Joint Spatial Modeling Bridges the Gap Between Disparate Disease Surveillance and Population Monitoring Efforts Informing Conservation of At-risk Bat Species

Christian STRATTON, Kathryn M. IRVINE, Katharine M. BANNER, Emily S. ALMBERG, Dan BACHEN, and Kristina SMUCKER

White-Nose Syndrome (WNS) is a wildlife disease that has decimated hibernating bats since its introduction in North America in 2006. As the disease spreads westward, assessing the potentially differential impact of the disease on western bat species is an urgent conservation need. The statistical challenge is that the disease surveillance and species response monitoring data are not co-located, available at different spatial resolutions, non-Gaussian, and subject to observation error requiring a novel extension to spatially misaligned regression models for analysis. Previous work motivated by epidemiology applications has proposed two-step approaches that overcome the spatial misalignment while intentionally preventing the human health outcome from informing estimation of exposure. In our application, the impacted animals contribute to spreading the fungus that causes WNS, motivating development of a joint framework that exploits the known biological relationship. We introduce a Bayesian, joint spatial modeling framework that provides inferences about the impact of WNS on measures of relative bat activity and accounts for the uncertainty in estimation of WNS presence at non-surveyed locations. Our simulations demonstrate that the joint model produced more precise estimates of disease occurrence and unbiased estimates of the association between disease presence and the count response relative to competing two-step approaches. Our statistical framework provides a solution that leverages disparate monitoring activities and informs species conservation across large landscapes. Stan code and documentation are provided to facilitate access and adaptation for other wildlife disease applications.

Key Words: Bayesian modeling; Data integration; Hierarchical modeling; Population monitoring; Spatial modeling.

C. Stratton (✉) · K. M. Banner, Department of Mathematical Sciences, Montana State University, Bozeman, MT, USA (E-mail: *christianstratton@montana.edu*). K. M. Irivne, U.S. Geological Survey, Northern Rocky Mountain Science Center, Bozeman, MT, USA. E.S. Almberg · K. Smucker, Montana Department of Fish, Wildlife, and Parks, Bozeman, MT, USA. D. Bachen, Montana Natural heritage Program, Montana State Library, Helena, MT, USA.

Journal of Agricultural, Biological, and Environmental Statistics
https://doi.org/10.1007/s13253-023-00593-8

Published online: 24 February 2024

1. INTRODUCTION

Conservation decision-making relies on species distribution models (SDMs) to provide spatially explicit predictions of species occurrence and inferences regarding species-habitat associations with uncertainty. Static environmental variables, such as elevation, are readily available as interpolated spatial surfaces with grid-based values and frequently included as predictors in SDMs. However, adjusting inferences about species occurrence or abundance relative to potential exposure to detrimental factors (e.g., disease vectors) often requires quantification of stressors at unsampled locations. When the stressor is quantified on an areal unit and the species response is measured at a point location, the different spatial resolutions present as a change-of-support problem. The spatial misalignment requires an approach that imputes or predicts the stressor variable at new areas while properly propagating the uncertainty in those predictions into the species response model. Spatially misaligned regression modeling (Banerjee et al. 2015, pp. 206–212), although more common in human health exposure applications (Warren et al. 2012; Lee et al. 2015; Cameletti et al. 2019), has not been used to assess a wildlife population response to a known disease. Our work is motivated by White-Nose Syndrome (WNS) which has decimated North American hibernating bats (U.S. Fish and Wildlife Service 2012) and resulted in two species being proposed for federal protection under the Endangered Species Act (U.S. Fish and Wildlife Service 2022b,a).

White-Nose Syndrome is a wildlife disease characterized by cutaneous infection during hibernation caused by the cold-adapted fungus *Pseudogymnoascus destructans* (*Pd*; Gargas et al. 2009; Lorch et al. 2011; Frick et al. 2015). The disease was first detected in North America in New York, USA, in 2006 (Turner et al. 2011) and has since spread to 40 states in the USA and eight Canadian provinces (Frank et al. 2019; WNS Response Team 2023), resulting in millions of fatalities (U.S. Fish and Wildlife Service 2012) and population declines exceeding 90% in several of the most susceptible species (Cheng et al. 2021). The disease presents as white fungal hyphae and lesions on the muzzle, wings, and ears of afflicted bats (Gargas et al. 2009; Chaturvedi et al. 2010) and causes mortality through disruption of torpor patterns during hibernation, leading to premature depletion of fat reserves and subsequent death due to starvation (Reeder et al. 2012; Frank et al. 2019). Studies documenting the decline in overwintering bat populations associated with WNS are based on count-based cave surveys jointly conducted with sampling for *Pd* and WNS on individual bats (Cheng et al. 2021). In order to measure *Pd* or document evidence of WNS, bats need to be captured and handled or visually inspected. The challenge in the western USA is that most of the susceptible bat species do not hibernate in known caves, mines, or other locations that are accessible for winter survey (Blejwas et al. 2023). Consequently, locating and capturing western bat species during hibernation, when their fungal loads are greatest, is extremely challenging. As a result, obtaining a sufficient number of samples to inform disease presence/absence at a given location is non-trivial (U.S. Geological Survey 2022).

In coordination with the US Fish and Wildlife Service's National White-Nose Syndrome Response Plan (U.S. Fish and Wildlife Service 2011), the US Geological Survey's National Wildlife Health Center designed and implemented a national surveillance program in 2012

to detect and monitor the spread of *Pd* across the USA. The original continental *Pd* sampling design was informed by a dynamic spatial diffusion model that identified high-risk areas where *Pd* was predicted to have spread in a given season (U.S. Geological Survey 2019, 2022). The associated sampling design results in intensive data collection along the predicted westward front of *Pd* spread. In the winter of 2021, a year after *Pd* was detected in Montana, Montana Fish Wildlife, and Parks supplemented the continental design by creating a 36-cell lattice to guide surveillance data collection efforts to ensure statewide coverage (Fig. 1). Also in response to WNS, a separate and distinct collaborative monitoring program was created in 2015 to track the status and temporal trajectories of bat populations across North America, known as the North American Bat Monitoring Program (NABat; Loeb et al. 2015). The NABat plan defines a lattice of 10 km × 10 km grid cells and an associated generalized random-tessellation stratified sampling design (GRTS, Stevens and Olsen 2003) to provide a spatially balanced master sample (Fig. 1). The NABat program employs passive acoustic recording units (ARUs) to measure relative bat activity during the pre-volant period (Loeb et al. 2015). A collective of Montana state agencies, federal agencies, non-government organizations, and tribal governments adopted the NABat design and have collaboratively conducted ARU-based surveys at 87 grid cells across the state in June and July annually since 2020 (Fig. 1). The best practices for integrating disease surveillance and population monitoring data at different spatial resolutions are not well understood, and previous attempts have struggled to appropriately propagate error and quantify uncertainty (Merkle et al. 2018). To our knowledge, this work is the first attempt to integrate data collected from a wildlife disease surveillance program and a population monitoring program through a joint spatial model to inform landscape conservation planning.

Previous data integration techniques in ecology constructed a joint likelihood composed of independent probability distributions for each response-type conditional on a common latent parameter(s), as in spatial data fusion or integrated SDM applications (Pacifici et al. 2017; Fletcher et al. 2019; Miller et al. 2019). For example, the latent species occurrence state of an areal unit can be jointly informed by capture and point-count surveys (Miller et al. 2019). In contrast, when the goal of inference is to assess how a disease impacts a population of interest, the parameter associated with the directional relationship between the exposure and response is the parameter of interest (e.g., a regression coefficient). Inference on how modeled quantities affect a response using regression tools is common in environmental health applications that investigate how exposure to a toxin or air pollutant affects a human population outcome (e.g., mortality or birth rates; Warren et al. 2012; Cameletti et al. 2019). If the exposure and response data are collected at different spatial resolutions, such as the *Pd* surveillance and acoustic monitoring of bats in Montana, techniques for spatially misaligned regression must be considered.

A common approach to misaligned regression modeling in the domain of human health is the so-called plug-in method (Lee and Shaddick 2010; Lee et al. 2015; Pannullo et al. 2016; Cameletti et al. 2019), in which the disease or exposure data are first modeled agnostic to the observed response. Then, the response data are regressed on a fixed one-number summary of exposure, such as the posterior mean or median. The "plug-in" technique presents a number of advantages: (1) It is easily accessible by practitioners and computationally efficient (Cameletti et al. 2019), (2) it accommodates varying degrees of missing data and

spatial misalignment between the disease and response variables, and (3) it models the exposure independent from the response. In human health applications, the lack of connection between the exposure and response models in the "plug-in" method is considered beneficial. For example, Warren et al. (2012) noted that a joint model for air pollution and preterm birth rates is not reasonable because human birth outcomes are not expected to influence the quantity or distribution of airborne contaminates. In contrast, allowing bat activity to inform the distribution of *Pd* is desirable because bats could influence the spread of the fungus to environmentally suitable areas by dispersing *Pd* within a roost and transporting the fungus to new roosts (Lorch et al. 2011; Warnecke et al. 2012; Wibbelt et al. 2010). Additionally, the "plug-in" method does not propagate any of the uncertainty in estimation of exposure through to the response model, resulting in potentially biased and overly precise estimates of the association between the exposure and response (Cameletti et al. 2019).

Bayesian techniques for spatially misaligned regression use informative prior distributions to incorporate the uncertainty in predicting exposure at unsurveyed locations into the response model. First the exposure data are modeled, which in our case is evidence of *Pd* or WNS at one of the 36 cells in Montana (Fig. 1). Then, the exposure variable is specified as a latent predictor within the response model with prior distributions informed by the posterior distribution from the exposure model (Warren et al. 2012; Powell and Lee 2013; Lee et al. 2016; Cameletti et al. 2019). For example, Cameletti et al. (2019) used estimated posterior distributions of air pollution exposure, represented as annual mean NO_2 concentrations, as prior distributions on the latent predictor variables in their response model describing hospitalizations. Similarly, Warren et al. (2012) used estimated posterior predictive distributions of climate variables as prior distributions when modeling air pollution. Related methods rely on fitting the response model multiple times, each time selecting a posterior sample from the disease model as a fixed covariate, then combining parameter estimates from the multiple fitted response models. For example, Cameletti et al. (2019) introduce the "feed-forward" approach, in which J posterior samples are selected from the exposure model and the response model is fit for each posterior sample. Similarly, Zhang et al. (2022) used multiple posterior samples from a COVID-19 exposure model as inputs in a mortality response model. Relying on prior distributions for latent parameters in the response model informed by the posterior distributions from the exposure model, hereafter the "prior" method, allows for propagation of the uncertainty in the estimation of the exposure model through to the response model.

When assessing the impact of *Pd* occurrence on Western bat populations, we propose that a better solution is to specify a single integrated modeling framework that completely captures the shared information between the *Pd* surveillance data and acoustic activity data. We compare the "plug-in" and "prior" methods to a joint model using a simulation study and empirical data in the state of Montana (Fig. 1). Specifically, our interest is whether a single integrated modeling framework that fully propagates the uncertainty in estimation of the wildlife disease model through to the ecological count model results in more accurate and precise estimates of disease occurrence and appropriate uncertainty in estimation of disease impacts on the ecological count response. In the following sections, we describe development of a joint model that accommodates non-Gaussian responses, imperfect detection, and spatial misalignment. Then, we describe our joint model extension that addresses

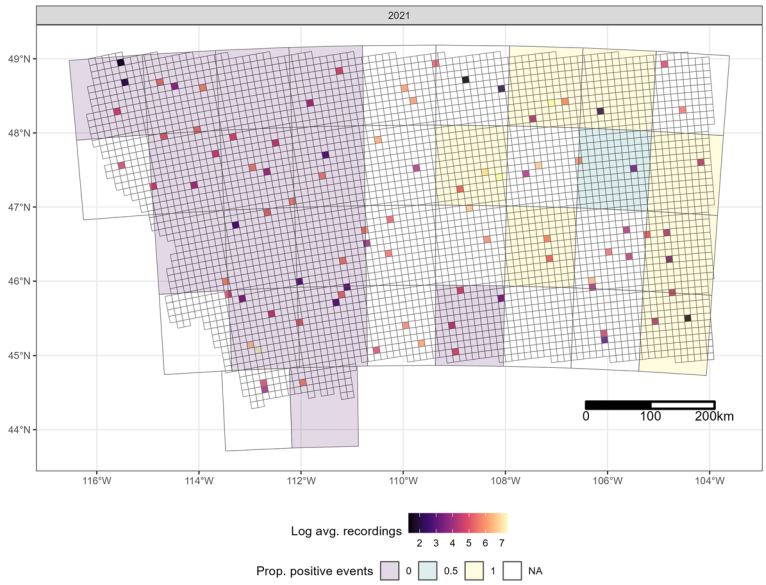

Figure 1. Data from both monitoring programs in 2021. The small grid cells represent sample locations from the NABat acoustic monitoring program (Loeb et al. 2015) and are colored by the log of average count of recordings per site-night within the cell. The large polygons represent the 36-cell lattice developed by MFWP and are colored by the proportion of survey events positive for *Pseudogymnoascus destructans* (*Pd*) within the polygon. Most *Pd* surveillance polygons received only a single sampling event, though some received as many as five. White fill represents unsampled grid cells and polygons.

the nuances of the data available in Montana, including multiple years of data collection and within-season temporal correlation among acoustic-based survey events. We provide Stan code (Stan Development Team 2023b) and documentation (Stratton et al. 2023), such that our method is accessible and adaptable for other ecological disease applications where leveraging joint information shared by the exposure and the response could be advantageous.

2. METHODS

2.1. BAT POPULATION MONITORING USING ACOUSTIC-BASED SURVEYS

Between 2020 and 2022, Montana Fish, Wildlife, and Parks (MFWP), the Montana Natural Heritage Program (MTNHP), and other partner agencies conducted acoustic monitoring of bats consistent with guidance from the North American Bat Monitoring Program (NABat, Loeb et al. 2015). The survey design defines the spatial domain of interest as the state of Montana, and an associated subset from the NABat master sample provides a spatially balanced sample from the NABat grid (Fig. 1). Between 2020 and 2022, 87 grid cells were sampled each year. Within each grid cell, an average of four ARUs (detectors) was deployed for an average of four nights each; detectors were placed sufficiently far apart to minimize spatial dependence among recorded calls from separate detectors (Loeb et al. 2015). Each detector recorded echolocating bats between sunset and sunrise nightly, resulting in an average of 16 detector-night combinations (hereafter, site-nights) per grid cell.

Acoustic recordings were assigned a species label using the SonoBat acoustic classification software (Szewczak 2023); proposed species assignments and call sequence attributes were derived from one of three regional classifiers to account for variations in community composition across the state. In order to account for the potential false-positive and false-negative detections that can result from using automated classification software (Chambert et al. 2018), auto-classified species labels are often manually reviewed by experts prior to analysis (Banner et al. 2018; Reichert et al. 2018). For the acoustic data collected in Montana, manually verified species labels were not available for all call sequences as only a subset of the calls collected were reviewed to confirm species presence at each detector location and many call sequences were not identified to species by the classifier. To account for the lack of robust species identifications, recordings were classified by whether the frequency fell within the range of WNS-susceptible species, and analyses were conducted on the recordings attributed to WNS-susceptible species in an attempt to remove false-positive detections and restrict inference on the impact of WNS to only WNS-susceptible species. In Montana, the majority of WNS-susceptible species are of the genus *Myotis* (Bachen et al. 2018). Attributes of call sequences from individuals of known species, archived within the MTNHP's bat call library, were used to determine a mean characteristic frequency threshold that separated *Myotis* bats from bats not susceptible to WNS. The frequency threshold, 34 kHz, was then used to restrict to recordings from only WNS-susceptible bats prior to analysis.

2.2. *Pseudogymnoascus destructans* (*Pd*) Surveillance Data

In 2019, MFWP began collaborating with the National Wildlife Health Center to implement *Pd* sampling (U.S. Geological Survey 2019, 2022), and *Pd* was first detected in the eastern part of the state during the winter of 2020–2021. Beginning in 2021, MFWP broadened its surveillance to include locations across the entire state in order to tie regional *Pd* status to summertime trends in bat activity informed by acoustic monitoring. To facilitate statewide monitoring within the constraint of finite resources, a 36-cell lattice of large rectangular polygons (hereafter, "*Pd* polygons") was superimposed over the state, and the state aimed to conduct at least one sampling event within each polygon each year (Fig. 1). However, 12, 22, and 29 polygons were sampled in 2020, 2021, and 2022, respectively. In Table 1, we categorize each of the 3983 acoustic monitoring grid cells in terms of the degree of overlap between the acoustic and *Pd* data sources.

Within each polygon, local biologist expertise was used to identify hibernacula, spring emergence mist-net sites, or maternity roost sites for sampling; attempts were made to evenly distribute the survey type, including hibernacula surveys, live animal trapping, or pooled guano and environmental sampling, across the state. Hibernacula surveys involved swabbing hibernating bats, cave substrates, or collecting soil and guano. Live animal trapping involved early season mist-netting or trapping bats emerging from bat boxes between April and June. Pooled guano surveys were conducted by collecting fresh guano at early season roost sites in buildings, beneath bridges, or in bat boxes during spring emergence. Samples from all survey types were then assessed for presence of *Pd* using polymerase chain reaction testing (PCR) at either the National Wildlife Health Center or Oregon Veterinary Diagnostic Laboratory.

Table 1. Count of NABat grid cells categorized by whether data were observed within the cell

	Acoustic data available	Pd data observed	Count
2020	No	No	2500
	No	Yes	1396
	Yes	No	59
	Yes	Yes	28
2021	No	No	1416
	No	Yes	2480
	Yes	No	30
	Yes	Yes	57
2022	No	No	768
	No	Yes	3128
	Yes	No	13
	Yes	Yes	74

Note that grid cells are not directly sampled for *Pseudogymnoascus destructans* (*Pd*); if a grid cell is contained within a *Pd* polygon that is sampled for *Pd*, the grid cell is considered sampled for *Pd*

A survey event was considered positive if at least one sample from the survey event tested positive for *Pd*.

2.3. SPATIALLY MISALIGNED REGRESSION MODELS

We consider three methods for estimating the association between *Pd* occurrence and WNS-susceptible species bat relative activity: (1) the "plug-in," (2) the "prior," and (3) the "joint." To facilitate our discussion of these methods, we assume the following notation. Let $\mathcal{A}_k$, $k = 1, \ldots, K$, denote the polygons defined for monitoring of *Pd* (i.e., 36-cell lattice covering MT), $x_k^{\mathcal{A}}$ denote the probability of *Pd* occurrence in polygon $\mathcal{A}_k$, and $i[k]$ index NABat grid cell i within polygon $\mathcal{A}_k$; NABat grid cells located on the boundaries of *Pd* polygons were attributed to a single polygon by the area of greatest overlap. Let y_k denote the number of positive *Pd* sampling events out of n_k events in polygon $\mathcal{A}_k$, $Z_{i[k]}$ denote the latent occurrence state of WNS-susceptible species in grid cell i within polygon $\mathcal{A}_k$, and $c_{ij[k]}$ denote the observed count of recordings classified as any WNS-susceptible species at grid cell i during visit j within polygon $\mathcal{A}_k$. We assume a zero-inflated negative binomial sampling model for WNS-susceptible species bat activity:

$$\begin{aligned} Z_{i[k]} &\sim \text{Bernoulli}(\psi_{i[k]}), \\ c_{ij[k]} | Z_{i[k]} = z_{i[k]} &\sim \text{Negative binomial}(z_{i[k]}\mu_{ij[k]}, \xi) \end{aligned} \tag{1}$$

where $\text{logit}(\psi_{i[k]}) = \boldsymbol{x}_{i[k]}^{(z)}\boldsymbol{\beta}^{(z)}$, $\log(\mu_{ij[k]}) = \boldsymbol{x}_{ij[k]}^{(c)}\boldsymbol{\beta}^{(c)} + \alpha x_k^{\mathcal{A}}$, the negative binomial distribution is parameterized by its mean and variance (Stan Development Team 2023b), and $\boldsymbol{x}_{i[k]}^{(z)}$ and $\boldsymbol{x}_{i[k]}^{(c)}$ denote vectors of covariates associated with occupancy and relative activity at a site-night, respectively. We use superscripts on the coefficients $\left(\text{e.g.,}\boldsymbol{\beta}^{(c)}\right)$ and associated covariates $\left(\text{e.g.,}\boldsymbol{x}_{ij[k]}^{(c)}\right)$ to distinguish the responses with which they are affiliated. The zero-inflation component accounts for potential false negatives arising from

no recordings classified to WNS-susceptible species on any site-night even though at least one species truly occurred within a grid cell (MacKenzie et al. 2002).

The ecological response model described in Eq. (1) requires specification of $x_k^{\mathcal{A}}$, the probability of *Pd* occurrence in polygon $\mathcal{A}_k$, in order to estimate α, the association between the probability of *Pd* occurrence and the log-mean WNS-susceptible bat activity. While the "plug-in," "prior," and "joint" methods differ in their propagation of the uncertainty in estimation of $x_k^{\mathcal{A}}$ through to the ecological response model, they all rely on the same model structure for evidence of *Pd* occurrence. Let y_k denote the observed count of positive *Pd* survey events from n_k trials in polygon $\mathcal{A}_k$ and $\boldsymbol{s}_k$ denote the centroid of the k^{th} observed *Pd* polygon. The model for *Pd* occurrence is:

$$\begin{aligned} y_k &\sim \text{Binomial}\left(n_k, x_k^{\mathcal{A}}\right), \\ \text{logit}(x_k^{\mathcal{A}}) &= \boldsymbol{x}_k^{(y)}\boldsymbol{\beta}^{(y)} + \eta_k, \\ \boldsymbol{\eta} &\sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma_{11}}\right) \end{aligned} \tag{2}$$

where the i^{th} row and j^{th} column of $\boldsymbol{\Sigma}_{11}$, denoted $\boldsymbol{\Sigma}_{11}^{ij}$, is given by $\boldsymbol{\Sigma}_{11}^{ij} = \sigma^2 \exp\left\{-\frac{1}{2\phi^2}||\boldsymbol{s}_i - \boldsymbol{s}_j||^2\right\}$. To obtain predictions of $x_k^{\mathcal{A}}$ in unsampled *Pd* polygons, we consider the spatial random effects from observed polygons, $\boldsymbol{\eta}$, and unobserved polygons, $\boldsymbol{\eta}^*$, as a multivariate normal process. Let $\boldsymbol{s}_k^*$ denote the centroids of the unobserved *Pd* polygons. Then,

$$\begin{bmatrix} \boldsymbol{\eta} \\ \boldsymbol{\eta}^* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}\right) \tag{3}$$

where $\boldsymbol{\Sigma}_{12}^{ij} = \sigma^2 \exp\left\{-\frac{1}{2\phi^2}||\boldsymbol{s}_i - \boldsymbol{s}_j^*||^2\right\}$, $\boldsymbol{\Sigma}_{22}^{ij} = \sigma^2 \exp\left\{-\frac{1}{2\phi^2}||\boldsymbol{s}_i^* - \boldsymbol{s}_j^*||^2\right\}$, and $\boldsymbol{\Sigma}_{21} = \boldsymbol{\Sigma}_{12}^T$. Conditioning on the observed spatial random effects yields a multivariate normal distribution for unobserved random effects, with covariance matrix $\boldsymbol{\Sigma}_{2|1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{21}^T$ and mean vector $\boldsymbol{\mu}_{2|1} = \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\eta}$ (Eaton 1983, pp. 116–117).

In order to fully propagate the uncertainty in estimation of the disease model through to the ecological response model and leverage shared information between the two data sources, the proposed joint modeling framework estimates the ecological response model described in Eq. (1) and the spatial model for disease occurrence described in Eq. (2) simultaneously in a single hierarchical framework. Conversely, the "plug-in" method relies on a two-step process. First, the disease surveillance data are modeled according to Eq. (2) to obtain posterior summaries, such as the mean or median, of the probability of *Pd* occurrence within polygon $\mathcal{A}_k$, denoted $\bar{x}_k^{\mathcal{A}}$. Then, those summaries are included as a fixed covariate in the ecological response model, $\log(\mu_{ij[k]}) = \boldsymbol{x}_{ij[k]}^{(c)}\boldsymbol{\beta}^{(c)} + \alpha\bar{x}_k^{\mathcal{A}}$.

The "prior" method also relies on a two-step process, first modeling the disease occurrence according to Eq. (2) to obtain posterior distributions for $x_k^{\mathcal{A}}$. However, rather than using posterior summaries of $x_k^{\mathcal{A}}$ as a fixed covariate in the ecological response model, the "prior" method treats these probabilities as unknown and specifies informative prior distributions based on the posterior distributions of $x_k^{\mathcal{A}}$ obtained from the model fit to the *Pd* surveillance data. In the case of probabilities, one may use the mean-variance parameteri-

zation of the beta distribution (Ferrari and Cribari-Neto 2004) as a prior distribution for the latent probabilities. That is, let $\log(\mu_{ij[k]}) = \boldsymbol{x}^{(c)}_{ij[k]}\boldsymbol{\beta}^{(c)} + \alpha x^{\mathcal{A}}_k$ where

$$x^{\mathcal{A}}_k \sim \text{beta}(a_k, b_k), \tag{4}$$

where $a_k = \mu_k\phi_k, b_k = (1-\mu_k)\phi_k, \phi_k = \frac{\mu_k(1-\mu_k)}{\tau^2_k} - 1$, and μ_k and τ^2_k are the posterior mean and variance of $x^{\mathcal{A}}_k$ from the disease model, respectively (Eq. 2). Similar to the joint modeling framework, this approach allows for propagation of the uncertainty in the estimation of $x^{\mathcal{A}}_k$ through to the ecological response model.

2.4. SIMULATION STUDY

We conducted a simulation study to compare the three models ("plug-in," "prior," and "joint") with respect to estimation error and uncertainty in the posterior distributions for two outcomes of interest: *Pd* occurrence probabilities $\left(x^{\mathcal{A}}_k\right)$ and the log-linear association between *Pd* occurrence and relative activity of WNS-susceptible bat species (α). We considered three scenarios representing varying magnitudes for the negative association between the evidence of the fungus and our measure of relative bat activity: "no effect" ($\alpha = 0$), "moderate effect" ($\alpha = -1$), or "strong effect" ($\alpha = -3$). For each of the three scenarios, 50 data sets were generated from the joint model described by Eqs. (1), (2), and (3) assuming that nine polygons were randomly sampled for *Pd* and 100 grid cells were randomly selected for bat population monitoring using acoustic recording devices. Selecting the joint model as the generating mechanism allowed for investigating the impact of model choice on parameter estimates regardless of whether the *Pd* occurrence and count processes were associated by varying the value of α. The remaining assumed parameter values were based on empirical estimates obtained from fitting a simplified version of the fully specified "joint" model to the Montana data assuming only one year of sampling and excluding the nested random effect structure (see Sect. 2.5).

For each simulated data set, the "plug-in," "prior," and "joint" models were fit and posterior mean estimates, 95% credibility intervals, whether those credibility intervals captured the data generating values, and the squared-error (SE) were tracked. For a posterior mean probability of $\hat{x}^{\mathcal{A}}_k$ and true generating probability of $x^{\mathcal{A}}_k$, the SE was calculated as $\left(x^{\mathcal{A}}_k - \hat{x}^{\mathcal{A}}_k\right)^2$. When summarizing the SE across all simulations, the square root of the mean of SEs, RMSE, was considered. For the two-step approaches, the *Pd* occurrence model described by Eqs. (2) and (3) was first fit to the simulated *Pd* data. Then, the ecological response model described in Eq. (1) was fit to the simulated acoustic data, incorporating posterior summaries of the disease occurrence probabilities in the linear predictor of the count response. The "plug-in" method used posterior mean estimates of disease occurrence probability as a covariate in the model, while the "prior" method assumed these probabilities were latent, relying on informed priors as described in Eq. (4). All models were fit using the probabilistic programming language Stan (Stan Development Team 2023a) and assessed for convergence with the Gelman–Rubin statistic (Brooks and Gelman 1998) and through visual inspection of trace plots.

2.5. DATA ANALYSIS

In order to estimate the association between *Pd* occurrence and bat activity between 2020 and 2022, the "plug-in," "prior," and "joint" modeling frameworks were fit to the data described in Sects. 2.1 and 2.2. For both of the two-step modeling approaches, the same iterative process described in Sect. 2.4 was used to incorporate estimates of *Pd* occurrence probabilities into the ecological response models. All three methods included the same covariates for the *Pd* $\left(x_k^{(y)}\right)$ and bat activity $\left(x_{ij[k]}^{(c)}, x_{i[k]}^{(z)}\right)$ processes, allowing for direct comparison between the methods.

The zero-inflation portion of each model included the mean elevation of the grid cell (meters), the average annual precipitation in the grid cell (millimeters), and the average annual temperature (degrees Celsius). A combination of bioclimatic and site-night specific covariates consistent with previous analyses of bat acoustic data was included in each negative binomial count model (e.g., Wright et al. 2018; Rodhouse et al. 2019; Stratton et al. 2022). Bioclimatic variables were obtained from the Parameter-elevation Regressions on Independent Slopes Model (PRISM, PRISM Climate Group 2022) at a 4 km resolution and included maximum nightly temperature (degrees Celsius). Site-night specific covariates included the log of the count of detections from the previous night, and an indicator for whether the detector was located by lentic water systems, lotic water systems, positioned in a flyway, positioned near a roosting structure, or other, and the first-order interaction between the log of the count of detections from the previous night and the site-type indicator. The linear predictor for *Pd* occurrence included the latitude, longitude, their interaction, and second and third degree polynomial effects. All continuous covariates were scaled to have a mean of zero and a standard deviation of one prior to estimating the models.

In order to account for potential temporal correlation in bat activity or *Pd* occurrence over the three surveyed years, first-order auto-regressive terms were included in both the negative binomial count and *Pd* occurrence portions of each model. Additionally, to account for the hierarchical nesting structure (consecutive survey nights at a detector location and multiple detector locations nested within a grid cell), nested detector-specific random intercepts were included in the activity portion of each model. Specifically, letting i index the NABat grid cell, d index the detector within grid cell i, and t index the year of survey, the random intercepts, γ_{idt}, were drawn from the following hierarchical formulation:

$$\begin{aligned} \mu_t &\sim N(0, \sigma_\mu^2) \\ \theta_{it} &\sim N(\mu_t, \sigma_\theta^2) \\ \gamma_{idt} &\sim N(\theta_{it}, \sigma_\gamma^2) \end{aligned} \tag{5}$$

All models were fit using Stan (Stan Development Team 2023a), with three independent chains of 10,000 MCMC iterations each; each model was assessed for convergence visually and through the Gelman–Rubin statistic (Brooks and Gelman 1998). A complete description of the code used to fit the model is provided in online Appendix (Stratton et al. 2023).

3. RESULTS

3.1. SIMULATION STUDY

The precision of the posterior distribution and RMSE for the probability of *Pd* occurrence, $x_k^{\mathcal{A}}$, depended on the choice of model, the assumed strength of association between *Pd* and bat activity (α), and whether both data sources were observed within the *Pd* polygon (Figs. 2 and 3). Generally, there were greater precision and lower RMSE in estimating $x_k^{\mathcal{A}}$ if *Pd* data were observed within $\mathcal{A}_k$ across all methods and scenarios. Additionally, the "joint" method resulted in the greatest precision and lowest RMSE when estimating $x_k^{\mathcal{A}}$ on average, followed by the "prior" and "plug-in" methods, respectively. However, in cases where the disease and ecological processes were independent ($\alpha = 0$), all three methods yielded approximately equivalent estimates of $x_k^{\mathcal{A}}$ with similar precision and RMSE (Figs. 2 and 3, right panels).

The "plug-in" method resulted in nearly equivalent precision and RMSE in $x_k^{\mathcal{A}}$ across the three values of α, and regardless of whether count data were observed within $\mathcal{A}_k$ (Figs. 2 and 3). Conversely, the "joint" method resulted in increasing precision, and decreasing RMSE, as α increased in magnitude regardless of whether count data were observed within $\mathcal{A}_k$, though the increase in precision and decrease in RMSE were greater in polygons where count data were observed (Figs. 2 and 3). Together, these results suggest that the "joint" method results in both greater precision and greater overall accuracy than the "plug-in" method when estimating the probability of *Pd* occurrence, so long as the disease and count processes are related. For example, consider the predicted probabilities in unsurveyed *Pd* polygons that contained count data. In the case where *Pd* had a strong effect on the count process, the RMSE of predicted probabilities was greater for the "plug-in" method than for the "joint" method, on average (mean RMSE for "joint" of 0.0476, mean RMSE for "plug-in" of 0.2410). In the case where *Pd* had a moderate effect on the count process, the RMSE of predicted probabilities was again greater for the "plug-in" method than for the "joint" method on average, though the discrepancy was lesser (mean RMSE for "joint" of 0.0996, mean RMSE for "plug-in" of 0.2310).

For the "prior" method, the impact of increasing the magnitude of α on the precision and RMSE of $x_k^{\mathcal{A}}$ depended on whether count data were observed within $\mathcal{A}_k$. If count data were observed within $\mathcal{A}_k$, the precision of $x_k^{\mathcal{A}}$ increased, and RMSE decreased, as α increased in magnitude, similar to the "joint" method, though the increase was lesser. If count data were not observed within $\mathcal{A}_k$, increasing the magnitude of α had minimal impact on the precision and RMSE of $x_k^{\mathcal{A}}$, similar to the "plug-in" method. These results combined suggest that in terms of precision and accuracy, the "prior" method results in *Pd* occurrence probability estimates that are more similar to the "joint" method if count data are observed in a *Pd* polygon, but more similar to the "plug-in" method if count data are not observed in a *Pd* polygon. For example, consider the predicted probabilities in unsurveyed *Pd* polygons in the scenarios where there was a strong impact of *Pd* on the count process. The RMSE of predicted probabilities where count data were observed was greater for the "prior" method than for the "joint" method, on average (mean RMSE for "joint" of 0.0476, mean RMSE for "prior" of 0.0795). However in the case where count data were not observed, the RMSE for the "prior" method was much greater than for the "joint" method, on average (mean RMSE

for "joint" of 0.0720, mean RMSE for "prior" of 0.218). See Appendix A for visualizations of the posterior mean probability of *Pd* occurrence for a subset of simulations and scenarios.

When estimating α, all three methods resulted in nearly equivalent estimates with similar precision, yielding credibility intervals that achieved nominal coverage if the ecological and disease processes were independent (Fig. 4, bottom panel). In the case where the magnitude of α was moderate, the "prior" and "joint" methods resulted in similar estimates of α, with the "prior" method resulting in slightly more precision than the "joint" method, and both methods yielding approximately nominal coverage (Fig. 4, middle panel). Conversely, the "plug-in" method resulted in credibility intervals that did not achieve nominal coverage, an effect that was exacerbated as the magnitude of α increased (Fig. 4, top panel). In the case where there was a strong effect of *Pd* on the ecological response, the "prior" method again resulted in narrower credibility intervals than the "joint" method, though this increased precision came at the cost of bias as the credibility intervals from the "prior" method did not provide nominal coverage and resulted in lesser coverage than the "joint" method (Fig. 4, top panel).

3.2. Data Analysis

Posterior mean estimates of the probability of *Pd* occurrence over time were similar for all three models and strongly suggest that the fungus has spread westward since 2020. Posterior uncertainty for the estimated probability of *Pd* occurrence was greatest in 2020, when the fewest polygons were sampled. Additionally, uncertainty in the posterior of the estimated probability tended to be greater in the middle of the state along the forefront of *Pd* spread within each year for all models considered. *Pd* occurrence probability uncertainty tended to be less for the "joint" model relative to the "plug-in" and "prior" methods, particularly in 2020 when the fewest *Pd* polygons were sampled (Fig. 5), consistent with the results of the simulation study. The discrepancy in standard deviations of the *Pd* occurrence probabilities is lesser in 2022 for both comparisons in Fig. 5 ("joint"–"prior" and "joint"–"plug-in"). The apparent gain in precision for the "prior" and "plug-in" models is a result of the additional *Pd* surveillance data that are available in 2022; as a result, less spatial prediction is required, benefiting the disjoint modeling frameworks. Even still, the "joint" model results in smaller standard deviations for the probabilities of *Pd* occurrence in 2022, though the discrepancy is lesser than in 2020 and 2021.

Ninety-five percent credibility intervals for a subset of regression coefficients from the log-linear predictor of the negative binomial counts are provided in Fig. 6. After accounting for the other site-level and bioclimatic variables, there is evidence of a strong, positive association between log-mean bat activity and whether the site was located near lentic water systems or lotic water systems, relative to the baseline category of "other." Furthermore, the association is stronger for lentic water systems than for lotic water systems, consistent with previous analyses of bat activity (Blakey et al. 2018). Conditional on the other predictors in the count model, there is also evidence of a strong positive association between the log-mean bat activity and the log of the number of detections from the previous night. After accounting for the site-level and bioclimatic variables, there was evidence of a negative

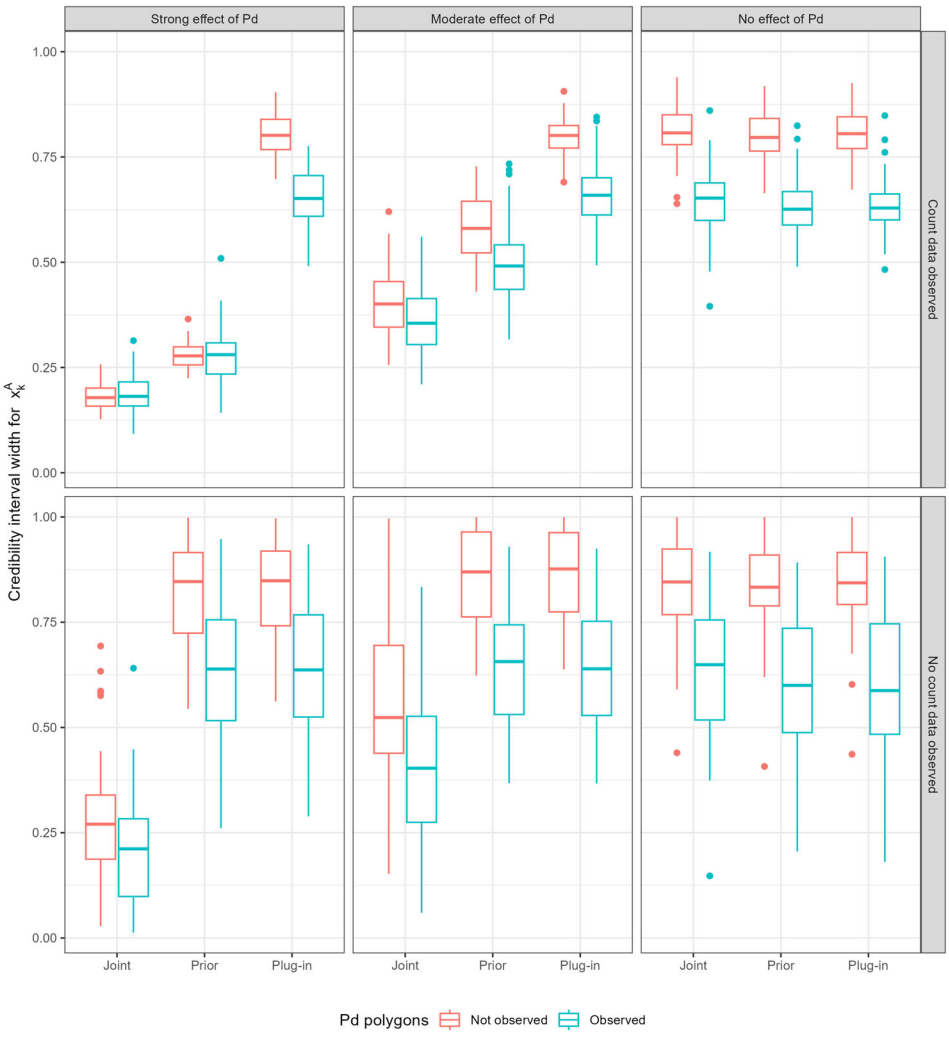

Figure 2. Box plots of 95% credibility interval widths for *Pseudogymnoascus destructans* (*Pd*) occurrence probabilities colored by whether the estimated occurrence probability was for a polygon that contained *Pd* data and paneled by the assumed strength of the effect of *Pd* on the ecological response (columns) and whether count data were observed within the polygon (rows). Each individual value within a box plot represents the mean credibility interval width for all *Pd* probabilities from a single simulated data set that are consistent with the faceting conditions. Generally, the "joint" method results in the greatest precision, followed by the "prior" and "plug-in" methods, respectively. The effect of the strength of *Pd* on the precision from the "prior" method depends on whether count data were observed within the polygon.

association between the probability of *Pd* occurrence and bat activity across all three years, though there was a high degree of uncertainty in each of these estimates (Fig. 6).

4. DISCUSSION

We explored three approaches to spatially misaligned regression modeling that allow for estimating the association between a known stressor (in our application, *Pd*, a non-native

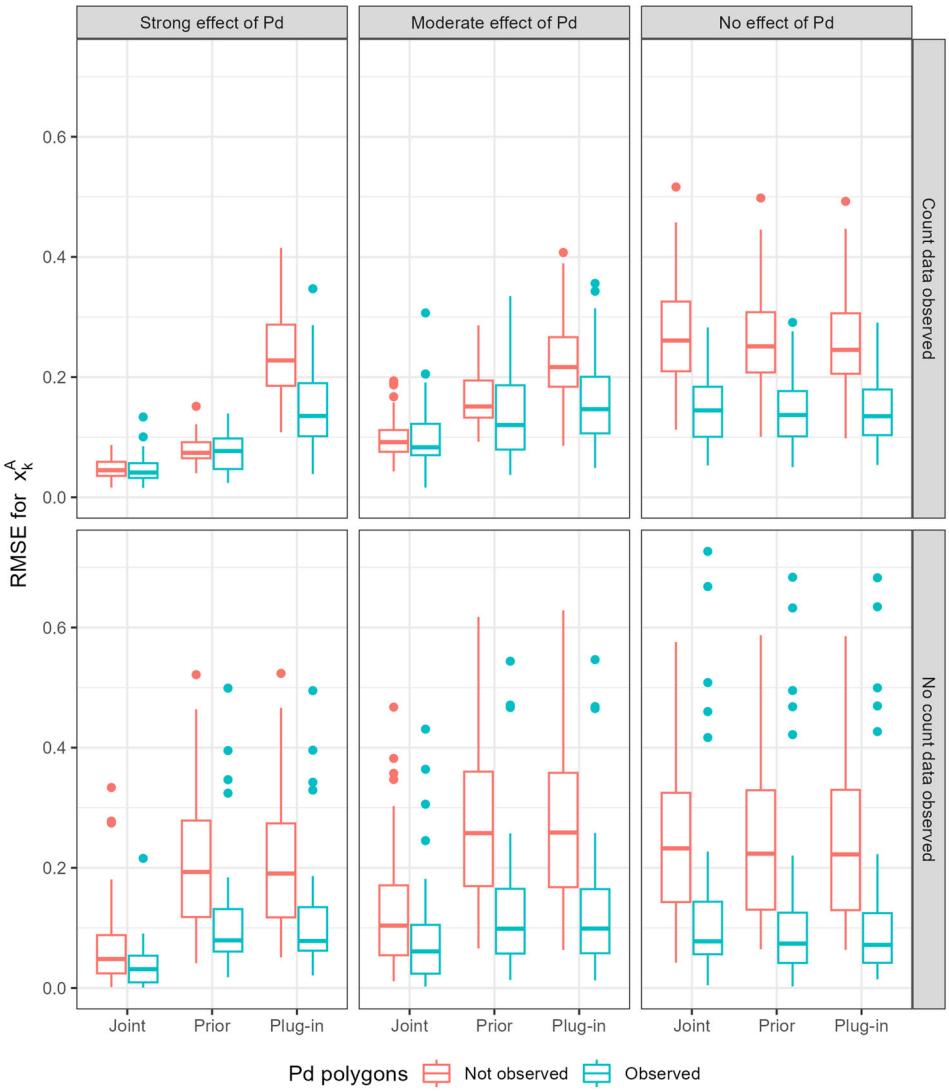

Figure 3. Box plots of RMSE for *Pseudogymnoascus destructans* (*Pd*) occurrence probabilities colored by whether the estimated occurrence probability was for a polygon that contained *Pd* data and paneled by the assumed strength of the effect of *Pd* on the ecological response (columns) and whether count data were observed within the polygon (rows). Each individual value within a box plot represents the square root of the mean of the squared errors for all *Pd* probabilities from a single simulated data set that are consistent with the faceting conditions. Generally, the "joint" method results in the lowest RMSE, followed by the "prior" and "plug-in" methods, respectively.

fungus that causes WNS) and a species response. Our simulation study demonstrated the importance of propagating the uncertainty in estimating a modeled quantity that is then included as a predictor within an ecological response model. In the scenarios where there was a moderate effect of *Pd* on mean relative bat activity, the "plug-in" method resulted in an estimated effect of the disease (α) that did not achieve nominal coverage rates. In the case where the disease had a strong impact on mean relative activity, both the "plug-in" and "prior" methods resulted in overprecise estimates of α, consistent with previous findings (Cameletti

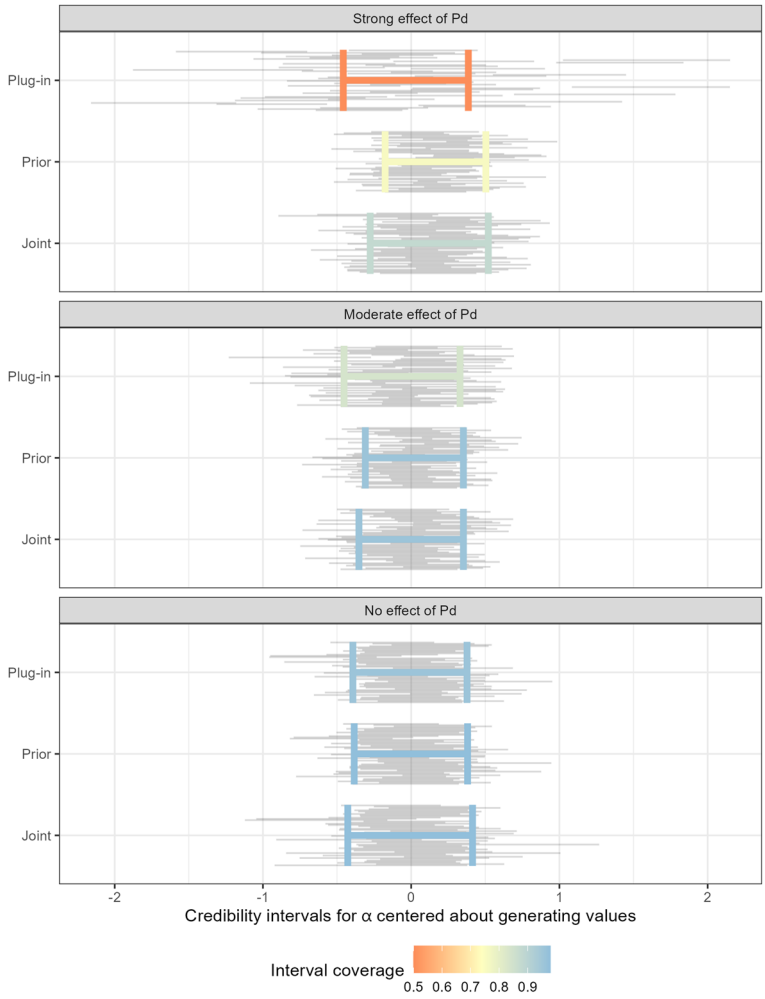

Figure 4. Line plots of 95% credibility intervals for the effect of *Pseudogymnoascus destructans* (*Pd*) presence on the ecological response, paneled by the strength of the effect of *Pd* on the ecological response. The larger error bars represent average credibility intervals from each method and are colored by the proportion of credibility intervals that captured the data generating values. As the strength of the effect of *Pd* on the ecological response increases, the "prior" and "plug-in" methods fail to achieve nominal coverage.

et al. 2019), with lower coverage rates than our "joint" modeling framework. For the Montana data sets, the posterior distributions for the estimated association between evidence of *Pd* and summertime acoustic activity for WNS-susceptible species were nearly equivalent among the three models. However, the estimated magnitude of α was relatively small for the three models, which was closer to our "no effect" and "moderate effect" simulation scenarios, so the similarity in posterior estimates was consistent with our simulation findings.

Our simulation study revealed an interesting practical benefit of jointly modeling two biologically linked processes, as is likely in our application, in that the response data informed the modeled predictor even when the two data types were not co-located. On average, our "joint" model resulted in less posterior uncertainty when estimating *Pd* occurrence than

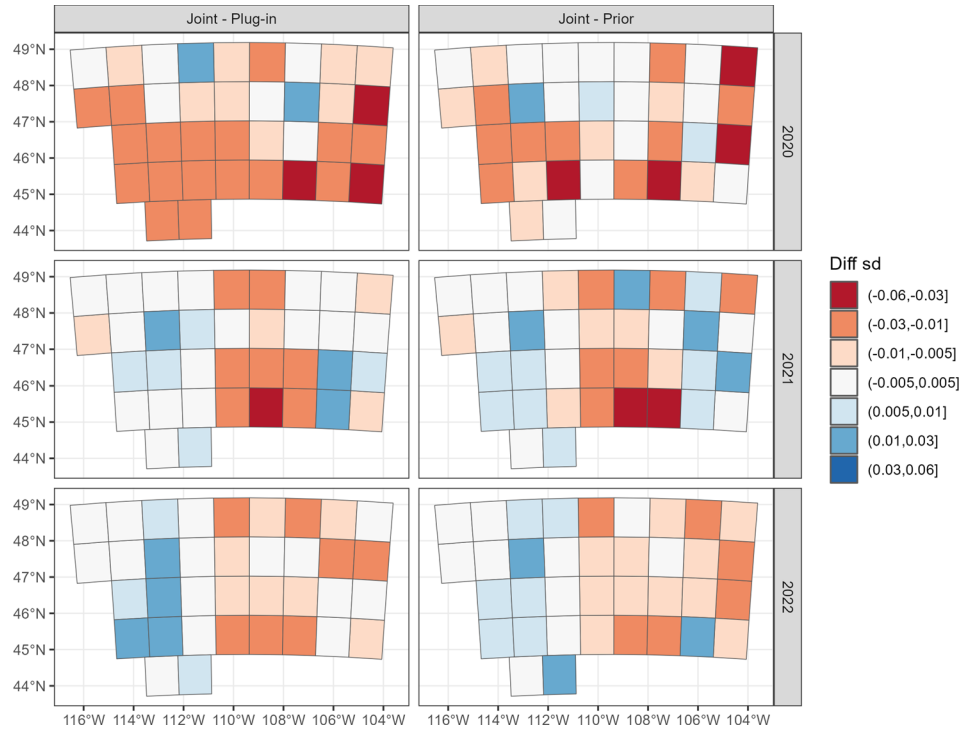

Figure 5. Map of the difference in posterior standard deviations of probabilities of *Pseudogymnoascus destructans Pd* occurrence between the (left) "joint" modeling approach and "plug-in" modeling approach and (right) "joint" modeling approach and "prior" modeling approach over time. Red fill denotes areas where the "joint" model resulted in greater precision.

did either the "plug-in" or "prior" methods. Similarly, in our empirical data analysis we found less uncertainty in the predicted *Pd* occurrence probabilities for the "joint" model relative to the "plug-in" and "prior" models, on average. Our simulations showed precision increased for estimated disease occurrence probabilities with increasing magnitude of the association between disease and the ecological response (α). In general, the simulation study and empirical data analysis suggested that the joint modeling framework resulted in the largest reduction of uncertainty in disease occurrence probabilities when either of the two data sources was sparsely observed or the data sources were not co-located. Together, these results demonstrate the value of the joint modeling framework for spatially misaligned regression modeling when integrating costly data sources. In such cases, the joint modeling framework can be used to improve the precision of parameter estimates when collecting large volumes of data is difficult or unfeasible.

While the joint modeling framework provides more precise estimates of disease occurrence and fully propagates the uncertainty in estimation of disease occurrence through to the ecological response, some care must be taken when implementing the "joint" model. Notably, by modeling the disease occurrence and relative bat activity jointly, information about relative activity can inform estimation of disease occurrence (Warren et al. 2012). While this is a desirable property when modeling *Pd* occurrence and bat activity due to

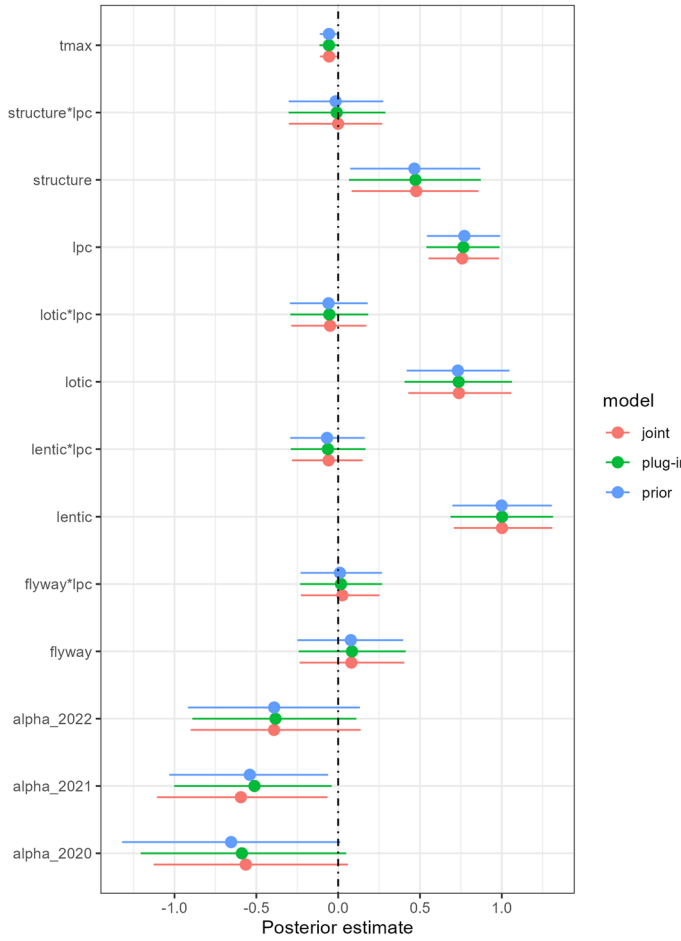

Figure 6. Ninety-five percent credibility intervals for a subset of regression coefficients in the count model; intercept terms, temporal random effects, and detector-level random effects are omitted, and "lpc" denotes the log of the count of detections from the preceding night. There are positive point estimates associated with sites located near water, roosting structures, or within a flyway, the maximum nightly temperature, and the activity from the previous night. There are negative effects associated with occurrence of *Pseudogymnoascus destructans* and some interactions between site type and activity from the previous night.

the transmission mechanics of WNS, this property may not be desirable for all applications; the implication of considering the disease and ecological processes jointly is a model assumption and should be carefully assessed prior to analysis. Additionally, by considering disease occurrence and relative activity jointly, misspecification of the disease occurrence model can lead to biased estimates of the impact of the disease on the ecological response (Cameletti et al. 2019). If the two data sources are spatially misaligned and spatial predictions are required, misspecification of either model could also lead to biased predictions of the probability of *Pd* occurrence, potentially resulting in biased estimates of the impact of *Pd* on the log-mean count. Consequently, it is especially important to perform rigorous model assessment when implementing the joint model. Future work could consider the severity of the impact of model misspecification on the predicted probabilities of *Pd* occurrence.

Additional levels of complexity could be added to both the species distribution and spatial components of the model. For example, random intercepts could be included in the linear predictors for ψ and μ to account for spatial correlation in the occupancy and count processes, respectively. If species-specific classifications are available for the acoustic recordings, multi-species count models can be implemented to improve model fit (Wright et al. 2020). Similarly, consideration of additional complexity in the *Pd* occurrence model, such as allowing for imperfect detection or differing detection probabilities by survey type (e.g., environmental, guano, or tissue) could be a useful next step to improve model fit (e.g., Campbell Grant et al. 2023). Additional temporal dynamics may also be included in the *Pd* occurrence model to account for potential lagged impacts of *Pd* occurrence. Finally, considering spatially varying coefficients for α could reveal more subtle relationships between WNS and summertime bat populations (e.g., Hastie and Tibshirani 1993; Fan and Zhang 1999; Finley 2011). Development of novel models for monitoring of wildlife disease remains an area of active research (Hefley et al. 2017; Hicks et al. 2020; Watsa, M. and Wildlife Disease Surveillance Focus Group 2020; Wiens and Thogmartin 2022), and as the statistical methodology evolves, our joint modeling framework can be easily updated accordingly.

For the empirical analysis outlined in this paper, point-referenced spatial information was obfuscated due to sensitive locations used for *Pd* surveillance. However, the joint modeling framework described in this paper is well suited to handle point-referenced disease occurrence data; alternatively, the geostatistical spatial process may be replaced by techniques for areal data, including a Gaussian Markov random field (Banerjee et al. 2015). While we did not encounter any computational challenges during our empirical analysis due to the relatively small number of *Pd* polygons, re-considering the surveillance data as point-referenced could increase the computational burden, particularly when expanding to larger extents. The spatial Gaussian process (GP) specified in the "joint" model is prone to scalability issues; notably, cubic complexity with the number of spatially indexed observations (Liu et al. 2018). Recently, locally approximated GPs (Gramacy and Apley 2015) and nearest neighbor GPs (Datta et al. 2016) have been proposed as solutions to the computational burden of GP modeling for large spatial data sets. Additionally, data augmentation strategies could be explored to afford computationally efficient Gibbs updates of regression coefficients for the negative binomial sampling model (Polson et al. 2013), or MCMC techniques may be abandoned for computationally efficient alternatives, such as the integrated nested Laplace approximation (Rue et al. 2009). Such improvements would allow for comparison of additional modeling techniques, including the "feed-forward" approach (Cameletti et al. 2019). Future exploration into optimization of the joint modeling framework for large spatial domains would directly benefit wildlife conservation activities coordinated across species ranges.

Future work could also focus on evaluating the impact of co-locating the *Pd* surveillance locations within the acoustic monitoring grid cells and consider optimal sampling designs for estimation of the "joint" model. Investigations into optimal designs for spatial modeling typically focus on guidance for selecting locations to measure one variable (e.g., relative bat activity) to minimize prediction errors or estimation errors for the spatial parameters (Zimmerman 2006; Irvine et al. 2007). However, exploring how to distribute sampling effort for both data sources within the joint modeling framework would be valuable. Optimal sam-

pling designs for the *Pd* surveillance process would ideally result in reduced uncertainty in predicted *Pd* occurrence at unsurveyed locations, a property of maximum-entropy sampling designs (Shewry and Wynn 1987; Wang et al. 2020). However, the optimal design must also accommodate the nuances of ecological sampling. Co-location is often either impractical or impossible because of the biology of the species, as in our application, or constrained by organizational barriers. The practical challenge is that one organization may be responsible for collecting and managing the pathogen or disease data and a different organization could be responsible for coordinating the surveys for the potentially impacted wildlife species throughout its range. Our proposed joint modeling framework provides a means to integrate disparate biosurveillance surveys with expansive wildlife population monitoring efforts without requiring co-location. Importantly, the joint modeling framework is a statistical solution to integrating, related but parallel, monitoring activities to better inform species conservation across large landscapes.

Authors' contributions C. Stratton: Conceptualization (lead), Formal analysis, Methodology (lead), Software, Validation, Visualization, Writing - Original draft, Writing - Review and editing; K.M. Irvine: Conceptualization, Methodology, Supervision, Writing - Review and editing; K.M. Banner: Conceptualization, Methodology, Writing - Review and editing; E.S. Almberg: Data curation, Writing - Review and editing; D. Bachen: Data curation, Writing - Review and editing; K. Smucker: Data curation, Writing - Review and editing.

Data Availability All data and model code used in this manuscript are published and freely available online (Stratton et al. 2023).

Funding Funding was provided by U.S. Geological Survey (Grant Number: G20AC00406-02).

Declarations
Conflict of interest All authors declared that they have no conflict of interest.

[Received June 2023. Revised October 2023. Accepted December 2023.]

APPENDIX A: ADDITIONAL SIMULATION RESULTS

In Figs. 7, 8, and 9, we provide maps of the difference between the posterior mean probability of *Pd* occurrence and the simulation generating values for a subset of simulated data sets for strong, moderate, and no impact of *Pd* on the log-mean count, respectively. On average, the "joint" model resulted in the smallest differences for the scenarios with strong and moderate impacts of *Pd* on the log-mean count (Figs. 7 and 8). For simulated data

sets with no impact of *Pd* on the log-mean count, all models resulted in nearly equivalent differences (Fig. 9). In general, the "joint" model tends to perform better than the "prior" and "plug-in" models as the strength of the impact of *Pd* on the log-mean count increases, consistent with the summary of results provided in the main body of the paper.

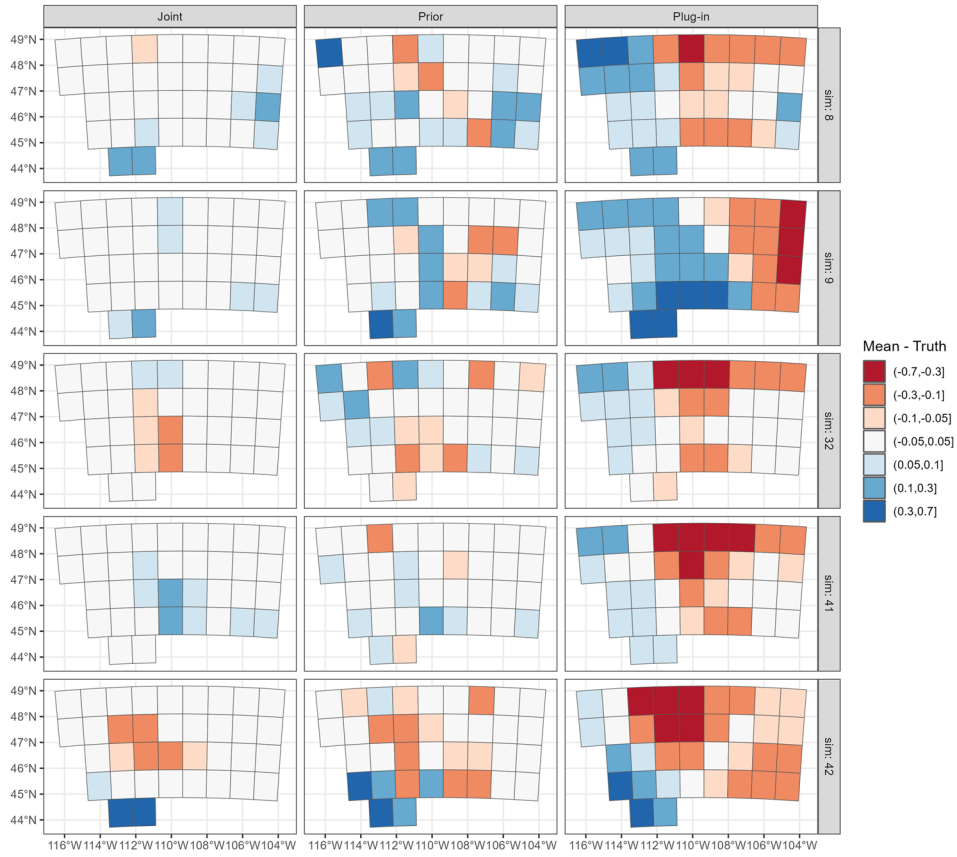

Figure 7. Difference between the posterior mean probability of *Pseudogymnoascus destructans* (*Pd*) occurrence $(\hat{x}_k^A)$ and true generating probability (x_k^A) for a subset of simulated data sets assuming a strong impact of *Pd* occurrence on the log-mean count. On average, the "joint" model resulted in the smallest differences, followed by the "prior" and "plug-in" models.

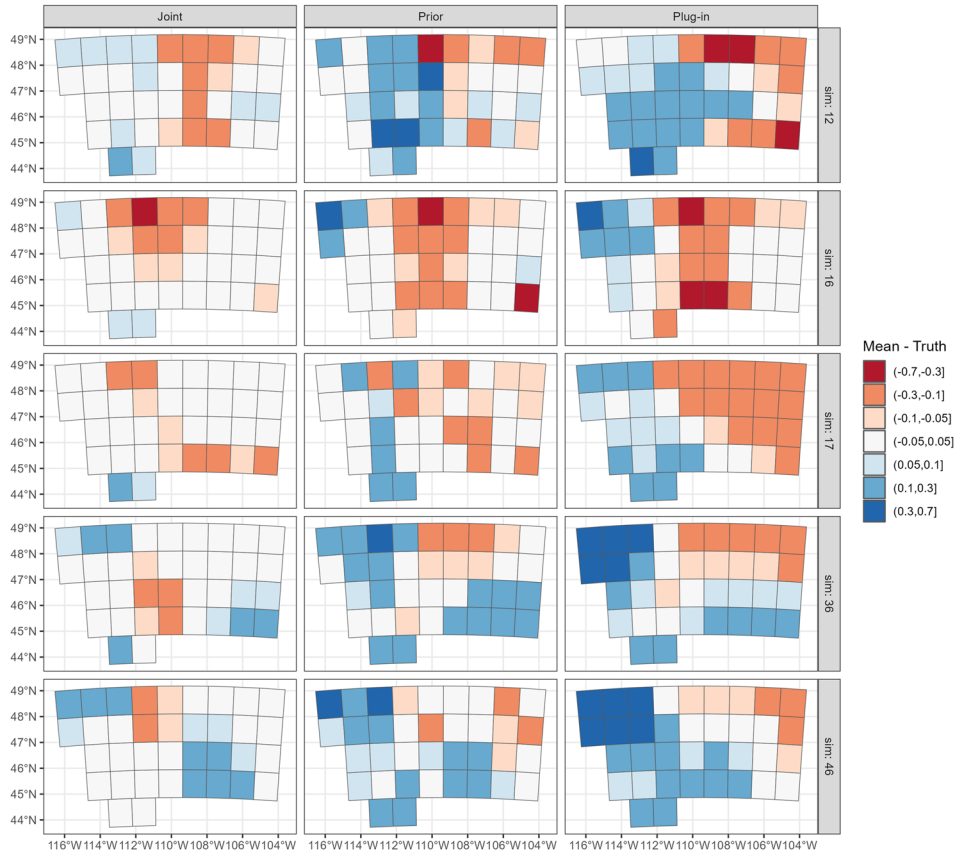

Figure 8. Difference between the posterior mean probability of *Pseudogymnoascus destructans* (*Pd*) occurrence $(\hat{x}_k^A)$ and true generating probability (x_k^A) for a subset of simulated data sets assuming a moderate impact of *Pd* occurrence on the log-mean count. On average, the "joint" model resulted in the smallest differences, followed by the "prior" and "plug-in" models.

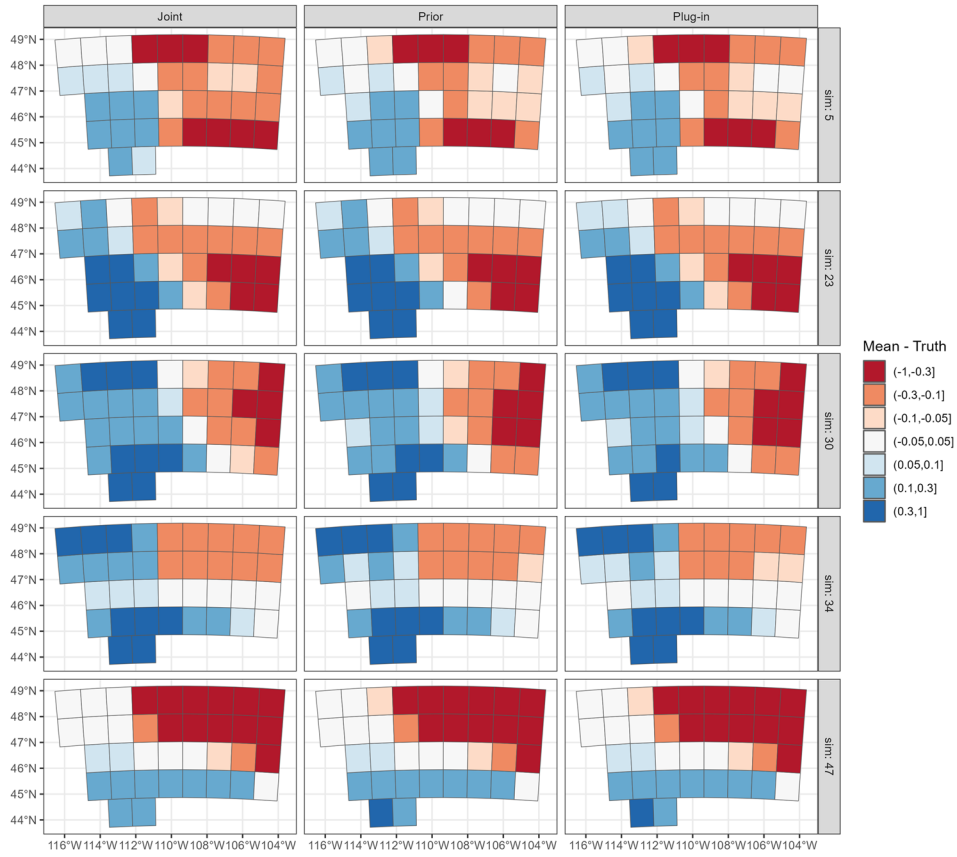

Figure 9. Difference between the posterior mean probability of *Pseudogymnoascus destructans* (*Pd*) occurrence $(\hat{x}_k^A)$ and true generating probability (x_k^A) for a subset of simulated data sets assuming no impact of *Pd* occurrence on the log-mean count. All models resulted in nearly equivalent differences.

REFERENCES

Bachen D, McEwan A, Burkholder B, Hilty S, Blurn S, Maxwell B (2018) Bats of Montana: identification and natural history. Report to Montana Department of Environmental Quality. Montana Natural Heritage Program, Helena, Montana

Banerjee S, Carlin BP, Gelfand AE (2015) Hierarchical modeling and analysis for spatial data, 2nd edn. Chapman & Hall, Boca Raton

Banner KM, Irvine KM, Rodhouse TJ, Wright WJ, Rodriguez RM, Litt AR (2018) Improving geographically extensive acoustic survey designs for modeling species occurrence with imperfect detection and misidentification. Ecol Evol 8(12):6144–6156

Blakey R, Law B, Straka T, Kingsford R, Milne D (2018) Importance of wetlands to bats on a dry continent: a review and meta-analysis. Hystrix Ital J Mammal 29:1–12

Blejwas K, Beard L, Buchanan J, Lausen CL, Neubaum D, Tobin A, Weller TJ (2023) Could White-nose syndrome manifest differently in *Myotis lucifugus* in western versus eastern regions of North America? A review of factors. J Wildl Dis 59(3):381–397

Brooks S, Gelman A (1998) General methods for monitoring convergence of iterative simulations. J Comput Graph Stat 7(4):434–455

Cameletti M, Gómez-Rubio V, Blangiardo M (2019) Bayesian modelling for spatially misaligned health and air pollution data through the INLA-SPDE approach. Spat Stat 31:100353

Campbell Grant EH, Mummah RO, Mosher BA, Evans J, DiRenzo GV (2023) Inferring pathogen presence when sample misclassification and partial observation occur. Methods Ecol Evol 14(5):1299–1311

Chambert T, Waddle JH, Miller DAW, Walls SC, Nichols JD (2018) A new framework for analysing automated acoustic species detection data: occupancy estimation and optimization of recordings post-processing. Methods Ecol Evol 9(3):560–570

Chaturvedi V, Springer DJ, Behr MJ, Ramani R, Li X, Peck MK, Ren P, Bopp DJ, Wood B, Samsonoff WA, Butchkoski CM, Hicks AC, Stone WB, Rudd RJ, Chaturvedi S (2010) Morphological and molecular characterizations of psychrophilic fungus *Geomyces destructans* from New York bats with White-nose syndrome (WNS). PLoS ONE 5(5):1–12

Cheng TL, Reichard JD, Coleman JTH, Weller TJ, Thogmartin WE, Reichert BE, Bennett AB, Broders HG, Campbell J, Etchison K, Feller DJ, Geboy R, Hemberger T, Herzog C, Hicks AC, Houghton S, Humber J, Kath JA, King RA, Loeb SC, Massé A, Morris KM, Niederriter H, Nordquist G, Perry RW, Reynolds RJ, Sasse DB, Scafini MR, Stark RC, Stihler CW, Thomas SC, Turner GG, Webb S, Westrich BJ, Frick WF (2021) The scope and severity of White-nose syndrome on hibernating bats in North America. Conserv Biol 35(5):1586–1597

Datta A, Banerjee S, Finley AO, Gelfand AE (2016) Hierarchical nearest-neighbor gaussian process models for large geostatistical datasets. J Am Stat Assoc 111(514):800–812

Eaton M (1983) Multivariate statistics: a vector space approach. John Wiley and Sons, Hoboken

Fan J, Zhang W (1999) Statistical estimation in varying coefficient models. Ann Stat 27(5):1491–1518

Ferrari S, Cribari-Neto F (2004) Beta regression for modelling rates and proportions. J Appl Stat 31(7):799–815

Finley AO (2011) Comparing spatially-varying coefficients models for analysis of ecological data with non-stationary and anisotropic residual dependence. Methods Ecol Evol 2(2):143–154

Fletcher RJ Jr, Hefley TJ, Robertson EP, Zuckerberg B, McCleery RA, Dorazio RM (2019) A practical guide for combining data to model species distributions. Ecology 100(6):e02710

Frank C, Davis A, Herzog C (2019) The evolution of a bat population with White-nose syndrome (WNS) reveals a shift from epizootic to an enzootic phase. Front Zool 16(40):1–9

Frick WF, Puechmaille SJ, Hoyt JR, Nickel BA, Langwig KE, Foster JT, Barlow KE, Bartonička T, Feller D, Haarsma A-J, Herzog C, Horáček I, van der Kooij J, Mulkens B, Petrov B, Reynolds R, Rodrigues L, Stihler CW, Turner GG, Kilpatrick AM (2015) Disease alters macroecological patterns of North American bats. Glob Ecol Biogeogr 24(7):741–749

Gargas A, Trest M, Christensen M, Volk T, Blehert D (2009) *Geomyces destructans* sp. nov. associated with bat White-nose syndrome. Mycotaxon 108(8):147–154

Gramacy RB, Apley DW (2015) Local gaussian process approximation for large computer experiments. J Comput Graph Stat 24(2):561–578

Hastie T, Tibshirani R (1993) Varying-coefficient models. J R Stat Soc Ser B Methodol 55(4):757–796

Hefley TJ, Hooten MB, Russell RE, Walsh DP, Powell JA (2017) When mechanism matters: Bayesian forecasting using models of ecological diffusion. Ecol Lett 20(5):640–650

Hicks LL, Schwab NA, Homyack JA, Jones JE, Maxell BA, Burkholder BO (2020) A statistical approach to White-nose syndrome surveillance monitoring using acoustic data. PLoS ONE 15(10):1–16

Irvine K, Gitelman A, Hoeting J (2007) Spatial designs and properties of spatial correlation: effects on covariance estimation. J Agric Biol Environ Stat 12:450–469

Lee A, Szpiro A, Kim S, Sheppard L (2015) Impact of preferential sampling on exposure prediction and health effect inference in the context of air pollution epidemiology. Environmetrics 26(4):255–267

Lee D, Mukhopadhyay S, Rushworth A, Sahu SK (2016) A rigorous statistical framework for spatio-temporal pollution prediction and estimation of its long-term impact on health. Biostatistics 18(2):370–385

Lee D, Shaddick G (2010) Spatial modeling of air pollution in studies of its short-term health effects. Biometrics 66(4):1238–1246

Liu H, Ong Y, Shen X, Cai J (2018) When Gaussian process meets big data: a review of scalable GPs. IEEE Trans Neural Netw Learn Syst 31:4405–4423

Loeb S, Rodhouse T, Ellison L, Lausen C, Reichard J, Irvine K, Ingersoll T, Coleman J, Thogmartin W, Sauer J, Francis C, Bayless M, Stanley T, Johnson D (2015) A plan for the North American Bat Monitoring Program (NABat). Gen. Tech. Report. SRS-208. U.S. Department of Agriculture Forest Service, Southern Research Station, Asheville

Lorch JM, Meteyer CU, Behr MJ, Boyles JG, Cryan PM, Hicks AC, Ballmann AE, Coleman JT, Redell DN, Reeder DM, Blehert DS (2011) Experimental infection of bats with *Geomyces destrucans* causes White-nose syndrome. Nature 480:376–378

MacKenzie DI, Nichols JD, Lachman GB, Droege S, Andrew Royle J, Langtimm CA (2002) Estimating site occupancy rates when detection probabilities are less than one. Ecology 83(8):2248–2255

Merkle JA, Cross PC, Scurlock BM, Cole EK, Courtemanch AB, Dewey SR, Kauffman MJ (2018) Linking spring phenology with mechanistic models of host movement to predict disease transmission risk. J Appl Ecol 55(2):810–819

Miller DAW, Pacifici K, Sanderlin JS, Reich BJ (2019) The recent past and promising future for data integration methods to estimate species' distributions. Methods Ecol Evol 10(1):22–37

Pacifici K, Reich BJ, Miller DAW, Gardner B, Stauffer G, Singh S, McKerrow A, Collazo JA (2017) Integrating multiple data sources in species distribution modeling: a framework for data fusion. Ecology 98(3):840–850

Pannullo F, Lee D, Waclawski E, Leyland A (2016) How robust are the estimated effects of air pollution on health? Accounting for model uncertainty using Bayesian model averaging. Spat Spatio Temporal Epidemiol 18:53–62

Polson NG, Scott JG, Windle J (2013) Bayesian inference for logistic models using Pólya-Gamma latent variables. J Am Stat Assoc 108(504):1339–1349

Powell H, Lee D (2013) Modelling spatial variability in concentrations of single pollutants and composite air quality indicators in health effects studies. J R Stat Soc Ser A Stat Soc 177(3):607–623

PRISM Climate Group (2022) Parameter-elevation Regressions on Independent Slopes Model. Oregon State University. https://prism.oregonstate.edu. Accessed 12 Dec 2022

Reeder DM, Frank CL, Turner GG, Meteyer CU, Kurta A, Britzke ER, Vodzak ME, Darling SR, Stihler CW, Hicks AC, Jacob R, Grieneisen LE, Brownlee SA, Muller LK, Blehert DS (2012) Frequent arousal from hibernation linked to severity of infection and mortality in bats with White-nose syndrome. PLoS ONE 7(6):1–10

Reichert B, Lausen C, Loeb S, Weller T, Allen R, Britzke E, Hohoff T, Siemers J, Burkolder B, Herzog C and Verant M (2018) A guide to processing bat acoustic data for the Northern American Bat Monitoring Program (NABat): U.S. Geological Survey open-file report 2018-1068, p 33. https://doi.org/10.3133/ofr20181068

Rodhouse TJ, Rodriguez RM, Banner KM, Ormsbee PC, Barnett J, Irvine KM (2019) Evidence of region-wide bat population decline from long-term monitoring and Bayesian occupancy models with empirically informed priors. Ecol Evol 9(19):11078–11088

Rue H, Martino S, Chopin N (2009) Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J. R. Stat. Soc. Ser. B Stat. Methodol. 71(2):319–392

Shewry MC, Wynn HP (1987) Maximum entropy sampling. J Appl Stat 14(2):165–170

Stan Development Team (2023a) RStan: the R interface to Stan. R package version 2.26.11. https://mc-stan.org/

Stan Development Team (2023b) Stan modeling language users guide and reference manual, 2.31. https://mc-stan.org

Stevens DL Jr, Olsen AR (2003) Variance estimation for spatially balanced samples of environmental resources. Environmetrics 14(6):593–610

Stratton C, Banner K, Irvine K (2023) Code vignette for application of the spatially misaligned regression model described in "Joint spatial modeling bridges the gap between disparate disease surveillance and population monitoring efforts informing conservation of at-risk bat species." Geological Survey software release, U.S. https://doi.org/10.5066/P9WQGCSN

Stratton C, Irvine KM, Banner KM, Wright WJ, Lausen C, Rae J (2022) Coupling validation effort with in situ bioacoustic data improves estimating relative activity and occupancy for multiple species with cross-species misclassifications. Methods Ecol Evol 13(6):1288–1303

Szewczak J (2023) SonoBat acoustic classification software, version 4.1. http://sonobat.com

Turner G, Reeder D, Coleman J (2011) A five-year assessment of mortality and geographic spread of White-nose syndrome in North American bats and a look to the future. Bat Res News 52(2):13–27

U.S. Fish and Wildlife Service (2011) A national plan for assisting states, federal agencies, and tribes in managing White-nose syndrome in bats

U.S. Fish and Wildlife Service (2012) U.S. Fish & Wildlife Service news release: North American bat death toll exceeds 5.5 million from White-nose syndrome. https://s3.us-west-2.amazonaws.com/prod-is-cms-assets/wns/prod/0c8dc8c0-8d41-11e9-93b7-61a86857b722-wns_mortality_2012_nr_final_0.pdf

U.S. Fish and Wildlife Service (2022a) Proposed rule: endangered species status for tricolored bat, 87 FR 56381, pp 56381–56393. https://www.federalregister.gov/documents/2022/09/14/2022-18852/endangered-and-threatened-wildlife-and-plants-endangered-species-status-for-tricolored-bat

U.S. Fish and Wildlife Service (2022b) Rule: endangered species status for northern long-eared bat, 87 FR 73488, pp 73488–73504. https://www.federalregister.gov/documents/2022/11/30/2022-25998/endangered-and-threatened-wildlife-and-plants-endangered-species-status-for-northern-long-eared-bat

U.S. Geological Survey (2019) U.S. Geological survey factsheet: designed surveillance for the White-nose syndrome fungus

U.S. Geological Survey (2022) National Wildlife Health Center protocol for non-lethal sampling methods for *Pseudogymnoascus destructans* (Pd) Surveillance

Wang Y, Le ND, Zidek JV (2020) Approximately optimal spatial design: how good is it? Spat Stat 37:100409

Warnecke L, Turner JM, Bollinger TK, Lorch JM, Misra V, Cryan PM, Wibbelt G, Blehert DS, Willis CK (2012) Inoculation of bats with European *Geomyces destructans* supports the novel pathogen hypothesis for the origin of White-nose syndrome. Proc Natl Acad Sci 109(18):6999–7003

Warren J, Fuentes M, Herring A, Langlois P (2012) Spatial-temporal modeling of the association between air pollution exposure and preterm birth: identifying critical windows of exposure. Biometrics 68(4):1157–1167

Watsa M, Wildlife Disease Surveillance Focus Group (2020) Rigorous wildlife disease surveillance. Science 369(6500):145–147

Wibbelt G, Kurth A, Hellmann D, Weishaar M, Barlow A, Veith M, Prüger J, Görföl T, Grosche L, Bontadina F, Zöphel U, Seidl H, Blehert D (2010) White-nose syndrome fungus (*Geomyces destructans*) in bats, Europe. Emerg Infect Dis 16(8):1237–1243

Wiens AM, Thogmartin WE (2022) Gaussian process forecasts *Pseudogymnoascus destructans* will cover coterminous United States by 2030. Ecol Evol 12(11):e9547

WNS Response Team (2023) Where is WNS now? White-nose syndrome response team spread map. https://www.whitenosesyndrome.org/where-is-wns. Accessed 16 June 2023

Wright WJ, Irvine KM, Almberg ES, Litt AR (2020) Modelling misclassification in multi-species acoustic data when estimating occupancy and relative activity. Methods Ecol Evol 11(1):71–81

Wright W, Litt A, Irvine K, Almberg E (2018) Estimating occupancy for Montana bat species prior to the arrival of White-nose syndrome. Montana Fish, Wildlife and Parks Report

Zhang Y, Chang HH, Iuliano A, Reed C (2022) Application of Bayesian spatial-temporal models for estimating unrecognized COVID-19 deaths in the United States. Spat Stat 50:100584

Zimmerman DL (2006) Optimal network design for spatial prediction, covariance parameter estimation, and empirical prediction. Environmetrics 17(6):635–652