Ecology, 100(6), 2019, e02713

Innovations in data integration for modeling populations

Elise F. Zipkin,[1,4] Brian D. Inouye[2], and Steven R. Beissinger[3]

Manuscript received 21 November 2018; accepted 2 January 2019. Corresponding Editor: Donald R. Strong.
[1]Department of Integrative Biology and Ecology, Evolutionary Biology, and Behavior Program, Michigan State University, East Lansing, Michigan 48824 USA.
[2]Biological Science, Florida State University, Tallahassee, Florida 32306 USA.
[3]Department of Environmental Science, Policy & Management, and Museum of Vertebrate Zoology, University of California, Berkeley, California 94720 USA.
[4] E-mail: ezipkin@msu.edu

Citation: Zipkin, E. F., B. D. Inouye, and S. R. Beissinger. 2019. Innovations in data integration for modeling populations. Ecology 100(6):e02713. 10.1002/ecy.2713

Key words: coupled human and natural systems; integral projection models; integrated models; integrated population models; population dynamics; species distribution models.

Assessing the status of species in the Anthropocene requires an understanding of basic ecological processes affecting population dynamics and the impacts of ongoing climate and environmental changes. Various combinations of forces and their feedbacks drive species distributions and abundances across large spatial and temporal scales, which can be a challenge to analyze with traditional ecological tools. Evaluating the relative effects of these forces on population dynamics is often difficult because data are insufficient and/or comprised of disparate types. A fundamental challenge of population ecology is to detect species trends, extrapolate inference across spatiotemporal scales, and create credible projections of dynamics and viability. Modeling approaches that incorporate multiple, interacting processes using dissimilar data types are thus needed, and indeed, the development of such models is a rapidly growing area of research within ecology.

Editors' Note: Papers in this Special Feature are linked online in a virtual table of contents at: www.wiley.com/go/ecologyjournal.

A meaning of "integrated" is to bring separate or disparate things together into a cohesive whole and this is the definition we use for the Special Feature. Our Special Feature contains papers illustrating ways to use multiple data types and modeling approaches, integrating them for the common goal of building robust methods to assess and project population distribution and abundance. For example, citizen science represents a relatively new and growing source of data for ecologists, but often data from these projects are not the same type of data that are collected from focused experiments or monitoring by experts. Integrating two or more data types, collected with multiple techniques and/or on different aspects of a study system, can yield better information about an ecological process of interest than use of a single data type. Integration of multiple data types can also facilitate understanding of the mechanisms that influence population dynamics and trends. The primary advantages of integrated models are (1) the ability to compensate for variability in data collection by reducing biases inherent in a single data set; (2) improved precision of estimates of demographic rates compared to those obtained from separate analyses; and (3) production of models suited to accommodate a broad suite of environmental processes through spatial methods and human interactions.

Two recent advances for combining disparate types of information offer opportunities to improve the ability to quantify and project population dynamics. First, *integrated distribution/population models* provide a powerful approach for addressing issues of sparse data by integrating multiple data sources into a single, unified, dynamic model of a target population. The second advance broadens the scope of study to incorporate humans directly into the study system from multiple perspectives. This approach has been characterized as *coupled human and natural systems* or socioecological systems research. When incorporated directly into models, it provides an opportunity to integrate multiple processes, as well as multiple data types, across scales to provide a holistic understanding of the factors driving population dynamics.

The goal of this Special Feature is to highlight emerging methods and challenges for integrated modeling approaches, particularly for studies that aim to quantify and project demographic rates, distributions, and dynamics of populations. Developments and applications of integrated models have led to critical advances in studies of population dynamics and greater understanding of the drivers of population change, but there is still much to be done. Individual papers in the Special Feature illustrate state-of-the-art methods, discuss important areas for near-term reach, and will hopefully

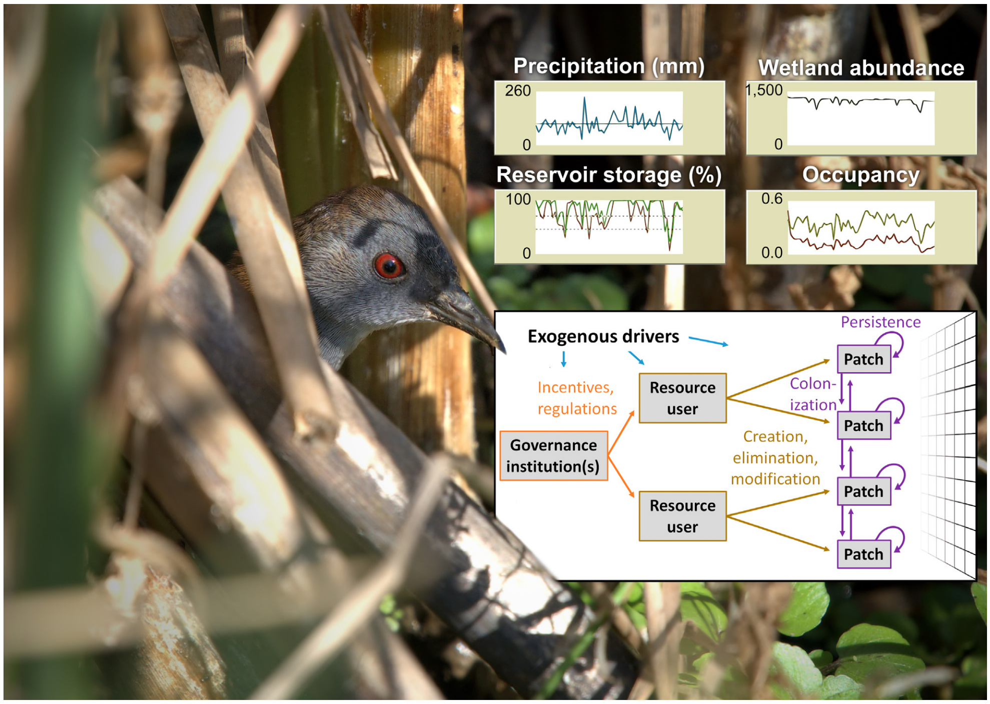

Fig. 1. Van Schmidt et al. (2019) combine five interdisciplinary data sets to model the metapopulation dynamics of the threatened Black Rail (*Laterallus jamaicensis*) in a coupled human and natural system, the irrigated rangelands of the foothills of California's Sierra Nevada. Their paper provides an example of integrating social and ecological data by coupling agent-based models of land-use change with a stochastic patch occupancy model of metapopulation dynamics. The inset plots highlight data sources and a conceptual model diagram from a single iteration of the resulting simulation. The coupled human and natural systems model illustrates that precipitation was an overwhelmingly strong driver of metapopulation dynamics, far more than would be apparent from models excluding linkages between precipitation and human-driven changes in irrigation behavior. See Van Schmidt et al. (2019) for details.

serve to stimulate further expansions of methods and applications to new questions in population and conservation ecology.

The papers by Saunders et al. (2019) and Plard et al. (2019) focus on *integrated population models* (one kind of IPMs), a specific subset of integrated models in which the joint likelihoods of multiple data sets (typically including capture–recapture and population count data) are used to estimate life history demographic rates (e.g., survival, fecundity, and immigration). Employing this approach, IPMs can estimate spatiotemporal population abundance using a mechanistic framework and identify the relative contribution of life stages or environmental factors (via covariates) to population trends. While IPMs are typically used because of deficiencies in various data sets, Saunders et al. (2019) highlight how IPMs can also be used to resolve discrepancies in inferences from individual analysis of independent data sets. Plard et al. (2019) discuss how individual-level processes can be incorporated into IPMs to examine the effects of density-dependence on population dynamics while also incorporating continuous trait data, unifying IPMs with integral projection models (the other IPM, integrating demographic functions over continuous individual traits).

Fletcher et al. (2019) and Pacifici et al. (2019) discuss opportunities and challenges for integrating multiple types of data (typically count, presence–absence, and/or presence-only) to estimate species distributions. They highlight a relatively new class of models, *integrated distribution models* (IDMs), which estimate the occurrence or abundance of species across spatial extents. Fletcher et al. (2019) provide a guide for combining opportunistic and planned survey data into a single modeling framework and illustrate how different quantities of available data can bias model results. Pacifici et al. (2019) show how statistical techniques can be used to resolve issues of spatial-temporal mismatch of available data, i.e., the "change of support" problem.

Van Schmidt et al. (2019) tackle the integration of social and ecological data via an agent-based modeling approach to understand metapopulation dynamics in coupled human and natural systems (Fig. 1). They use a multidisciplinary framework to identify and model the key actors, processes, and feedbacks that drive landscape change and, in turn, metapopulation dynamics. Their paper highlights the importance of considering the heterogeneity of human motivations and resultant activities that affect ecological processes, especially when considering appropriate management actions.

The development of integrated models is well underway, and such approaches have already been applied successfully to a variety of systems, taxa, and data types to examine ecological and conservation questions. As a result, it is clear that integrated models are an emerging tool for the study of population dynamics and trends, allowing researchers to rigorously test theories by analyzing multiple data types with a unified framework. The set of papers in this Special Feature provide an overview of integrated modeling approaches and their application to population dynamics questions, including assessing interacting environmental stressors, incorporating multiple sources of uncertainty, evaluating potential conservation actions, modeling metapopulation dynamics in coupled human–natural systems, and combining citizen science and planned survey data. The approaches to data integration presented here highlight important advances for research programs addressing questions across spatial and temporal scales and help to provide a holistic understanding of the factors driving population dynamics.

Ecologists increasingly see the benefits of linking previously disparate approaches to quantifying and modeling population dynamics into more powerful, unified tools. As these tools continue to be developed for utilizing multiple data types, coupling human and natural systems, and tracking individual trait distributions as well as total population abundance, the uses of “integrated,” “integrative,” and “integral” will also continue to grow. Over the coming decade, we anticipate that both modeling and software advances will make the application of integrated models a standard tool in population ecology.

Literature Cited

Fletcher Jr., R. J., T. Hefley, E. P. Robertson, B. Zuckerberg, R. A. McCleery, and R. M. Dorazio. 2019. A practical guide for combining data to model species distributions. Ecology 100: e02710.

Pacifici, K., B. Reich, D. Miller, and B. Pease. 2019. Resolving misaligned spatial data with integrated distribution. Ecology 100:e02709.

Plard, F., R. Fay, M. Kéry, A. Cohas, and M. Schaub. 2019. Integrated population models: powerful methods to embed individual processes in population dynamics models. Ecology 100:e02715.

Saunders, S. P., M. T. Farr, A. D. Wright, C. A. Bahlai, J. W. Ribeiro Jr., S. Rossman, A. L. Sussman, T. W. Arnold, and E. F. Zipkin. 2019. Disentangling data discrepancies with integrated population models. Ecology 100:e02714.

Van Schmidt, N. D., T. Kovach, A. M. Kilpatrick, J. L. Oviedo, L. Huntsinger, T. V. Hruska, N. L. Miller, and S. R. Beissinger. 2019. Integrating social and ecological data to model metapopulations in coupled human and natural systems. Ecology 100:e02711.