

Integrated community models: A framework combining multispecies data sources to estimate the status, trends and dynamics of biodiversity

Elise F. Zipkin¹  | Jeffrey W. Doser¹  | Courtney L. Davis^{1,2}  | Wendy Leuenberger¹ |
 Samuel Ayebare¹ | Kayla L. Davis¹ 

¹Department of Integrative Biology; Ecology, Evolutionary Biology, and Behavior Program, Michigan State University, East Lansing, Michigan, USA

²Cornell Lab of Ornithology, Cornell University, Ithaca, New York, USA

Correspondence

Elise F. Zipkin
 Email: ezipkin@msu.edu

Funding information

National Science Foundation, Grant/Award Number: 1954406

Handling Editor: Thierry Boulinier

Abstract

1. Data deficiencies among rare or cryptic species preclude assessment of community-level processes using many existing approaches, limiting our understanding of the trends and stressors for large numbers of species. Yet evaluating the dynamics of whole communities, not just common or charismatic species, is critical to understanding and the responses of biodiversity to ongoing environmental pressures.
2. A recent surge in both public science and government-funded data collection efforts has led to a wealth of biodiversity data. However, these data collection programmes use a wide range of sampling protocols (from unstructured, opportunistic observations of wildlife to well-structured, design-based programmes) and record information at a variety of spatiotemporal scales. As a result, available biodiversity data vary substantially in quantity and information content, which must be carefully reconciled for meaningful ecological analysis.
3. Hierarchical modelling, including single-species integrated models and hierarchical community models, has improved our ability to assess and predict biodiversity trends and processes. Here, we highlight the emerging 'integrated community modelling' framework that combines both data integration and community modelling to improve inferences on species- and community-level dynamics.
4. We illustrate the framework with a series of worked examples. Our three case studies demonstrate how integrated community models can be used to extend the geographic scope when evaluating species distributions and community-level richness patterns; discern population and community trends over time; and estimate demographic rates and population growth for communities of sympatric species. We implemented these worked examples using multiple software methods through the R platform via packages with formula-based interfaces and through development of custom code in JAGS, NIMBLE and Stan.

5. Integrated community models provide an exciting approach to model biological and observational processes for multiple species using multiple data types and sources simultaneously, thus accounting for uncertainty and sampling error within a unified framework. By leveraging the combined benefits of both data integration and community modelling, integrated community models can produce valuable information about both common and rare species as well as community-level dynamics, allowing for holistic evaluation of the effects of global change on biodiversity.

KEY WORDS

community analysis, data integration, hierarchical models, population dynamics

1 | INTRODUCTION

The consequences of global change on animal communities remain unclear for all but the most abundant taxa because of data limitations (Breiner et al., 2015; Kindvater et al., 2018). As biodiversity continues to decline, it is critical to assess the status and dynamics of whole communities, and not just those common or charismatic species with large amounts of data. While many sampling schemes target multiple species simultaneously (e.g. breeding birds via point counts, small mammals via trapping), traditional approaches to evaluate community-level processes require a large number of observations (Hamilton et al., 2015; Sor et al., 2017), precluding assessments of rare or data-deficient species. One in six species has been classified as data deficient by the International Union for the Conservation of Nature (Bland et al., 2017), leading to gaps in our understanding of the trends and stressors for a large proportion of species, often within taxonomically related communities. Furthermore, data sets collected by individual researchers are generally restricted in spatiotemporal scope, resulting in narrow and inadequate assessments of the biotic and abiotic factors influencing species' trends when the effects of environmental factors vary across space and/or time (Rollinson et al., 2021).

Fortunately, there has been a surge of data collection programmes in recent decades, including hundreds of public (citizen) science (e.g. eBird, iNaturalist) and government-funded (e.g. National Ecological Observation Network [NEON]) programmes that provide a wealth of fine- to broad-scale data on multiple species simultaneously (Barnett et al., 2019; Thornhill et al., 2016). Incorporating these distinct data sources into rigorous analyses is an ongoing challenge because of variations in data type, quality, sampling protocols and geographic coverage (Chandler et al., 2017; Moussy et al., 2021). Yet, understanding and evaluating both species- and community-level processes, including responses to environmental change, is critical to maintaining biodiversity, and hence, a key objective of the ecological research community (Johnson et al., 2017).

Hierarchical modelling has significantly advanced the use of such ecological data because it allows for the separation of biological and observation processes (Kéry & Royle, 2016) and can thus account for variations in the types and information content of specific data

sources. Two recent advancements in hierarchical modelling are critical to the development of rapid, comprehensive assessments of biodiversity: (1) single-species integrated models and (2) hierarchical community models.

Single-species integrated models combine multiple data sets into a single model on a target species, generally by defining a joint likelihood among available data sources (Miller et al., 2019; Schaub & Kéry, 2021). A key advantage of integrated modelling (sometimes referred to as data fusion) is the ability to merge multiple data types, regardless of collection method and spatiotemporal scope (Zipkin et al., 2021). Single-species integrated models improve inferences through increased precision of parameter estimates (Mosnier et al., 2015), the estimation of parameters for which no explicit data are available (Oppel et al., 2014), and by accounting for uncertainties and correlations among data sets (Lee et al., 2015). Single-species integrated modelling has primarily taken two forms: integrated population models and integrated distribution models (Zipkin et al., 2019). Integrated population models allow for a mechanistic understanding of species-level processes primarily by combining demographic data with time-series population- or site-level count data (Brown & Collropy, 2013; Saunders et al., 2018). Integrated distribution models synthesize presence-only, detection–nondetection and/or count data to estimate species distribution patterns and the effects of covariates on occurrence or abundance (Fletcher Jr. et al., 2019; Gratarola et al., 2023).

Hierarchical community models link together occurrence or abundance parameters of individual species (estimated using a single, multispecies detection–nondetection or count data set), through community-level distributions (Devarajan et al., 2020; Dorazio et al., 2006; Sollmann et al., 2016). This is done either by treating species-level parameters (i.e. intercepts or covariate effects) as random variables arising from normal distributions, characterized by a community-level mean and variance across species (e.g. multispecies occurrence models; Zipkin et al., 2010; Guillera-Arroita, 2017), or by using multivariate distributions to explicitly model species associations (e.g. joint species distribution models; Ovaskainen et al., 2017; Warton et al., 2015). Hierarchical community models improve biological inferences by accounting for both species-level effects and the

aggregated effects of covariates on a community as a whole, leading to increased precision in species parameter estimates, even for those species that were observed infrequently (Zipkin et al., 2009), and the ability to estimate biodiversity metrics such as richness and composition (Guillera-Arroita et al., 2019).

Despite recent advances in single-species integrated models and hierarchical community models, there has been comparatively less research focused on modelling multiple species using more than one data source. Several applications of multispecies integrated population models in which count data are combined with demographic data have examined interactions between species, including competition (Perón & Koons, 2012), synchrony in survival and reproduction rates (Lahoz-Monfort et al., 2017) and predator–prey dynamics (Barraquand & Giménez, 2019; Clark, 2021; Paquet & Barraquand, 2022; Quéróu et al., 2021). A key feature of these models is that species' demographic rates are assumed to be influenced by the dynamics of one or two other species such that the models account for both abiotic effects (via covariates) and biotic interactions among species. There are also examples of integrated community occupancy models that combine multiple detection–nondetection data sets to estimate species co-occurrence patterns using two or more data sets (Doser, Leuenberger, et al., 2022; Lauret et al., 2023). Clark et al. (2017) developed a generalized joint attribute model to estimate the distribution and abundance of multiple species using combinations of detection–nondetection and various types of count data. In such

models, species abundances are assumed to come from a community level, multivariate normal distribution, thereby accounting for associations among related species within the predefined community. Recent efforts integrating multispecies data sets only scratch the surface of what can be achieved by uniting available data sources on multiple species and whole communities (Rapacciulo & Blois, 2019), but highlight the growing relevance of, and need for, a generalizable modelling framework.

In this paper, we highlight the class of models broadly categorized as integrated community models. Although the term integrated community modelling has been used with several meanings, here we define it as a framework that links multiple data sources on multiple sympatric species through a hierarchical community component. We outline the general approach of the method, demonstrate three specific applications with different biological and observation model structures using empirical data and simulations and offer a discussion about various factors to consider before using integrated community models. We conclude by providing suggestions for future developments.

2 | EXPLANATION OF THE METHOD

We consider integrated community models broadly as the combination of single-species integrated models and hierarchical community models (Figure 1). Each of the different data sources (e.g.

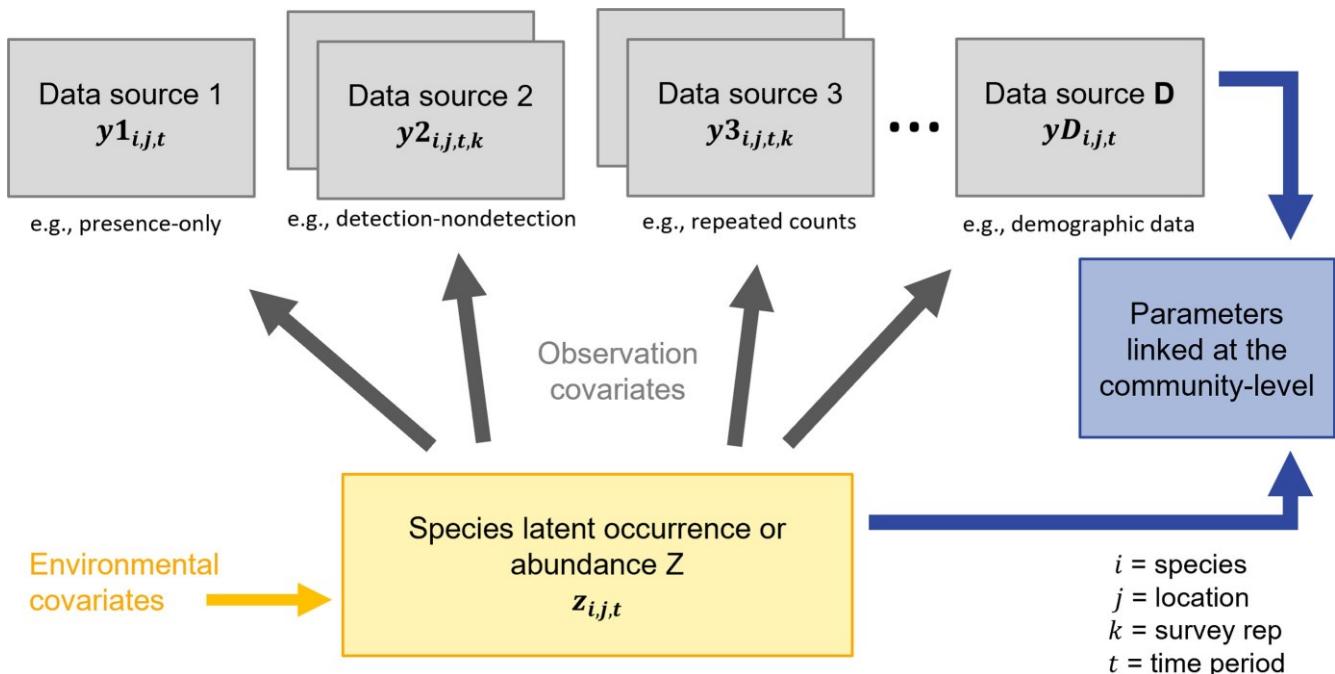


FIGURE 1 Integrated community models enable the estimation of the status, trends and/or population dynamics of multiple species simultaneously by combining disparate data sources within a unified analysis framework. The schematic diagram outlines the hierarchical approach of the integrated community modelling process: the underlying species-level biological process (yellow; e.g. abundance, $z_{i,j,t}$) is modelled using relevant environmental covariates and/or demographic rates (not shown). Each data source ($y_1, y_2, y_3, \dots, y_D$) provides a piece of information on the biological process and is connected to the biological parameters via an observation process model (grey). Species-specific parameters are linked through community-level distributions to share information across species.

detection–nondetection, count, demographic) can be used to inform various components of the underlying biological process model through hierarchical, observation models linked together with a joint likelihood or through simpler approaches that account for variation across data sources through covariates and/or random effects (Wikle & Berliner, 2007; Zipkin et al., 2021). Covariates can be included using link functions to incorporate relevant biotic and abiotic factors that influence species dynamics and detection across space and time. The biological process models for species (Figure 1, yellow shading) can range from simple (e.g. estimates of species occurrence; Miller et al., 2019) to complex (e.g. estimates of species survival, reproduction and abundance; Schaub & Abadi, 2011), and depend on both the life history and ecology of the taxonomic group and the quantity and type of available data.

Parameters in the biological process model are linked to the survey data through observation process models (Figure 1, grey shading). As with the biological process, the observation models can range from simple (e.g. fixed and/or random effects on individual surveys or data sources) to increasingly mechanistic (e.g. modelling detection probabilities based on covariates and recorded auxiliary sampling information), primarily depending on the type and quantity of available data (Dorazio, 2014; Kéry & Royle, 2016). Commonly collected wildlife data types span a spectrum of collection effort and information content including (from low to high): presence-only data (observations of a species in a given location at a given time), detection–nondetection data (records that also contain locations that were surveyed but the species was not observed), count data (total number of individuals observed at a location within a specific time frame) and demographic data (e.g. capture–recapture, productivity data; to inform survival, reproduction, immigration etc.). Typically, the higher the information content, the more difficult it is to collect the data and likely that the geographic coverage is limited. In addition to survey type, we can classify presence-only, detection–nondetection and count data as structured, semi-structured or unstructured. Structured data are collected systematically with a design protocol (e.g. NEON), while unstructured data are collected opportunistically and without a defined purpose (e.g. iNaturalist). Semi-structured data are typically collected in volunteer-based monitoring programmes (e.g. eBird) that have some degree of structure (e.g. collection of auxiliary sampling information, checklists) that helps mitigate observational and sampling biases present in unstructured data, but lack the rigorous sampling design of structured data (Altweg & Nichols, 2019; Kelling et al., 2019).

The various available data sources provide unique or complementary information (in terms of spatiotemporal location, life stage and/or demography) on the biological processes of interest (Saunders, Farr, et al., 2019). For example, available data sources may consist of multiple different data types (e.g. one count and one demographic data set) or may all be the same type (e.g. two detection–nondetection data sets) but collected under distinct protocols with different types of observation and/or sampling biases. Thus, the value of combining available data sources can either be inference on mechanistic biological processes (as is done in integrated population

models; Schaub & Kéry, 2021) or expansion of the spatiotemporal scope of inference (as is done in integrated distribution models; Isaac et al., 2020). When modelled together, the information from each data source can be used to jointly estimate biological parameters, improving accuracy on inferences (Pacifici et al., 2017), and in many cases, allowing estimation of a greater number of parameters than is possible through independent analyses of the various data sets (Plard et al., 2019; Schaub et al., 2007).

The species-level biological and observation process models are joined by allowing parameters to come from common, community-level distributions (Figure 1, blue shading), which can be univariate (e.g. Farr et al., 2019) or multivariate (e.g. Thorson et al., 2016), depending on whether estimation of covariances among species is desirable. This facilitates information sharing among species in the community, allowing parameter estimation for all species and not just those with large sample sizes (Zipkin et al., 2009). Additionally, this approach produces estimates of community-level parameters (mean across all species and variance among species), which can be used to summarize community responses to relevant covariates and environmental stressors (Threlfall et al., 2017).

In theory, integrated community models could be analysed with either a frequentist or Bayesian approach. In most cases, Bayesian analysis will be more practical to simultaneously estimate the biological and observation parameters at species and community levels. Bayesian inference allows for maximum flexibility in terms of hierarchical model structure and convergence (Fordyce et al., 2011) and enables straightforward calculation of derived biodiversity metrics (e.g. site-level richness, evenness, composition and turnover; Dorazio, 2016; Gelman & Hill, 2007). In our case studies, parameters are all estimated using Bayesian analysis with Beyer several different approaches in the R software platform (R Core Team, 2022), including built-in packages and custom code that implements Markov chain Monte Carlo (MCMC) algorithms via JAGS (Plummer, 2003), NIMBLE (de Valpine et al., 2017) and Stan (Stan Development Team, 2022).

3 | WORKED EXAMPLES

Here, we show three case studies to provide practical examples of the integrated community modelling framework that (1) extend the geographic scope of inference to evaluate species distributions and effects of habitat for a specialized bird community, (2) discern the population trends of open-habitat-associated butterfly species over a recent decade and (3) estimate demographic rates and population growth for communities of sympatric species. Our worked examples demonstrate approaches to combine structured and semi-structured data including detection–nondetection, site- and population-level count and demographic data types. We show how the integrated community modelling framework is capable of producing inferences for both individual species (rare and common) and community-level metrics of interest. We present a general overview and the basic methods of each case study in the main text with details including mathematical equations and the

specifics of implementation described in associated supplemental materials. The complete code for the worked examples is available at <https://zipkinlab.github.io/#icm2023Z> and is also archived at Zipkin et al. (2023).

3.1 | Case study 1: Spatial distributions of forest birds across the Northeastern United States

3.1.1 | Background and motivation

Communities of species with specialized habitat requirements, such as grassland or interior forest obligate birds, are particularly vulnerable to global change because of limited availability of habitat. Furthermore, species in specialized communities tend to be rare, resulting in few observations within large-scale monitoring programmes and low precision of species distribution estimates and effects of environmental drivers (Lomba et al., 2010). In this case study, we use an integrated community occupancy model (Doser, Leuenberger, et al., 2022) to estimate the spatial variation in species richness of a community of 27 interior forest obligate bird species, most of which are rarely detected, across 11 states in the Northeastern United States. The integrated community occupancy model combines multiple replicated and/or nonreplicated detection–nondetection data sources in a community occupancy modelling framework to provide inferences on species-specific and community-level occurrence patterns across space (and time). By sharing information across species and data sources, this approach can yield increased precision and accuracy of species-specific occurrence patterns compared to single data source models and single species models (Doser, Leuenberger, et al., 2022). Additionally, the community component enables estimation of aggregated biodiversity metrics, such as site-level species richness, that allows for local-scale assessment of available habitat use by the community.

3.1.2 | Data

We integrated detection–nondetection data on 27 interior forest obligate bird species from 356 routes in the North American Breeding Bird Survey (BBS; Pardieck et al., 2020) and 10,383 eBird checklists (Sullivan et al., 2009) across 11 states in the Northeastern United States, all collected during the breeding season of 2017 (Supplemental Information S1). Many species were detected infrequently, with 52% and 96% of species detected at less than 30% of the spatial locations in the BBS and eBird data sets respectively. For the BBS data, which is a structured road-side survey, observers recorded all birds seen within a 0.4-km radius at 50 locations (i.e. stops) along each ~39.2 km roadside survey (i.e. route). We assigned each route to a 5 × 5 km grid cell based on its midpoint location and aggregated the 50 stops per route into five spatial replicates of 10 stops each (i.e. we assume each route was sampled

five times, but the specific sampling locations occur at different spots along the route). We summarized the detection (1) or non-detection (0) of each species at each of the five spatial replicates. For the eBird data, which is an opportunistic, semi-structured data source, we assigned each complete checklist during a 3-week period in early June (to match the time frame of the BBS data) to a specific grid cell, and used multiple checklists within each cell as repeat surveys for the occupancy modelling framework. To mitigate preferential sampling biases, we used multiple spatiotemporal filtering criteria following standard recommendations for using eBird data in occupancy models (Johnston et al., 2021; Strimas-Mackey et al., 2020).

3.1.3 | Modelling approach

We modelled the data using an integrated community occupancy model that consisted of individual observation models (i.e. likelihoods) for the BBS and eBird data sets, which shared a species-specific biological process model. The biological process model described how species-specific occurrence varies across space as a function of elevation, forest cover and five bioclimatic variables, with individual species-specific effects treated as random variables arising from common community-level normal distributions (Dorazio & Royle, 2005). Both observation models explicitly accounted for imperfect detection by allowing detection probability to vary across species, space and replicate surveys. For the BBS observation component, we modelled detection probability as a function of day of the year. For the eBird observation component, we modelled eBird detection probability as a function of day of the year, time of day of collection, time spent, distance travelled and the number of observers. Similar to the biological process model, we assumed that each of the species-specific parameters in the observation models arise from community-level distributions. We estimated species richness of the community in the 5 × 5 km grid cells as a derived quantity across the Northeastern United States. We fit the model in a Bayesian framework using the *spOccupancy* R package (Doser, Finley, et al., 2022). We also fit the model with custom code by calling NIMBLE through R to provide potential users with additional implementation resources (code provided but results not shown). See Supplemental Information S1 for full model details.

3.1.4 | Results

Species richness of the interior forest bird community varied substantially across the northeast, with high richness across the Appalachian and Adirondack mountains and low richness in urban areas (Figure 2a). Spatial variation in richness and species-specific occurrence probabilities were largely driven by forest cover, with all species in the community showing a positive relationship to the amount of local-level forest cover (Figure 2b).

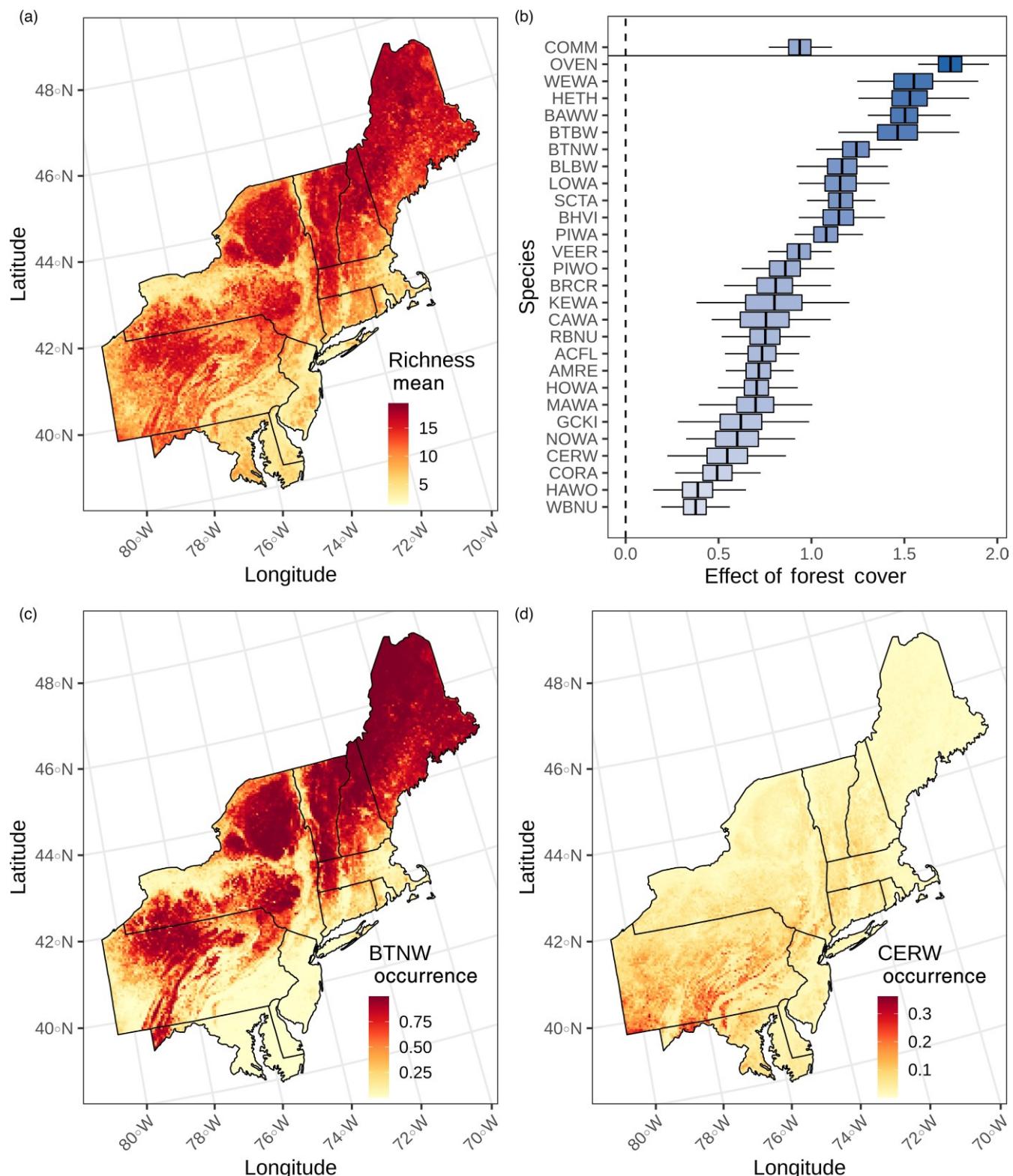


FIGURE 2 Estimates from an integrated community occupancy model for a community of 27 interior forest obligate bird species across the Northeastern United States using a 5×5 km grid. Panel (a) shows estimated mean species richness across the region while panel (b) shows the estimated mean (dark line), 50% credible interval (box) and 95% credible interval (whiskers) for the effect of forest cover on the overall community (COMM) and individual species (see Table S1.1 for species codes). Panel (c) shows estimated mean occurrence probabilities for Black-throated Green Warblers (*Setophaga virens*; BTNW), a common species of least concern, while panel (d) shows mean occurrence probabilities for Cerulean Warbler (*Setophaga cerulea*; CERW), a rare species that is classified as near threatened.

The integrated community occupancy model shares information across individual species and data sources, which allowed us to estimate reasonably precise occurrence probabilities and covariate effects throughout a broad spatial extent for common as well as rare species (e.g. Figure 2c,d). Given that the species most vulnerable to global change are often rarely observed in large-scale monitoring programmes, integrated community modelling frameworks that can leverage comparable data types (e.g. structured or semi-structured occurrence or count data) from multiple monitoring programmes and multiple species can provide a richer understanding of vulnerable communities and the specific species within them.

3.2 | Case study 2: Temporal trends of butterflies in the Midwestern United States

3.2.1 | Background and motivation

Insect communities face numerous threats, with a variety of anthropogenic stressors contributing to population and diversity declines (Forister et al., 2021; Hallmann et al., 2017; Wepprich et al., 2019). Rigorously quantifying insect population trends is notoriously difficult due to their complex life histories (e.g. seasonal variation in activity; Saunders, Ries, et al., 2019), rarity and elusiveness of species, biases in long-term data sets (e.g. natural history collections; Davis et al., 2023; Ries et al., 2019) and unbalanced sampling across space and time in volunteer monitoring programmes (Dennis et al., 2013). Integrating multiple data sources in a community modelling framework can mitigate many of these data complexities by increasing the amount of data available, sharing information across species and accommodating sampling biases within individual data sources. In this case study, we use an integrated community model to quantify relative abundance trends in 10 open-habitat-associated butterfly species in the Midwestern United States over a recent decade. Our model explicitly accounts for variation at multiple spatial (i.e. site, county) and observational levels (i.e. survey) while simultaneously accounting for variation in expected counts between different data sources due to differences in survey protocols and observer skills (Zylstra et al., 2021).

3.2.2 | Data

We integrated count data from five volunteer-based monitoring programmes to assess early-summer butterfly trends (June through July) from 2008 to 2017 across six Midwestern US states (Iowa, Wisconsin, Illinois, Indiana, Michigan, Ohio). We focused our analysis on 10 species that are year-round residents, active during summer, inhabit open areas, multivoltine, relatively easy to detect and adequately sampled by all monitoring programmes (Supplemental Information S2). Four data sets come from statewide, structured butterfly surveys: (1) Illinois Butterfly Monitoring Network (266

spatial locations); (2) Iowa Butterfly Survey Network (61 spatial locations); (3) Michigan Butterfly Network (133 spatial locations); and (4) Ohio Lepidopterists (118 spatial locations). In each programme, data are collected following a Pollard transect protocol in which trained volunteers walk ~1 km transects weekly or biweekly throughout the summer and record every butterfly detected (Pollard, 1977), although survey protocols differ slightly among states. While each of these four data sets has substantial, repeated temporal sampling across a given summer, they are restricted in spatial extent to only a single state. Our fifth data set comprises semi-structured count data from the North American Butterfly Association (NABA), in which volunteer observers extensively survey a 25-km diameter circle once a year, recording all butterflies observed, by species (<https://www.naba.org>). We used data from 85 sites that fell within our study region and survey time period. Alone, the NABA data may not provide reliable estimates of species temporal trends since only one count is performed at a given site in each year, which often does not adequately represent temporal variation in butterfly counts due to high variation in seasonal activity periods of species (Dennis et al., 2013; Zylstra et al., 2021). However, NABA data have a much larger spatial extent (i.e. North America) compared to the statewide butterfly surveys, and are thus able to provide critical information across the full study region.

3.2.3 | Modelling approach

Our integrated community model is based on a negative binomial hierarchical model, adapted from an analysis on monarch butterflies (Zylstra et al., 2021). We modelled the mean expected count for each butterfly species in a given week during a given year as a function of multiple covariates and random effects. To account for variation in survey effort across datasets, we included a fixed categorical variable of survey protocol (five levels) and a (log) linear effect of survey effort (i.e. the total number of surveys performed at a given spatial location in a given week). We further included a linear and quadratic effect of week to account for species-specific seasonal variation in activity, a linear effect of year to estimate any temporal trends in butterfly abundance, and random effects of county, site and year to account for additional variation. We allowed the species-specific effects of week to vary by year to account for differences in butterfly phenology across the decade. We treated each of the species-specific parameters (intercept and covariate effects) as random variables that come from common, parameter-specific, community-level normal distributions. We estimated a derived annual relative abundance index for each species as the expected number of individuals counted in a single survey (i.e. using the species-specific intercept, linear trend of year and year random effect) to create an average across sites and weeks. We fit the model in a Bayesian framework using the *spAbundance* R package (Doser, 2023). We also fit the model with custom code by calling Stan through R (code provided but results not shown). See Supplemental Information S2 for full model details.

3.2.4 | Results

The integrated community model revealed varying support for linear trends in relative abundance across butterfly species. The 95% credible intervals of all trend estimates overlapped zero, indicating uncertainty in population changes over this short time period (Figure 3), which is not entirely unexpected for butterfly species as insects tend to have high variations in abundance from year to year (Didham et al., 2020; Wagner et al., 2021). Seven of the 10 species had negative average trend estimates, with the most support for declines in Eastern Tiger Swallowtail (*Papilio glaucus*; 0.87), Cabbage White (*Pieris rapae*; probability negative trend = 0.83), Peck's Skipper (*Polites peckius*; 0.82) and Spring/summer Azure (*Celastrina ladon*; 0.80). There was mild support (69% probability) for a declining trend across the community (log-scale mean = -0.04) although with high uncertainty ($SD = 0.08$). In our case study, the integrated community modelling framework allowed us to take advantage of the within-season repeated sampling of the statewide structured monitoring data sets while simultaneously leveraging the large spatial extent of the NABA data to generate trend estimates of multiple butterfly species across the

Midwestern United States. As global climate and land-use change continue to pose threats to animal communities, modelling frameworks that leverage all available data sources on multiple species can provide critical insights on which species are most vulnerable and which may be less susceptible.

3.3 | Case study 3: Estimating species- and community-level demographic rates and population growth

3.3.1 | Background and motivation

Comprehensive monitoring and evaluation of biodiversity requires data not only on species distribution and abundance patterns but also on their demographic rates including survival and reproduction (Beyer & Manica, 2020). However, collecting multiple data types is resource-intensive for individual species, and orders of magnitudes harder for entire ecological communities. Thus, analytical frameworks that can combine very different types of data, that derive from a variety of sources and protocols,

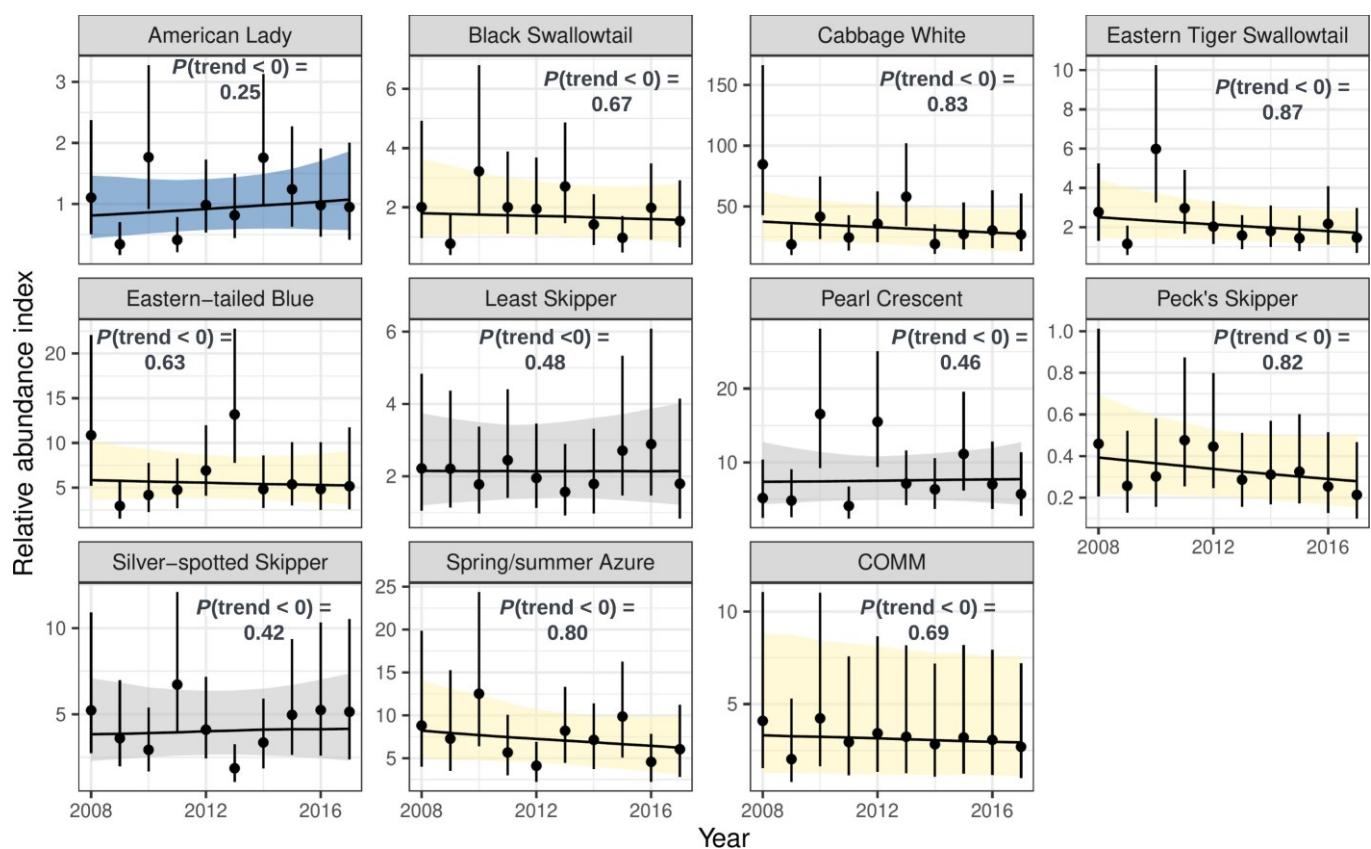


FIGURE 3 Relative abundance trends of 10 butterfly species and the community (COMM) in the Midwestern United States from 2008 to 2017 using an integrated community model. The probability that the linear trend estimate is less than zero is also included in each panel (i.e. that abundance is decreasing through time). Points (posterior medians with 95% credible interval lines) show the derived annual relative abundance index for each species. Black trend lines represent the posterior median trend estimate with the shaded area denoting the 95% credible interval. Species that indicate a possible decline over the 10-year study period are shaded in yellow (i.e. probability of a negative trend >0.6), while species that are likely to have increased (i.e. probability of negative trend <0.4) are shaded in grey. All others are shaded in grey.

are especially valuable for biodiversity assessments. In this case study, we develop a multispecies integrated population model that extends integrated population models (Besbeas et al., 2002; Schaub & Kéry, 2021) to a community level within a single analytical framework. Previous work has leveraged the single-species integrated population model to estimate interspecific interactions within a multispecies context (e.g. Quéróué et al., 2021), but only for a small number (i.e. two or three) of species. The approach we outline here is capable of making inferences on tens of species simultaneously by treating species-specific demographic rates as random effects from shared, community-level distributions (Iknayan et al., 2014). By combining multiple data types (e.g. population-level counts, productivity, capture–recapture data) on multiple sympatric species, this integrated community modelling approach has the potential to improve parameter identifiability for species and/or demographic rates for which only limited data are available (Zipkin & Saunders, 2018). We demonstrate the potential utility of the multispecies integrated population modelling framework using a data simulation approach.

3.3.2 | Data and approach

To construct the multispecies integrated population model, we first defined a biological process model that incorporates both demographic rates and population sizes for each of 10 hypothetical species that we assume are all part of the same community. We linked species' population sizes with their demographic rates using a female-based, age-structured matrix model (Caswell, 2000) with two age classes (juvenile and adult) and a pre-breeding census for each species (Figure 4a). We assume that species-specific fecundity, juvenile survival and adult survival are each derived from parameter-specific, normal distributions with a community-level mean and variance (Dorazio & Royle, 2005). We used the biological process model to simulate 100 independent, annual population-level counts (i.e. census), productivity (i.e. number of juveniles produced per adult) and capture–mark–recapture data sets for the hypothetical community of 10 species over a 10-year time period. While we simulated all three types of data (census, productivity and capture–recapture data) for all 10 species, species differed in the relative amount of data depending on their population size. For example, rare and declining species naturally had less capture–recapture and productivity data because there were fewer individuals available for sampling. We further assumed that all data sources were collected via design-based, structured sampling protocols, such that the data sources are representative of each species' population but may also contain sampling error. We then estimated parameters from a joint likelihood of the three independent data sets to make inferences on the species- and community-level demographic rates as well as derived parameters, including annual population sizes and growth rates (Kéry & Royle, 2016; Schaub & Kéry, 2021). We evaluated the performance of the model by calculating the relative bias $\frac{\text{posterior mean} - \text{true value}}{\text{true value}}$ of estimated parameters at both the species and community levels.

We fit our model by developing custom code in R and JAGS with the *jagsUI* package (Kellner, 2021). See *Supplemental Information S3* for full model details.

3.3.3 | Results

The multispecies integrated population model was able to recover the true biological parameters at both the species and community levels with little bias, and with especially high accuracy and precision of community-level demographic rates (Figure 4b). Mean population growth rates across the 10 years were also highly accurate for all species, as well as the community (Figure 4c). However, estimates of population growth rates tended to be less precise for declining species as compared to those species whose growth rates were positive (Figure 4c,d), likely because declining species tend to have smaller population sizes and thus less data available for analysis. In conducting this simple simulation, we demonstrated that combining multispecies demographic and population-level count data provides a viable solution for quantifying species-specific and community-level dynamics and growth rates, which can ultimately aid in biodiversity monitoring and assessments from local to regional scales. Applications of this approach may be particularly beneficial in cases where data are limited for some species within a community (but plentiful for others), or when one or more data types are not collected for all species in every year (e.g. multispecies amphibian survey data, waterfowl banding data, mist net capture data of birds). As these types of models are fairly new, additional work is necessary to understand the full inferential benefits—and potential biases—under scenarios that vary parameter values (e.g. low vs. high detection probability and/or sampling error) and include an exploration as to how the amounts of various data types influence model estimates (e.g. when all three types of data are not available for all species in all years).

4 | THINGS TO CONSIDER BEFORE USING THIS METHOD

Potential users should consider several points before developing an integrated community model for their system. Initial steps in determining the utility and value of combining multispecies data sources should focus on the specific information that could be gained from an integrated community model and the quantity and types of available data. Data integration approaches are becoming increasingly popular (Zipkin et al., 2021), and for good reason, as they have immense potential to expand inferential and predictive capabilities from available, yet imperfect, data (Zylstra & Zipkin, 2021). However, data integration is not without its challenges and limitations. In some cases, combining multispecies data sources—rather than estimating parameters separately for individual species or data sets—may be orders of magnitude more complicated with high computational burdens (Gotway & Young, 2002; Pacifici et al., 2019), or may not substantially help to answer the particular question of

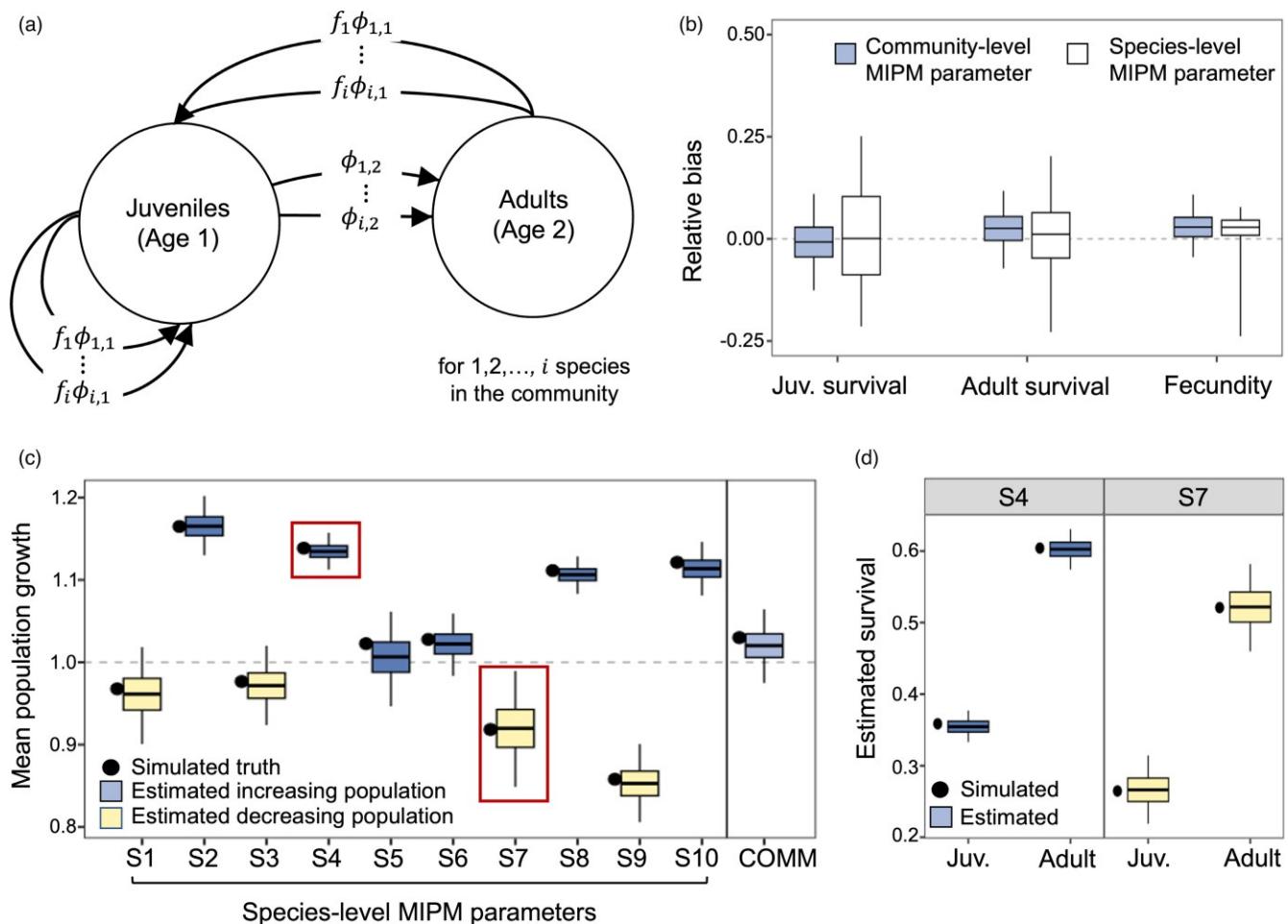


FIGURE 4 Multispecies integrated population models (MIPM) enable unbiased inference on species-specific and community-level demographic parameters as well as mean population growth rates over time, as shown by our simple simulation study. Panel (a) describes the female-based age-structured model with two age classes (juvenile and adult) and a prebreeding census that was used to link population sizes with demographic rates with $i = 10$ species in our hypothetical community. Species-specific demographic rates of fecundity f_i , juvenile survival $\phi_{i,1}$ and adult survival $\phi_{i,2}$ were derived from shared distributions with a community-level mean and variance (not shown). Panel (b) shows the mean relative bias in estimated demographic parameters (dark line) with the 50% (box) and 95% (whiskers) credible intervals at both the community (blue) and species (white) levels compared to a simulated truth. Population growth rates were also estimated from the MIPM with high accuracy and precision, as shown in panel (c), where black dots show the true mean population growth rates for each species (S1–S10) and dark lines show the mean estimated values with 50% (box) and 95% credible interval (whiskers). Community-level (COMM) growth rates were derived as the geometric mean across species. Estimated population growth rates were less precise for declining species with low survival, as shown in panel (d) for two species (S4 and S7) with the same expected fecundity.

interest. Although integrating different data types generally helps with parameter identifiability and precision (Doser et al., 2021; Farr et al., 2021), data integration alone cannot correct for biases in unstructured data or problems with collinearity among environmental variables, especially when the drivers of such biases are unknown or cannot be incorporated within the observation model (Simmonds et al., 2020). Simulations and model assessments can help establish the inferential value of integrated community models as compared to simpler alternatives for specific study systems. Furthermore, the quantity and types of data available have important implications for the potential structure and complexity of the biological process model. Although the incorporation of mechanistic processes within models is a clear goal within ecological research, if data are unavailable to estimate detailed demographic parameters, no amount of

integration will be able to rectify the situation (Plard et al., 2021; Riecke et al., 2019).

Several papers have focused on the individual challenges of both single-species data integration (Isaac et al., 2020; Miller et al., 2019; Zipkin et al., 2021) and hierarchical community modelling (Guillera-Arroita, 2017; Iknayan et al., 2014). Successful data integration models require consideration of the spatial extent and scale of the various data sources to resolve mismatches in the collection grain (e.g. through change of support; Farr et al., 2021; Pacifici et al., 2019; Zipkin et al., 2017), spatial biases in unstructured and semi-structured data due to preferential sampling (e.g. by including site selection within the model [Conn et al., 2017] or spatially correlated random effects [Hefley et al., 2017]), and issues related to unbalanced quantities of various data sources

(e.g. through subsampling or downweighting high volume, low information-content data [Johnston et al., 2018] or modelling biases within the likelihood [Fer et al., 2018; Tang et al., 2021]). Similarly, the value of inferences within community models will depend on how the community is defined and the proportion of species in the community that are rare or undersampled. Statistically speaking, including all species that were observed during a multispecies data collection event within a community model is legitimate, as the result will simply draw species-level parameters towards community averages (Dorazio & Royle, 2005). In practice, however, researchers will want to have an ecological justification for including individual species within a community modelling analysis. Additionally, when a high proportion of species are rare or infrequently observed, it is difficult to achieve convergence in community models and parameter estimates are likely to be exceedingly imprecise for many species (Zipkin et al., 2020), potentially rendering such analyses less informative.

Within the context of integrated community models, the challenges of both data integration and community modelling are likely to be present and may be exacerbated. For example, problems associated with unbalanced data in integrated community models may be acute when the spatial extent of the study is broad (Zipkin et al., 2021) or the goal is to understand mechanistic processes (Campbell et al., 2018). This is especially true if data quantities are highly uneven among species or many species are underrepresented or absent within a specific data source, either due to sampling constraints or cryptic behaviours that vary among species within the community. Combining replicated and non-replicated data sets typically requires more complex ways of accounting for errors in detection (Doser, Leuenberger, et al., 2022). In many hierarchical community models, which focus solely on estimating occurrence or abundance rates, defining the community is done rather loosely and may simply be 'all species' (all observed species, or both observed and unobserved species if using data augmentation; Royle et al., 2007). In the context of an integrated community model, it may be more important to have a clear definition of, and strong biological justification for, the target community because of the additional parameters that are estimated (e.g. demographic rates). Otherwise, inferences on community parameters—as well as for rare species—may not be biologically meaningful.

Finally, we expect there to be computational issues for many types of integrated community models, as this is a concern for both single-species integrated models and hierarchical community models. Many hierarchical models are analysed using Bayesian methods, such as MCMC approaches that can take an exceedingly long time to run and may be difficult to troubleshoot. Typically, such models are fit using common Bayesian software packages such as JAGS, NIMBLE and Stan, which provide flexibility for defining specialized models, but require a substantial amount of programming knowledge and custom coding. Thus, potential users should consider if it is worth developing a complicated, but more adaptable, model or if qualitatively similar inferences can be achieved using built-in software (e.g. R packages). Fortunately, there has been

substantial development on creating user-friendly and computationally efficient software for both single-species integrated models and community models. Simple forms of integrated community models, like the hierarchical negative binomial model used in the butterfly case study, can be fit using user-friendly R packages that can accommodate a variety of random effect structures (e.g. brms [Bürkner, 2017], spAbundance [Doser, 2023]), while the integrated community occupancy model used in the bird case study can be fit with the spOccupancy R package (Doser, Finley, et al., 2022). Continued development of specialized and computationally efficient software is an important avenue for future research that can more readily facilitate the implementation of integrated community models. However, for now, many complex biological and observational model structures are likely to require custom coding using Bayesian programming languages.

5 | CONCLUSIONS

Integrated community models are an exciting framework to assess biodiversity dynamics using unified approaches that can combine data from multiple different sources that vary in information content. Such methods can leverage public science, government-funded and traditional, scientist-collected data sources to expand the spatial and temporal scope of inference for the many data-deficient species on which little is known. Integrated community models take advantage of the collective benefits of data integration and community modelling frameworks, while accounting for both biological processes and observation errors. Thus, these models can improve the accuracy and precision on both species-level parameters and community-level metrics, enhancing understanding about the variations in species' response to local environmental factors and global climate and land-use change. Future research in integrated community modelling could focus on approaches to assess model fit and compare among competing model structures, incorporating wider varieties of structured and unstructured data types including opportunistic presence-only data and expanding inferences to communities with greater numbers of species. The development of integrated models also has important implications for the design of new data collection activities. Forthcoming sampling and monitoring programmes should consider what data sets are already available on particular species and communities in order to obtain maximally beneficial data (Chandler et al., 2017; Moussy et al., 2021). For example, targeting areas with limited data or collecting data on unknown demographic rates could be particularly advantageous within an integrated community modelling framework. While integrated community models are still at an early stage of development, given the exponentially growing use of both single-species integrated models and hierarchical community models over the last decade, we expect there to be much activity in this area over the next decade. Such approaches are leading to increasingly efficient and useful assessments of biodiversity, which is critically important as the world's environment is changing rapidly.

AUTHOR CONTRIBUTIONS

Elise F. Zipkin conceived of the study, lead the team and wrote the initial draft of the paper with input from Jeffrey W. Doser and Courtney L. Davis. Jeffrey W. Doser, Wendy Leuenberger and Courtney L. Davis led the development and analysis of case studies one, two and three, respectively, with input from Samuel Ayebare, Kayla L. Davis and Elise F. Zipkin. All authors reviewed the manuscript and provided edits during revisions.

ACKNOWLEDGEMENTS

We thank Leslie Ries and the North American Butterfly Association (NABA) for providing access to the data used in case study two. We thank the editors and reviewers for many helpful comments on earlier drafts, especially Rob Salguero-Gómez for extensive feedback at various stages of the writing process. This study was funded by the United States National Science Foundation (NSF) with award DEB-1954406.

CONFLICT OF INTEREST STATEMENT

The authors have no conflict of interest.

DATA AVAILABILITY STATEMENT

The code and data used in the integrated community modelling case studies can be found on the Zipkin Lab Code Archive (<https://zipkinlab.github.io/>) and are also permanently archived on Zenodo at <https://doi.org/10.5281/zenodo.8361425> (Zipkin et al., 2023).

ETHICS STATEMENT

Not applicable.

ORCID

Elise F. Zipkin  <https://orcid.org/0000-0003-4155-6139>

Jeffrey W. Doser  <https://orcid.org/0000-0002-8950-9895>

Courtney L. Davis  <https://orcid.org/0000-0002-6467-4288>

Kayla L. Davis  <https://orcid.org/0000-0002-9360-828X>

REFERENCES

Altweig, R., & Nichols, J. D. (2019). Occupancy models for citizen-science data. *Methods in Ecology and Evolution*, 10, 8–21. <https://doi.org/10.1111/2041-210X.13090>

Barnett, D. T., Duffy, P. A., Schimel, D. S., Krauss, R. E., Irvine, K. M., Davis, F. W., Gross, J. E., Azuaje, E. I., Thorpe, A. S., Gudex-Cross, D., & Patterson, M. (2019). The terrestrial organism and biogeochemistry spatial sampling design for the National Ecological Observatory Network. *Ecosphere*, 10, e02540. <https://doi.org/10.1002/ecs2.2540>

Barraquand, F., & Gimenez, O. (2019). Integrating multiple data sources to fit matrix population models for interacting species. *Ecological Modelling*, 411, 108713. <https://doi.org/10.1016/j.ecolmodel.2019.06.001>

Besbeas, P., Freeman, S. N., Morgan, B. J. T., & Catchpole, E. A. (2002). Integrating mark-recapture-recovery and census data to estimate animal abundance and demographic parameters. *Biometrics*, 58, 540–547. <https://doi.org/10.1111/j.0006-341x.2002.00540.x>

Beyer, R. M., & Manica, A. (2020). Historical and projected future range sizes of the world's mammals, birds, and amphibians. *Nature Communications*, 11, 5633. <https://doi.org/10.1038/s41467-020-19455-9>

Bland, L. M., Bielby, J., Kearney, S., Orme, C. D. L., Watson, J. E., & Collen, B. (2017). Toward reassessing data-deficient species. *Conservation Biology*, 3, 531–539. <https://doi.org/10.1111/cobi.12850>

Breiner, F. T., Guisan, A., Bergamini, A., & Nobis, M. P. (2015). Overcoming limitations of modelling rare species by using ensembles of small models. *Journal of Animal Ecology*, 6, 1210–1218. <https://doi.org/10.1111/2041-210X.12403>

Brown, J. L., & Collopy, M. W. (2013). Immigration stabilizes a population of threatened cavity-nesting raptors despite possibility of nest box imprinting. *Journal of Avian Biology*, 44, 141–148. <https://doi.org/10.1111/j.1600-048X.2012.05728.x>

Bürkner, P.-C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, 80(1), 1–28. <https://doi.org/10.18637/jss.v080.i01>

Campbell, S. P., Zylstra, E. R., Darst, C. R., Averill-Murray, R. C., & Steidl, R. J. (2018). A spatially explicit hierarchical model to characterize population viability. *Ecological Applications*, 28, 2055–2065. <https://doi.org/10.1002/eaap.1794>

Caswell, H. (2000). *Matrix population models: Construction, analysis, and interpretation* (2nd ed.). Sinauer Associates.

Chandler, M., See, L., Copas, K., Bonde, A. M., López, B. C., Danielsen, F., Legind, J. K., Masinde, S., Miller-Rushing, A. J., Newman, G., & Rosemartin, A. (2017). Contribution of citizen science towards international biodiversity monitoring. *Biological Conservation*, 213, 280–294. <https://doi.org/10.1016/j.biocon.2016.09.004>

Clark, J. S., Nemergut, D., Seyednasrollah, B., Turner, P. J., & Zhang, S. (2017). Generalized joint attribute modeling for biodiversity analysis: Median-zero, multivariate, multifarious data. *Ecological Monographs*, 87, 34–56. <https://doi.org/10.1002/ecm.1241>

Clark, T. J. (2021). *Large carnivore recolonization reshapes population and community dynamics: Implications for harvest management*. Dissertation. University of Montana.

Conn, P. B., Thorson, J. T., & Johnson, D. S. (2017). Confronting preferential sampling when analysing population distributions: Diagnosis and model-based triage. *Methods in Ecology and Evolution*, 8, 1535–1546. <https://doi.org/10.1111/2041-210X.12803>

Davis, C. L., Guralnick, R. P., & Zipkin, E. F. (2023). Challenges and opportunities for using natural history collections to estimate insect population trends. *Journal of Animal Ecology*, 92, 237–249. <https://doi.org/10.1111/1365-2656.13763>

de Valpine, P., Turek, D., Paciorek, C. J., Anderson-Bergman, C., Temple Lang, D., & Bodik, R. (2017). Programming with models: Writing statistical algorithms for general model structures with NIMBLE. *Journal of Computational and Graphical Statistics*, 26, 403–413. <https://doi.org/10.1080/10618600.2016.1172487>

Dennis, E. B., Freeman, S. N., Brereton, T., & Roy, D. B. (2013). Indexing butterfly abundance whilst accounting for missing counts and variability in seasonal pattern. *Methods in Ecology and Evolution*, 4, 637–645. <https://doi.org/10.1111/2041-210X.12053>

Devarajan, K., Morelli, T. L., & Tenan, S. (2020). Multi-species occupancy models: Review, roadmap, and recommendations. *Ecography*, 43, 1612–1624. <https://doi.org/10.1111/ecog.04957>

Didham, R. K., Basset, Y., Collins, C. M., Leather, S. R., Littlewood, N. A., Menz, M. H., Müller, J., Packer, L., Saunders, M. E., Schönrogge, K., Stewart, A. J. A., Yanoviak, S. P., & Hassall, C. (2020). Interpreting insect declines: Seven challenges and a way forward. *Insect Conservation and Diversity*, 13, 103–114. <https://doi.org/10.1111/icad.12408>

Dorazio, R. M. (2014). Accounting for imperfect detection and survey bias in statistical analysis of presence-only data. *Global Ecology and Biogeography*, 23, 1472–1484. <https://doi.org/10.1111/geb.12216>

Dorazio, R. M. (2016). Bayesian data analysis in population ecology: Motivations, methods, and benefits. *Population Ecology*, 58, 31–44. <https://doi.org/10.1007/s10144-015-0503-4>

Dorazio, R. M., & Royle, J. A. (2005). Estimating size and composition of biological communities by modeling the occurrence of species. *Journal of the American Statistical Association*, 100, 389–398. <https://doi.org/10.1198/016214505000000015>

Dorazio, R. M., Royle, J. A., Söderström, B., & Glimskär, A. (2006). Estimating species richness and accumulation by modeling species occurrence and detectability. *Ecology*, 87, 842–854. [https://doi.org/10.1890/0012-9658\(2006\)87\[842:ESRAAB\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2006)87[842:ESRAAB]2.0.CO;2)

Doser, J. W. (2023). "spAbundance: An R package for univariate and multivariate spatially-explicit abundance models". R package version v0.1.0. <https://github.com/doserjef/spAbundance>

Doser, J. W., Finley, A. O., Kéry, M., & Zipkin, E. F. (2022). spOccupancy: An R package for single-species, multi-species, and integrated spatial occupancy models. *Methods in Ecology and Evolution*, 13, 1670–1678. <https://doi.org/10.1111/2041-210X.13897>

Doser, J. W., Finley, A. O., Weed, A. S., & Zipkin, E. F. (2021). Integrating automated acoustic vocalization data and point count surveys for estimation of bird abundance. *Methods in Ecology and Evolution*, 12(6), 1040–1049. <https://doi.org/10.1111/2041-210X.13578>

Doser, J. W., Leuenberger, W., Sillett, T. S., Hallworth, M. T., & Zipkin, E. F. (2022). Integrated community occupancy models: A framework to assess occurrence and biodiversity dynamics using multiple data sources. *Methods in Ecology and Evolution*, 13, 919–932. <https://doi.org/10.1111/2041-210X.13811>

Farr, M. T., Green, D. S., Holekamp, K. E., Roloff, G. J., & Zipkin, E. F. (2019). Multispecies hierarchical modeling reveals variable responses of African carnivores to management alternatives. *Ecological Applications*, 29, e01845. <https://doi.org/10.1002/eaap.1845>

Farr, M. T., Green, D. S., Holekamp, K. E., & Zipkin, E. F. (2021). Integrating distance sampling and presence-only data to estimate abundance. *Ecology*, 102(1), e03204. <https://doi.org/10.1002/ecy.3204>

Fer, I., Kelly, R., Moorcroft, P. R., Richardson, A. D., Cowdery, E. M., & Dietze, M. C. (2018). Linking big models to big data: Efficient ecosystem model calibration through Bayesian model emulation. *Biogeosciences*, 15, 5801–5830. <https://doi.org/10.5194/bg-15-5801-2018>

Fletcher, R. J., Jr., Hefley, T. J., Robertson, E. P., Zuckerberg, B., McCleery, R. A., & Dorazio, R. M. (2019). A practical guide for combining data to model species distributions. *Ecology*, 100, e02710. <https://doi.org/10.1002/ecy.2710>

Fordyce, J. A., Gompert, Z., Forister, M. L., & Nice, C. C. (2011). A hierarchical Bayesian approach to ecological count data: A flexible tool for ecologists. *PLoS ONE*, 6, e26785. <https://doi.org/10.1371/journal.pone.0026785>

Forister, M. L., Halsch, C. A., Nice, C. C., Fordyce, J. A., Dilts, T. E., Oliver, J. C., Prudic, K. L., Shapiro, A. M., Wilson, J. K., & Glassberg, J. (2021). Fewer butterflies seen by community scientists across the warming and drying landscapes of the American West. *Science*, 371, 1042–1045. <https://doi.org/10.1126/science.abe5585>

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Gotway, C. A., & Young, L. J. (2002). Combining incompatible spatial data. *Journal of the American Statistical Association*, 97, 632–648.

Grattarola, F., Bowler, D., & Keil, P. (2023). Integrating presence-only and presence-absence data to model changes in species geographic ranges: An example in the Neotropics. *Journal of Biogeography*, 50, 1561–1675. <https://doi.org/10.1111/jbi.14622>

Guillera-Arroita, G. (2017). Modelling of species distributions, range dynamics and communities under imperfect detection: Advances, challenges and opportunities. *Ecography*, 40, 281–295. <https://doi.org/10.1111/ecog.02445>

Guillera-Arroita, G., Kéry, M., & Lahoz-Monfort, J. J. (2019). Inferring species richness using multispecies occupancy modeling: Estimation performance and interpretation. *Ecology and Evolution*, 9, 780–792. <https://doi.org/10.1002/ece3.4821>

Hallmann, C. A., Sorg, M., Jongejans, E., Siepel, H., Hofland, N., Schwan, H., Stenmans, W., Müller, A., Sumser, H., Hörren, T., Goulson, D., & de Kroon, H. (2017). More than 75 percent decline over 27 years in total flying insect biomass in protected areas. *PLoS ONE*, 12, e0185809. <https://doi.org/10.1371/journal.pone.0185809>

Hamilton, S. H., Pollino, C. A., & Jakeman, A. J. (2015). Habitat suitability modelling of rare species using Bayesian networks: Model evaluation under limited data. *Ecological Modelling*, 299, 64–78. <https://doi.org/10.1016/j.ecolmodel.2014.12.004>

Hefley, T. J., Hooten, M. B., Hanks, E. M., Russell, R. E., & Walsh, D. P. (2017). Dynamic spatio-temporal models for spatial data. *Spatial Statistics*, 20, 206–220. <https://doi.org/10.1016/j.spasta.2017.02.005>

Iknayan, K. J., Tingley, M. W., Furnas, B. J., & Beissinger, S. R. (2014). Detecting diversity: Emerging methods to estimate species diversity. *Trends in Ecology & Evolution*, 29, 97–106. <https://doi.org/10.1016/j.tree.2013.10.012>

Isaac, N. J., Jarzyna, M. A., Keil, P., Dambly, L. I., Boersch-Supan, P. H., Browning, E., Freeman, S. N., Golding, N., Guillera-Arroita, G., Henrys, P. A., & Jarvis, S. (2020). Data integration for large-scale models of species distributions. *Trends in Ecology & Evolution*, 35, 56–67. <https://doi.org/10.1016/j.tree.2019.08.006>

Johnson, C. N., Balmford, A., Brook, B. W., Buettel, J. C., Galetti, M., Guangchun, L., & Wilmshurst, J. M. (2017). Biodiversity losses and conservation responses in the Anthropocene. *Science*, 356, 270–275. <https://doi.org/10.1126/science.aam93>

Johnston, A., Fink, D., Hochachka, W. M., & Kelling, S. (2018). Estimates of observer expertise improve species distributions from citizen science data. *Methods in Ecology and Evolution*, 9, 88–97. <https://doi.org/10.1111/2041-210X.12838>

Johnston, A., Hochachka, W. M., Strimas-Mackey, M. E., Ruiz Gutierrez, V., Robinson, O. J., Miller, E. T., Auer, T., Kelling, S. T., & Fink, D. (2021). Analytical guidelines to increase the value of community science data: An example using eBird data to estimate species distributions. *Diversity and Distributions*, 27, 1265–1277. <https://doi.org/10.1111/ddi.13271>

Kelling, S., Johnston, A., Bonn, A., Fink, D., Ruiz-Gutierrez, V., Bonney, R., Fernandez, M., Hochachka, W. M., Julliard, R., Kraemer, R., & Guralnick, R. (2019). Using semi-structured surveys to improve citizen science data for monitoring biodiversity. *Bioscience*, 69, 170–179. <https://doi.org/10.1093/biosci/biz010>

Kellner, K. (2021). *JagUI: A wrapper around 'rjags' to streamline JAGS analyses*. R package version 1.5.2, <https://CRAN.R-project.org/package=jagsUI>

Kéry, M., & Royle, J. A. (2016). *Applied hierarchical modeling in ecology: Analysis of distribution, abundance, and species richness in R and BUGS*. Academic Press.

Kindvater, H. K., Dulvy, N. K., Horswill, C., Juan-Jordá, M. J., Mangel, M., & Matthiopoulos, J. (2018). Overcoming the data crisis in biodiversity conservation. *Trends in Ecology & Evolution*, 33, 676–688. <https://doi.org/10.1016/j.tree.2018.06.004>

Lahoz-Monfort, J. J., Harris, M. P., Wanless, S., Freeman, S. N., & Morgan, B. J. (2017). Bringing it all together: Multi-species integrated population modelling of a breeding community. *Journal of Agricultural, Biological and Environmental Statistics*, 22, 140–160. <https://doi.org/10.1007/s13253-017-0279-4>

Lauret, V., Labach, H., David, L., Authier, M., & Gimenez, O. (2023). Using integrated multispecies occupancy models to map co-occurrence between bottlenose dolphins and fisheries in the Gulf of lion, French Mediterranean Sea. *Oikos*, e10270. <https://doi.org/10.1111/oik.10270>

Lee, A. M., Bjørkvoll, E. M., Hansen, B. B., Albon, S. D., Stien, A., Sæther, B.-E., Engen, S., Veiberg, V., Loe, L. L., & Grøtan, V. (2015). An integrated population model for a long-lived ungulate: More efficient

data use with Bayesian methods. *Oikos*, 124, 806–816. <https://doi.org/10.1111/oik.01924>

Lomba, A., Pellissier, L., Randin, C., Vicente, J., Moreira, F., Honrado, J., & Guisan, A. (2010). Overcoming the rare species modelling paradox: A novel hierarchical framework applied to an Iberian endemic plant. *Biological Conservation*, 143, 2647–2657. <https://doi.org/10.1016/j.biocon.2010.07.007>

Miller, D. A., Pacifici, K., Sanderlin, J. S., & Reich, B. J. (2019). The recent past and promising future for data integration methods to estimate species' distributions. *Methods in Ecology and Evolution*, 10, 22–37. <https://doi.org/10.1111/2041-210X.13110>

Mosnier, A., Doniol-Valcroze, T., Gosselin, J.-F., Lesage, V., Measures, L. N., & Hammill, M. O. (2015). Insights into processes of population decline using an integrated population model: The case of the St. Lawrence estuary beluga (*Delphinapterus leucas*). *Ecological Modelling*, 314, 15–31. <https://doi.org/10.1016/j.ecolmodel.2015.07.006>

Moussy, C., Burfield, I. J., Stephenson, P. J., Newton, A. F., Butchart, S. H., Sutherland, W. J., Gregory, R. D., McRae, L., Bubb, P., & Roesler, I. (2021). A quantitative global review of species population monitoring. *Conservation Biology*, 36, e1372. <https://doi.org/10.1111/cobi.13721>

Oppel, S., Hilton, G., Ratcliffe, N., Fenton, C., Daley, J., Gray, G., Vickery, J., & Gibbons, D. (2014). Assessing population viability while accounting for demographic and environmental uncertainty. *Ecology*, 95, 1809–1818. <https://doi.org/10.1890/13-0733.1>

Ovaskainen, O., Tikhonov, G., Norberg, A., Guillaume Blanchet, F., Duan, L., Dunson, D., Roslin, T., & Abrego, N. (2017). How to make more out of community data? A conceptual framework and its implementation as models and software. *Ecology Letters*, 20, 561–576. <https://doi.org/10.1111/ele.12757>

Pacifici, K., Reich, B. J., Miller, D. A., Gardner, B., Stauffer, G., Singh, S., McKerrow, A., & Collazo, J. A. (2017). Integrating multiple data sources in species distribution modeling: A framework for data fusion. *Ecology*, 98, 840–850. <https://doi.org/10.1002/ecy.1710>

Pacifici, K., Reich, B. J., Miller, D. A., & Pease, B. S. (2019). Resolving misaligned spatial data with integrated species distribution models. *Ecology*, 100, e02709. <https://doi.org/10.1002/ecy.2709>

Paquet, M., & Barraquand, F. (2022). Assessing species interactions using integrated predator-prey models. *EcoEvoRxiv* <https://doi.org/10.32942/X2RC7W>

Pardieck, K., Ziolkowski, D., Jr., Lutmerding, M., Aponte, V., & Hudson, M. A. (2020). *North American breeding bird survey dataset 1966–2019*. U.S. Geological Survey Data Release. <https://doi.org/10.5066/P9J6QUF6>

Perón, G., & Koons, D. N. (2012). Integrated modeling of communities: Parasitism, competition, and demographic synchrony in sympatric ducks. *Ecology*, 93, 2456–2464. <https://doi.org/10.1890/11-1881.1>

Plard, F., Fay, R., Kery, M., Cohas, A., & Schaub, M. (2019). Integrated population models: Powerful methods to embed individual processes in population dynamics models. *Ecology*, 100(6), e02715. <https://doi.org/10.1002/ecy.2715>

Plard, F., Turek, D., & Schaub, M. (2021). Consequences of violating assumptions of integrated population models on parameter estimates. *Environmental and Ecological Statistics*, 28(3), 667–695. <https://doi.org/10.1007/s10651-021-00507-2>

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling.

Pollard, E. (1977). A method for assessing changes in the abundance of butterflies. *Biological Conservation*, 12, 115–134. [https://doi.org/10.1016/0006-3207\(77\)90065-9](https://doi.org/10.1016/0006-3207(77)90065-9)

Quéroué, M., Barbraud, C., Barraquand, F., Turek, D., Delord, K., Pacourea, N., & Gimenez, O. (2021). Multispecies integrated population model reveals bottom-up dynamics in a seabird predator-prey system. *Ecological Monographs*, 91, e01459. <https://doi.org/10.1002/ecm.1459>

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org>

Rapacciulo, G., & Blois, J. L. (2019). Understanding ecological change across large spatial, temporal and taxonomic scales: Integrating data and methods in light of theory. *Ecography*, 42, 1247–1266. <https://doi.org/10.1111/ecog.04616>

Riecke, T. V., Williams, P. J., Behnke, T. L., Gibson, D., Leach, A. G., Sedinger, B. S., Street, P. A., & Sedinger, J. S. (2019). Integrated population models: Model assumptions and inference. *Methods in Ecology and Evolution*, 10(7), 1072–1082. <https://doi.org/10.1111/2041-210X.13195>

Ries, L., Zipkin, E. F., & Guralnick, R. P. (2019). Tracking trends in monarch abundance over the 20th century is currently impossible using museum records. *Proceedings of the National Academy of Sciences of the United States of America*, 116, 13745–13748. <https://doi.org/10.1073/pnas.1904807116>

Rollinson, C. R., Finley, A. O., Alexander, M. R., Banerjee, S., Hamil, K. D., Koenig, L. E., Locke, D. H., Peterson, M., Tingley, M. W., Wheeler, K., Youngflesh, C., & Zipkin, E. F. (2021). Working across space and time: Nonstationarity in ecological research and application. *Frontiers in Ecology and the Environment*, 19, 66–72. <https://doi.org/10.1002/fee.2298>

Royle, J. A., Dorazio, R. M., & Link, W. A. (2007). Analysis of multinomial models with unknown index using data augmentation. *Journal of Computational and Graphical Statistics*, 16(1), 67–85.

Saunders, S. P., Cuthbert, F. J., & Zipkin, E. F. (2018). Evaluating population viability and efficacy of conservation management using integrated population models. *Journal of Applied Ecology*, 55, 1380–1392. <https://doi.org/10.1111/1365-2664.13080>

Saunders, S. P., Farr, M. T., Wright, A. D., Bahlai, C. A., Ribeiro, J. W., Rossman, S., Sussman, A. L., Arnold, T. W., & Zipkin, E. F. (2019). Disentangling data discrepancies with integrated population models. *Ecology*, 100, e02714. <https://doi.org/10.1002/ecy.2714>

Saunders, S. P., Ries, L., Neupane, N., Ramírez, M. I., García-Serrano, E., Rendón-Salinas, E., & Zipkin, E. F. (2019). Multi-scale seasonal factors drive the size of winter monarch colonies. *Proceedings of the National Academy of Sciences of the United States of America*, 116, 8609–8614. <https://doi.org/10.1073/pnas.1805114116>

Schaub, M., & Abadi, F. (2011). Integrated population models: A novel analysis framework for deeper insights into population dynamics. *Journal of Ornithology*, 152, 227–237.

Schaub, M., Gimenez, O., Sierro, A., & Arlettaz, R. (2007). Use of integrated modeling to enhance estimates of population dynamics obtained from limited data. *Conservation Biology*, 21, 945–955. <https://doi.org/10.1111/j.1523-1739.2007.00743.x>

Schaub, M., & Kéry, M. (2021). *Integrated population models: Theory and ecological applications with R and JAGS*. Academic Press.

Simmonds, E. G., Jarvis, S. G., Henrys, P. A., Isaac, N. J., & O'Hara, R. B. (2020). Is more data always better? A simulation study of benefits and limitations of integrated distribution models. *Ecography*, 43, 1413–1422. <https://doi.org/10.1111/ecog.05146>

Sollmann, R., Gardner, B., Williams, K. A., Gilbert, A. T., & Veit, R. R. (2016). A hierarchical distance sampling model to estimate abundance and covariate associations of species and communities. *Methods in Ecology and Evolution*, 7, 529–537. <https://doi.org/10.1111/2041-210X.12518>

Sor, R., Park, Y. S., Boets, P., Goethals, P. L., & Lek, S. (2017). Effects of species prevalence on the performance of predictive models. *Ecological Modelling*, 354, 11–19. <https://doi.org/10.1016/j.ecolmodel.2017.03.006>

Stan Development Team. (2022). Stan modeling language users guide and reference manual, Version 2.31. <https://mc-stan.org>

Strimas-Mackey, M., Hochachka, W. M., Ruiz-Gutiérrez, V., Robinson, O. J., Miller, E. T., Auer, T., Kelling, S., Fink, D., & Johnston, A. (2020). *Best practices for using eBird data. Version 1.0*. Cornell Lab

of Ornithology. <https://doi.org/10.5281/zenodo.3620739>; <https://cornellabofornithology.github.io/ebird-best-practices/>

Sullivan, B. L., Wood, C. L., Iliff, M. J., Bonney, R. E., Fink, D., & Kelling, S. (2009). eBird: A citizen-based bird observation network in the biological sciences. *Biological Conservation*, 142, 2282–2292. <https://doi.org/10.1016/j.biocon.2009.05.006>

Tang, B., Clark, J. S., & Gelfand, A. E. (2021). Modeling spatially biased citizen science effort through the eBird database. *Environmental and Ecological Statistics*, 28, 609–630. <https://doi.org/10.1007/s10651-021-00508-1>

Thornhill, I., Loiselle, S., Lind, K., & Ophof, D. (2016). The citizen science opportunity for researchers and agencies. *Bioscience*, 66, 720–721. <https://doi.org/10.1093/biosci/biw089>

Thorson, J. T., Ianelli, J. N., Larsen, E. A., Ries, L., Scheuerell, M. D., Szuwalski, C., & Zipkin, E. F. (2016). Joint dynamic species distribution models: A tool for community ordination and spatio-temporal monitoring. *Global Ecology and Biogeography*, 25, 1144–1158. <https://doi.org/10.1111/geb.12464>

Threlfall, C. G., Mata, L., Mackie, J. A., Hahs, A. K., Stork, N. E., Williams, N. S., & Livesley, S. J. (2017). Increasing biodiversity in urban green spaces through simple vegetation interventions. *Journal of Applied Ecology*, 54, 1874–1883. <https://doi.org/10.1111/1365-2664.12876>

Wagner, D. L., Grames, E. M., Forister, M. L., Berenbaum, M. R., & Stopak, D. (2021). Insect decline in the Anthropocene: Death by a thousand cuts. *Proceedings of the National Academy of Sciences of the United States of America*, 118, e2023989118. <https://doi.org/10.1073/pnas.2023989118>

Warton, D. I., Blanchet, F. G., O'Hara, R. B., Ovaskainen, O., Taskinen, S., Walker, S. C., & Hui, F. K. C. (2015). So many variables: Joint modeling in community ecology. *Trends in Ecology & Evolution*, 30, 766–779. <https://doi.org/10.1016/j.tree.2015.09.007>

Wepprich, T., Adrion, J. R., Ries, L., Wiedmann, J., & Haddad, N. M. (2019). Butterfly abundance declines over 20 years of systematic monitoring in Ohio, USA. *PLoS ONE*, 14, e0216270. <https://doi.org/10.1371/journal.pone.0216270>

Wikle, C. K., & Berliner, L. M. (2007). A Bayesian tutorial for data assimilation. *Physica D*, 230, 1–16. <https://doi.org/10.1016/j.physd.2006.09.017>

Zipkin, E. F., DeWan, A., & Royle, J. A. (2009). Impacts of forest fragmentation on species richness: A hierarchical approach to community modelling. *Journal of Applied Ecology*, 46, 815–822. <https://doi.org/10.1111/j.1365-2664.2009.01664.x>

Zipkin, E. F., DiRenzo, G. V., Ray, J. M., Rossman, S., & Lips, K. R. (2020). Tropical snake diversity collapses after widespread amphibian loss. *Science*, 367(6479), 814–816. <https://doi.org/10.1126/science.aay5733>

Zipkin, E. F., Doser, J. W., Davis, C. L., Leuenberger, W., Ayebare, S., & Davis, K. L. (2023). Data from: Integrated community models: A framework combining multi-species data sources to estimate the status, trends, and dynamics of biodiversity. *Zenodo*. <https://doi.org/10.5281/zenodo.8361425>

Zipkin, E. F., Inouye, B. D., & Beissinger, S. R. (2019). Innovations in data integration for modeling populations. *Ecology*, 100, e02713. <https://doi.org/10.1002/ecy.2713>

Zipkin, E. F., Rossman, S., Yackulic, C. B., Wiens, J. D., Thorson, J. T., Davis, R. J., & Grant, E. H. C. (2017). Integrating count and detection–nondetection data to model population dynamics. *Ecology*, 98, 1640–1650. <https://doi.org/10.1002/ecy.183>

Zipkin, E. F., Royle, J. A., Dawson, D. K., & Bates, S. (2010). Multi-species occurrence models to evaluate the effects of conservation and management actions. *Biological Conservation*, 143, 479–484. <https://doi.org/10.1016/j.biocon.2009.11.016>

Zipkin, E. F., & Saunders, S. P. (2018). Synthesizing multiple data types for biological conservation using integrated population models. *Biological Conservation*, 217, 240–250. <https://doi.org/10.1016/j.biocon.2017.10.017>

Zipkin, E. F., Zylstra, E. R., Wright, A. D., Saunders, S. P., Finley, A. O., Dietze, M. C., Itter, M., & Tingley, M. W. (2021). Addressing data integration challenges to link ecological processes across scales. *Frontiers in Ecology and the Environment*, 19, 30–38. <https://doi.org/10.1002/fee.2290>

Zylstra, E. R., Ries, L., Neupane, N., Saunders, S. P., Ramírez, M. I., Rendón-Salinas, E., Oberhauser, K. S., Farr, M. T., & Zipkin, E. F. (2021). Changes in climate drive recent monarch butterfly dynamics. *Nature Ecology and Evolution*, 5, 1441–1452. <https://doi.org/10.1038/s41559-021-01504-1>

Zylstra, E. R., & Zipkin, E. F. (2021). Accounting for sources of uncertainty when forecasting population responses to climate change. *Journal of Animal Ecology*, 90, 558–561. <https://doi.org/10.1111/1365-2656.13443>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Supplemental Information S1. Case study 1: Spatial distributions of forest birds across the Northeastern United States.

Supplemental Information S2. Case study 2: Temporal trends of butterflies in the Midwestern United States.

Supplemental Information S3. Case study 3: Estimating species- and community-level demographic rates and population growth.

How to cite this article: Zipkin, E. F., Doser, J. W., Davis, C. L., Leuenberger, W., Ayebare, S., & Davis, K. L. (2023). Integrated community models: A framework combining multispecies data sources to estimate the status, trends and dynamics of biodiversity. *Journal of Animal Ecology*, 00, 1–15. <https://doi.org/10.1111/1365-2656.14012>