

# Overcoming data gaps using integrated models to estimate migratory species' dynamics during cryptic periods of the annual cycle

Matthew T. Farr<sup>1</sup>  | Erin R. Zylstra<sup>1,2</sup>  | Leslie Ries<sup>3</sup> | Elise F. Zipkin<sup>1</sup> 

<sup>1</sup>Department of Integrative Biology;  
 Ecology, Evolution, and Behavior Program,  
 Michigan State University, East Lansing,  
 Michigan, USA

<sup>2</sup>Tucson Audubon Society, Tucson,  
 Arizona, USA

<sup>3</sup>Department of Biology, Georgetown  
 University, Washington, District of  
 Columbia, USA

**Correspondence**  
 Matthew T. Farr  
 Email: [farrmat1@msu.edu](mailto:farrmat1@msu.edu)

**Funding information**  
 Division of Environmental Biology, Grant/  
 Award Number: 1702179 and 1702635;  
 U.S. Geological Survey, Grant/Award  
 Number: G21AC10369-00; Division of  
 Biological Infrastructure, Grant/Award  
 Number: 1954406

**Handling Editor:** Nick Isaac

## Abstract

- Environmental and anthropogenic factors affect the population dynamics of migratory species throughout their annual cycles. However, identifying the spatiotemporal drivers of migratory species' abundances is difficult because of extensive gaps in monitoring data. The collection of unstructured opportunistic data by volunteer (citizen science) networks provides a solution to address data gaps for locations and time periods during which structured, design-based data are difficult or impossible to collect.
- To estimate population abundance and distribution at broad spatiotemporal extents, we developed an integrated model that incorporates unstructured data during time periods and spatial locations when structured data are unavailable. We validated our approach through simulations and then applied the framework to the eastern North American migratory population of monarch butterflies during their spring breeding period in eastern Texas. Spring climate conditions have been identified as a key driver of monarch population sizes during subsequent summer and winter periods. However, low monarch densities during the spring combined with very few design-based surveys in the region have limited the ability to isolate effects of spring weather variables on monarchs.
- Simulation results confirmed the ability of our integrated model to accurately and precisely estimate abundance indices and the effects of covariates during locations and time periods in which structured sampling are lacking. In our case study, we combined opportunistic monarch observations during the spring migration and breeding period with structured data from the summer Midwestern breeding grounds. Our model revealed a nonstationary relationship between weather conditions and local monarch abundance during the spring, driven by spatially varying vegetation and temperature conditions.
- Data for widespread and migratory species are often fragmented across multiple monitoring programs, potentially requiring the use of both structured and unstructured data sources to obtain complete geographic coverage. Our integrated

This is an open access article under the terms of the [Creative Commons Attribution](#) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Authors. *Methods in Ecology and Evolution* published by John Wiley & Sons Ltd on behalf of British Ecological Society.

model can estimate population abundance at broad spatiotemporal extents despite structured data gaps during the annual cycle by leveraging opportunistic data.

#### KEY WORDS

data integration, hierarchical modelling, integrated modelling, migratory species, monarch butterfly

## 1 | INTRODUCTION

Migratory species offer many ecosystem services that are highly valued by humans including nutrient cycling, pest control, seed dispersal, recreational opportunities, and food (Green & Elmberg, 2014; Mattsson et al., 2018; Thogmartin et al., 2022). Across taxonomic groups, migratory species have declined and face ongoing threats from myriad factors including climate change and habitat loss (Deinet et al., 2020; Zurell et al., 2018; Zylstra et al., 2022). Migratory species are among the most vulnerable species groups because they traverse large geographic areas throughout their annual cycle, making them susceptible to stressors across wide geographic regions including breeding grounds, nonbreeding grounds and throughout migratory pathways (Saunders et al., 2021; Zylstra et al., 2021). This poses unique challenges for disentangling the effects of various stressors and determining appropriate conservation actions.

Structured data, or data collected during systematic surveys based on consistent protocols, are paramount in identifying ecological processes that lead to declines in migratory species (Lindenmayer et al., 2012). However, structured data often require large investments of time and resources and thus tend to be limited in scope, often capturing only a subset of a species' range. This is especially true for migratory species, in that the bulk of data collection efforts often occur over a limited portion of their annual cycle. For instance, for North American birds, most monitoring occurs on breeding grounds when species are comparatively stationary, with less data collected on non-breeding grounds and during periods of migration (Marra et al., 2015; Rushing et al., 2016). Scaling up systematic monitoring efforts from well sampled areas to a species' full range during its annual cycle is challenging (but see Şekercioğlu, 2012; Suman et al., 2023), leaving spatiotemporal gaps in data, and hampering current analyses of migratory species ecology (Marra et al., 2015). As a result, knowledge of how environmental factors affect the abundance and distribution of migratory species is incomplete, especially during cryptic and hard-to-study periods of the annual cycle.

Collections of unstructured, opportunistic data via volunteer-based networks (e.g. iNaturalist, eBird) are providing a means to address geographic and temporal gaps in currently available structured data (Chandler et al., 2017). Unstructured data are those that are collected without a planned survey design or monitoring protocol, such as an opportunistic sighting of a given species. Though unstructured data can be plentiful, producing unbiased and meaningful inferences from these data can be challenging (Fithian et al., 2015).

Spatial sampling biases can lead to a disproportionate number of observations near urban areas (e.g. roads, city parks), where people are concentrated. Additionally, observers typically record when a species is present but not when it is absent, inhibiting estimation of occurrence probabilities or abundance indices when exclusively using presence-only data (Dorazio, 2014; Farr et al., 2021). Consequently, the value of unstructured data for population-level inferences is more limited.

Integrated modelling, in which multiple data types are analysed in a single, unified framework, can be a useful tool to address limitations encountered when using structured (i.e. limited extent and quantity) or unstructured (i.e. limited information content) data to model dynamics of animal populations (Fletcher et al., 2019; Isaac et al., 2020; Miller et al., 2019; Schindler et al., 2022; Zipkin et al., 2021; Zipkin & Saunders, 2018). For migratory species, integrated modelling can resolve discrepancies in estimated population trends from separate periods of the annual cycle by linking seasonal dynamics (Saunders, Farr, et al., 2019). Integrated models have also been used to evaluate potential environmental drivers of population trends across a species' full annual cycle (Rushing et al., 2017; Rushing et al., 2021; Zylstra et al., 2021). However, an approach has yet to be developed that formally incorporates unstructured data to explore population dynamics during time periods and spatial domains when structured data are unavailable.

Here, we develop an integrated model to estimate spatiotemporal population abundance for a migratory species when structured data gaps exist within the annual cycle. Underlying our integrated framework is a Poisson point process model that describes variation in abundance using relevant covariates and spatial random effects (Dorazio, 2014; Fithian et al., 2015). We integrate structured count data and unstructured presence-only observations using a joint-likelihood, which shares parameters between data types within the point process model (Fletcher et al., 2019). The joint-likelihood provides a pathway to lend information on abundance from structured data to periods of the annual cycle when only unstructured data are available. We demonstrate the validity of this framework using a simulation study and then apply our model to a case study of monarch butterflies in eastern North America.

Our case study was motivated by our work on the effects of factors influencing adult monarch butterfly abundances during the spring breeding season when structured data are unavailable (Oberhauser et al., 2015). The eastern population of monarch butterflies in North America has declined by as much as 80% over the

last 25 years (Brower et al., 2012; Thogmartin et al., 2017). The population undertakes a spectacular four-generation annual migration from their wintering grounds in central Mexico, through the spring breeding corridor in eastern Texas and Oklahoma, and then up to the summer breeding grounds in the northern half of the U.S. and southern Canada east of the Rocky Mountains before the last generation returns to the wintering grounds in Mexico (Brower et al., 2012). Analyses of recent data from the Midwestern summer breeding areas reveal that climate conditions during the spring and summer breeding seasons were more than seven times as important than other factors (e.g. size of the population at the end of winter, crop cover, and glyphosate application rates in the Midwest) in determining the peak summer population size between 2004 and 2018 (Zylstra et al., 2021). Despite the importance of the spring migratory and breeding area for eastern monarchs, few structured surveys have been conducted in this area. Thus, we lack direct evidence of the relationship between weather variables and the spatiotemporal dynamics of monarchs throughout the spring breeding season. We use our integrated model to estimate monarch abundance and associated effects of environmental conditions during their spring breeding period, allowing detailed inferences throughout this important, yet understudied and under sampled, period of the monarch's migratory cycle.

## 2 | MATERIALS AND METHODS

### 2.1 | Model description

The purpose of our integrated model is to generate spatially explicit estimates of population abundance within one or more seasons (hereafter, periods) of a species' annual migratory cycle when structured data are unavailable (Figure 1a,b). We define abundance,  $N_t$ , during periods of the annual cycle,  $t = 1, \dots, T$ , as the number of individuals in a corresponding spatial domain  $S_t$  (e.g. breeding range, non-breeding range, migratory pathway) occupied by the species during period  $t$ . Abundance is assumed to be a random variable arising from a Poisson process,  $N_t \sim \text{Pois}(\Lambda_t)$ , with expected abundance,  $\Lambda_t$ . We specify the spatial distribution of individuals within domain  $S_t$  during period  $t$  using a point process model. Individuals of the population are distributed within the domain according to an intensity function,  $\lambda_t(s)$ , which describes the number of individuals at location  $s$  ( $s \in S_t$ ) during period  $t$ . We link abundance to the point process by specifying expected abundance as the integral of the intensity function across the domain  $S_t$ :  $\Lambda_t = \int_{S_t} \lambda_t(s) ds$ .

#### 2.1.1 | Spatial heterogeneity in the distribution of individuals

Local abundance often changes as a function of one or more environmental factors. As these factors likely vary within the spatial domain,  $S_t$ , abundance is heterogeneous, and thus can be referred

to as an inhomogeneous point process. We begin by modelling local abundance (i.e. the intensity function),  $\lambda_t(s)$ , during a period of the annual cycle where both structured and unstructured data exist. In this description, we assume for simplicity that this occurs at  $t = 1$ . We model  $\lambda_{t=1}(s)$  with spatial covariates using an inhomogeneous point process:

$$\log(\lambda_{t=1}(s)) = \log(\lambda_0) + \beta' \mathbf{X}_{t=1}(s) + \omega(s), \quad (1)$$

where  $\mathbf{X}_{t=1}(s)$  is a vector of environmental covariates at location  $s$  during period,  $t = 1$ .  $\beta'$  is the vector of corresponding coefficients, and  $\lambda_0$  is the expected number of individuals at location  $s$  when covariates  $\mathbf{X}$  are at their mean values (standardized to mean of zero and standard deviation of 1). Both  $\lambda_{t=1}(s)$  and  $\lambda_0$  are estimated on the log scale as their support ranges from 0 to  $\infty$  while the support of covariates is from  $-\infty$  to  $\infty$ . We account for spatial variation not explained by environmental covariates with a spatial random effect,  $\omega(s)$ , using a Gaussian random field (i.e. the log-Gaussian Cox process; Møller et al., 1998). We implemented the Gaussian random field using a stochastic partial differential equation approximation via a triangulated spatial mesh (which may vary by time period) of  $k = 1, \dots, K_t$  nodes across  $S_t$  (Kraainski et al., 2018; Simpson et al., 2016; Supporting Information S1).

#### 2.1.2 | Seasonal change in abundance

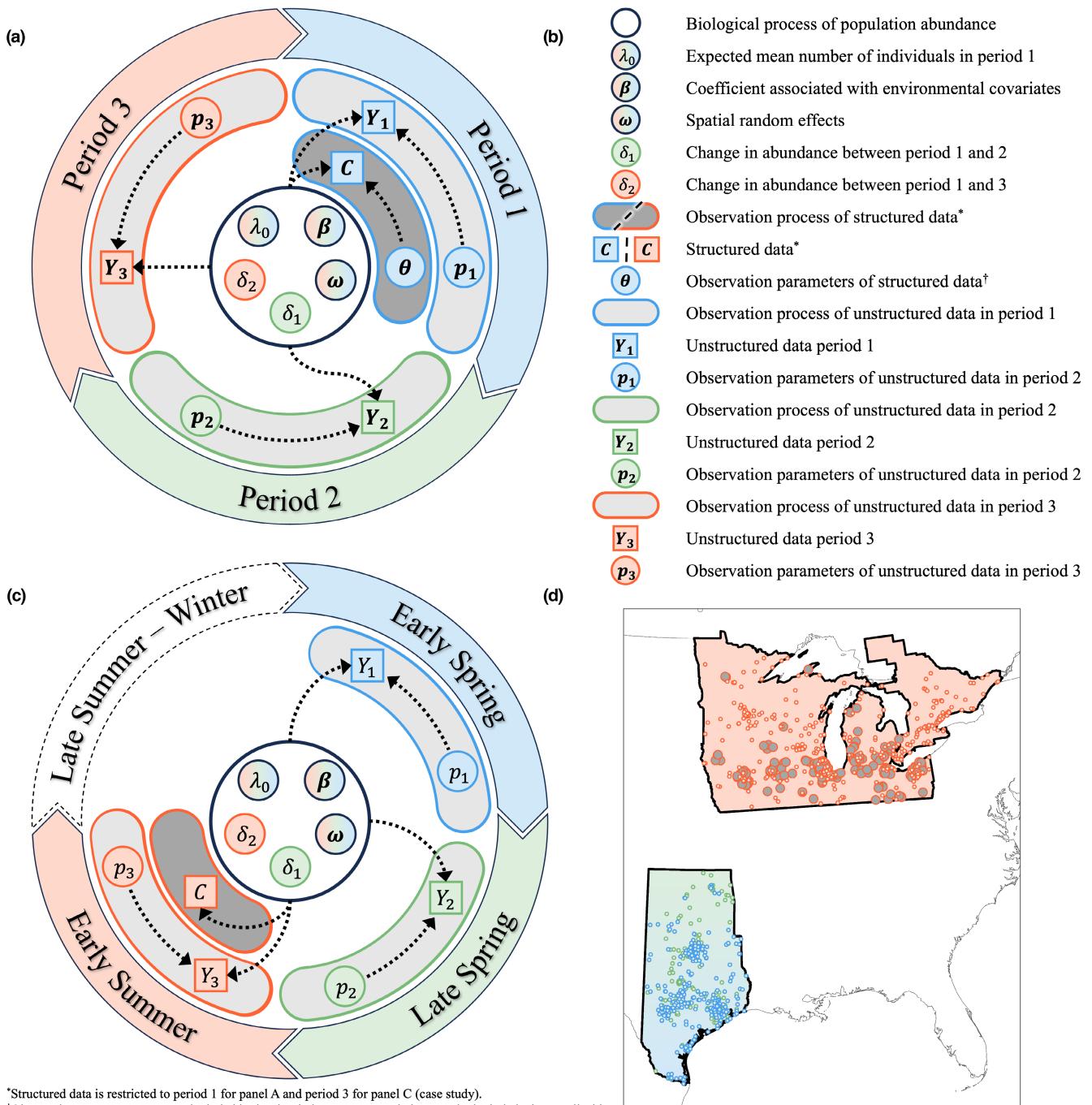
During subsequent periods of the annual cycle,  $t = 2, \dots, T$ , abundance may change because of demographic processes (i.e. mortality, recruitment, immigration, emigration). Various methods exist to describe population dynamics (Kéry & Royle, 2020), but here we simply describe changes in abundance between periods as a function of spatiotemporal variation in environmental factors along with a fixed effect for each subsequent period:

$$\log(\lambda_t(s)) = \log(\lambda_0) + \beta' \mathbf{X}_t(s) + \delta_t + \omega(s). \quad (2)$$

$\mathbf{X}_t(s)$  is a vector of environmental covariates for each period after the first,  $t = 2, \dots, T$ . Changes in abundance not captured by covariates are represented by  $\delta_t$ , which is the difference in the log of mean abundance between time periods  $t - 1$  and  $t$ .

#### 2.1.3 | Unstructured presence-only data

Presence-only data are typically the outcome of opportunistic, unstructured sampling where observations of the target species are recorded, whereas the absence of the species at a given location is not recorded. We modelled presence-only data resulting from opportunistic sampling using a thinned Poisson process,  $Y_t \sim \text{Pois}\left(\int_{S_t} \lambda_t(s)p_t(s)ds\right)$ , where  $Y_t$  is the number of presence-only observations during period  $t$  within spatial domain  $S_t$  (Dorazio, 2014). The thinning rate function,  $p_t(s)$ , represents the observation process that relates expected abundance at location  $s$ ,  $\lambda_t(s)$ , to the presence-only data (i.e. the number of observations at location  $s$ ). Here,  $p_t(s)$ , is the probability that an individual was



\*Structured data is restricted to period 1 for panel A and period 3 for panel C (case study).

†Observation parameters were not included in the simulation nor case study but may be included when applicable.

**FIGURE 1** An integrated modelling framework allows estimation of population abundance across space and time for species with spatiotemporal gaps in structured data. (a) Relationships between data and parameters for a general description of our integrated modelling framework are shown with a directed acyclic graph (DAG). The DAG also displays data availability across a species annual cycle (three spatiotemporal periods differentiated by colour [period 1, blue; period 2, green; period 3, orange]). The outer circle displays time while oblong shapes within the outer circle depict each data type (light grey shading indicates unstructured data; dark grey shading indicates structured data). The border colours of the oblong shapes match the period in which data were collected. The innermost circle represents the biological process for latent abundance that is shared between all data types. Data (squares) and parameters (circles) are shown with dashed lines displaying the flow of information. The border and shading of data and parameter objects indicate the period with which they are associated. Parameters in the abundance model with black borders and rainbow shading were estimated across periods. (b) Descriptions of parameters and data for each DAG (a, c) and the case study map (d) are shown within a legend. (c) Modifications of the integrated model for the case study (see footnotes) are shown within a separate DAG. Unstructured (i.e. presence-only) data exist in early and late spring (blue and green shading) along with early summer (orange shading). But, structured (i.e. single-visit counts) data are restricted to the early summer. Our analysis does not include data from late summer through the winter portion of monarch's annual cycle. (d) Location and timing of observation processes for each data type (unstructured data, smaller circles; structured data larger circles; colours match legend descriptions) are displayed within a map.

detected and is constrained to be between 0 and 1 where we assume double counting is trivial compared to imperfect detection. We account for spatiotemporal variation in imperfect detection and sampling bias by specifying a linear model with a logit link for  $p_t(s)$ , with an intercept ( $p_0$ ), covariates ( $\mathbf{W}_t(s)$ ) and their effects ( $\alpha$ ) (Dorazio, 2012).

#### 2.1.4 | Structured count data

We assume the structured count data,  $C_{tj}$ , occur at a set of sites (i.e. subsets of the spatial domain)  $j = 1, \dots, J_t$ , each with an associated area  $D_j$  where  $D_{1,\dots,J_t} \subset S_t$ . Count data are also described with a Poisson process:  $C_{tj} \sim \text{Pois}(\int_{D_j} \lambda_t(s) ds)$ . Count data, like many structured data types, cannot be used as direct measures of true abundance due to observation error. When counts are replicated or additional information is recorded, an observation process model can be included to account for imperfect detection. We use a single-visit count model that does not account for imperfect detection. Thus, we were only able to estimate relative abundance (assuming that detection does not vary over time and space), similar to many large-scale structured monitoring programs (e.g. Christmas Bird Count). Because counts of individuals are aggregated within a site (i.e. observations are not associated with a point location,  $s_i$ , for each individual  $i$ ), a change-of-support problem (i.e. mismatches in spatial scales of multiple data sources) must be addressed (Pacifici et al., 2019). Given that the expected sum of Poisson random variables is equivalent to the sum of their means, we specify an area offset and approximate the integral over each site  $j$  as  $D_j \bar{\lambda}_{tj} \approx \int_{D_j} \lambda_t(s) ds$ . This assumes a homogeneous process within each site (i.e.  $\bar{\lambda}_{tj}$  is the mean intensity for  $D_j$ ) and a linear relationship between area and the number of individuals (i.e. no density dependence). Counts at each period  $t$  and site  $j$  can be modelled as:  $C_{tj} \sim \text{Pois}(D_j \bar{\lambda}_{tj})$ , where  $\log(\bar{\lambda}_{tj}) = \log(\lambda_0) + \beta' \mathbf{X}_{tj} + \delta_t + \omega_j$ . Covariates,  $\mathbf{X}_{tj}$ , are summarized across each  $D_j$ , and a projection matrix,  $A_{kj}$ , is needed to interpolate the random effects,  $\omega(s)$ , estimated at each node  $k$  in the spatial mesh to each site  $j$  ( $\omega_j = \omega_k \cdot A_{kj}$ ; Krainski et al., 2018).

#### 2.1.5 | Spatiotemporal data integration

Within our modelling framework, structured data provide an anchor point for estimating the intercept of the intensity function,  $\lambda_0$ , which is unidentifiable in a standalone presence-only data analysis (Dorazio, 2014; Farr et al., 2021). We assume that structured data exist during one or more periods within a species' annual cycle but are not available during every period, whereas unstructured data are available during periods with and without structured data. Structured data are used to estimate  $\lambda_0$ , which also appears in the model for unstructured data. This allows us to estimate spatially explicit abundance during periods when structured data are unavailable. A critical assumption of this framework is that changes in abundance between time periods are explained by covariates or captured as deviations

from abundance in  $t = 1$  (i.e. through a time period effect,  $\delta_t$ ). To create the integrated model, a joint-likelihood is formed using the product of the likelihoods of the structured (e.g. count) and unstructured (e.g. presence-only) data (Supporting Information S1).

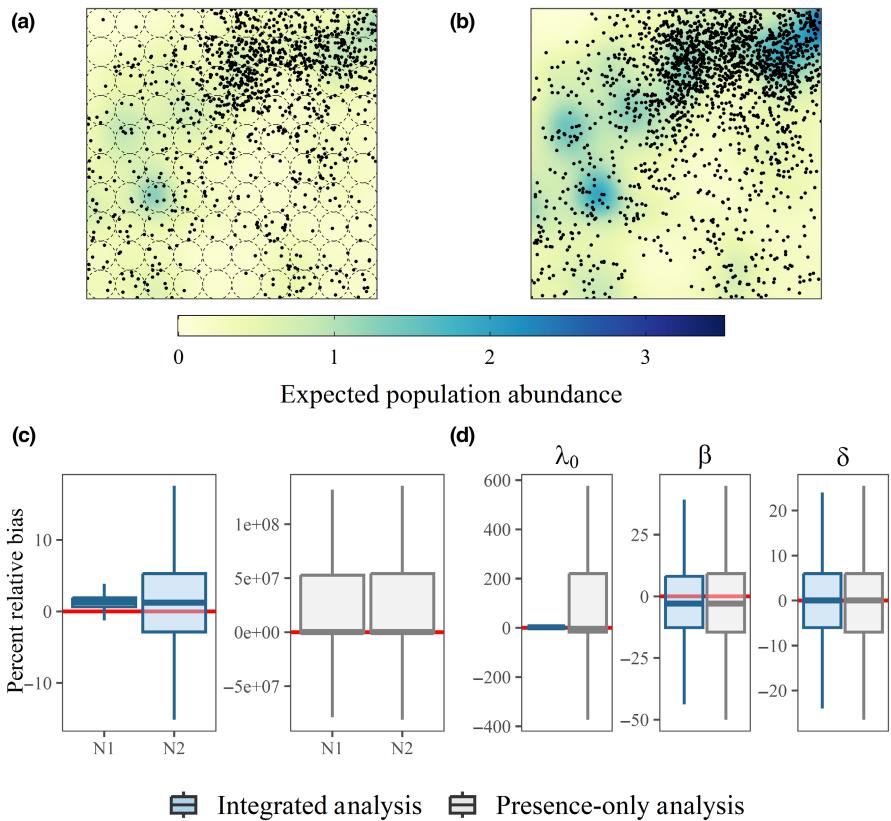
#### 2.2 | Simulation study

The objective of our simulation study was to assess whether the integrated model, which relies on a shared intensity function (Equation 2) between structured and unstructured data collected over multiple time periods, can produce unbiased estimates of unknown parameters. We evaluated our simulation study's objective by measuring the accuracy and precision of estimated parameters, including the expected number of individuals,  $\lambda_0$ , and the change in the log of mean abundance between time periods,  $\delta_t$ . We simulated abundance in two periods within a spatial domain,  $S$  (Figure 2a,b). Both unstructured and structured data were available during the first period, while only unstructured data were available during the second period. We simulated presence-only observations for the unstructured data and single-visit counts for the structured data.

We specified  $\lambda_0$  and  $\delta$  to generate abundances near 3600 and 9900 individuals within  $S$  for periods 1 and 2, respectively (Figure 2a,b). To create a spatially varying landscape, we specified an inhomogeneous point process via a single environmental covariate that varied over space and between periods,  $X_t(s)$ , with unmeasured spatial heterogeneity,  $\omega(s)$ . We specified a single covariate effect  $\beta$  to establish a relationship between the environmental covariate and local abundance (i.e. intensity) that was constant between periods (Figure 1a,b; Supporting Information S2). The spatial random effect,  $\omega(s)$ , was simulated using a Gaussian random field with precision ( $\tau$ ) and scale ( $\kappa$ ) hyperparameters (Supporting Information S1 and S2). We then simulated the true number of individuals at locations for each season using a Poisson point process and the corresponding intensity function.

We generated presence-only data for both periods (Figure 1a,b) where the thinning process incorporated effect  $\alpha$  and a single covariate that did not vary with period,  $W(s)$ , representing sampling bias (simulated from a Gaussian random field). We standardized the covariate by subtracting the minimum value and dividing by the mean, which forced areas with low sampling intensity (i.e. lower covariate values) towards the intercept,  $p_0$ . We set  $p_0 = \text{logit}(0.01)$ ; thus, locations at the lower end of the covariate range represent low or no sampling effort (i.e.  $p_t \approx 0$ ). We generated single-visit count data at 100 sites (i.e. circles with uniform areas) for the first period only (Figure 2a), where detections of individuals in a site were summed to generate a single count. For simplicity, we specified that all individuals present were detected during each survey (i.e. counts were perfect).

We simulated both true abundance and observed datasets 1000 times, and for each simulation, estimated population abundance to evaluate our ability to leverage structured data over space and time using the integrated model. The majority of fixed



**FIGURE 2** Visualization of our simulation study showing latent abundance and the amount of both structured and unstructured data within domain  $S$  for (a) period 1 and (b) period 2. The background gradient indicates the expected abundance at location  $s$  (reported as individuals per unit area). (a) Period 1 contains both presence-only data (black dots) as well as 100 sites with single-visit count data (dashed circles). (b) Period 2 exclusively contains presence-only data (black dots). (c, d) Percent relative bias associated with estimates from the integrated and presence-only models (red horizontal lines at zero indicate no bias in estimated parameters; positive values are overestimates; negative values are underestimates; boxes represent the interquartile range of estimates from 1000 simulations; centre lines are the median value; whiskers are values within 1.5 times the interquartile range). (c) Percent relative bias in estimates of abundance for period 1 (N1) and period 2 (N2) using the integrated model (left) and in an analysis of just the presence-only data (right). (d) Percent relative bias in estimates of the intensity function intercept ( $\lambda_0$ ), covariate effect ( $\beta$ ), and change in abundance between periods ( $\delta$ ) as estimated using the integrated model (blue) and in an analysis of the presence-only data (grey).

effect parameters ( $\lambda_0$ ,  $\delta$ ,  $\kappa$ ,  $\tau$ ,  $p_0$ ,  $\alpha$ ), were constant across simulations. We allowed the environmental effect parameter,  $\beta$ , to vary between 0.5 and 1.5 on the log-scale to evaluate the accuracy and precision of estimates under a range of conditions. The covariates  $X_t(s)$  and  $W(s)$ , the spatial random effects,  $\omega(s)$ , and individual animal locations were redrawn for each simulation (see [Supporting Information S2](#)). Parameters were estimated using maximum likelihood with R-INLA, Template Model Builder, and R (Kristensen et al., 2016; Lindgren & Rue, 2015; R Core Team, 2020). To assess convergence (i.e. negative log likelihood at a minimum), we checked that the absolute values of the final gradient for each parameter were near 0 and checked that the Hessian matrix was positive definite (Skaug & Fournier, 2006). To assess performance, we calculated percent relative bias for the mean estimate of each parameter:  $\left(\frac{\text{estimated mean} - \text{true}}{\text{true}}\right) \times 100$ . We also derived estimates of abundance and associated uncertainty to compare with true abundance. Finally, we estimated abundance for each of the 1000 simulated datasets using only the presence-only data to compare with estimates from our integrated framework. We did not estimate

parameters using only the count data as there was no structured sampling during the second period.

### 2.3 | Case study

The eastern North American migratory population of monarch butterflies has a multi-generational migratory cycle. Adult monarchs overwinter (December–February) in large clusters blanketing patches of high-elevation Oyamel fir (*Abies religiosa*) forests in central Mexico. In early spring (March–April), monarchs migrate north and arrive on the breeding grounds in and around eastern Texas, where they lay eggs, producing the first new generation of the year (Brower et al., 2012). This first generation continues the northerly migration to the summer breeding grounds, with most adults concentrated across the Midwestern and northeastern U.S. and southeastern Canada. About three more generations are produced on the summer breeding grounds (May–August); the final generation of the year enters reproductive diapause and commences the southerly

migration in autumn (September–October) to the overwintering grounds in Mexico (arriving by mid-December). Monitoring of the migration varies dramatically across the annual cycle (Oberhauser et al., 2015). Structured monitoring occurs annually on the overwintering grounds in central Mexico (where researchers delineate the number of hectares occupied by monarch aggregations) and throughout the Midwestern summer breeding grounds (where volunteer observers count butterflies on established plots or transects). However, structured monitoring has only recently begun during their spring breeding and autumn migration. Further compounding monitoring challenges, the population is at its smallest size during the spring phase of migration and generally occurs at low densities (Ries & Oberhauser, 2015).

Multiple hypotheses have been proposed to explain the decline of monarchs in eastern North America (Thogmartin et al., 2017). Recent retrospective analyses demonstrate the importance of weather conditions during the spring breeding season, highlighting a critical period of population growth (Saunders et al., 2016, 2018). The peak size of the monarch population on the summer breeding grounds, which is highly correlated with the size of the population that ultimately returns to Mexico each winter, is strongly influenced by weather conditions on the spring breeding grounds (Zylstra et al., 2021). Spring temperature and precipitation affect the recruitment of monarchs by impacting development and survival as well as host plant resources (Zalucki, 1982), but the importance of these and other factors are only understood indirectly through analyses of count data collected on the summer breeding grounds (Saunders et al., 2018; Zipkin et al., 2012).

With spring conditions potentially becoming less favourable for monarch recruitment (Neupane et al., 2022; Zylstra et al., 2022), a deeper understanding of the direct links between local weather conditions and monarch abundance can aid management and policy decisions for eastern monarchs, a population of substantial conservation concern. Fortunately, multiple volunteer data-collection networks compile opportunistic observations of adult monarchs in the spring and early summer. Using our integrated framework, we leveraged these opportunistic data to estimate population abundance of monarchs during spring breeding and arrival on the summer breeding grounds. We also assessed the effects of weather and greenness (i.e. a proxy for host and nectar resources) on monarch abundance during this critical period of the migratory cycle.

### 2.3.1 | Data

Our analysis focuses on spring migrations from 2016 to 2018. Within a given year, we partitioned spring migration into three distinct spatiotemporal periods. We indexed years with  $r$  ( $R = 3$ ) and used  $S_t$  ( $T = 3$ ) to denote each spatiotemporal period during spring migration in a given year (Figure 1d). The spatiotemporal periods of the spring migration were constant across years and defined similarly to previous studies (Saunders, Ries, et al., 2019; Zylstra et al., 2021). Early spring,  $S_1$ , encompassed eastern Texas and Oklahoma (93.5°W

to 100°W, 25.8°N to 37°N) between 8 March and 4 April and was specified to include adult monarchs arriving from Mexico as they lay eggs that become the first generation of the year. Late spring,  $S_2$ , covered the same spatial extent but occurred between 19 April and 2 May, which primarily included sightings of adults that are part of the first generation produced that year. Early summer,  $S_3$ , captured these adults shortly after they arrive on the summer breeding grounds, between 3 May and 6 June. We used data from an area that spans the Midwestern U.S. and Ontario, Canada (74.3°W to 97.2°W, 39.8°N to 49.4°N). Although monarchs disperse outside of this area during summer, we focused on the Midwest because the majority of individuals that arrive on the overwintering grounds in Mexico originate from this region (Flockhart et al., 2017). This early summer period is also important as it contains both structured and unstructured data, whereas we did not have access to structured data during  $S_1$  and  $S_2$ .

We used opportunistic, presence-only data from five data collection networks: North American Butterfly Association's "Butterflies I've Seen" and "Recent Sightings" programs, iNaturalist, eButterfly, Butterflies and Moths of North America, and Journey North (iNaturalist Community, 2019; Lotts & Naberhaus, 2019; Oberhauser et al., 2015; Prudic et al., 2017; Sheehan & Weber-Grullon, 2021). Each of these networks allow members of the general public to report butterfly sightings (e.g. date, location, species, count) to online databases that are available publicly or through data requests. From each program, we extracted monarch sightings (location and date) during our study period and converted them to presence-only data (denoted as a presence at a specific location and date if  $\geq 1$  adult monarch was observed). The average number of observations reported each year by period was:  $n_{S_1} = 196$ ,  $n_{S_2} = 157$ , and  $n_{S_3} = 166$ . Prior to modelling, we pooled unstructured data into a single dataset rather than modelling each source with a separate submodel as protocols and data types were similar.

The structured monarch data collected during early summer (3 May–6 June) come from five different monitoring programs. The first program we used was the North American Butterfly Association's Fourth of July (NFJ) surveys. These structured surveys are conducted annually in summer, with optimal timing depending on local conditions at specified sites throughout the summer breeding range. Volunteers survey areas within a fixed 25-km diameter circle and record the number of adult butterflies observed, by species (Oberhauser et al., 2015). We summed monarch observations among volunteers during each survey, resulting in a single count for each site and year  $C_{ij}$ . The other four monitoring programs were 'Pollard walk' surveys from state-level butterfly monitoring networks (BMNs) in Illinois, Ohio, Iowa, and Michigan (Oberhauser et al., 2015; Pollard, 1977). Volunteers walked fixed transects and counted the number of adult butterflies observed. Unlike NFJ surveys, volunteers surveyed Pollard transects multiple times each year; however, we only used the first survey of the year because our focus was estimating monarch abundance during early summer. Data from the NFJ and Pollard transects thus contain a single count for each site and year. The average number of

counts each year in  $S_3$  was:  $n_{S_3} = 229$ . Similar to the unstructured data, we pooled the structured data into a single dataset rather than modelling each with a separate submodel as similar protocols led to a common data type.

We included environmental covariates that are likely to influence monarch recruitment and survival, and hence spatiotemporal abundance. Monarch larvae are obligate feeders on milkweed (*Asclepias* spp.; Pleasants & Oberhauser, 2013); thus, milkweed availability is likely to drive monarch recruitment. Because data on milkweed distribution and abundance are lacking, we used the normalized difference vegetation index (NDVI), a measure of landscape greenness, as a proxy for milkweed distribution during the spring and early summer (Flockhart et al., 2013; Lemoine, 2015). Weather variables are also vital to the recruitment of monarchs as they influence the rates of development and survival from eggs to adults (Zalucki, 1982). Similar to other monarch models, we used the number of growing degree days (GDD) to describe thermal conditions on the spring and summer breeding grounds (Saunders et al., 2016, 2018; Zipkin et al., 2012; Zylstra et al., 2021). GDD is the accumulation of heat within a specific temperature range that allows for monarch development (McMaster & Wilhelm, 1997). We calculated GDD values as the heat accumulated at a given location over a 14-day period immediately preceding each observation or survey. NDVI and GDD variables were calculated for each observation, survey, and mesh node (Didan, 2015; Thornton et al., 2020, Supporting Information S3).

We also included environmental covariates in our model that were likely to influence the number of monarchs observed during sampling. For the presence-only data, we used two covariates in the thinning function to account for spatiotemporal variation in sampling intensity. First, we used observations of non-monarch butterflies (i.e. number of detections for each species) within the superfamily *Papilioidea* in iNaturalist (iNaturalist Community, 2019), assuming high numbers of butterfly observations correlate with high sampling intensity. We used human population density as another proxy for sampling intensity with the assumption that the number of people in the surrounding area is positively correlated with the number of local observers (Center for International Earth Science Information Network-CIESIN-Columbia University, 2018; Geldmann et al., 2016; Supporting Information S3). For the structured data, the amount of effort expended searching for monarchs on each survey is likely to affect the number of individuals observed. Because the area searched during each structured survey (e.g. length of each BMN transect or area covered during each NFJ survey) was not reported, we converted the number of person hours spent surveying into area sampled based on the average human walking speed (5-km/hour; Browning et al., 2006) and assumed that butterflies within 1-m of the observer were detected.

### 2.3.2 | Analysis

We used the integrated modelling framework described above to estimate abundance of monarchs in spring and early summer in

each year from 2016 to 2018 with a few case-specific modifications (Figure 1b,c). We selected a 100-m<sup>2</sup> baseline resolution for presence-only data and converted the area offset ( $D$ ) accordingly. In addition to the fixed effects describing population change between spatiotemporal domains within a single annual cycle ( $\delta_t$ ), we included a fixed effect to capture population change between years ( $\gamma_r$ ). We allowed the effects of covariates to vary by period and included quadratic effects for both NDVI and GDD to account for peaks in optimal conditions. Effects were fixed across years (i.e. no interaction between covariates and years). Thus, the updated equation for abundance (modified from Equation 2) is:

$$\log(\lambda_{tr}(s)) = \log(\lambda_0) + \beta_{1,t} \cdot \text{NDVI}_{tr}(s) + \beta_{2,t} \cdot \text{NDVI}_{tr}(s)^2 + \beta_{3,t} \cdot \text{GDD}_{tr}(s) + \beta_{4,t} \cdot \text{GDD}_{tr}(s)^2 + \delta_t + \gamma_r + \omega(s) \quad (3)$$

We report abundance as the expected number of monarchs per 100-m<sup>2</sup>. Parameters were estimated using maximum likelihood with R-INLA, Template Model Builder, and R (Kristensen et al., 2016; Lindgren & Rue, 2015; R Core Team, 2020). See Supporting Information S3 for additional implementation details.

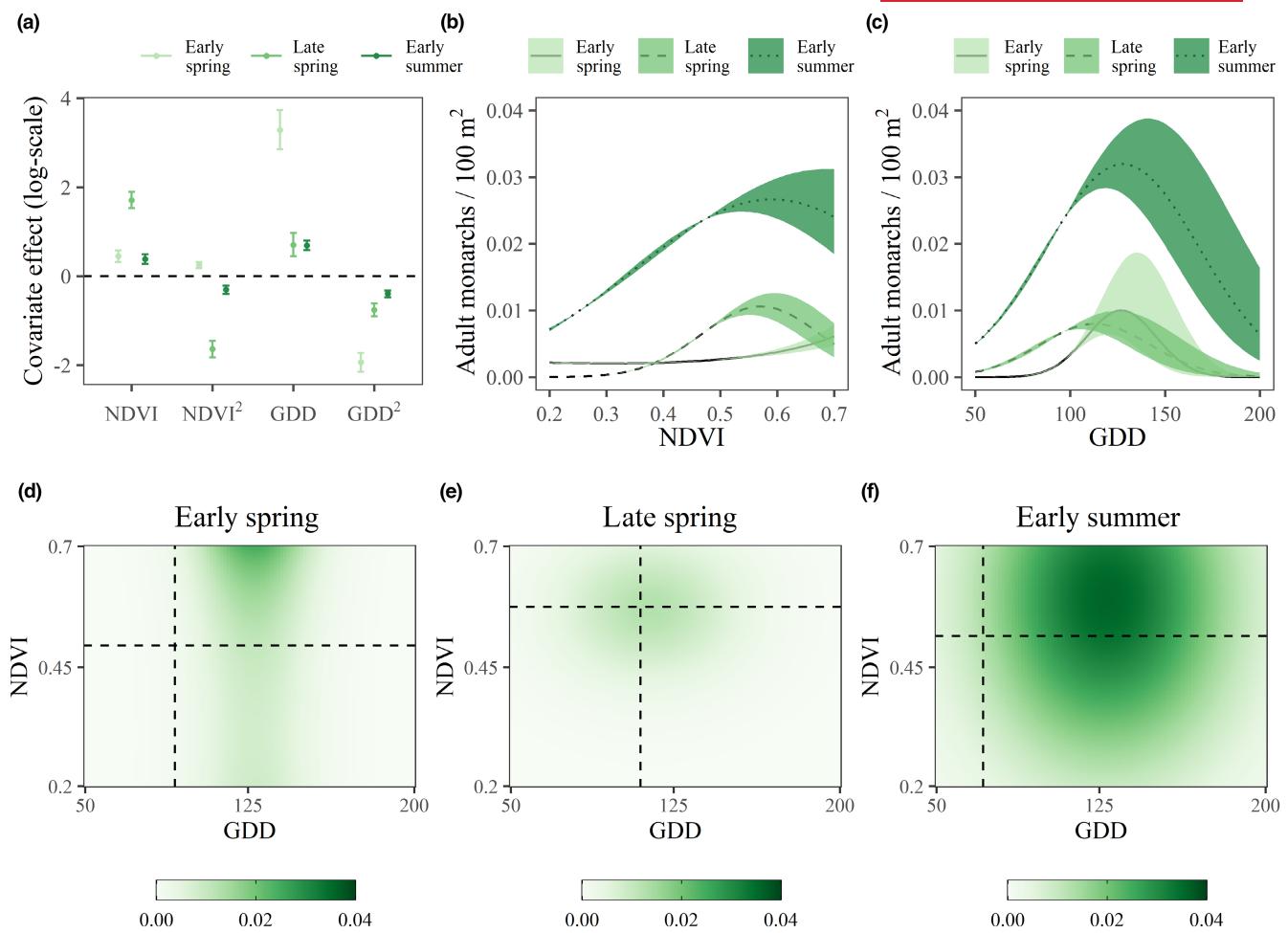
## 3 | RESULTS

### 3.1 | Simulation study

The integration of structured data from the first time period with unstructured data from both periods allowed for nearly unbiased estimates of abundance (Figure 2c, left), even in the second period during which no structured data were available (1.3% and 1.2% relative biases in periods 1 and 2, respectively). In contrast, abundance could not be estimated when relying exclusively on either presence-only data (Figure 2c, right; Dorazio, 2014) or structured data that were limited to a single time period. Precision of estimates was also higher for the integrated model as compared to a presence-only analysis (Supporting Information S2: Table S2.1). Our integrated modelling framework successfully identified the intercept of the intensity function,  $\lambda_0$  (1.7% relative bias), effects of an environmental covariate on local abundance,  $\beta$  (-3.0% relative bias; simulated values between 0.5 and 1.5 on the log-scale), and the change in abundance between periods,  $\delta$  (0.0% relative bias; Figure 2d; see Supporting Information S2: Table S2.1 for full results). The thinning parameters,  $p_0$  and  $\alpha$ , were also estimated with low bias (-0.4% and -1.2% relative bias, respectively). Hyperparameters of the spatial random effect,  $\tau$  and  $\kappa$ , had larger biases across simulations (-37.0% for  $\tau$  and 14.8% for  $\kappa$ ), which may be a result of the generic spatial mesh used for analyses (Dambly et al., 2023; Jullum, 2020).

### 3.2 | Case study

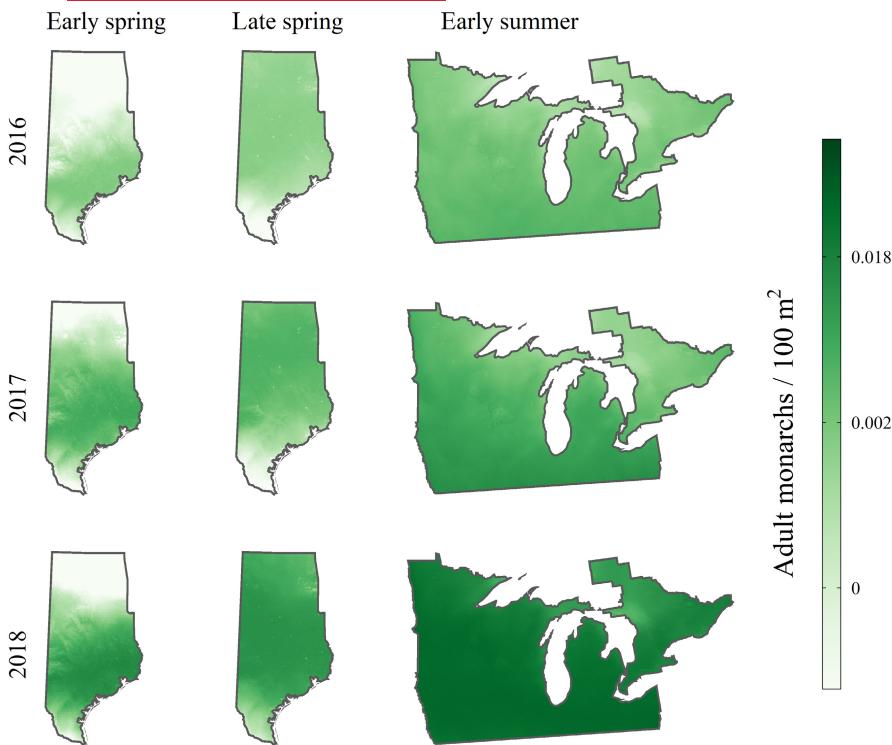
Environmental conditions in the eastern Texas region during the spring breeding season, particularly in March, strongly influenced the overall size of the monarch population and the distribution of



**FIGURE 3** Effects of NDVI and GDD on monarch abundance (reported as the number of adults/100 m<sup>2</sup>). (a) Parameter estimates (mean and 95% confidence intervals) for the linear and quadratic effects of NDVI and GDD during each period: early spring, late spring, and early summer. Dashed line at zero indicates no covariate effect on monarch abundance. (b, c) Estimated marginal effects (mean and 95% confidence intervals) of NDVI and GDD, respectively, on monarch population abundance during each period. Panels in the bottom row depict the marginal effect (mean) of both NDVI and GDD on monarch abundance in (d) early spring, (e) late spring, and (f) early summer. Dashed lines indicate the mean values of NDVI and GDD summarized across 2016–2018 for each period's spatial domain (at 100-km<sup>2</sup> resolution) and date range.

individuals across the breeding area (Figure 3a). We identified non-linear effects of both NDVI and GDD during all three time periods (i.e. early spring, late spring, and early summer; Figure 2b,c). The relationships between environmental variables and monarch abundance changed throughout the spring and early summer breeding seasons (Figure 3d–f). Highest monarch abundances in the early spring and summer were associated with above average NDVI and temperature values (i.e. 100-km<sup>2</sup> resolution values averaged across 2016–2018 for each period's domain; Figure 3d,f), while the highest abundances in late spring were associated with average NDVI and temperature values (Figure 3e). Estimates for the GDD linear and quadratic effects (Figure 3a,c,d) reveal that peak monarch abundance in the early spring breeding period was constrained to a tight range of values (especially compared to summer; Figure 3d versus Figure 3f), suggesting that a precise range of cumulative temperature conditions at the start of the annual breeding season may be necessary for large population growth.

Using the integrated model, we estimated spatially explicit monarch abundances (reported as adults per 100 m<sup>2</sup>) across the spring and summer breeding areas (Figure 4), providing the first such estimates for eastern monarchs in spring, a time period when only unstructured, opportunistic data are available. In each year, monarch abundance increased over the course of the spring breeding season, with the lowest expected population sizes in early spring and the highest in early summer (Figure 4). These results were expected based on monarch life history patterns that reflect the ongoing breeding of subsequent generations from the beginning of spring until the end of the summer breeding period. The annual size of the eastern monarch population peaks in late summer just before the final generation enters diapause and initiates the southerly migration (Ries et al., 2015). Yearly estimates of spring and early summer population indices are consistent with estimates of overwintering population size in Mexico later that same year, with considerably higher abundance in 2018 than in the



**FIGURE 4** Predicted abundance of monarchs (reported as the number of adults/100 m<sup>2</sup>) from 2016 to 2018 during the spring in eastern Texas and Oklahoma (left two columns) and during the early summer breeding season in the midwestern U.S. and southern Ontario, Canada (right column). Abundance was calculated using parameter estimates and values of NDVI and GDD averaged across each period: early spring (March 8–April 4), late spring (April 19–May 2), and early summer (May 3–June 6) at a given location and year. Rows and columns correspond to years and periods, respectively.

previous 2 years ([Supporting Information S3: Table S3.1](#)). These results further highlight the importance of the spring breeding season and suggest that annual variations in the overall size of the monarch population can be seen as early as the first generation of the year ([Figure 4](#)).

## 4 | DISCUSSION

We developed an integrated model using a Poisson point process to estimate seasonal abundances of migratory species with unstructured, opportunistic data when structured data are sparse or unavailable during portions of the annual cycle. Integration of unstructured data with information-rich structured data can improve inferences on migratory populations relative to independent analyses of either data source ([Figure 2](#)). Analyses relying exclusively on structured data may result in inferences that are limited to certain periods of a species annual cycle (e.g. breeding), while analyses using just presence-only data cannot estimate absolute abundance due to unidentifiable parameters. By addressing data gaps and parameter identifiability through integrated modeling, our framework can be used to estimate relationships between environmental factors and population abundance and elucidate dynamics that occur across multiple periods of a species' migratory or annual cycle.

We validated our model through a simulation study where we demonstrated that abundance or relative abundance (i.e. when imperfect detection is unmeasured or ignored) can be estimated for time periods in which structured data are unavailable. The simulations also confirmed that our modelling framework can yield

unbiased estimates of overall abundance and the effects of spatially varying covariates. However, estimates of spatial hyperparameters were somewhat biased, suggesting uncertainty in the model's ability to parse spatial heterogeneity beyond that explained by covariates. We used a non-customized mesh across simulations, which likely contributed to biases. For specific applications, such as our monarch case study ([Supporting Information S3](#)), tuning the spatial mesh to the system at hand can minimize such biases (Dambly et al., [2023](#); Jullum, [2020](#); Krainski et al., [2018](#)). Krainski et al. ([2018](#)) provides in-depth instruction on how to build triangulated meshes using the R-INLA package while Dambly et al. ([2023](#)) documents tradeoffs encountered when specifying a spatial mesh.

We demonstrated the merit of our model by generating spatiotemporal estimates of relative monarch abundance during their spring breeding season when structured data are largely unavailable and densities are low. The spring breeding season has been identified as a critical period for the eastern monarch population, as evidenced by consistent and strong links between spring weather conditions and the abundance of monarchs in late summer and early winter (Saunders et al., [2016, 2018](#); Zylstra et al., [2021](#)). Our model corroborates these results from long-term analyses, despite the use of different data sources and a different spatiotemporal resolution than has been used in previous work. Further, our results revealed nonstationary effects of environmental variables on monarch abundances during the spring and early summer (Rollinson et al., [2021](#)). We observed a shift in the effects of NDVI (a metric of landscape greenness) on monarch abundance, where peak abundances were associated with the greenest conditions in early spring and summer and slightly lower NDVI values in late spring ([Figure 3b](#)). Although there was some

variation, the highest monarch abundances were associated with cumulative 14-day GDD values between 100 and 150 throughout the spring and early summer.

As demonstrated in the simulation and case studies, our integrated model offers a powerful approach for estimating abundance during periods and spatial locations that lack design-based monitoring data. However, we recommend that potential users evaluate the appropriateness of our framework for their systems prior to implementation. For example, some taxa may violate the assumption of a shared intensity function across space and time because they contain distinct subpopulations with unique demographic processes in different parts of their range. The amount and type of data available could also render data integration futile if low amounts of structured data preclude the estimation of parameters or if the amount of structured data is sufficient to estimate parameters across a species full annual cycle without the use of unstructured data. Additionally, a critical assumption of our integrated framework is that the relationship between unstructured data and species abundance is constant over space and time (i.e. parameters in [Equation 3](#) are stationary), with the Poisson thinning rate accurately capturing variation in the observation process of unstructured data. Violations of this assumption may lead to increased biases as estimation of abundance moves further away in space and time from where structured data were used to estimate  $\lambda_0$ .

We advise that practitioners develop their own simulations to assess the feasibility of this type of integrated modelling, as case specific considerations may require further model complexity (see either [Supporting Information S2](#) and [S3](#), Zenodo [DOI: <https://doi.org/10.5281/zenodo.8433370>], or GitHub [[https://github.com/zipkinlab/Farr\\_et al\\_2024\\_MEE](https://github.com/zipkinlab/Farr_et al_2024_MEE)] for code used in our analyses). Observation processes, such as imperfect detection of individuals during structured data collection, may reduce parameter accuracy and precision if not properly accounted for within the model structure. Uncertainty in covariate values along with potential correlations between environmental and observation covariates should be evaluated to minimize errors and collinearity ([Dorazio, 2014](#)). Practitioners should also consider if large quantities of unstructured data in an area are likely to reflect high quality habitat (i.e. environmental covariate) or simply spatial biases from variable data collection effort (i.e. observation covariate). Finally, improper covariate selection may lead to parameter biases ([Kéry & Royle, 2020](#)), and for many species and systems, relevant observation covariates may be hard to identify or measure ([Kelling et al., 2019](#); [Troudet et al., 2017](#)).

Nearly one sixth of examined species are classified as data deficient by the IUCN, and far more have little to no structured data available throughout much of their range ([IUCN, 2022](#)). Migratory species, in particular, often lack data across large portions of their geographic range despite being vulnerable to wide ranging threats ([Marra et al., 2015](#)). A shift in biodiversity monitoring from structured to unstructured data collection protocols is providing new sources of information to mitigate data gaps for migratory species. However, new analytical approaches, such as integrated modelling,

are needed to overcome observation biases in unstructured data and realize the potential of this emerging data type to help address and mitigate ongoing species declines.

## AUTHOR CONTRIBUTIONS

Matthew T. Farr, Erin R. Zylstra, Leslie Ries and Elise F. Zipkin conceived of the research. Leslie Ries contributed data. Matthew T. Farr, Erin R. Zylstra, and Elise F. Zipkin constructed the model. Matthew T. Farr ran analyses. Matthew T. Farr, Erin R. Zylstra and Elise F. Zipkin wrote the first drafts of the paper. All authors contributed critically to the drafts and gave final approval for publication.

## ACKNOWLEDGEMENTS

We thank the many public scientists and volunteers who contributed to data collection as this research would not be possible without their efforts. The North American Butterfly Association data were provided by Jeffrey Glassberg, Journey North data by Elizabeth Howard, eButterfly data by Max Larrivee, and Butterflies and Moths of North America data by Thomas Naberhaus. We also thank the Pollard program directors for data from Illinois (Doug Taron), Iowa (Nathan Brockman), Michigan (Ronda Spink), and Ohio (Jerry Wiedmann). This work was supported by NSF grant nos. DBI-1954406 (E.F.Z.), DEB-1702635 (E.F.Z.), and DEB-1702179 (L.R.). This work was also supported by grant no. G21AC10369-00 from the United States Geological Survey (USGS). Its contents are solely the responsibility of the authors and do not necessarily represent the views of the Midwest Climate Adaptation Science Center or the USGS. This manuscript is submitted for publication with the understanding that the United States Government is authorized to reproduce and distribute reprints for Governmental purposes.

## CONFLICT OF INTEREST STATEMENT

None of the authors have a conflict of interest.

## 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.14282>.

## DATA AVAILABILITY STATEMENT

Monarch data from iNaturalist and covariate data are available on either Zenodo (DOI: <https://doi.org/10.5281/zenodo.8433370>) or GitHub ([https://github.com/zipkinlab/Farr\\_et al\\_2024\\_MEE](https://github.com/zipkinlab/Farr_et al_2024_MEE)) along with code for both simulation and case studies. Monarch data from the other monitoring programs are proprietary and were obtained from the North American Butterfly Association (<https://www.naba.org/>), Journey North (<https://journeynorth.org/>), eButterfly (<https://www.e-butterfly.org/>), Butterflies and Moths of North America (<https://www.butterfliesandmoths.org/>), the Illinois Butterfly Monitoring Network (<https://ibmn.org/>), the Iowa Butterfly Survey Network (<https://reimangardens.com/insects#IBSN>), the Michigan Butterfly Network (<http://naturecenter.org/michiganbutterfly/>), and the Ohio Lepidopterists (<http://>

[www.ohiolepidopterists.org/](http://www.ohiolepidopterists.org/)). These data may be available upon reasonable request from the authors (LR) and with permission from the aforementioned programs.

## ORCID

Matthew T. Farr  <https://orcid.org/0000-0003-1011-6851>

Erin R. Zylstra  <https://orcid.org/0000-0002-2536-0403>

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

## REFERENCES

- Brower, L. P., Taylor, O. R., Williams, E. H., Slayback, D. A., Zubieta, R. R., & Ramírez, M. I. (2012). Decline of monarch butterflies overwintering in Mexico: Is the migratory phenomenon at risk? *Insect Conservation and Diversity*, 5, 95–100.
- Browning, R. C., Baker, E. A., Herron, J. A., & Kram, R. (2006). Effects of obesity and sex on the energetic cost and preferred speed of walking. *Journal of Applied Physiology*, 100, 390–398.
- Center for International Earth Science Information Network-CIESIN-Columbia University. (2018). *Gridded population of the world, version 4 (GPWv4): Population density, revision 11*. NASA Socioeconomic Data and Applications Center (SEDAC). <https://doi.org/10.7927/H49C6VHW>
- Chandler, M., See, L., Copas, K., Bonde, A. M. Z., López, B. C., Danielsen, F., Legind, J. K., Masinde, S., Miller-Rushing, A. J., Newman, G., Roesmartin, A., & Turak, E. (2017). Contribution of citizen science towards international biodiversity monitoring. *Biological Conservation*, 213, 280–294.
- 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 parameterization. *Ecography*, 2023, e06391.
- Deinet, S., Scott-Gatty, K., Rotton, H., Twardek, W. M., Marconi, V., McRae, L., Baumgartner, L. J., Brink, K., Claussen, J. E., Cooke, S. J., Darwall, W., Eriksson, B. K., Garcia de Leaniz, C., Hogan, Z., Royte, J., Silva, L. G. M., Thieme, M. L., Tickner, D., Waldman, J., ... Berkhuysen, A. (2020). *The living planet index (LPI) for migratory freshwater fish—Technical report*. World Fish Migration Foundation.
- Didan, K. (2015). MOD13A2 MODIS/Terra vegetation indices 16-day L3 global 1 km SIN grid V006 [Data set]. NASA EOSDIS Land Processes DAAC <https://doi.org/10.5067/MODIS/MOD13A2.006>
- Dorazio, R. M. (2012). Predicting the geographic distribution of species from presence-only data subject to detection errors. *Biometrics*, 68, 1302–1312.
- 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.
- 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.
- Fithian, W., Elith, J., Hastie, T., & Keith, D. A. (2015). Bias correction in species distribution models: Pooling survey and collection data for multiple species. *Methods in Ecology and Evolution*, 6, 424–438.
- 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.
- Flockhart, D. T. T., Brower, L. P., Ramirez, M. I., Hobson, K. A., Wassenaar, L. I., Altizer, S., & Norris, D. R. (2017). Regional climate on the breeding grounds predicts variation in the natal origin of monarch butterflies overwintering in Mexico over 38 years. *Global Change Biology*, 23, 2565–2576.
- Flockhart, D. T. T., Wassenaar, L. I., Martin, T. G., Hobson, K. A., Wunder, M. B., & Norris, D. R. (2013). Tracking multi-generational colonization of the breeding grounds by monarch butterflies in eastern North America. *Proceedings of the Royal Society B: Biological Sciences*, 280, 20131087.
- Geldmann, J., Heilmann-Clausen, J., Holm, T. E., Levinsky, I., Markuseen, B., Olsen, K., Rahbek, C., & Tøttrup, A. P. (2016). What determines spatial bias in citizen science? Exploring four recording schemes with different proficiency requirements. *Diversity and Distributions*, 22, 1139–1149.
- Green, A. J., & Elmberg, J. (2014). Ecosystem services provided by waterbirds. *Biological Reviews*, 89, 105–122.
- iNaturalist Community. (2019). Observations of Papilioidea from North America observed between 2016–2018. Exported from <https://www.inaturalist.org> on 2019-06-12.
- 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., Prescott, 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.
- IUCN. (2022). The IUCN red list of threatened species. Version 2022-2. <https://www.iucnredlist.org>
- Jullum, M. (2020). Investigating mesh-based approximation methods for normalization constant in the log Gaussian Cox process likelihood. *Stat*, 9, e285.
- 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 semistructured surveys to improve citizen science data for monitoring biodiversity. *Bioscience*, 69, 170–179.
- Kéry, M., & Royle, J. A. (2020). *Applied hierarchical modeling in ecology: Analysis of distribution, abundance and species richness in R and BUGS: Volume 2: Dynamic and advanced models*. Academic Press.
- Krainski, E. T., Gómez-Rubio, V., Bakka, H., Lenzi, A., Castro-Camilo, D., Simpson, D., Lindgren, F., & Rue, H. (2018). *Advanced spatial modeling with stochastic partial differential equations using R and INLA*. Chapman and Hall.
- Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., & Bell, B. M. (2016). TMB: Automatic differentiation and Laplace approximation. *Journal of Statistical Software*, 70, 1–21.
- Lemoine, N. P. (2015). Climate change may alter breeding ground distributions of eastern migratory monarchs via range expansion of Asclepias host plants. *PLoS One*, 10, e0118614.
- Lindenmayer, D. B., Likens, G. E., Andersen, A., Bowman, D., Bull, C. M., Burns, E., Dickman, C. R., Hoffmann, A. A., Keith, D. A., Liddell, M. J., Lowe, A. J., Metcalfe, D. J., Phinn, S. R., Russell-Smith, J., Thurgate, N., & Wardle, G. M. (2012). Value of long-term ecological studies. *Animal Ecology*, 37, 745–757.
- Lindgren, F., & Rue, H. (2015). Bayesian spatial modelling with R-INLA. *Journal of Statistical Software*, 63, 1–25.
- Lotts, K., & Naberhaus, T. (2019). Butterflies and moths of North America. Data Set. <http://www.butterfliesandmoths.org/>
- Marra, P. P., Cohen, E. B., Loss, S. R., Rutter, J. E., & Tonra, C. M. (2015). A call for full annual cycle research in animal ecology. *Biology Letters*, 11, 20150552.
- Mattsson, B. J., Dubovsky, J. A., Thogmartin, W. E., Bagstad, K. J., Goldstein, J. H., Loomis, J. B., Diffendorfer, J. E., Semmens, D. J., Wiederholt, R., & López-Hoffman, L. (2018). Recreation economics to inform migratory species conservation: Case study of the northern pintail. *Journal of Environmental Management*, 206, 971–979.
- McMaster, G. S., & Wilhelm, W. W. (1997). Growing degree-days: One equation, two interpretations. *Agricultural and Forest Meteorology*, 87, 291–300.
- 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.

- Møller, J., Syversveen, A. R., & Waagepetersen, R. P. (1998). Log Gaussian cox processes. *Scandinavian Journal of Statistics*, 25, 451–482.
- Neupane, N., Zipkin, E. F., Saunders, S. P., & Ries, L. (2022). Grappling with uncertainty in ecological projections: A case study using the migratory monarch butterfly. *Ecosphere*, 13(1), e03874.
- Oberhauser, K. S., Ries, L., Altizer, S., Batalden, R. V., Kudell-Ekstrum, J., Garland, M., Howard, E., Jepsen, S., Lovett, J., Monroe, M., Morris, G., Rendón-Salinas, E., RuBino, R. G., Ryan, A., Taylor, O. R., Treviño, R., Villalblanca, F. X., & Walton, D. (2015). Contributions to monarch biology and conservation through citizen science. In K. S. Oberhauser, K. S. Nail, & S. Altizer (Eds.), *Monarchs in a changing world: Biology and conservation of an iconic butterfly* (pp. 13–30). Cornell University Press.
- Pacifici, K., Reich, B. J., Miller, D. A. W., & Pease, B. S. (2019). Resolving misaligned spatiotemporal data with integrated distribution models. *Ecology*, 100, e02709.
- Pleasants, J. M., & Oberhauser, K. S. (2013). Milkweed loss in agricultural fields because of herbicide use: Effect on the monarch butterfly population. *Insect Conservation and Diversity*, 6, 135–144.
- Pollard, E. (1977). A method for assessing changes in the abundance of butterflies. *Biological Conservation*, 12, 115–134.
- Pradic, K. L., McFarland, K. P., Oliver, J. C., Hutchinson, R. A., Long, E. C., Kerr, J. T., & Larrivée, M. (2017). eButterfly: Leveraging massive online citizen science for butterfly conservation. *Insects*, 8, 53.
- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Ries, L., & Oberhauser, K. S. (2015). A citizen army for science: Quantifying the contributions of citizen scientists to our understanding of monarch butterfly biology. *Bioscience*, 65, 419–430.
- Ries, L., Taron, D. J., & Rendón-Salinas, E. (2015). The disconnect between summer and winter monarch trends for the eastern migratory population: Possible links to differing drivers. *Annals of the Entomological Society of America*, 108, 691–699.
- Rollinson, C. R., Finley, A. O., Alexander, M. R., Banerjee, S., Hamil, K. D., Koenig, L. E., Locke, D. H., DeMarche, M. L., Tingley, M. W., Wheeler, K., Youngflesh, C., & Zipkin, E. F. (2021). Working across space and time: Nonstationarity in ecological research and application. *Macrosystems Biology*, 19, 66–72.
- Rushing, C. S., Hostetler, J. A., Sillett, T. S., Marra, P. P., Rotenberg, J. A., & Ryder, T. B. (2017). Spatial and temporal drivers of avian population dynamics across the annual cycle. *Ecology*, 98, 2837–2850.
- Rushing, C. S., Ryder, T. B., & Marra, P. P. (2016). Quantifying drivers of population dynamics for a migratory bird throughout the annual cycle. *Proceedings of the Royal Society B: Biological Sciences*, 283, 20152846.
- Rushing, C. S., Tatenhove, A. M. V., Sharp, A., Ruiz-Gutierrez, V., Freeman, M. C., Sykes, P. W., Given, A. M., & Sillett, T. S. (2021). Integrating tracking and resight data enables unbiased inferences about migratory connectivity and winter range survival from archival tags. *Ornithological Applications*, 123, 1–14.
- 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 and deficiencies with integrated population models. *Ecology*, 100, e02714.
- Saunders, S. P., Piper, W., Farr, M. T., Bateman, B. L., Michel, N. L., Westerkam, H., & Wilsey, C. B. (2021). Interrelated impacts of climate and land-use change on a widespread waterbird. *Journal of Animal Ecology*, 90, 1165–1176.
- Saunders, S. P., Ries, L., Neupane, N., Ramírez, M. I., García-Serrano, E., Rendón-Salinas, E., & Zipkin, E. F. (2019). Multiscale 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.
- Saunders, S. P., Ries, L., Oberhauser, K. S., Thogmartin, W. E., & Zipkin, E. F. (2018). Local and cross-seasonal associations of climate and land use with abundance of monarch butterflies. *Ecography*, 41, 278–290.
- Saunders, S. P., Ries, L., Oberhauser, K. S., & Zipkin, E. F. (2016). Evaluating confidence in climate-based predictions of population change in a migratory species. *Global Ecology and Biogeography*, 25, 1000–1012.
- 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.
- Sekercioğlu, Ç. H. (2012). Promoting community-based bird monitoring in the tropics: Conservation, research, environmental education, capacity-building, and local incomes. *Biological Conservation*, 151, 69–73.
- Sheehan, N., & Weber-Grullon, L. (2021). Journey north-monarch butterfly and milkweed observations by volunteer community scientists across central and North America (1996–2020) ver 1. Environmental Data Initiative <https://doi.org/10.6073/pasta/f7d7bef57f94b33b8a18a26954252412>
- Simpson, D., Illian, J. B., Lindgren, F., Sørbye, S. H., & Rue, H. (2016). Going off grid: Computationally efficient inference for log-Gaussian cox processes. *Biometrika*, 103, 49–70.
- Skaug, H. J., & Fournier, D. A. (2006). Automatic approximation of the marginal likelihood in non-Gaussian hierarchical models. *Computational Statistics and Data Analysis*, 51, 699–709.
- Suman, A., Ravikanthachari, N., & Kunte, K. (2023). A comparison between time-constrained counts and line transects as methods to estimate butterfly diversity and monitor populations in tropical habitats. *Insect Conservation and Diversity*. <https://doi.org/10.1111/icad.12693>
- Thogmartin, W. E., Haefele, M. A., Diffendorfer, J. E., Semmens, D. J., Derbride, J. J., Lien, A., Huang, T., & López-Hoffman, L. (2022). Multi-species, multi-county analysis reveals North Americans are willing to pay for transborder migratory species conservation. *People and Nature*, 4, 549–562.
- Thogmartin, W. E., Wiederholt, R., Oberhauser, K. S., Drum, R. G., Diffendorfer, J. E., Altizer, S., Taylor, O. R., Pleasants, J., Semmens, D., Semmens, B., Erickson, R., Libby, K., & López-Hoffman, L. (2017). Monarch butterfly population decline in North America: Identifying the threatening processes. *Royal Society Open Science*, 4, 170760.
- Thornton, M. M., Shrestha, R., Wei, Y., Thornton, P. E., Kao, S., & Wilson, B. E. (2020). Daymet: Daily surface weather data on a 1-km grid for North America, version 4. ORNL DAAC. <https://doi.org/10.3334/ORNLDAAC/1840>
- Troudet, J., Grandcolas, P., Blin, A., Vignes-Lebbe, R., & Legendre, F. (2017). Taxonomic bias in biodiversity data and societal preferences. *Scientific Reports*, 7, 9132.
- Zalucki, M. P. (1982). Temperature and rate of development in *Danaus plexippus* L. and *D. chrysippus* L. (Lepidoptera: Nymphalidae). *Journal of the Australian Entomological Society*, 21, 241–246.
- Zipkin, E. F., Ries, L., Reeves, R., Regetz, J., & Oberhauser, K. S. (2012). Tracking climate impacts on the migratory monarch butterfly. *Global Change Biology*, 18, 3039–3049.
- Zipkin, E. F., & Saunders, S. P. (2018). Synthesizing multiple data types for biological conservation using integrated population models. *Biological Conservation*, 217, 240–250.
- Zipkin, E. F., Zylstra, E. R., Wright, A. D., Saunders, S. P., Finely, A. O., Dietze, M. C., Itter, M. S., & Tingley, M. S. (2021). Addressing data integration challenges to link ecological processes across scales. *Frontiers in Ecology and the Environment*, 19, 30–38.
- Zurell, D., Graham, C. H., Galien, L., Thuiller, W., & Zimmermann, N. E. (2018). Long-distance migratory birds threatened by multiple independent risks from global change. *Nature Climate Change*, 8, 992–996.

Zylstra, E. R., Neupane, N., & Zipkin, E. F. (2022). Multi-season climate projections forecast declines in migratory monarch butterflies. *Global Change Biology*, 28, 6135–6151.

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 & Evolution*, 5, 1441–1452. <https://doi.org/10.1038/s41559-021-01504-1>

## SUPPORTING INFORMATION

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

**Supplementary S1.** Additional modeling details.

**Supplementary S2.** Simulation study code and output.

**Supplementary S3.** Case study details, code, and output.

**How to cite this article:** Farr, M. T., Zylstra, E. R., Ries, L., & Zipkin, E. F. (2024). Overcoming data gaps using integrated models to estimate migratory species' dynamics during cryptic periods of the annual cycle. *Methods in Ecology and Evolution*, 00, 1–14. <https://doi.org/10.1111/2041-210X.14282>