C., Tank, D.C., Sullivan, J., and Espandola, A. (2018) Predicting plant conservation priorities on a global scale. *Proceedings of the National Academy of Sciences of the United States of America* 115: 13027–13032.

Reichstein, M., Camps-Valls, G., and Stevens, B., et al. (2019) Deep learning and process understanding for data-driven Earth system science. *Nature* 566: 195–204.

Verner, J., Morrison, M.L., Ralph, C.J. (eds.) (1986) *Wildlife 2000: Modeling habitat relationships of terrestrial vertebrates*. University of Wisconsin Press, Madison, WI.

Whitlock, M.C. and Schluter, D. (2008) The analysis of biological data. Roberts and Company Publishers, Greenwood Village, CO.

Wolfram, S. (2002) A new kind of science. Wolfram Media, Champaign, IL.

Relevant Websites

–CyVerse https://www.cyverse.org
–Data Portal | NSF NEON www.neonscience.org
–Netlogo https://www.ccl.northwestern.edu/netlogo/
–NIMBioS: National Institute for Mathematical and Biological https://www.NIMBioS.org