Check for updates

Received: 20 December 2023 | Accepted: 31 May 2024

DOI: 10.1111/2041-210X.14368

Methods in Ecology and Evolution Open Access BRITISH ECOLOGICAL SOCIETY

RESEARCH ARTICLE

# Integrating counts from rigorous surveys and participatory science to better understand spatiotemporal variation in population processes

**Qing Zhao** | **Quresh S. Latif** | **Bryan L. Nuse** | **David C. Pavlacky Jr.** | **Christopher L. Kilner** | **T. Brandt Ryder** | **Christopher E. Latimer**

Bird Conservancy of the Rockies,
Fort Collins, Colorado, USA

**Correspondence**
Qing Zhao
Email: qing.zhao@birdconservancy.org

**Funding information**
Colorado Parks and Wildlife; Environment and Climate Change Canada; National Center for Ecological Analysis and Synthesis; Canadian Wildlife Service

**Handling Editor:** Res Altwegg

## Abstract

1. Knowledge of variation in population processes (e.g. population growth) across broad spatiotemporal scales is fundamental to population ecology and critical for conservation decision-making. Count data from rigorous surveys (e.g. surveys with probabilistic sampling design and distance sampling information) can inform population processes but are often limited in space and time. Participatory science data cover broader spatiotemporal extents but are prone to bias due to limited to no sampling design and lack of distance sampling information, hindering their capability of informing population processes.
2. Here, we developed an integrated dynamic N-mixture model that jointly analyses rigorous survey and participatory science data to inform population growth at broad spatiotemporal extents. The model contains a flexible scaling parameter that allows fixed and random effects to account for biases and errors in participatory science data. We conducted simulations to evaluate the inference performance of this model across a broad range of spatial and temporal overlap between rigorous survey and participatory science data. We also conducted a case study of Baird's Sparrow (*Centronyx bairdii*), a species of conservation concern, to illustrate the application of the integrated model with rigorous survey data from the Integrated Monitoring in Bird Conservation Regions programme and participatory science North American Breeding Bird Survey and eBird data.
3. Simulations showed that the integrated model improved precision without biasing parameter estimates, in comparison with a model informed by rigorous survey data alone. The case study further demonstrated the utility of the integrated model for quantifying range-wide, long-term population processes and environmental drivers despite limited spatiotemporal extent of rigorous survey data. In particular, we found that population growth rate peaked under medium temperature, which were only apparent in the integrated model.

---

4. The integrated model developed in this study is useful for understanding wildlife population processes at broad spatiotemporal scales with count data. The flexible structure of this model, in particular the scaling parameter, makes it highly adaptable to a broad range of ecological systems and survey procedures. These properties make this modelling approach highly relevant for both population ecology and conservation practice.

**KEYWORDS**
citizen science, data integration, distance sampling, integrated distribution model, N-mixture model, opportunistic data, population dynamics, structured survey

## 1 | INTRODUCTION

Understanding spatial and temporal variation in population processes (i.e. population growth and vital rates) beyond simply describing population trends is fundamental to population ecology and critical for developing conservation strategies for wildlife populations under increasing climate and land use change (Benton et al., 2006; Davis, Saunders, et al., 2023; Gamelon et al., 2017; Lewis et al., 2023). In particular, accelerating population declines elevate the value of understanding range-wide, long-term population processes of wildlife species, which is critical for evaluating and predicting their responses to various threats and conservation actions (Paniw et al., 2023; Zhao et al., 2019). Traditional approaches that use capture-based data (e.g. capture–recapture; Pollock, 1991) to inform population processes, however, are logistically and financially intractable to implement across the broad spatial and temporal extents for many species.

Recent advances in dynamic N-mixture models (Bellier et al., 2016; Dail & Madsen, 2011; Hostetler & Chandler, 2015) allow estimation of parameters representing population processes (population growth, apparent survival and recruitment) from counts of unmarked populations. Reliable estimation of these parameters requires count data from rigorous surveys, that is surveys with a probabilistic sampling design (Albert et al., 2010; Williams & Brown, 2019) and auxiliary information (e.g. distance sampling and removal sampling; Seber, 1982, Royle et al., 2004, Sollmann et al., 2016). While probabilistic sampling design allows unbiased representation of the study populations, auxiliary information is important for separating observation errors from ecological processes (Knape et al., 2018; Zhao & Royle, 2019). The requirements increase the time and financial resources needed for data collection, which limits the spatiotemporal extent of rigorous survey data and thus their ability to represent range-wide, long-term population processes.

Conversely, participatory science (Zoellick et al., 2012; also known as citizen science or community science) data can cover broad spatial and temporal extents and thus have been used broadly in ecological research (Sullivan et al., 2014) such as those related to species distributions (Bradsworth et al., 2017; Clark et al., 2017; Johnston et al., 2020; Steen et al., 2021), migration (Cox et al., 2023; Sharma et al., 2022) and community structure (Davis, Bai, et al., 2023; Reif et al., 2022). Participatory science data, however, are often collected under limited to no sampling design and without strong auxiliary information (e.g. distance sampling), making them prone to sampling biases and observations errors that are difficult to separate from ecological processes (Bowler et al., 2022; Hochachka et al., 2012; Johnston et al., 2023; Kamp et al., 2016; Lukyanenko et al., 2016). Due to such limitations, researchers have mainly used participatory science data to inform population trends (Fink et al., 2023; Horns et al., 2018; Phalan et al., 2019; Sauer & Link, 2011). Using participatory science data in dynamic N-mixture models leads to biased parameter estimates and thus inaccurate understanding of population processes (Hostetler & Chandler, 2015).

Integrated distribution models (IDMs) provide a way to leverage the strengths while overcoming the weaknesses of multiple data sets from unmarked populations (count, detection/non-detection and presence-only; Dorazio, 2014, Pacifici et al., 2017, Fletcher et al., 2019, Miller et al., 2019, Isaac et al., 2020), sometimes involving participatory science data (Di Febbraro et al., 2023; Farr et al., 2021; Pagel et al., 2014; Robinson et al., 2020; Schindler et al., 2022; Stillman et al., 2023). Given the relative ease with which participatory science data can be obtained across broad spatiotemporal extents, IDMs can thus provide new opportunities for understanding relationships between ecological patterns and global change drivers (Di Febbraro et al., 2023; Doser et al., 2021; Grüss et al., 2023).

While numerous studies have developed IDMs, it remains challenging to integrate participatory science with rigorous survey data without inducing bias in parameter estimates (Di Febbraro et al., 2023; Hochachka et al., 2012; Johnston et al., 2023; Pacifici et al., 2017). Correctly accounting for biases and errors in participatory science data is particularly important for dynamic N-mixture models due to the sensitivity of these models to assumption violations (Barker et al., 2018; Hostetler & Chandler, 2015; Link et al., 2018; Zhao & Royle, 2019). One potential solution is to introduce a parameter that scales the estimates based on participatory science data to the same level as the estimates based on rigorous survey data (Schindler et al., 2022; Stillman et al., 2023). This scaling parameter, however, has only been used in models that inform population trends (Schindler et al., 2022, Stillman et al., 2023), but not in models that inform population processes. Furthermore, this scaling parameter has not been developed to account for both sampling biases and observation errors in participatory science data. The robustness of this scaling parameter under varying extents of spatiotemporal overlap between rigorous survey and participatory science data also remains to be evaluated.

In this study, we developed an integrated dynamic N-mixture model that jointly analyses rigorous survey and participatory science data to estimate the spatiotemporal variation in abundance, population growth and environmental drivers. The model contains a flexible scaling parameter that allows fixed and random effects to account for potential sampling biases and observation errors in participatory science data. We first used simulations to evaluate whether integrating participatory science data with rigorous survey data improved precision without biasing parameter estimates, while considering a broad range of spatiotemporal overlap between these data. We then illustrated the application of the integrated model with a case study of Baird's Sparrow (*Centronyx bairdii*), a species of conservation concern. In the case study, we jointly analysed rigorous survey data from the Integrated Monitoring in Bird Conservation Regions (IMBCR) programme (Pavlacky et al., 2017) and participatory science data from North American Breeding Bird Survey (BBS; Sauer et al., 2017) and eBird (Sullivan et al., 2009), to quantify range-wide, long-term population processes and environmental drivers of Baird's Sparrow population dynamics. Finally, we provide guidance for the applications of the integrated model in population ecology and conservation.

## 2 | METHODS

### 2.1 | Motivating data

Our model development was motivated by one data set from a rigorous survey, the IMBCR programme and two data sets from participatory science, BBS and eBird. The IMBCR programme has been designed and implemented by Bird Conservancy of the Rockies across 13 US states nested in eight Bird Conservation Regions through collaborations with federal and state agencies, non-governmental organizations and universities. The IMBCR programme is the largest land bird monitoring programme in North America with a rigorous survey (Pavlacky et al., 2017). Due to its probabilistic sampling design (Pavlacky et al., 2017; Stevens & Olsen, 2004) and inclusion of distance and removal sampling information, IMBCR data allow unbiased inference of population abundance and underlying processes. An IMBCR site is 1 km × 1 km in area that contains maximum 16 evenly spaced point count plots as spatial replicates to further facilitate the estimation of observation errors.

BBS is a long-term, volunteer-based, international monitoring program of North American bird species administered by the US Geological Survey and Canadian Wildlife Service that has been operating annually during peak bird breeding season (May and June) since 1966 (Sauer et al., 2017; Ziolkowski Jr. et al., 2022). BBS data have been collected annually at >2500 routes across the United States and Canada (Robbins et al., 1986). While these routes are widespread across landscapes, their vicinity to road can lead to biased representation of local population size. Each route is approximately 40 km long with 50 stops spaced 0.8 km apart; at each stop, surveyors implement a 3-min point count during which they record all species seen or heard. Along with count data, BBS provides information distinguishing individual surveyors and whether they were new to BBS. Potential sampling biases and observer-specific observation errors in BBS data can be represented by random effects, and new observer-specific observation errors can be represented by a fixed effect.

eBird is one of the largest participatory science projects globally and is managed by the Cornell Lab of Ornithology. eBird enables volunteer observers to contribute opportunistic observations to a publicly available database through Cornell's website (Sullivan et al., 2009; https://ebird.org/about/download-ebird-data-products). eBird data are often filtered to minimize effects of sampling biases using the 'auk' package in program R (Strimas-Mackey et al., 2021). Specifically, we censored data to only include complete checklists where observers recorded counts of all species detected to reduce the effects of preferential species reporting; we also filtered data for observations that started between 5 and 11 am, with a ≤6-h duration, ≤10-km distance and ≤10 observers (Strimas-Mackey et al., 2021). We further include all non-zero counts; for counts of 0s, we kept one observation for each location and year to reduce spatial clustering of the data (Robinson et al., 2018). Even with these filtering efforts, eBird data are still prone to sampling biases and observer-specific observation errors that can be represented by random effects and other observation errors such as those related to observation duration that can be represented by fixed effects. Permission was not needed for collecting IMBCR, BBS and eBird data.

### 2.2 | Model description

We developed an integrated dynamic N-mixture model that jointly analyses count data from rigorous surveys and participatory science. We generalized the characteristics of IMBCR, BBS and eBird data to three data sets, one rigorous survey (RS) data and two participatory science (PS1, PS2) data so that the modelling approach is applicable for other data sets. More specifically, PS1 and PS2 data contained biases and errors represented by random and fixed effects that were broadly applicable for most participatory science data (Bird et al., 2014; Johnston et al., 2023). Due to the flexibility of our model in analysing a varying number of data sets, we considered a model that jointly analyses RS, PS1 and PS2 data, a model that jointly analyses RS and PS1 data, and a model that analyses RS data only.

#### 2.2.1 | Process sub-model

The process sub-model specifies how abundance varies across space and time. More specifically, the process sub-model follows the model described by Hostetler and Chandler (2015) such that

$$\left.\begin{array}{c} N_{i,t}\text{-Poisson}\left(\lambda_{i,t}\right) \\ \log\left(\lambda_{i,t=1}\right)=\beta_0^{[0]}+\beta_1^{[0]}\times x_{i,1,1}+\ \ldots\ +\beta_k^{[0]}\times x_{i,1,k} \\ \lambda_{i,t\geq 2}=N_{i,t-1}\times\rho_{i,t-1}+\delta \\ \log\left(\rho_{i,t-1}\right)=\beta_0^{[\rho]}+\beta_D^{[\rho]}\times N_{i,t-1}+\beta_1^{[\rho]}\times x_{i,t,1}+\ \ldots\ +\beta_k^{[\rho]}\times x_{i,t,k} \end{array}\right\}, \tag{1}$$

in which $N_{i,t}$ is the true abundance at site $i$ in year $t$ that follows a Poisson distribution, and $\lambda_{i,t}$ is its expectation. In the first year, $\lambda_{i,t=1}$ is a function of $k$ environmental covariates $x$'s with intercept and slope parameters $\beta^{[0]}$. For subsequent years (i.e. $t \geq 2$), $\lambda_{i,t\geq 2}$ is calculated from population size in the previous year $N_{i,t-1}$, population growth rate $\rho_{i,t-1}$ and expected number of immigrants $\delta$. Population growth rate $\rho_{i,t-1}$ is further a function of density and environmental covariates with intercept and slope parameters $\beta^{[\rho]}$.

### 2.2.2 | Observation sub-models

The first observation sub-model links RS data with the true abundance while utilizing distance and removal sampling information (Amundson et al., 2014; Royle et al., 2004; Seber, 1982) such that

$$\left.\begin{array}{c} y_{i,t,j,k,s}\text{-Multinomial}\left(N_{i,t},\pi_{i,t,j,k,s}\right) \\ \pi_{i,t,j,k,s}=\dfrac{2\times\sigma_{i,t,j}^2\times\left[exp\left(\frac{-D_{k,L}^2}{2\times\sigma_{i,t,j}^2}\right)-exp\left(\frac{-D_{k,U}^2}{2\times\sigma_{i,t,j}^2}\right)\right]}{r^2}\times\left[\left(1-\theta_{i,t,j}\right)^{s-1}\times\theta_{i,t,j}\right] \\ \log\left(\sigma_{i,t,j}\right)=\alpha_0^{[\sigma]}+\alpha_1^{[\sigma]}\times w_{i,t,j,1}+\ \ldots\ +\alpha_l^{[\sigma]}\times w_{i,t,j,l} \\ \text{logit}\left(\theta_{i,t,j}\right)=\alpha_0^{[\theta]}+\alpha_1^{[\theta]}\times w_{i,t,j,1}+\ \ldots\ +\alpha_l^{[\theta]}\times w_{i,t,j,l} \end{array}\right\}, \tag{2}$$

in which $y_{i,t,j,k,s}$ is the RS count at site $i$ in year $t$ at point $j$ in distance bin $k$ and time interval $s$, and is assumed to follow a multinomial distribution with detection probability $\pi_{i,t,j,k,s}$, $\sigma_{i,t,j}$ is the standard deviation for a half-normal distance sampling function describing the decline in perceptibility with distance from the surveyor, $D_{k,L}$ and $D_{k,U}$ are the distances of the lower and upper bounds of the $k$th distance bin, respectively, $r$ is the radius of the area within which observations are conducted and $\theta_{i,t,j}$ is availability within one time interval. Further, $\sigma_{i,t,j}$ and $\theta_{i,t,j}$ are functions of the same RS-specific covariates $w$'s with intercept and slope parameters $\alpha^{[\sigma]}$ and $\alpha^{[\theta]}$, respectively, to capture potential heterogeneity in perceptibility and availability across sites, years, and points. Further, the mean $\sigma_{i,t,j}$ and $\theta_{i,t,j}$ can be expressed as $\overline{\sigma} = exp\left(\alpha_0^{[\sigma]}\right)$ and $\overline{\theta} = logit^{-1}\left(\alpha_0^{[\theta]}\right)$, respectively, since $w$'s are standardized to have 0 mean.

The second observation sub-model links PS1 data with abundance while considering potential sampling biases and observation errors represented by random and fixed effects such that

$$\left.\begin{array}{c} c_{i,t,j,o}\text{-Poisson}\left(N_{i,t}\times\chi_{i,t,j,o}\right) \\ \log\left(\chi_{i,t,j,o}\right)=\alpha_0^{[\chi]}+\tau_o+\alpha_1^{[\chi]}\times z_{i,t,j,1}+\ \ldots\ +\alpha_l^{[\chi]}\times z_{i,t,j,l} \end{array}\right\}, \tag{3}$$

where $c_{i,t,j,o}$ is the PS1 count at site $i$ in year $t$ at point $j$ of cluster (e.g. observations collected by the same observer) $o$, $\chi_{i,t,j,o}$ is a scaling parameter that represents potential sampling biases and observation errors with mean $\overline{\chi} = exp\left(\alpha_0^{[\chi]}\right)$, a cluster-specific random effect $\tau_o$ with mean 0 and standard deviation $\sigma^{[\chi]}$ and fixed effects of PS1-specific covariates $z$'s with slope parameters $\alpha^{[\chi]}$.

The third observation sub-model links PS2 data with abundance while considering potential sampling biases and observation errors represented by fixed effects such that

$$\left.\begin{array}{c} e_{i,t,j}\text{-Poisson}\left(N_{i,t}\times\omega_{i,t,j}\right) \\ \log\left(\omega_{i,t,j}\right)=\alpha_0^{[\omega]}+\alpha_1^{[\omega]}\times v_{i,t,j,1}+\ \ldots\ +\alpha_l^{[\omega]}\times v_{i,t,j,l} \end{array}\right\}, \tag{4}$$

where $e_{i,t,j}$ is the PS2 count at site $i$ in year $t$ at point $j$, and $\omega_{i,t,j}$ is a scaling parameter with mean $\overline{\omega} = exp\left(\alpha_0^{[\omega]}\right)$ and driven by PS2-specific covariates $v$'s with slope parameters $\alpha^{[\omega]}$. By specifying separate observation sub-models for PS1 and PS2 data, our approach enables estimating separate scaling parameters for each data set to achieve flexibility.

## 2.3 | Simulation study

We conducted simulations to detect any bias in parameter estimates. The simulations also aimed to compare the precision of parameter estimates between the three models. During the simulations, we considered a broad range of spatial and temporal overlap between RS, PS1 and PS2 data that went beyond the situation for IMBCR, BBS and eBird data so that our study can guide integration of other data.

We considered 100 sites for RS, 150 sites for PS1 and 200 sites for PS2 in our simulations to represent the idea that participatory science data often cover a larger number of sites than rigorous survey data. Initial abundance and population growth were functions of a covariate representing an environmental gradient with a quadratic form, while the covariate was random generated from a standard Normal distribution. We set $\overline{\chi}$ to 1.5 and $\overline{\omega}$ to 5 to allow substantial bias in PS1 and PS2 data. Further, $\sigma$, $\theta$, $\chi$ and $\omega$ contained variation that was explained by random and/or fixed effects. More specifically, we considered two fixed effects for $\sigma$, two fixed effects for $\theta$, one random effect and three fixed effects for $\chi$ and three fixed effects for $\omega$, using the approach described in Equations 2–4.

The first simulation aimed to evaluate the performance of the models under different temporal availability of RS data due to the concern that rigorous surveys often cover a shorter time period than participatory science and may not be implemented in consecutive years (Hochachka et al., 2012; Johnston et al., 2023). In this simulation, we assumed that all three data sets covered the entire study area, and 50% of the sites from RS overlapped with sites from PS1 and PS2 (Table S1). In the first scenario, all three surveys covered the entire study period of 20 years. In the second scenario, RS covered the last 10 years, PS1 covered the entire 20 years and PS2 covered the last 15 years. In the third scenario, RS covered the 10 odd years (i.e. 1, 3, ..., 19), PS1 covered the entire 20 years and PS2 covered the last 15 years.

The second simulation aimed to evaluate the performance of the models while assuming that RS data covered different proportion of the environmental gradient due to the concern that rigorous surveys often cover limited spatial extents and thus environmental gradient (Hochachka et al., 2012, Johnston et al., 2023). In this simulation, we assumed that 50% of the sites from RS overlapped with sites from PS1 and PS2. We also assumed that RS covered the last 10 years, PS1 covered the entire 20 years and PS2 covered the last 15 years. In the three scenarios, we assumed that RS covered the upper 50%, 25% and 10% proportion of the environmental gradient, respectively, while PS1 and PS2 covered the entire environmental gradient.

The third simulation aimed to evaluate the performance of the models while assuming varied numbers of overlapped sites between RS and PS1/PS2 due to the concern that rigorous surveys and participatory science often are not coordinated and thus may not cover the same sites (Hochachka et al., 2012, Johnston et al., 2023). In this simulation, we assumed that RS covered the last 10 years, PS1 covered the entire 20 years and PS2 covered the last 15 years. We also assumed that RS covered the upper 50% proportion of the environmental gradient, while the PS1 and PS2 covered the entire environmental gradient. In the three scenarios, we assumed that 50%, 25% and 0% of RS sites overlapped with PS1 and PS2 sites, respectively (Table S1).

Because $\overline{\chi}$ and $\overline{\varpi}$ represent a combined effect of sampling biases and observation errors, they can be greater (data overrepresent abundance) or less than 1 (data underrepresent abundance). To evaluate the performance of the integrated model when data underrepresent abundance, we conducted a complementary simulation in which we set $\overline{\chi}$ to 0.1 and $\overline{\varpi}$ to 0.2.

We simulated 100 data sets for each scenario. Posterior samples of all 100 simulations for each situation were pooled to create violin plots. Each violin plot represents the posterior distribution of the parameter. Additionally, a violin plot also contains a boxplot that represents key statistics of the posterior distribution including the median, 50% credible interval and 95% credible interval. We considered a lack of evidence for bias in parameter estimates if the 50% credible interval of the posterior distribution covers the true value of the parameter used in data simulation (Bellier et al., 2016; Doser et al., 2021). Precision was also visualized from the width of the violin plots (Figures 1–3, Figures S1–S3).

## 2.4 | Case study

### 2.4.1 | Baird's Sparrow ecology and conservation status

One-third of bird populations in North America has been lost during the last half century, among which grassland bird

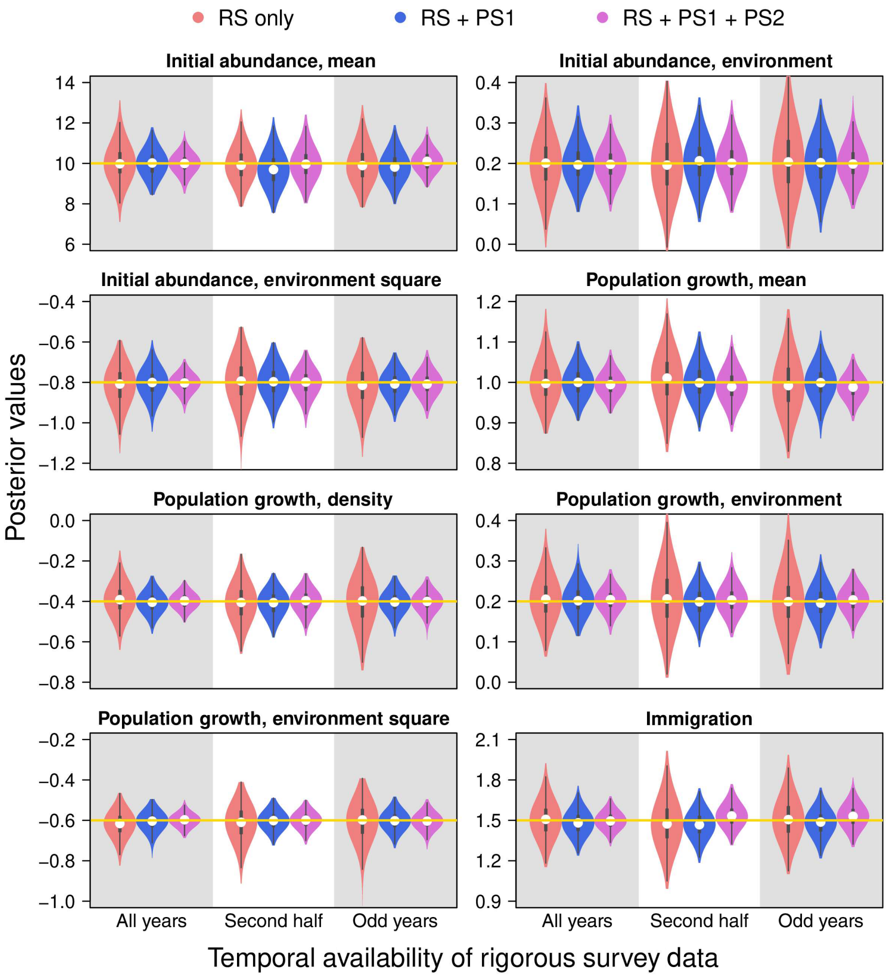


**FIGURE 1** Posterior distributions (violin) and key statistics (median: white dot, 50% credible interval: box, 95% credible interval: whiskers) of process parameters in dynamic N-mixture models that use rigorous survey (RS) data only (red), RS and the first participatory science (PS1) data (blue), or RS, PS1 and the second participatory science (PS2) data (purple) when RS data cover the entire study period of 20 years, the second half of the study period, or odd years of the study period in the simulation study. A horizontal yellow line represents the true value of the corresponding parameter used in data simulation.

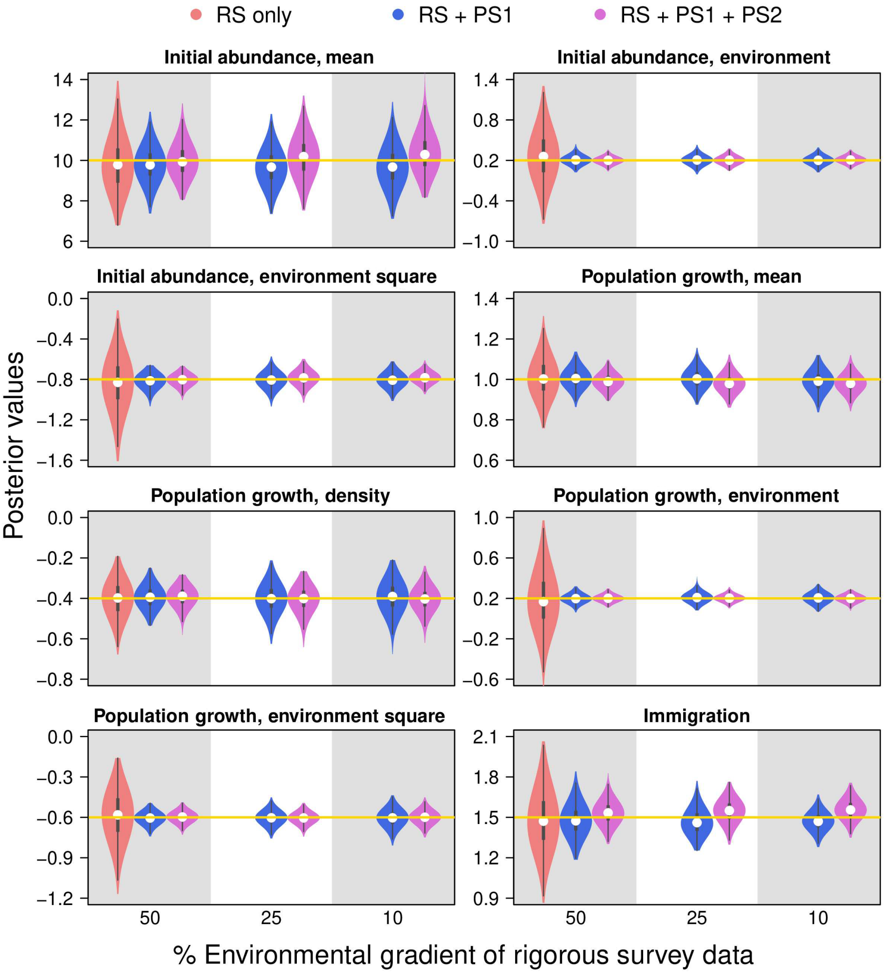


**FIGURE 2** Posterior distributions (violin) and key statistics (median: white dot, 50% credible interval: box, 95% credible interval: whiskers) of process parameters in dynamic N-mixture models that use rigorous survey (RS) data only (red), RS and the first participatory science (PS1) data (blue), or RS, PS1 and the second participatory science (PS2) data (purple) when RS data cover the upper 50%, 25% or 10% proportion of the environmental gradient in the simulation study. The RS-only models did not converge under 25% and 10% coverages, and thus are not shown. A horizontal yellow line represents the true value of the corresponding parameter used in data simulation.

populations have experienced the steepest decline with some species having lost 70%–90% of their historic populations (Rosenberg et al., 2019). Baird's Sparrow (*Centronyx bairdii*) is a grassland bird species of conservation concern due to the fact that their populations have declined by 57% throughout their range from 1970 to 2019 (Rosenberg et al., 2019). Therefore, the species is currently listed as a species of greatest conservation need in multiple state wildlife action plans, bird of conservation concern for US Fish and Wildlife Service, and a tipping point species by Road to Recovery (https://r2rbirds.org/tipping-point-species). As a species that breeds in intact grasslands with tall grass structure and little shrub cover, understanding the effects of grassland habitat loss and climate change on its population processes is essential for effective conservation of this species.

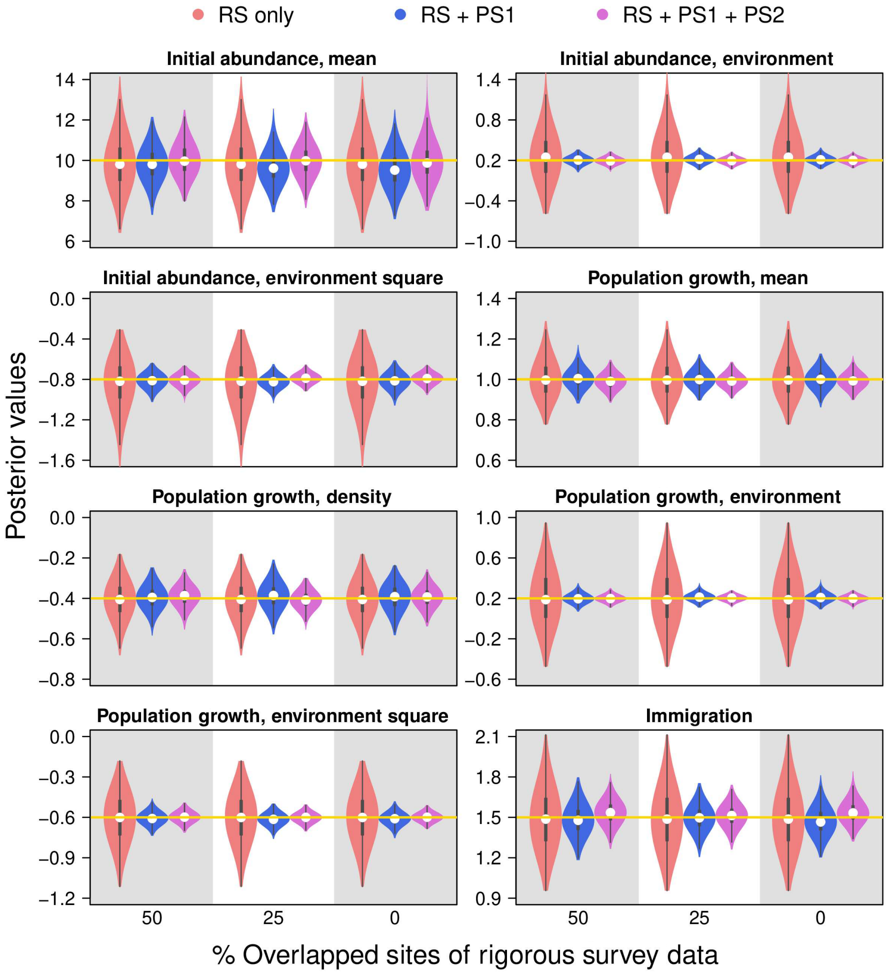
### 2.4.2 | Data preparation

The study area covers the breeding range of Baird's Sparrow that composes two Bird Conservation Regions, Prairie Pothole and Badlands and Prairies (Figure 4). The study area was delineated to 5km×5km grid cells to allow adequate spatial heterogeneity. All three data sets are associated with the grid cells if an IMBCR site, the starting point of a BBS route, or an eBird location falls into the grid cell. We used IMBCR data from years 2010 through 2021 due to consistency in its protocol during this period of time. We used BBS data from 2002 through 2021 because the numbers of routes surveyed are relatively consistent during this period of time. We used eBird data from 2002 through 2021 to be consistent with BBS data. For environmental covariates, we obtained land cover data from the MODIS product MCD12Q1 v061 (Friedl & Sulla-Menashe, 2022) and mean monthly maximum temperatures from the gridded daily surface weather data product for North America, Daymet V4 (Thornton et al., 2022). Grassland coverage and mean temperature of May–July were then calculated for each grid cell and year to represent the spatiotemporal variation of key environmental drivers of the species (Figures S4 and S5).

### 2.4.3 | Model specification

The IMBCR only model covered the southeastern portion of the study area and Years 2010–2021 for which IMBCR data are available. We ran the two integrated models (i.e. IMBCR+BBS,

**FIGURE 3** Posterior distributions (violin) and key statistics (median: white dot, 50% credible interval: box, 95% credible interval: whiskers) of process parameters in dynamic N-mixture models that use rigorous survey (RS) data only (red), RS and the first participatory science (PS1) data (blue), or RS, PS1 and the second participatory science (PS2) data (purple) when RS sites overlap with 50%, 25% or 0% of the PS1 and PS2 sites in the simulation study. A horizontal yellow line represents the true value of the corresponding parameter used in data simulation.



IMBCR + BBS + eBird) for two purposes. First, to evaluate the inference consistency between the models, we limited BBS and eBird data to the same spatiotemporal extent as IMBCR data for the integrated models. Second, to illustrate the capability of the integrated models in range-wide, long-term inference, we extended BBS and eBird data to the entire study area and Years 2002–2021. We also ran a model that used BBS and eBird, but not IMBCR data, to illustrate the importance of IMBCR data in the integrated model.

The model structure overall follows the description above. For the process sub-model, we considered zero inflation for the true abundance with a probability of $1-\zeta$ of having 0s, and included grassland coverage, temperature and temperature square as covariates. For IMBCR data, we assumed constant $\sigma$ and $\theta$ due to consistency in survey effort. For BBS data, we considered random effects of route and observer, and a fixed effect of new observer on the scaling parameter (Sauer & Link, 2011). Because location coordinates are only available for the first stop of each BBS route, all stops of a given route were considered observations of the same grid where the first stop falls in. However, because later stops are more likely to fall out of the grid than earlier stops, we included an error term to describe the difference between the stop-level count and the true abundance of the grid while allowing increasing variance of the error term to represent increasing spatial discrepancy of the stops along each route. For eBird data, we considered random effects of location (with mean 0 and standard deviation $\sigma^{[\omega]}_{\text{location}}$) and observer (with mean 0 and standard deviation $\sigma^{[\omega]}_{\text{observer}}$) and fixed effects of the type (stationary or travelling), duration, distance and number of observers of the observations on the scaling parameter. We also allowed zero inflation (with a probability of $1-\zeta^{[\omega]}$) and false-positive observation errors (with a rate of $\delta^{[\omega]}$) in eBird data.

We conducted posterior predictive checks for the models, using chi-squared statistics as discrepancy measures (Conn et al., 2018; Gelman et al., 2014). A posterior predictive *p*-value between 0.1 and 0.9 was considered indicator of adequate goodness-of-fit of the model to the data (Gelman et al., 2014). We calculated tail probability (*p*-tail) for the slope parameters that represent the effects of grassland coverage, temperature and temperature square on initial population size and population growth rate. The tail probability was defined as the proportion of posterior samples that are below 0 for a positive effect or above 0 for a negative effect. We considered to have a strongly supported effect when a *p*-tail was <0.05, a moderately supported effect when a *p*-tail was 0.05–0.20 and no effect when a *p*-tail was >0.20 (Zhao et al., 2023).

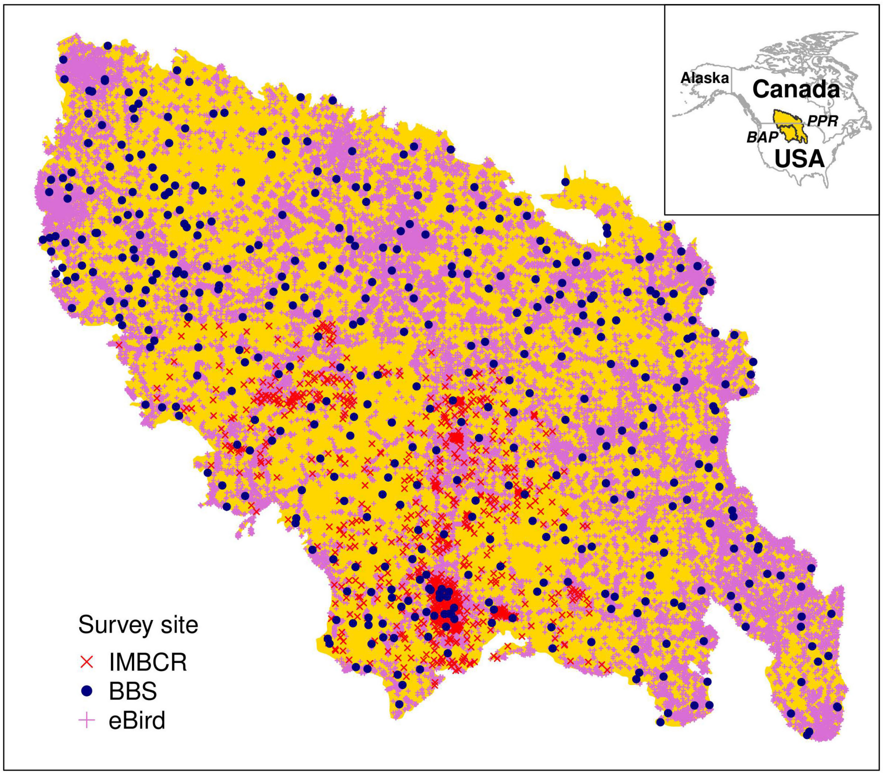


**FIGURE 4** Location of the study area encompassing the Prairie Pothole Region (PPR) and Badlands and Prairies (BAP) in North America (inner panel), and the locations of IMBCR, BBS and eBird survey sites in the study area.

## 2.5 | Model implementation

We conducted Markov chain Monte Carlo (MCMC) computing in Nimble (de Valpine et al., 2017) through the R programming language (R Development Core Team, 2019). We used vague priors in both simulation and case studies. To generate posterior samples, we used three chains and 5000 iterations including 4000 burn-in without thinning for the simulation studies, and three chains and 100,000 iterations including 40,000 burn-in without thinning for the case study. The convergence of the MCMC computing was checked using the diagnostic plot of the posterior samples and Brooks–Gelman–Rubin Diagnostics, where $\hat{R} \leq 1.05$ was considered indicative of convergence (Brooks & Gelman, 1998).

# 3 | RESULTS

## 3.1 | Simulation study

We found no evidence for bias in any of the models under any scenario we considered because the true values of parameters used in data simulations were always covered by the 50% credible interval of the posterior samples of the corresponding parameter (Figures 1–3, Figures S1–S3). For the first simulation study that allowed different temporal availability of rigorous survey data, the integrated models had higher precision in the estimates of process parameters than the model that used rigorous survey data only (Figure 1). For the second simulation study that allowed rigorous survey data to cover different proportions of the environmental gradient, the model that used rigorous survey data only converged when the data covered 50% of the environmental gradient but did not converge when the data covered 25% or 10% of the environmental gradient. The integrated models always converged and had higher precision in the estimates of process parameters than the model that used rigorous survey data only, and the improvement in precision was particularly high for parameters related to environmental effects on initial population size and population growth rate (Figure 2). For the third simulation study that allowed different levels of overlapped sites between rigorous survey and participatory science data, the integrated models had higher precision in the estimates of process parameters than the model that used rigorous survey data only (Figure 3). When considering values of $\overline{\chi}$ and $\overline{\varpi}$ that are less than 1, the integrated model provided improved precision for slope parameters of environmental effects on initial population size and population growth rate (Figure S6).

## 3.2 | Case study

Posterior predictive checks showed adequate goodness-of-fit of all three models to IMBCR, BBS and eBird data (Table 1). The integrated models (i.e. IMBCR + BBS, IMBCR + BBS + eBird) had higher precision in the estimates of process parameters than the IMBCR-only model (Figure 5; Table S2). When limiting BBS and eBird data to the same spatiotemporal extent as IMBCR, the three models were highly consistent in their parameter estimates (Figure 5; Table S2). When extending BBS and eBird data to the

**TABLE 1** Posterior predictive *p* values for IMBCR, BBS and eBird data in a model that analyses IMBCR data only, a model that jointly analyses IMBCR and BBS data, and a model that jointly analyses IMBCR, BBS and eBird data. BBS and eBird data are either limited to the same spatiotemporal extent as IMBCR data, or extended to the entire study area and period.

| | IMBCR-only | IMBCR + BBS, limited | IMBCR + BBS, extended | IMBCR + BBS + eBird, limited | IMBCR + BBS + eBird, extended |
|---|---|---|---|---|---|
| IMBCR data | 0.256 | 0.257 | 0.272 | 0.259 | 0.275 |
| BBS data | | 0.506 | 0.508 | 0.507 | 0.510 |
| eBird data | | | | 0.206 | 0.248 |

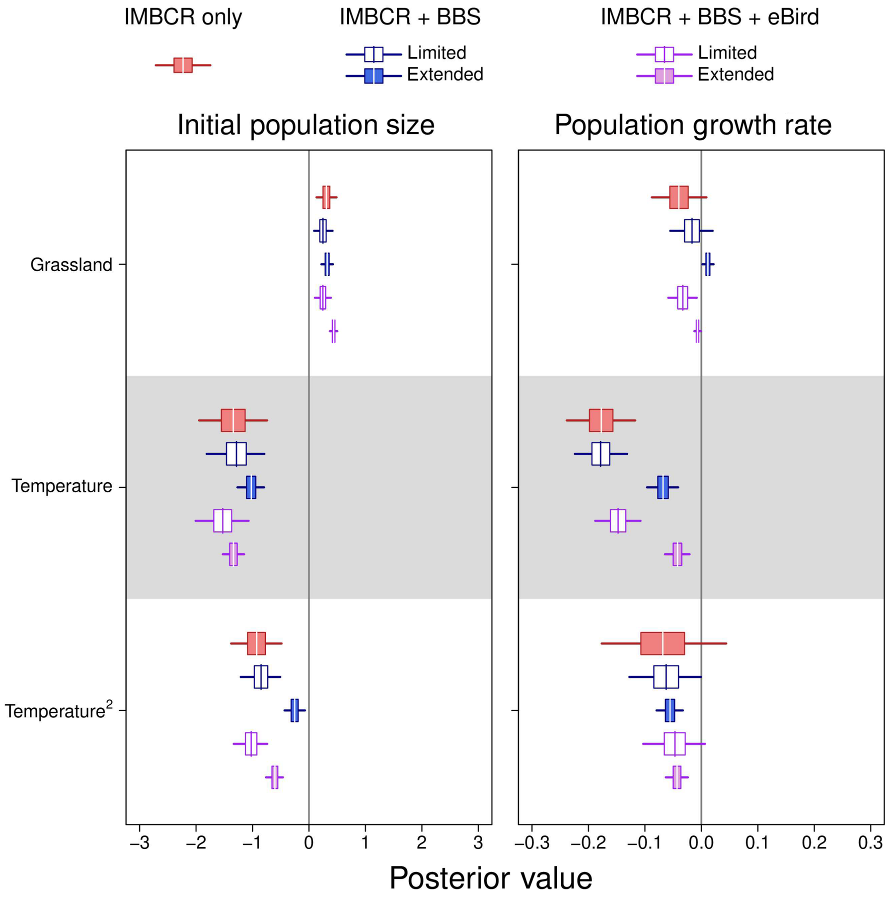


**FIGURE 5** Posterior summaries (white line: median, box: 50% CI, whiskers: 95% CI) of the effect of grassland coverage, effect of temperature, and effect of temperature square on initial population size and population growth rate in a model that analyses IMBCR data only, a model that jointly analyses IMBCR and BBS data, and a model that jointly analyses IMBCR, BBS, and eBird data. BBS and eBird data are either limited to the same spatiotemporal extent as IMBCR data, or extended to the entire study area and period. The vertical grey line represents 0, that is no effect.

entire study area and period, the three models agreed that initial population size peaked under high grassland coverage and medium temperature, but disagreed in the effects of grassland coverage and temperature on population growth rate (Figure 5; Table S2). More specifically, the IMBCR-only model estimated a negative effect, but the IMBCR + BBS model and IMBCR + BBS + eBird model estimated weak effects of grassland coverage on population growth rate (Figure 5; Table S2). While the IMBCR + BBS model and IMBCR + BBS + eBird model agreed in a quadratic relationship, the IMBCR-only model supported a negative relationship but was uncertain about a quadratic relationship between population growth rate and temperature (Figures 5 and 6; Table S2). The model that used BBS and eBird but not IMBCR data did not converge (Figure S7).

## 4 | DISCUSSION

In this study, we introduced a novel modelling approach of integrating count data from rigorous surveys and participatory science to understand wildlife population processes at broad spatial and temporal scales. The flexibility of the scaling parameter allows fixed and random effects to account for sampling biases and observation errors in participatory science data. Simulations demonstrated that integrating participatory science data with rigorous survey data leads to improved precision without biasing parameter estimates under a broad range of spatiotemporal overlap between these data, in comparison with a model that uses rigorous survey data only. A case study further demonstrated the utility of the integrated model for quantifying range-wide, long-term population processes and environmental drivers despite limited spatiotemporal extent of rigorous survey data.

### 4.1 | Model characteristics

Our modelling approach addresses a major challenge in statistical ecology that using participatory science data in integrated models may bias parameter estimates (Di Febbraro et al., 2023; Hochachka et al., 2012; Johnston et al., 2023; Pacifici et al., 2017). We addressed this challenge by using a parameter in the integrated model that scale the estimates of participatory science data to the same level of the estimates of rigorous survey data (Schindler et al., 2022; Stillman et al., 2023). In this way, we successfully used participatory science data in a dynamic N-mixture model to inform population growth, while previous studies that integrated participatory science data were mostly constrained to static (Conn et al., 2022; Dambly et al., 2023; Farr et al., 2021; Gelfand & Shirota, 2019; Koshkina et al., 2017; Robinson et al., 2020) or trend models (Schindler et al., 2022; Stillman et al., 2023). Considering the fast development of dynamic N-mixture models to allow inferences of demographic rates (Dail & Madsen, 2011), disease structures (DiRenzo et al., 2019), spatial dependency among sites (Howell et al., 2020) or interspecific interactions (Zhao et al., 2022), there are great potentials to extend our model to understand various ecological processes at broad spatiotemporal scales represented by participatory science data.

Our approach is highly generalizable because of the flexibility of the scaling parameter. In particular, we developed the scaling parameter to represent a variety of factors that may influence sampling and observation using a combination of random and fixed effects. Studies have shown that observation errors in participatory science data can be represented by random and fixed effects (Bird et al., 2014; Johnston et al., 2023). For instance, Sicacha-Parada et al. (2021) used a combination of random and fixed effects to address errors related to accessibility in a participatory science data set of moose. Here, we extended these ideas to account for spatiotemporal variations in both biases and errors in participatory science data, allowing the applications of our approach to go beyond the data sets and taxon considered in the current study.

Furthermore, our simulations demonstrated the robustness of the integrated model under a broad range of spatial and temporal overlaps between rigorous survey and participatory science data. For example, the integrated model works well when rigorous survey data are only available in odd years, meaning that, when combined with participatory science data, these data are useful for inferring population processes even if they are not collected in consecutive years. The integrated model also works well when rigorous survey data only cover a small proportion of the environmental gradient. Considering that participatory science data are broadly available across space and time, this means even limited rigorous survey data can be valuable. Lastly, we showed that rigorous survey data do not need to be collected at the same sites as participatory science data to make the integrated model work, which further broadens the application of the integrated model.

Due to the vast variation in life history characteristics (e.g. slow life history, group living) and sampling and observation processes, further investigations are needed to explore additional model structures such as zero-inflation and over-dispersion in the scaling parameter. Recent advances in Bayesian model selection (Hooten & Hobbs, 2015) can play an important role in identifying the optimal structure and predictors of this parameter.

### 4.2 | Importance of both rigorous and participatory science data

Rigorous survey data are indeed key for the success of this integrated model because the scaling parameter is estimable only when rigorous survey data are used. Without rigorous survey data to allow separating observation errors from ecological processes, the scaling parameter will be confounded with the abundance parameter because multiple combinations of abundance and scaling parameter can yield the same count value (Figure S7; also see Dorazio, 2014). While studies have shown the possibility of estimating abundance indices by using only participatory science data in N-mixture models (Goldstein & de Valpine, 2022), it provides inaccurate understanding about population processes when biases and errors are not separated from abundance (Hostetler & Chandler, 2015). Considering the importance of understanding population processes in population ecology and conservation, we highly recommend researchers to collect rigorous survey data.

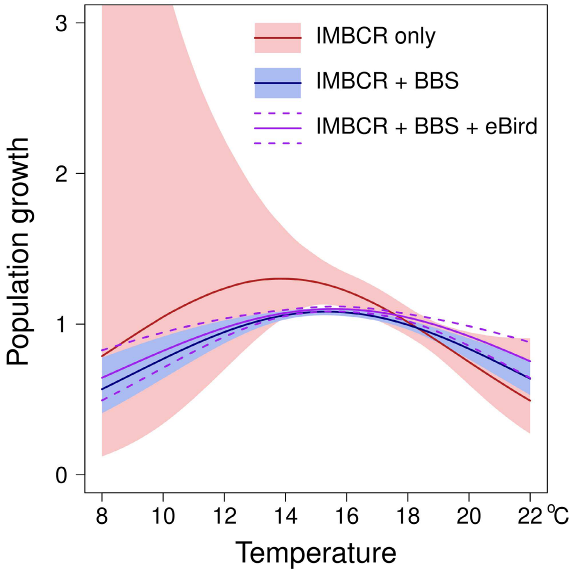


**FIGURE 6** Predicted population growth rate across temperature gradient using parameter estimates from a model that analyses IMBCR data only, a model that jointly analyses IMBCR and BBS data, and a model that jointly analyses IMBCR, BBS and eBird data. BBS and eBird data cover the entire study area and period.

Using a case study, we demonstrated the value of participatory science data in range-wide, long-term inferences of population processes. In particular, the integrated model showed a stronger support for the quadratic effect of temperature on population growth than the model that uses rigorous survey data only, due to the higher precision of the former (Figure 6). This pattern is consistent with theoretical expectation of an optimal temperature in the range of a species (Corkrey et al., 2019), which could be challenging to detect if the inference is geographically limited to a portion of temperature gradient. Taken together, both rigorous survey and participatory science data contain unique values, and jointly analysing them provides a promising direction for leveraging their strength.

## 5 | CONCLUSION

Understanding range-wide, long-term population processes requires data that allow the separation of biases and errors from ecological processes and cover broad spatial and temporal extents, yet, no single data set can satisfy both requirements. The integrated model developed in this study can leverage the strength and overcome the weakness of count data from rigorous surveys and participatory science for the inference of population processes. The flexible structure of this model, in particular the scaling parameter, makes it adaptable to a broad range of ecological systems and survey procedures, and thus highly relevant to population ecology and conservation practice.

### AUTHOR CONTRIBUTIONS

Qing Zhao, Quresh S. Latif. and Christopher E. Latimer conceived the idea, Qing Zhao conducted the analysis and wrote the manuscript, and all coauthors provided critical revision for the manuscript.

### ACKNOWLEDGEMENTS

This work resulted from the Central Grassland Bird Working Group conducted at the National Center for Ecological Analysis and Synthesis at the University of California, Santa Barbara. Two anonymous reviewers provided valuable comments.

### CONFLICT OF INTEREST STATEMENT

The authors have no conflict of interest to declare.

### PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/2041-210X.14368.

### DATA AVAILABILITY STATEMENT

The data and code used in this study are archived in Zenodo (Zhao, 2024; https://zenodo.org/records/11475799).

### ORCID

*Qing Zhao* https://orcid.org/0000-0003-2759-1114
*Quresh S. Latif* https://orcid.org/0000-0003-2925-5042
*T. Brandt Ryder* https://orcid.org/0000-0002-5517-6607

### REFERENCES

Albert, C. H., Yoccoz, N. G., Edwards, T. C., Graham, C. H., Zimmermann, N. E., & Thuiller, W. (2010). Sampling in ecology and evolution—Bridging the gap between theory and practice. *Ecography*, *33*, 1028–1037.

Amundson, C. L., Royle, J. A., & Handel, C. M. (2014). A hierarchical model combining distance sampling and time removal to estimate detection probability during avian point counts. *The Auk*, *131*, 476–494.

Barker, R. J., Schofield, M. R., Link, W. A., & Sauer, J. R. (2018). On the reliability of N-mixture models for count data. *Biometrics*, *74*, 369–377.

Bellier, E., Kéry, M., & Schaub, M. (2016). Simulation-based assessment of dynamic N-mixture models in the presence of density dependence and environmental stochasticity. *Methods in Ecology and Evolution*, *7*, 1029–1040.

Benton, T. G., Plaistow, S. J., & Coulson, T. N. (2006). Complex population dynamics and complex causation: Devils, details and demography. *Proceedings of the Royal Society B: Biological Sciences*, *273*, 1173–1181.

Bird, T. J., Bates, A. E., Lefcheck, J. S., Hill, N. A., Thomson, R. J., Edgar, G. J., Stuart-Smith, R. D., Wotherspoon, S., Krkosek, M., Stuart-Smith, J. F., Pecl, G. T., Barrett, N., & Frusher, S. (2014). Statistical solutions for error and bias in global citizen science datasets. *Biological Conservation*, *173*, 144–154.

Bowler, D. E., Callaghan, C. T., Bhandari, N., Henle, K., Benjamin Barth, M., Koppitz, C., Klenke, R., Winter, M., Jansen, F., Bruelheide, H., & Bonn, A. (2022). Temporal trends in the spatial bias of species occurrence records. *Ecography*, *2022*, e06219.

Bradsworth, N., White, J. G., Isaac, B., & Cooke, R. (2017). Species distribution models derived from citizen science data predict the fine scale movements of owls in an urbanizing landscape. *Biological Conservation*, *213*, 27–35.

Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, *7*, 434–455.

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.

Conn, P. B., Johnson, D. S., Williams, P. J., Melin, S. R., & Hooten, M. B. (2018). A guide to Bayesian model checking for ecologists. *Ecological Monographs*, *88*, 526–542.

Conn, P. B., Ver Hoef, J. M., McClintock, B. T., Johnson, D. S., & Brost, B. (2022). A GLMM approach for combining multiple relative abundance surfaces. *Methods in Ecology and Evolution*, *13*, 2236–2247.

Corkrey, R., Macdonald, C., & McMeekin, T. (2019). The biokinetic spectrum for temperature and optimal Darwinian fitness. *Journal of Theoretical Biology*, *462*, 171–183.

Cox, A. R., Frei, B., Gutowsky, S. E., Baldwin, F. B., Bianchini, K., & Roy, C. (2023). Sixty-years of community-science data suggest earlier fall migration and short-stopping of waterfowl in North America. *Ornithological Applications*, *125*, 1–14.

Dail, D., & Madsen, L. (2011). Models for estimating abundance from repeated counts of an open metapopulation. *Biometrics*, *67*, 577–587.

Dambly, L. I., Isaac, N. J. B., Jones, K. E., Boughey, K. L., & O'Hara, R. B. (2023). Integrated species distribution models fitted in INLA are sensitive to mesh parameterisation. *Ecography*, *2023*, e06391.

Davis, C., Bai, Y., Chen, D., Robinson, O., Ruiz-Gutierrez, V., Gomes, C., & Fink, D. (2023). Deep learning with citizen science data enables estimation of species diversity and composition at continental extents. *Ecology*, *104*, e4175.

Davis, K. L., Saunders, S. P., Beilke, S., Ford, E. R., Fuller, J., Landgraf, A., & Zipkin, E. F. (2023). Breeding season management is unlikely to improve population viability of a data-deficient migratory species in decline. *Biological Conservation*, *283*, 110104.

de Valpine, P., Turek, D., Paciorek, C. J., Anderson-Bergman, C., Lang, D. T., & Bodik, R. (2017). Programming with models: Writing statistical algorithms for general model structures with NIMBLE. *Journal of Computational and Graphical Statistics in Press*, *26*, 403–413.

Di Febbraro, M., Bosso, L., Fasola, M., Santicchia, F., Aloise, G., Lioy, S., Tricarico, E., Ruggieri, L., Bovero, S., Mori, E., & Bertolino, S. (2023). Different facets of the same niche: Integrating citizen science and scientific survey data to predict biological invasion risk under multiple global change drivers. *Global Change Biology*, *29*, 5509–5523.

DiRenzo, G. V., Che-Castaldo, C., Saunders, S. P., Campbell Grant, E. H., & Zipkin, E. F. (2019). Disease-structured N-mixture models: A practical guide to model disease dynamics using count data. *Ecology and Evolution*, *9*, 899–909.

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.

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*, 1040–1049.

Farr, M. T., Green, D. S., Holekamp, K. E., & Zipkin, E. F. (2021). Integrating distance sampling and presence-only data to estimate species abundance. *Ecology*, *102*, e03204.

Fink, D., Johnston, A., Strimas-Mackey, M., Auer, T., Hochachka, W. M., Ligocki, S., Oldham Jaromczyk, L., Robinson, O., Wood, C., Kelling, S., & Rodewald, A. D. (2023). A double machine learning trend model for citizen science data. *Methods in Ecology and Evolution*, *14*, 2435–2448.

Fletcher, R. J., 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.

Friedl, M., & Sulla-Menashe, D. (2022). MODIS/Terra+Aqua land cover type yearly L3 global 500m SIN grid V061 [data set]. NASA EOSDIS Land Processes Distributed Active Archive Center. in press.

Gamelon, M., Grøtan, V., Nilsson, A. L. K., Engen, S., Hurrell, J. W., Jerstad, K., Phillips, A. S., Røstad, O. W., Slagsvold, T., Walseng, B., Stenseth, N. C., & Sæther, B. E. (2017). Interactions between demography and environmental effects are important determinants of population dynamics. *Science Advances*, *3*, e1602298.

Gelfand, A. E., & Shirota, S. (2019). Preferential sampling for presence/absence data and for fusion of presence/absence data with presence-only data. *Ecological Monographs*, *89*, e01372.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis*. Chapman & Hall/CRC.

Goldstein, B. R., & de Valpine, P. (2022). Comparing N-mixture models and GLMMs for relative abundance estimation in a citizen science dataset. *Scientific Reports*, *12*, 12276.

Grüss, A., Charsley, A. R., Thorson, J. T., Anderson, O. F., O'Driscoll, R. L., Wood, B., Breivik, O. N., & O'Leary, C. A. (2023). Integrating survey and observer data improves the predictions of New Zealand spatio-temporal models. *ICES Journal of Marine Science*, *80*, 1991–2007.

Hochachka, W. M., Fink, D., Hutchinson, R. A., Sheldon, D., Wong, W. K., & Kelling, S. (2012). Data-intensive science applied to broad-scale citizen science. *Trends in Ecology & Evolution*, *27*, 130–137.

Hooten, M. B., & Hobbs, N. T. (2015). A guide to Bayesian model selection for ecologists. *Ecological Monographs*, *85*, 3–28.

Horns, J. J., Adler, F. R., & Şekercioğlu, Ç. H. (2018). Using opportunistic citizen science data to estimate avian population trends. *Biological Conservation*, *221*, 151–159.

Hostetler, J. A., & Chandler, R. B. (2015). Improved state-space models for inference about spatial and temporal variation in abundance from count data. *Ecology*, *96*, 1713–1723.

Howell, P. E., Hossack, B. R., Muths, E., Sigafus, B. H., & Chandler, R. B. (2020). Informing amphibian conservation efforts with abundance-based metapopulation models. *Herpetologica*, *76*, 240–250.

Isaac, N. J. B., 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., Lahoz-Monfort, J., Pagel, J., Pescott, O. L., Schmucki, R., Simmonds, E. G., & O'Hara, R. B. (2020). Data integration for large-scale models of species distributions. *Trends in Ecology & Evolution*, *35*, 56–67.

Johnston, A., Matechou, E., & Dennis, E. B. (2023). Outstanding challenges and future directions for biodiversity monitoring using citizen science data. *Methods in Ecology and Evolution*, *14*, 103–116.

Johnston, A., Moran, N., Musgrove, A., Fink, D., & Baillie, S. R. (2020). Estimating species distributions from spatially biased citizen science data. *Ecological Modelling*, *422*, 108927.

Kamp, J., Oppel, S., Heldbjerg, H., Nyegaard, T., & Donald, P. F. (2016). Unstructured citizen science data fail to detect long-term population declines of common birds in Denmark. *Diversity and Distributions*, *22*, 1024–1035.

Knape, J., Arlt, D., Barraquand, F., Berg, Å., Chevalier, M., Pärt, T., Ruete, A., & Żmihorski, M. (2018). Sensitivity of binomial N-mixture models to overdispersion: The importance of assessing model fit. *Methods in Ecology and Evolution*, *9*, 2102–2114.

Koshkina, V., Wang, Y., Gordon, A., Dorazio, R. M., White, M., & Stone, L. (2017). Integrated species distribution models: Combining presence-background data and site-occupany data with imperfect detection. *Methods in Ecology and Evolution*, *8*, 420–430.

Lewis, W. B., Cooper, R. J., Chandler, R. B., Chitwood, R. W., Cline, M. H., Hallworth, M. T., Hatt, J. L., Hepinstall-Cymerman, J., Kaiser, S. A., Rodenhouse, N. L., Sillett, T. S., Stodola, K. W., Webster, M. S., & Holmes, R. T. (2023). Climate-mediated population dynamics of a migratory songbird differ between the trailing edge and range core. *Ecological Monographs*, *93*, e1559.

Link, W. A., Schofield, M. R., Barker, R. J., & Sauer, J. R. (2018). On the robustness of N-mixture models. *Ecology*, *99*, 1547–1551.

Lukyanenko, R., Parsons, J., & Wiersma, Y. F. (2016). Emerging problems of data quality in citizen science. *Conservation Biology*, *30*, 447–449.

Miller, D. A. W., 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.

Pacifici, K., Reich, B. J., Miller, D. A. W., 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.

Pagel, J., Anderson, B. J., O'Hara, R. B., Cramer, W., Fox, R., Jeltsch, F., Roy, D. B., Thomas, C. D., & Schurr, F. M. (2014). Quantifying range-wide variation in population trends from local abundance surveys and widespread opportunistic occurrence records. *Methods in Ecology and Evolution*, *5*, 751–760.

Paniw, M., García-Callejas, D., Lloret, F., Bassar, R. D., Travis, J., & Godoy, O. (2023). Pathways to global-change effects on biodiversity: New opportunities for dynamically forecasting demography and species interactions. *Proceedings of the Royal Society B: Biological Sciences*, *290*, 20221494.

Pavlacky, D. C., Lukacs, P. M., Blakesley, J. A., Skorkowsky, R. C., Klute, D. S., Hahn, B. A., Dreitz, V. J., George, T. L., & Hanni, D. J. (2017). A statistically rigorous sampling design to integrate avian monitoring and management within Bird conservation regions. *PLoS One*, *12*, e0185924.

Phalan, B. T., Northrup, J. M., Yang, Z., Deal, R. L., Rousseau, J. S., Spies, T. A., & Betts, M. G. (2019). Impacts of the northwest forest plan on forest composition and bird populations. *Proceedings of the National Academy of Sciences of the United States of America*, *116*, 3322–3327.

Pollock, K. H. (1991). Review papers: Modeling capture, recapture, and removal statistics for estimation of demographic parameters for fish and wildlife populations: Past, present, and future. *Journal of the American Statistical Association*, *86*, 225–238.

R Development Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. in press.

Reif, J., Vermouzek, Z., Voříšek, P., Romportl, D., & Morelli, F. (2022). Birds' ecological characteristics differ among habitats: An analysis based on national citizen science data. *Community Ecology*, *23*, 173–186.

Robbins, C. S., Bystrak, D., & Geissler, P. H. (1986). *The breeding bird survey: its first fifteen years, 1965-1979*. Resource Publication-US Fish & Wildlife Service 157.

Robinson, O. J., Ruiz-Gutierrez, V., & Fink, D. (2018). Correcting for bias in distribution modelling for rare species using citizen science data. *Diversity and Distributions*, *24*, 460–472.

Robinson, O. J., Ruiz-Gutierrez, V., Reynolds, M. D., Golet, G. H., Strimas-Mackey, M., & Fink, D. (2020). Integrating citizen science data with expert surveys increases accuracy and spatial extent of species distribution models. *Diversity and Distributions*, *26*, 976–986.

Rosenberg, K. V., Dokter, A. M., Blancher, P. J., Sauer, J. R., Smith, A. C., Smith, P. A., Stanton, J. C., Panjabi, A., Helft, L., Parr, M., & Marra, P. P. (2019). Decline of the north American avifauna. *Science (1979)*, *366*, 120–124.

Royle, J. A., Dawson, D. K., & Bates, S. (2004). Modeling abundance effects in distance sampling. *Ecology*, *85*, 1951–1957.

Sauer, J. R., & Link, W. A. (2011). Analysis of the north American breeding Bird survey using hierarchical models. *The Auk*, *128*, 87–98.

Sauer, J. R., Pardieck, K. L., Ziolkowski, D. J., Smith, A. C., Hudson, M. A. R., Rodriguez, V., Berlanga, H., Niven, D. K., & Link, W. A. (2017). The first 50 years of the North American breeding Bird survey. *Condor*, *119*, 576–593.

Schindler, A. R., Cunningham, S. A., Schafer, T. L. J., Sinnott, E. A., Clements, S. J., DiDonato, F. M., Mosloff, A. R., Walters, C. M., Shipley, A. A., Weegman, M. D., & Zhao, Q. (2022). Joint analysis of structured and semi-structured community science data improves precision of relative abundance but not trends in birds. *Scientific Reports*, *12*, 20289.

Seber, G. A. F. (1982). *The estimation of animal abundance and related parameters*. Charles Griffin and Company Limited.

Sharma, S., Andrus, R., Bergeron, Y., Bogdziewicz, M., Bragg, D. C., Brockway, D., Cleavitt, N. L., Courbaud, B., Das, A. J., Dietze, M., Fahey, T. J., Franklin, J. F., Gilbert, G. S., Greenberg, C. H., Guo, Q., Lambers, J. H. R., Ibanez, I., Johnstone, J. F., Kilner, C. L., … Clark, J. S. (2022). North American tree migration paced by climate in the West, lagging in the East. *Proceedings of the National Academy of Sciences of the United States of America*, *119*, e2116691118.

Sicacha-Parada, J., Steinsland, I., Cretois, B., & Borgelt, J. (2021). Accounting for spatial varying sampling effort due to accessibility in Citizen Science data: A case study of moose in Norway. *Spatial Statistics*, *42*, 100446.

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.

Steen, V. A., Tingley, M. W., Paton, P. W. C., & Elphick, C. S. (2021). Spatial thinning and class balancing: Key choices lead to variation in the performance of species distribution models with citizen science data. *Methods in Ecology and Evolution*, *12*, 216–226.

Stevens, D. L., & Olsen, A. R. (2004). Spatially balanced sampling of natural resources. *Journal of the American Statistical Association*, *99*, 262–278.

Stillman, A. N., Howell, P. E., Zimmerman, G. S., Bjerre, E. R., Millsap, B. A., Robinson, O. J., Fink, D., Stuber, E. F., & Ruiz-Gutierrez, V. (2023). Leveraging the strengths of citizen science and structured surveys to achieve scalable inference on population size. *Journal of Applied Ecology*, *60*, 2389–2399.

Strimas-Mackey, M., Hochachka, W. M., Ruiz-Gutierrez, V., Robinson, O. J., Miller, E. T., Auer, T., Kelling, S., Fink, D., & Johnston, A. (2021). *Best practices for using eBird data*. Version 1.0. Cornell Lab of ornithology.

Sullivan, B. L., Aycrigg, J. L., Barry, J. H., Bonney, R. E., Bruns, N., Cooper, C. B., Damoulas, T., Dhondt, A. A., Dietterich, T., Farnsworth, A., Fink, D., Fitzpatrick, J. W., Fredericks, T., Gerbracht, J., Gomes, C., Hochachka, W. M., Iliff, M. J., Lagoze, C., La Sorte, F. A., … Kelling, S. (2014). The eBird enterprise: An integrated approach to development and application of citizen science. *Biological Conservation*, *169*, 31–40.

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.

Thornton, M. M., Shrestha, R., Wei, Y., Thornton, P. E., Kao, S., & Wilson, B. E. (2022). *Daymet: Daily surface weather data on a 1-km grid for North America, version 4 R1*. Oak Ridge National Laboratory Distributed Active Archive Center. in press.

Williams, B. K., & Brown, E. D. (2019). Sampling and analysis frameworks for inference in ecology. *Methods in Ecology and Evolution*, *10*, 1832–1842.

Zhao, Q. (2024). Code and data for "Integrating counts from rigorous surveys and participatory science to better understand spatiotemporal variation in population processes". *Zenodo*. https://doi.org/10.5281/zenodo.11475799

Zhao, Q., Boomer, G. S., & Royle, J. A. (2019). Integrated modeling predicts shifts in waterbird population dynamics under climate change. *Ecography*, *42*, 1470–1481.

Zhao, Q., Devries, J. H., Clark, R. G., & Weegman, M. D. (2023). Causes and consequences of demography in continent-scale, full-annual-cycle population dynamics under global change. *Global Ecology and Conservation*, *43*, e02461.

Zhao, Q., Fuller, A. K., & Royle, J. A. (2022). Spatial dynamic N-mixture models with interspecific interactions. *Methods in Ecology and Evolution*, *13*, 2209–2221.

Zhao, Q., & Royle, J. A. (2019). Dynamic N-mixture models with temporal variability in detection probability. *Ecological Modelling*, *393*, 20–24.

Ziolkowski, D. J., Jr., Lutmerding, M., Aponte, V. I., & Hudson, M. A. R. (2022). North American Breeding Bird Survey dataset 1966–2021: U.S. Geological Survey data release.

Zoellick, B., Nelson, S. J., & Schauffler, M. (2012). Participatory science and education: Bringing both views into focus. *Frontiers in Ecology and the Environment*, *10*, 310–313.

## SUPPORTING INFORMATION

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

**Table S1.** Number of sites covered by rigorous survey (RS), the first participatory science (PS1), and the second participatory science (PS2) when 50%, 25% and 0% of RS sites overlap with PS1 or PS2 sites.

**Table S2.** Median, 95% Credible Interval, and *p*-tail statistics for parameters in a model that analyzes IMBCR data only, a model that jointly analyzes IMBCR and BBS data, and a model that jointly analyzes IMBCR, BBS, and eBird data.

**Figure S1.** Posterior distributions (violin) and key statistics (median: white dot, 50% credible interval: box, 95% credible interval: whiskers) of observation parameters in dynamic N-mixture models that use rigorous survey (RS) data only (red), RS and the first participatory science (PS1) data (blue), or RS, PS1 and the second participatory science (PS2) data (purple) when RS data cover the entire study period of 20 years, the second half of the study period, or odd years of the study period in the simulation study.

**Figure S2.** Posterior distributions (violin) and key statistics (median: white dot, 50% credible interval: box, 95% credible interval: whiskers) of observation parameters in dynamic N-mixture models that use rigorous survey (RS) data only (red), RS and the first participatory science (PS1) data (blue), or RS, PS1 and the second participatory science (PS2) data (purple) when RS data cover the upper 50%, 25% or 10% proportion of the environmental gradient in the simulation study.

**Figure S3.** Posterior distributions (violin) and key statistics (median: white dot, 50% credible interval: box, 95% credible interval: whiskers) of observation parameters in dynamic N-mixture models that use rigorous survey (RS) data only (red), RS and the first participatory science (PS1) data (blue), or RS, PS1 and the second participatory science (PS2) data (purple) when RS sites overlap with 50%, 25% or 0% of the PS1 and PS2 sites in the simulation study.

**Figure S4.** Spatial distributions and temporal variation of grassland coverage in the study area during the study period, 2002–2021.

**Figure S5.** Spatial distributions and temporal variation of temperature in the study area during the study period, 2002–2021.

**Figure S6.** Posterior distributions (violin) and key statistics (median: white dot, 50% credible interval: box, 95% credible interval: whiskers) of process and observation parameters in dynamic N-mixture models that use rigorous survey (RS) data only (red), RS and the first participatory science (PS1) data (blue), or RS, PS1 and the second participatory science (PS2) data (purple) when $\overline{\chi}$ and $\overline{\varpi}$ are less than 1.

**Figure S7.** Scatter plots showing the confounding between the estimates of mean abundance and mean scaling parameters for BBS (left panel) and eBird (right panel) data in a model that uses BBS and eBird, but not IMBCR data.

**How to cite this article:** Zhao, Q., Latif, Q. S., Nuse, B. L., Pavlacky, D. C. Jr., Kilner, C. L., Ryder, T. B., & Latimer, C. E. (2024). Integrating counts from rigorous surveys and participatory science to better understand spatiotemporal variation in population processes. *Methods in Ecology and Evolution*, *00*, 1–14. https://doi.org/10.1111/2041-210X.14368