ORIGINAL RESEARCH

Spatial Gaussian processes improve multi-species occupancy models when range boundaries are uncertain and nonoverlapping

Wilson J. Wright[1] | Kathryn M. Irvine[2] | Thomas J. Rodhouse[3] |
Andrea R. Litt[1]

[1]Department of Ecology, Montana State University, Bozeman, MT, USA

[2]U.S. Geological Survey, Northern Rocky Mountain Science Center, Bozeman, MT, USA

[3]National Park Service and Human and Ecosystem Resilience and Sustainability Lab, Oregon State University-Cascades, Bend, OR, USA

Correspondence
Kathryn M. Irvine, U.S. Geological Survey, Northern Rocky Mountain Science Center, Bozeman, MT, USA.
Email: kirvine@usgs.gov

Abstract

1. Species distribution models enable practitioners to analyze large datasets of encounter records and make predictions about species occurrence at unsurveyed locations. In omnibus surveys that record data on multiple species simultaneously, species ranges are often nonoverlapping and misaligned with the administrative unit defining the spatial domain of interest (e.g., a state or province). Consequently, some species display differentially restricted extents within a study area. Assuming hard boundaries based on expert opinion or published range maps to restrict species occurrence predictions implies a false sense of certainty in model-based inferences.

2. We propose a multi-species occupancy model with a spatial Gaussian process on site-specific effects for each species as a model-based solution. Specifying informative Bayesian hyperpriors on the spatial hyperparameters encapsulates broad-scale correlation among site occupancy probabilities for each species. We fit this model to acoustic detection/nondetection data collected with autonomous recording units during summer of 2016–2019 throughout Oregon and Washington, USA, on 15 bat species.

3. We found vast improvements in spatial predictions of spotted bat (*Euderma maculatum*), canyon bat (*Parastrellus hesperus*), and Brazilian free-tailed bat (*Tadarida brasiliensis*) when the available environmental predictors were insufficient for characterizing their restricted ranges within the region.

4. In contrast, widespread species (*Lasionycteris noctivagans*, *Myotis californicus*, *Myotis evotis*, *Myotis volans*) were appropriately modeled using only environmental predictors, such as percentage forest cover and cliff and canyon cover.

5. Utilizing spatial Gaussian processes within a community or multi-species model incorporates uncertainty in range boundaries and allows for simultaneous predictions for the entire faunal assemblage even if species have nonoverlapping or restricted ranges within a spatial domain of interest. Such modeling improvements

are essential if species distribution models are to accurately inform monitoring, species recovery plans, and other conservation efforts.

1 | INTRODUCTION

Modeling the distribution of species across large geographic extents has emerged as a cornerstone practice of ecology and conservation biology (Elith & Leathwick, 2009; Franklin, 2010; Guisan & Thuiller, 2005). Species distribution models (SDMs) enable practitioners to analyze large datasets of encounter records and predict species occurrence at unsurveyed locations. This capability is particularly helpful to address questions about rare and cryptic species of uncertain conservation status where more direct methods of counting individuals are infeasible. Advanced technologies, such as autonomous recording units (ARUs) and camera traps, facilitate efficient collection of occurrence data over expansive geographic extents and for multiple species concurrently (Gibb et al., 2019; Kucera & Barrett, 2011). A common inferential goal for such omnibus surveys is to estimate distributions for multiple species simultaneously within a community or regional faunal assemblage to support species conservation assessments. Multi-species datasets, however, present analytical challenges when estimating species distributions because extents of occurrences (sensu Gaston & Fuller, 2009), the exterior polygons of species' ranges, are often not aligned with one another or study regions. Ideally, inferences and predictions for each species are limited to be within range boundaries, but appropriately defining these boundaries is an impediment to distribution modeling because species ranges are uncertain.

Species distributions are often predicted across arbitrary extents (e.g., country, administrative, or ecoregional boundaries) that may not align with biological ranges. Failing to constrain inferences based on species' extents of occurrence could result in unrealistic predictions because the entire study region is assumed to be within the range of each assemblage member. Some approaches use published range maps, expert opinion, or estimated convex hulls to mask out areas thought to be beyond each species' range, where SDMs otherwise might predict a high probability of occupancy based on environmental variables (Graham & Hijmans, 2006; Ko et al., 2016; Meyer et al., 2017; Peterson et al., 2016; Rodhouse et al., 2015; VanDerWal et al., 2009). However, these approaches are unsatisfying because they result in hard boundaries for each species that are not biologically realistic and empirical data often show species are detected outside assumed ranges (e.g., Peterson et al., 2016). Furthermore, defining range boundaries becomes more onerous when considering that species are detected imperfectly and distributions are dynamic over time.

Alternatively, estimation of species range boundaries can be incorporated directly into SDMs using spatial covariance functions that utilize proximity among detections in addition to environmental predictors to allow for data-informed spatial correlation among occurrence probabilities. Spatial patterns in ecological data occur at multiple scales (Levin, 1992), and species ranges result in species distribution data having broad-scale spatial correlation. It is well established that SDMs can benefit by accounting for spatial correlation (Elith & Leathwick, 2009), and a variety of methods are available to model spatial patterns in species distributions (e.g., Johnson et al., 2013; Ovaskainen et al., 2016; Rushing et al., 2019). Incorporating spatial correlation can improve inferences by reducing bias and increasing the predictive performance of SDMs (Guélat & Kéry, 2018; Ovaskainen et al., 2016). However, spatial autocorrelation is more than just a nuisance to be accounted for; it is also a key source of ecological information relevant to estimating both extent of occurrence (the range) as well as the area of occupancy within the outer extent that is typically the subject of a SDM.

Occupancy models are widely used for analyzing encounter datasets because they explicitly account for imperfect detection while estimating species occurrence (MacKenzie et al., 2002). By relying on detection/nondetection observations, occupancy models provide a framework for analyzing data that are more readily collected over large geographic extents (MacKenzie et al., 2002; Noon et al., 2012) and for multiple species (Manley et al., 2005; Noon et al., 2012)—the types of datasets that require the consideration of uncertain and nonoverlapping species ranges. However, applications of spatial occupancy models typically focus on accounting for fine-scale spatial patterns (among clusters of adjacent sites) compared with those resulting from broader spatial processes manifesting in species range boundaries (substantial portions of the study area with no evidence of a species). For instance, conditional autoregressive models account for fine-scale spatial correlation (Lichstein et al., 2002) and are often used to incorporate spatial correlation among adjacent sites in occupancy models (Johnson et al., 2013). Gaussian processes, on the other hand, provide a flexible framework for analyzing spatial data (Gelfand & Schliep, 2016) and may be able to more directly model spatial patterns over large geographic extents because they estimate a length-scale (or "effective range") parameter that defines a variance–covariance matrix for the site effects based on the distances between them.

Recent applications of Gaussian processes within occupancy models use latent variable modeling to account for both spatial correlation and interactions among species (Ovaskainen et al., 2016; Tikhonov et al., 2020). However, some suggest caution in interpreting spatial patterns of co-occurrence as evidence of species interactions (Blanchet et al., 2020). Another interpretation is that

unmeasured covariates are related to the occupancy probabilities of each species and modeling these latent variables as Gaussian processes allow them to be spatially autocorrelated. These models simultaneously estimate values for the unmeasured covariates ("latent factors") at each site and parameters ("factor loadings") relating these to occupancy probabilities for each species (Ovaskainen et al., 2016). In general, spatial correlation in ecological data is often interpreted as resulting from unmeasured covariates but also arises because many ecological processes are inherently spatial (Guisan & Thuiller, 2005; Legendre & Fortin, 1989). Species ranges, for example, can reflect species dispersal characteristics and biogeographic barriers that may be better conceptualized by purely spatial processes in statistical models.

We explored how broad-scale spatial patterns, such as those resulting from species ranges, could be incorporated in occupancy models using Gaussian processes. While previous applications of Gaussian processes in occupancy models focused on modeling a single time period, our approach analyzes multiple years of data and includes temporal correlation in the spatial site-specific coefficients for each species. This harnesses the knowledge that broad-scale spatial patterns are similar over consecutive years but still allows for shifts in species distributions to be estimated. Our approach estimates the distributions of multiple species simultaneously while still allowing for unique spatial patterns, including range boundaries, for each species. The multi-species framework is advantageous because information can be shared across species when estimating regression (e.g., Dorazio & Royle, 2005) even though the extents of occurrences may differ across species. We were motivated to explore including broad-scale spatial patterns in occupancy models by our analyses of multi-species acoustic data for bats in Oregon and Washington, USA. These data were collected with ARUs from 2016 to 2019 following the protocols developed for the North American Bat Monitoring Program (NABat; https://www.nabatmonitoring.org; Loeb et al., 2015; Rodriguez et al., 2019) to estimate the distributions of bat species over broad spatial scales and monitor populations for changes in occupancy. We use these data to demonstrate how Gaussian processes can improve predictions and produce more realistic species distribution maps when fitting models for species with differentially restricted ranges over a study area.

2 | MATERIALS AND METHODS

2.1 | Data

We used detection/nondetection data from acoustic bat surveys conducted across Oregon and Washington, USA, during the summers of 2016–2019. The dataset was a product of NABat administered by a regional collaborative, the Northwestern Bat Hub (Rodriguez et al., 2019, https://osucascades.edu/HERS/northwestern-bat-hub), in which ARUs were deployed at $n = 202$ sample units following a spatially balanced random master sample of a grid-based sampling frame (Larsen et al., 2008; Stevens & Olsen, 2004).

Autonomous recording units passively record the echolocation calls made by bats, and these acoustic recordings are used to identify the bat species detected at each sample unit. Within each 10×10 km grid cell (sample unit), surveys were performed at four locations and a species was detected (1) or not (0) based on the nightly recordings at each location. These data were used to construct "detection histories" (MacKenzie et al., 2002) for each species at each sample unit. Detections were determined via a workflow that used a combination of automated bat call classification and manual verification (Banner et al., 2018; Rodriguez et al., 2019). We followed the Rodriguez et al. (2019) protocol to guide detector placement decisions and additional details about the field survey protocol can be found in that report.

In addition to recording bat calls, we also summarized local-scale conditions surrounding the recorder and during the survey period for inclusion in models as detectability covariates. We estimated vegetation clutter in the recording space, as clutter influences both bat microhabitat and the ability to record identifiable bat echolocation calls. We distinguished whether recording occurred at a water feature or not, as bats in our region often congregate around water for drinking and foraging but not all survey locations provided a water feature for ARU deployment. We also obtained estimates of average minimum temperature during the survey period from the 1-km resolution Daily Surface Weather and Climatological Summaries (DAYMET) dataset (Thornton et al., 2018), as summertime bat activity in our region declines during cold weather events. Clutter, an indicator for water feature, and temperature are outlined by NABat protocols as important predictors for bat detectability when performing acoustic surveys (Loeb et al., 2015).

Land cover estimates for each grid cell were used as predictors for occupancy. We estimated mean percent forest cover and cliff and canyon cover for each grid cell using the 30-m resolution GAP/LANDFIRE National Terrestrial Ecosystems dataset (Homer et al., 2015). We also summarized 30-year mean annual precipitation (PRISM Climate Group; Daly et al., 2008) for each grid cell as an additional occupancy covariate. These covariates were chosen because they were identified during previous studies (e.g., Rodhouse et al., 2015) as being important to regional bat biogeography. When analyzing data for multiple years, we assumed that the environmental predictors for occupancy were static over time.

2.2 | Bayesian spatial occupancy model

We analyzed these data using an occupancy model that included environmental predictors (mean forest cover, cliff/canyon cover, 30-year mean annual precipitation) and spatially correlated site-specific effects (analogous to frequentist "random effects"). First, we describe the model for a single year (i.e., "single-season") with a focus on accounting for residual spatial correlation, which could include the patterns associated with species' ranges. The next section expands on our spatial model to include multiple years of data with temporal correlation in the site effects over consecutive years. Let

$i = 1, \cdots, n$ denote the surveyed sites, which are NABat grid cells in our data example, and $k = 1, \cdots, K$ denote the different species. We modeled occupancy for each species as

$$Z_{ik} \sim \text{Bernoulli}\left(\psi_{ik}\right), \tag{1}$$

$$\Phi^{-1}(\psi_{ik}) = x_i\alpha_k + \eta_{ik}, \tag{2}$$

where Φ^{-1} denotes the probit link, x_i is a row vector containing environmental predictors for site i, α_k is a vector of occupancy-level predictor coefficients for species k, and η_{ik} is a site- and species-specific coefficient. We assumed the site-specific effects for each species were realizations from a spatial Gaussian process to add spatial correlation among occupancy probabilities. Therefore, the vector of site effects for a species, $\eta_k \equiv [\eta_{1k}, \eta_{2k}, ..., \eta_{nk}]^T$, was specified as

$$\eta_k \sim \text{MultivariateNormal}(0, \Sigma_k), \tag{3}$$

where each Σ_k is a $n \times n$ variance–covariance matrix defined by the intersite distances using the exponentiated quadratic covariance function with parameters for the marginal standard deviation, τ_k, and the length-scale, ρ_k. This defined the covariance between sites i and i' as $\tau_k^2 \exp(-(1/2\rho_k^2)(\|s_i - s_{i'}\|)^2)$, where s_i denotes the spatial coordinates of site i and $\|s_i - s_{i'}\|$ is the Euclidean distance between the two sites (p. 501 Gelman et al., 2013).

The marginal standard deviation (τ_k) and length-scale (ρ_k) control the magnitude and smoothness of the possible realizations from a spatial Gaussian process. We used Bayesian methods when fitting this model and specified hyperprior distributions for τ_k and ρ_k that reflected our interest in modeling broad-scale patterns that reflect species' ranges within the study area. The marginal standard deviation controls the magnitude of variability in the site effects, and we used Half-Normal(0, 1) hyperprior distributions for each τ_k. This places the most hyperprior density on values close to zero, allowing all site coefficients to be near zero if the data provide little evidence of residual spatial correlation. This hyperprior also restricts the posterior distributions for τ_k away from larger values that are not realistic based on the scale of the probit link. For the length-scale parameters, ρ_k, we assumed Inverse-Gamma (16.9, scale = 17.9) hyperprior distributions. This simultaneously restricts the posterior distribution for ρ_k from including values that are too small or too large based on the observed intersite distances and spatial patterns of interest. See Appendix S1 for how we justify these informative hyperprior distributions based on exploring visualizations of realizations from spatial Gaussian processes with different parameters.

The observation component of this model accounted for imperfect detection of each species at occupied sites. Because manual verification of acoustic files was used to exclude any ambiguous species detections, we assumed no false-positive detections were included in this analysis and only accounted for false negatives. Let $j = 1, \cdots, J_i$ denote the surveys for site i. We modeled detections as

$$\left[Y_{ijk} | Z_{ik} = 1\right] \sim \text{Bernoulli}(p_{ijk}), \tag{4}$$

$$\Phi^{-1}(p_{ijk}) = v_{ij}\beta_k, \tag{5}$$

where v_{ij} is a row vector of covariates for visit j to site i and β_k is a vector of detection coefficients. Covariates included nightly variables (minimum temperature, day length) and environmental characteristics in the immediate vicinity of the detector (measure of clutter, indicator for water body) that could influence bat activity and/or detection.

For each occupancy and detection environmental predictor coefficient, we assumed a hierarchical structure that allowed for partial pooling across species (Dorazio & Royle, 2005). That is, we let $\alpha_k \sim \text{MultivariateNormal}(\mu_\alpha, \text{I}\sigma_\alpha^2)$ and $\beta_k \sim \text{MultivariateNormal}(\mu_\beta, \text{I}\sigma_\beta^2)$ for each species. The parameters $\mu_\alpha, \mu_\beta, \sigma_\alpha$, and σ_β are vectors of the means and standard deviations for the coefficients. This structure allowed information to be borrowed across species for a specific coefficient (e.g., the relationship between occupancy and forest cover could be similar across species). We specified hyperprior distributions as Normal(0, 25) for the intercept means in μ_α and μ_β, Normal(0, 6.25) for the slope means in μ_α and μ_β, and Gamma (4, rate = 1) for each element of σ_α and σ_β.

2.3 | Adding temporal correlation

To incorporate data from multiple years, we expanded the model structure for the site-specific effects. While we expected similar broad-scale spatial patterns in occupancy for a species over consecutive years, we also wanted to estimate different site effects each year to allow for changes in a species' range or distribution. To capture these characteristics, we considered the site effects for each species as multiple, correlated realizations of a spatial Gaussian process (Gelfand & Schliep, 2016). We assumed an autoregressive correlation structure over time (specifically AR(1)) with correlation parameter ϕ_k for each species, using Uniform(0, 1) hyperprior distributions for these parameters. With $t = 1, \cdots, T$ denoting years, the full model became

$$Z_{ikt} \sim \text{Bernoulli}(\psi_{ikt}), \tag{6}$$

$$\Phi^{-1}(\psi_{ikt}) = x_i\alpha_k + \eta_{ikt}, \tag{7}$$

$$\left[Y_{ijkt} | Z_{ijkt} = 1\right] \sim \text{Bernoulli}(p_{ijkt}), \tag{8}$$

$$\Phi^{-1}(p_{ijkt}) = v_{ijt}\beta_k. \tag{9}$$

The vectors of site coefficients for each year, $\eta_{kt} \equiv [\eta_{1kt}, \eta_{2kt}, ..., \eta_{nkt}]^T$, were then modeled as

$$\begin{pmatrix} \eta_{k1} \\ \eta_{k2} \\ \vdots \\ \eta_{kT} \end{pmatrix} \sim \text{MultivariateNormal} \left[\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_k & \phi_k\Sigma_k & \cdots & \phi_k^{T-1}\Sigma_k \\ \phi_k\Sigma_k & \Sigma_k & \cdots & \phi_k^{T-2}\Sigma_k \\ \vdots & \vdots & \ddots & \vdots \\ \phi_k^{T-1}\Sigma_k & \phi_k^{T-2}\Sigma_k & \cdots & \Sigma_k \end{pmatrix} \right]. \tag{10}$$

If there is no temporal correlation (i.e., $\phi_k = 0$), the variance–covariance matrix in Equation (10) is block diagonal and results in independent realizations of the spatial Gaussian process each year. Incorporating the correlation parameters ϕ_k harnessed the fact that we expected similar broad-scale spatial patterns for a species over consecutive years. The detection component of this model (Equations 8 and 9) is unchanged except for the additional subscript for each year.

We assumed the occupancy and detection coefficients (α_k and β_k, respectively) were constant over time. Any differences in occurrence over time were modeled with the site effects from each year. The parameterization of our model did not explicitly estimate local extinction and colonization probabilities conditional on the occupancy state (Z_{ik}) from the previous year, as often done when fitting dynamic occupancy models (e.g., MacKenzie et al., 2003). A Markovian structured covariate for occupancy in years $t = 2, \cdots, T$ could be considered in future analyses as more data become available.

2.4 | Fitting models and summarizing results

We compared inferences from our analysis with a simpler occupancy model without the spatiotemporal Gaussian process. This "basic" model included the same environmental predictors for occupancy and detection but also included intercepts for occupancy that varied by year to account for any overall changes in occupancy over time. We did not include these varying intercepts in the spatiotemporal model because unique site coefficients were estimated each year and this allowed for differences in occupancy over time for each species. To compare the predictive performance for the two fitted models, we calculated the area under the receiver operating characteristic curve (AUC) using the posterior distributions of the occupancy probabilities (ψ_{ikt}) and latent occupancy states (Z_{ikt}) at surveyed sites (Zipkin et al., 2012). In general, accounting for residual spatial correlation can improve the predictive performance of occupancy models (e.g., Ovaskainen et al., 2016). However, the use of AUC for SDMs has been criticized (e.g., Lobo et al., 2008), so we also visually compared maps of predicted occupancy probabilities across the two models. These predictions were obtained for all 4,492 NABat grid cells (sites) in Oregon and Washington. For our spatiotemporal model, predictions are based on the site-level environmental variables and the site-specific coefficient (η_{ikt}) from the Gaussian process. See Appendix S2 for additional details on obtaining posterior draws for the η_{ikt} parameters at unsurveyed sites. For each species, we created maps showing the posterior mean occupancy probability and width of the 95% posterior interval for occupancy at each grid cell every year. These posterior summaries help visualize the patterns in occupancy over time and the associated uncertainty in the model predictions. We compare the predicted occurrence maps to species ranges from the U.S. Geological Survey (USGS; Gap Analysis Project, 2018).

We used R (version 3.6.3; R Core Team, 2020) and the tidyverse package (version 1.3.0; Wickham et al., 2019) to analyze these data.

Models were fit with (Carpenter et al., 2017) using the rstan package (version 2.19.3; Stan Development Team, 2020). We fit models with four independent chains of 4,000 iterations and discarded the first half of each chain as burn-in (leaving 2,000 iterations per chain for inference). We assessed model convergence by checking $\hat{R}$ values (Gelman & Rubin, 1992) and examining parameter traceplots for stationarity and mixing. Posterior draws for unsurveyed sites were obtained using functions coded in C++ and accessed in R using the Rcpp (version 1.0.4.6; Eddelbuettel & François, 2011) and RcppEigen (version 0.3.3.7.0; Bates & Eddelbuettel, 2013) packages. We created the figures displaying these maps using the ggplot2 (version 3.3.0; Wickham, 2016) and sf (version 0.9–2; Pebesma, 2018) packages. Our bat dataset and code to fit models are available at https://irma.nps.gov/DataStore/Reference/Profile/2275603.

3 | RESULTS

Our analysis included 4 years of data for 15 species, but we focus on the 2019 results for three species to highlight how incorporating a spatial Gaussian process can affect occupancy predictions depending on the spatial extent of species detections observed in the data. Maps showing the estimated probabilities of occupancy for each year and all species are available in Appendix S3. Two species we highlight here—the Brazilian free-tailed bat (*Tadarida brasiliensis*; TABR) and the western small-footed myotis (*Myotis ciliolabrum*; MYCI)—are assumed to have restricted ranges across the study area. The final species we focus on, the silver-haired bat (*Lasionycteris noctivagans*; LANO), is widespread.

Compared with the basic model, which excluded the site-specific effects, the spatial Gaussian process model strongly restricted the occupancy predictions for TABR to the southwest corner of Oregon, consistent with where this species is believed to occur within our study area (Gap Analysis Project, 2018, see Figure 1). While this pattern is evident in locations of species detections, the environmental predictors alone were not able to capture the restricted range of TABR and the basic model produced unrealistic predictions that were inconsistent with the observed data. The TABR occupancy predictions from the spatiotemporal analysis are consistent with the assumed range from USGS, but the spatial Gaussian process allows for gradual changes in occupancy instead of imposing an abrupt boundary in the predicted distribution (Figure 1). Additionally, in multiple years TABR was detected in grid cells outside its assumed range (Figure 1, Appendix S3).

Even though MYCI also has a restricted range in Oregon and Washington, there were less substantial differences in the predictions between the basic and spatiotemporal models (Figure 2). In this case, the environmental predictors were able to capture the broad-scale spatial pattern in species detections and appropriately limited the larger estimated probabilities of occupancy to the eastern portion of this region. The η parameters for MYCI reflected its assumed range but had little impact on the model-based predictions of occupancy (compare basic and spatiotemporal maps in Figure 2).
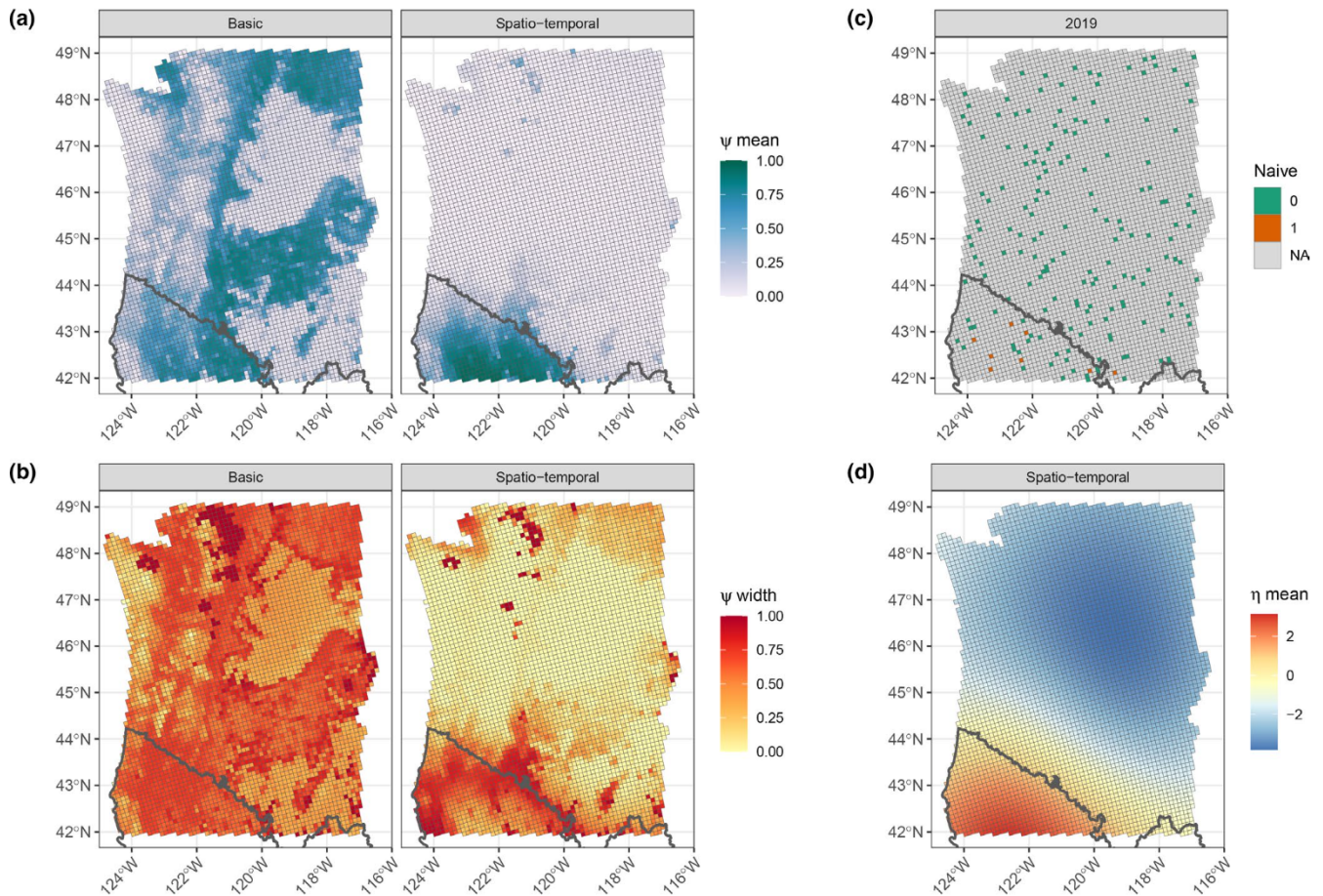
FIGURE 1 Comparison of 2019 occupancy predictions for the Brazilian free-tailed bat (*Tadarida brasiliensis*; TABR) from a model including only environmental predictors ("Basic") and one that also included a Gaussian process ("Spatiotemporal"). Occupancy posterior means (a) and 95% posterior interval width (b) are shown for both models. Naive occupancy at sampled grid cells (c) shows which grid cells had at least one detection in 2019. The posterior means of the spatially correlated site effects (η parameters from the spatiotemporal model) are shown in (d). The thick line in each panel shows the assumed species range from the USGS (Gap Analysis Project, 2018)

In contrast to the other two species, LANO is widespread throughout and does not have a restricted range in this region. For LANO, the occupancy predictions and associated uncertainty from the two fitted models were similar and there was little evidence for strong patterns of residual spatial correlation (Figure 3, note the scale of the η parameters). It may not be necessary to include the Gaussian process when estimating occupancy for LANO because there were no obvious differences in predictions across the two models (Figure 3).

Including the spatiotemporal Gaussian process increased the posterior mean AUC values for 14 of 15 species compared with the basic model fit (Appendix S3). The only species with a lower AUC was the Townsend's big-eared bat (*Corynorhinus townsendii*; COTO), which is generally difficult to detect acoustically and had the most posterior uncertainty for its AUC values. For the remaining species, the magnitude of increase in AUC for the spatiotemporal occupancy model varied. Many species had small AUC increases with the 95% posterior intervals for this metric mostly overlapping for the two models (e.g., LANO). Other species showed more substantial increases in AUC, indicating they had more residual spatial correlation when only the environmental predictors were used to model

occupancy probabilities. Many species with larger increases in AUC for the spatiotemporal occupancy model have restricted ranges in Oregon and Washington, including the spotted bat (*Euderma maculatum*; EUMA), the canyon bat (*Parastrellus hesperus*, PAHE), and TABR. The differences in AUC between the two models were most pronounced when the included environmental predictors did not adequately capture the spatial patterns in occupancy associated with species' restricted ranges.

The species with the largest AUC increases from the spatiotemporal occupancy model generally also had larger estimated marginal standard deviations (τ) for the associated Gaussian processes (Figure 4a shows our three focal species). For instance, the posterior distribution of τ for TABR includes values further from zero compared with that of LANO. In this way, posterior summaries for τ_k provide another way to evaluate the importance of the Gaussian process for each species. The posterior distributions of the length-scale (ρ) and temporal correlation (ϕ) parameters showed some variability across species but were often similar to the assumed hyperprior distributions (Figure 4b,c). The data were unable to inform the posterior distributions of these parameters when there was little residual spatial correlation (τ near zero) after accounting for the environmental
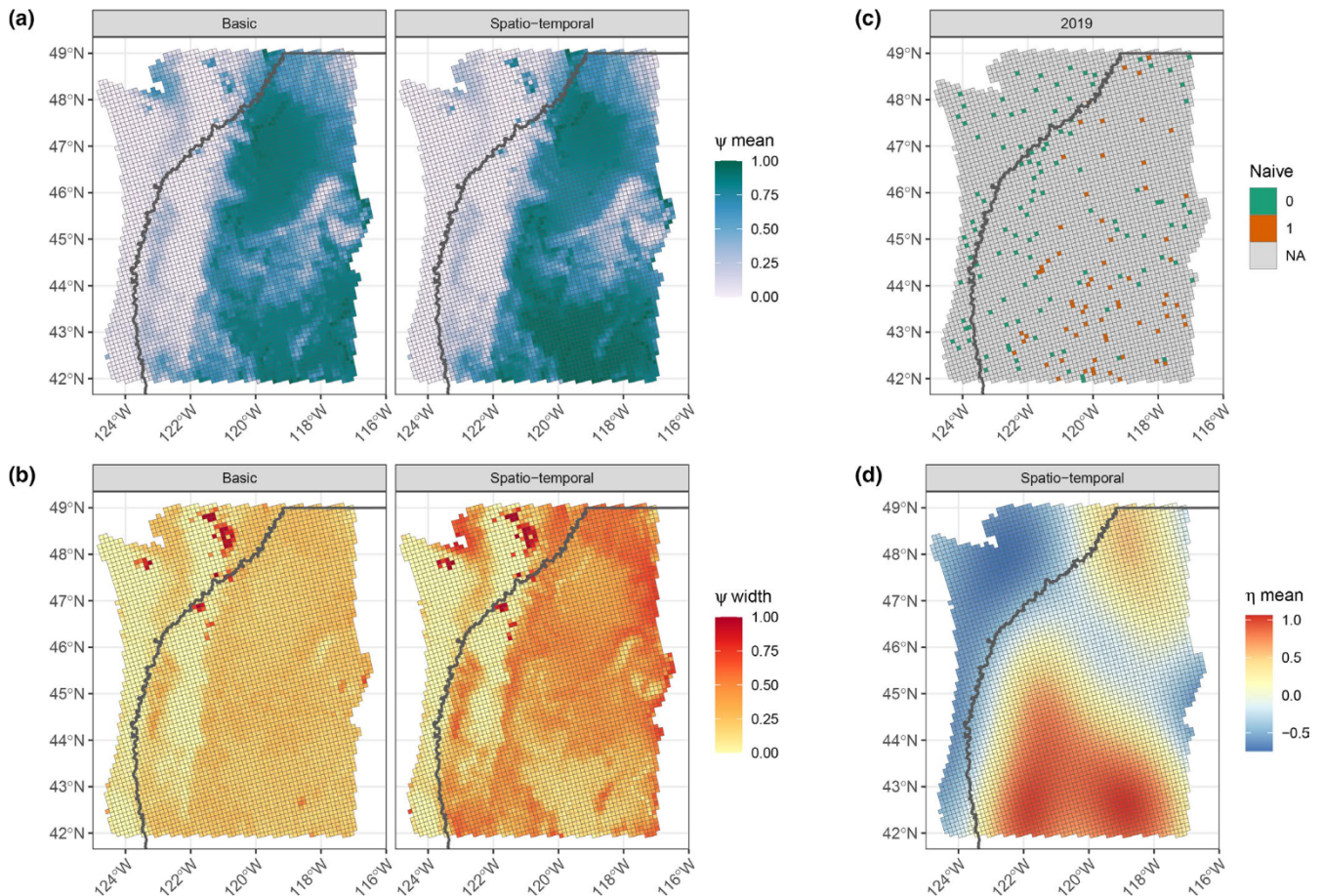
FIGURE 2 Comparison of 2019 occupancy predictions for the western small-footed myotis (*Myotis ciliolabrum*; MYCI) from a model including only environmental predictors ("Basic") and one that also included a Gaussian process ("Spatiotemporal"). Occupancy posterior means (a) and 95% posterior interval widths (b) are shown for both models. Naive occupancy at sampled grid cells (c) shows which grid cells had at least one detection in 2019. The posterior means of the spatially correlated site coefficients (η parameters from the spatiotemporal model) are shown in (d). The thick line in each panel shows the assumed species range the USGS (Gap Analysis Project, 2018)

predictors (e.g., LANO). When the Gaussian process was needed to capture more substantial residual spatial patterns in the data, the posterior distributions for ρ and ϕ deviated more from the assumed hyperprior distributions (e.g., TABR and MYCI). The length-scale parameters, ρ, were difficult to estimate in general, and their posterior distributions were strongly influenced by the hyperprior distribution. The posterior distributions for ϕ were concentrated near one for most species. As expected, this high temporal correlation resulted in similar spatial patterns in each species' estimated occupancy probabilities over time (see Appendix S3).

4 | DISCUSSION

The NABat data from Oregon and Washington exemplify the challenges associated with estimating distributions for multiple species across broad geographic extents. Collecting data and making inferences within administrative boundaries, in this case state borders, means model predictions will likely need to account for the restricted ranges of some species. We applied spatial Gaussian processes to

model these broad-scale spatial patterns and found that improvement for inferences on range-restricted species depended on how their extents of occurrence related to the environmental predictors available for modeling. The distribution of MYCI is limited to the eastern portion of Oregon and Washington, aligning with more arid areas, and the spatially explicit environmental variables were generally able to capture this geographic pattern. When the environmental predictors were not aligned with a species' restricted range, however, adding a spatial Gaussian process to account for these broad-scale spatial patterns improved predictions more substantially. For instance, the distribution limits of TABR in Oregon and Washington are not well understood and recent acoustic detections suggest the species occurs farther north than previously believed (Ommundsen et al.,2017). Because TABR is a generalist species and capable of long-distance dispersal, its range boundaries can shift rapidly and are difficult to define (Genoways et al., 2000; McCracken et al., 2018). These characteristics help explain why environmental predictors alone were inadequate when making predictions for TABR and we needed to also include the spatial Gaussian process to estimate its distribution. Overall, our example analysis highlights
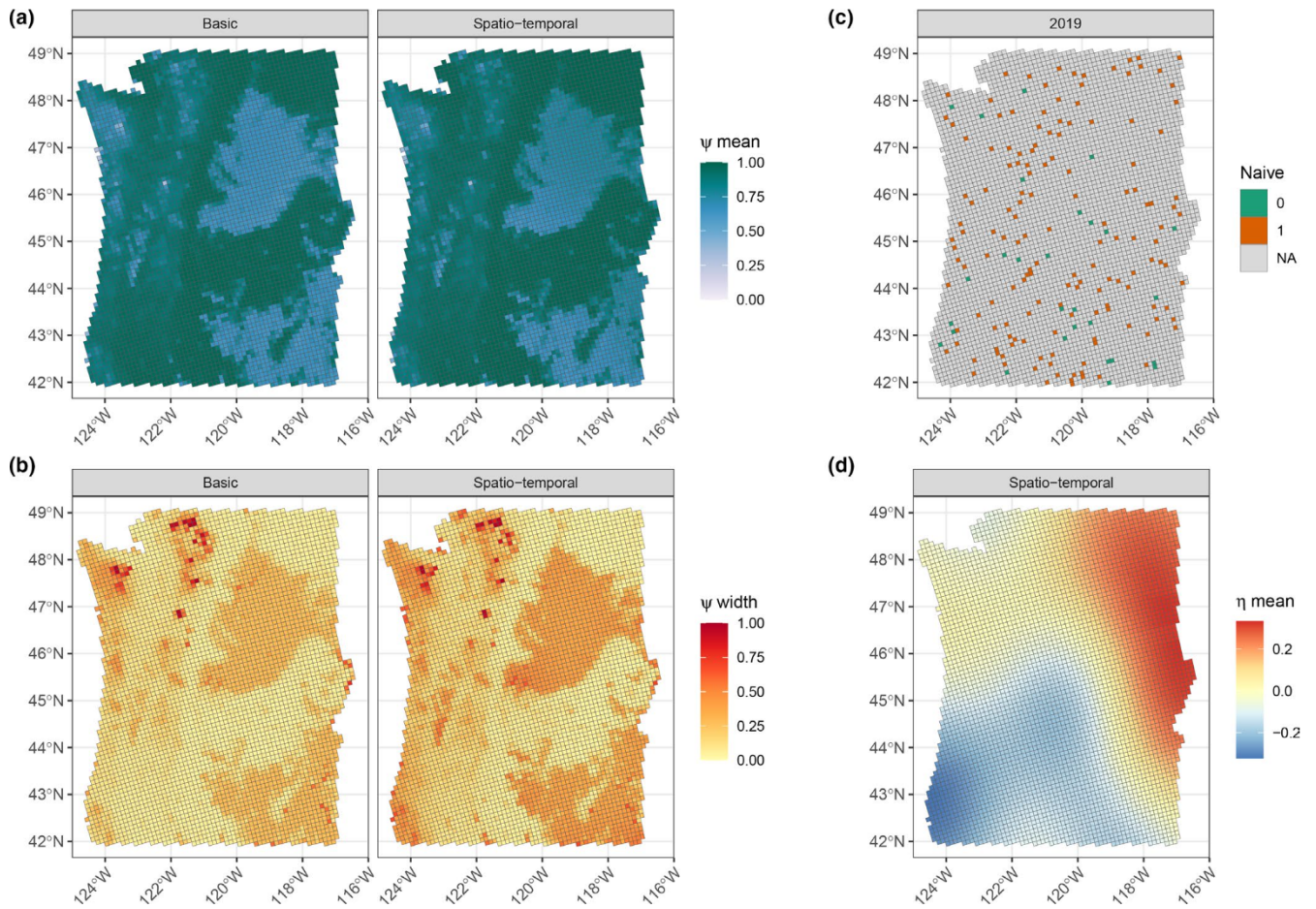
FIGURE 3 Comparison of 2019 occupancy predictions for the silver-haired bat (*Lasionycteris noctivagans*; LANO) from a model including only environmental predictors ("Basic") and one that also included a Gaussian process ("Spatiotemporal"). Occupancy posterior means (a) and 95% posterior interval widths (b) are shown for both models. Naive occupancy at sampled grid cells (c) shows which grid cells had at least one detection in 2019. The posterior means of the spatially correlated site effects (η parameters from the spatiotemporal model; d) are small, likely because LANO does not have a restricted range in this region

the importance of collecting meaningful environmental information and incorporating spatially explicit predictors into SDMs to account for heterogeneity in occupancy probabilities when possible. Spatial patterns may still persist, however, and Gaussian processes provide an approach for modeling the broad-scale spatial patterns that can result from species having differentially restricted ranges within a study area.

Some approaches rely on restricting inferences from SDMs to assumed boundaries based on published range maps, expert opinion, or convex hulls. Abrupt boundaries are biologically unrealistic because rates of extinction and turnover may be higher at the edges of species' ranges, compared with core areas (Brown et al., 1995; Curnutt et al., 1996; Doherty et al., 2003). In contrast, using a spatial Gaussian process does not assume a species' range is known and instead allows it to be estimated from observed data. Additionally, Gaussian processes allow for gradual declines in predicted occupancy probabilities and can result in more uncertainty around range borders (e.g., Figure 1 and TABR maps in Appendix S3). Relying on assumed range boundaries for our example analysis would have resulted in overly restrictive predictions because most species with

a previously published range map were detected outside these boundaries at least once (Appendix S3). This pattern of observing species outside their assumed ranges has been reported by others (e.g., Peterson et al., 2016) and is consistent with NABat data from South Carolina, USA (Neece et al., 2019). Spatial Gaussian processes allow these range boundaries to be informed by empirical observations and updated over time, without making strong assumptions on the potential spatial patterns in occupancy probabilities that may be observed for each species.

We focused on accounting for the broad-scale spatial patterns associated with species ranges, but other spatial occupancy models emphasize modeling fine-scale spatial patterns using conditional autoregressive models (e.g., Johnson et al., 2013). In some cases, ecological data may have spatial patterns across multiple scales simultaneously (Levin, 1992). We did not explore accounting for fine-scale spatial patterns because the spatially balanced sampling design used by NABat results in few adjacent grid cells being sampled. However, when the available data allow, models could be fit to account for both broad-scale spatial patterns using Gaussian processes and fine-scale patterns among adjacent sites (e.g., by also adding a
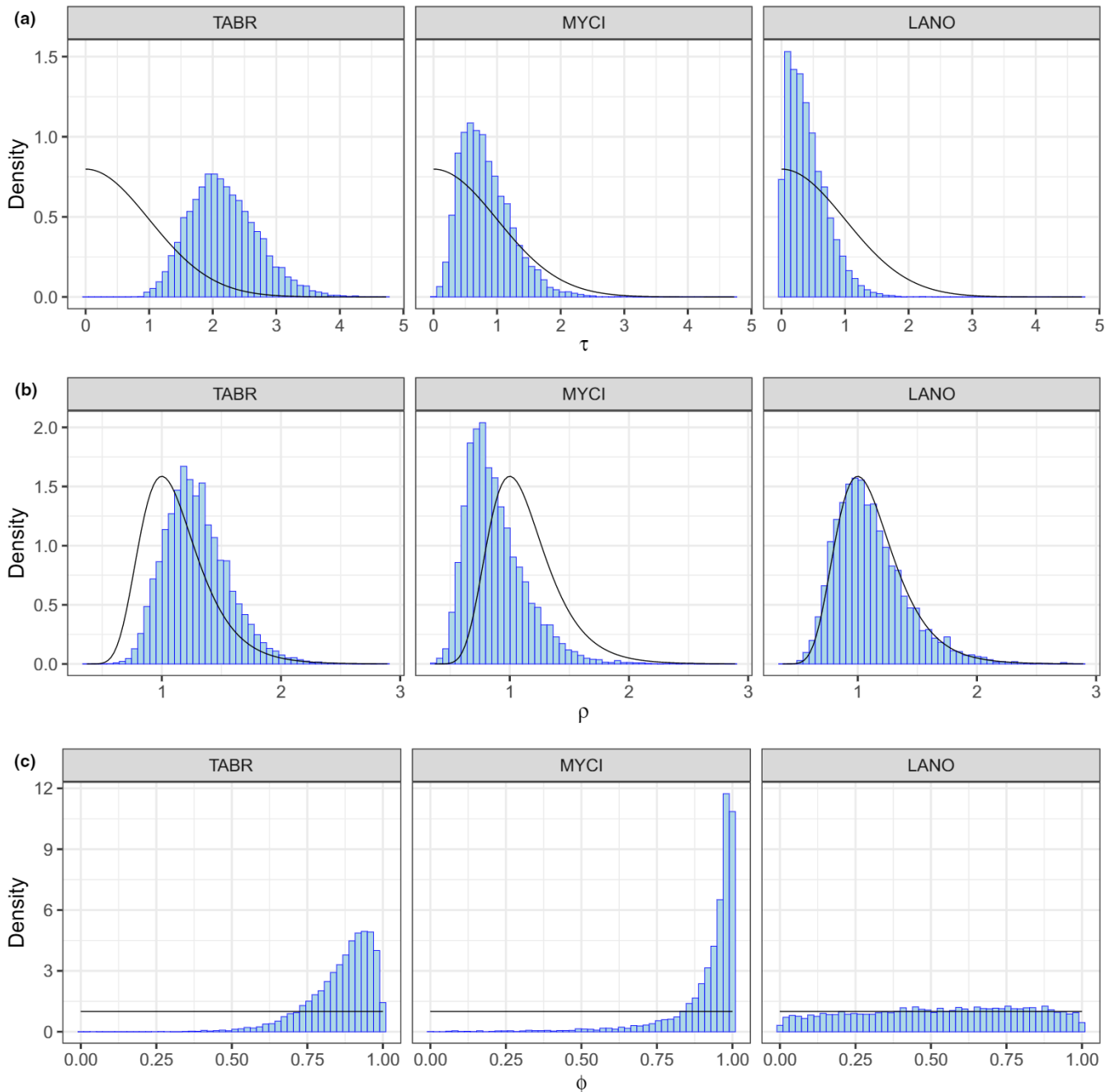
FIGURE 4 Histograms of the posterior draws for the parameters associated with the Gaussian processes—the marginal standard deviation (τ), length-scale (ρ), and temporal correlation (ϕ). The Brazilian free-tailed bat (*Tadarida brasiliensis*; TABR) and western small-footed myotis (*Myotis ciliolabrum*; MYCI) have restricted ranges in the study region while the silver-haired bat (*Lasionycteris noctivagans*; LANO) is widespread. The lines in each panel show the prior distributions for these parameters

conditional autoregressive model structure). Another option would be to include a second Gaussian process with a smaller length-scale parameter (ρ) to capture additional spatial patterns. This could be a convenient approach because covariance functions for Gaussian processes can be combined easily (Gelman et al., 2013).

    Our analysis included multiple years of data for each species by adding temporal correlation to the site-specific effects. We found this approach greatly improved the precision of the occupancy predictions for some species (e.g., TABR) compared with those from

analyzing 1 year of data (results not shown). In particular, analyzing multiple years of data resulted in much greater precision about areas where TABR was unlikely to occur (beyond its range) and this improved inferences about its extent of occurrence in Oregon and Washington. The broad-scale spatial patterns were similar across time for each species, as expected based on the relatively short time frame examined here and high site fidelity showed by many bat species (e.g., Barclay & Birgham, 2001; Lewis, 1995). However, our model structure for the Gaussian processes allowed the site

coefficients to vary over time while still leveraging the temporal correlation in broad-scale spatial patterns when analyzing multiple years of data. Future work should explore analyzing data collected over longer time periods, where larger range shifts are possible, and compare our approach to other dynamic occupancy models that explicitly estimate colonization and extinction probabilities. Because spatial Gaussian processes in occupancy models can be computationally intensive, methods to approximate inferences (see Tikhonov et al., 2020) may be useful if more years and/or more sites are incorporated into analyses. In particular, we suspect that Gaussian predictive processes (Banerjee et al., 2008; Tikhonov et al., 2020) would allow for reduced computation times without negatively impacting inferences for broad-scale spatial patterns.

Maps depicting species distributions and how these distributions are changing over time are invaluable to many conservation efforts. For many taxa, estimating these maps is now more feasible because the development and increased use of technology such as ARUs allow the requisite data to be efficiently collected over large spatial extents. Consequently, SDMs are increasingly relied upon for regional and continental scale monitoring (e.g., Adams et al., 2013; Cariveau et al., 2019; Grant & Bradbury, 2019; Loeb et al., 2015), assessing scenarios of climate change impacts (Thomas et al., 2004; Wiens et al., 2009), and informing endangered species policy and recovery plans (Cassini, 2011; Guisan et al., 2013). Improvements to the statistical methodologies used to estimate species distributions are crucial for accurately informing these conservation efforts.

CONFLICT OF INTEREST

None declared.

AUTHOR CONTRIBUTIONS

Wilson Jay Wright: Conceptualization (equal); Formal analysis (lead); Methodology (equal); Writing—original draft (lead); Writing—review and editing (equal). **Kathryn Irvine:** Conceptualization (equal); Formal analysis (supporting); Methodology (equal); Writing—original draft (equal); Writing—review and editing (equal). **Thomas J Rodhouse:** Conceptualization (equal); Data curation (lead); Methodology (equal); Writing—original draft (equal); Writing—review and editing (equal). **Andrea Litt:** Conceptualization (equal); Methodology (equal); Writing—original draft (equal); Writing—review and editing (equal).

DATA AVAILABILITY STATEMENT

The data are available through the National Park Service IRMA Portal at https://irma.nps.gov/DataStore/Reference/Profile/2275603.

ORCID

Wilson J. Wright https://orcid.org/0000-0003-4276-3850
Kathryn M. Irvine https://orcid.org/0000-0002-6426-940X
Thomas J. Rodhouse https://orcid.org/0000-0001-5953-9113
Andrea R. Litt https://orcid.org/0000-0002-6915-8047

REFERENCES

Adams, M. J., Miller, D. A. W., Muths, E., Corn, P. S., Grant, E. H. C., Bailey, L. L., Fellers, G. M., Fisher, R. N., Sadinski, W. J., Waddle, H., & Walls, S. C. (2013). Trends in amphibian occupancy in the United States. _PLoS One_, _8_(5), 1–5. https://doi.org/10.1371/journal.pone.0064347

Banerjee, S., Gelfand, A. E., Finley, A. O., & Sang, H. (2008). Gaussian predictive process models for large spatial data sets. _Journal of the Royal Statistical Society Series B_, _70_, 825–848.

Banner, K. M., Irvine, K. M., Rodhouse, T. J., Wright, W. J., Rodriguez, R. M., & Litt, A. R. (2018). Improving geographically extensive acoustic survey designs for modeling species occurrence with imperfect detection and misidentification. _Ecology and Evolution_, _8_, 6144–6156. https://doi.org/10.1002/ece3.416

Barclay, R. M. R., & Birgham, R. M. (2001). Year-to-year reuse of tree-roosts by California bats (Myotis californicus) in Southern British Columbia. _The American Midland Naturalist_, _146_(1), 80–85. https://doi.org/10.1674/0003-0031(2001)146[0080:YTYROT]2.0.CO;2

Bates, D., & Eddelbuettel, D. (2013). Fast and elegant numerical linear algebra using the RcppEigen package. _Journal of Statistical Software_, _52_(5), 1–24. http://www.jstatsoft.org/v52/i05/

Blanchet, F. G., Cazelles, K., & Gravel, D. (2020). Co-occurrence is not evidence of ecological interactions. _Ecology Letters_, _23_, 1050–1063. https://doi.org/10.1111/ele.13525

Brown, J. H., Mehlman, D. W., & Stevens, G. C. (1995). Spatial variation in abundance. _Ecology_, _76_(7), 2028–2043. https://doi.org/10.2307/1941678

Cariveau, A. B., Holt, H. L., Ward, J. P., Lukens, L., Kasten, K., Thieme, J., Caldwell, W., Tuerk, K., Baum, K. A., Drobney, P., Drum, R. G., Grundel, R., Hamilton, K., Hoang, C., Kinkead, K., McIntyre, J., Thogmartin, W. E., Turner, T., Weiser, E. L., & Oberhauser, K. (2019). The integrated monarch monitoring program: From design to implementation. _Frontiers in Ecology and Evolution_, _7_, 167. https://doi.org/10.3389/fevo.2019.00167

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. _Journal of Statistical Software_, _76_(1), 1–32.

Cassini, M. H. (2011). Ranking threats using species distribution models in the IUCN red list assessment process. *Biodiversity and Conservation*, 20(14), 3689–3692. https://doi.org/10.1007/s10531-011-0126-9

Curnutt, J. L., Pimm, S. L., & Maurer, B. A. (1996). Population variability of sparrows in space and time. *Oikos*, 131–144. https://doi.org/10.2307/3545755

Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., Curtis, J., & Pasteris, P. P. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International Journal of Climatology*, 28(15), 2031–2064. https://doi.org/10.1002/joc.1688

Doherty, P. F. Jr, Boulinier, T., & Nichols, J. D. (2003). Local extinction and turnover rates at the edge and interior of species' ranges. *Annales Zoologici Fennici*, 40, 145–153.

Dorazio, R. M., & Royle, J. A. (2005). Estimating size and composition of biological communities by modeling the occurrence of species. *Journal of the American Statistical Association*, 100(470), 389–398. https://doi.org/10.1198/016214505000000015

Eddelbuettel, D., & François, R. (2011). Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, 40(8), 1–18. https://doi.org/10.18637/jss.v040.i08

Elith, J., & Leathwick, J. R. (2009). Species distribution models: Ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, 40, 677–697. https://doi.org/10.1146/annurev.ecolsys.110308.120159

Franklin, J. (2010). *Mapping species distributions: Spatial inference and prediction*. Cambridge University Press.

Gaston, K. J., & Fuller, R. A. (2009). The sizes of species' geographic ranges. *Journal of Applied Ecology*, 46, 1–9. https://doi.org/10.1111/j.1365-2664.2008.01596.x

Gelfand, A. E., & Schliep, E. M. (2016). Spatial statistics and Gaussian processes: A beautiful marriage. *Spatial Statistics*, 18, 86–104. https://doi.org/10.1016/j.spasta.2016.03.006

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*, 3rd ed. CRC Press/Taylor; Francis Group.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4), 457–472. https://doi.org/10.1214/ss/1177011136

Genoways, H. H., Freeman, P. W., & Grell, C. (2000). Extralimital records of the Mexican free-tailed bat (Tadarida brasiliensis mexicana) in the central United States and their biological significance. *Transactions of the Nebraska Academy of Sciences*, 26, 85–96.

Gibb, R., Browning, E., Glover-Kapfer, P., & Jones, K. E. (2019). Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. *Methods in Ecology and Evolution*, 10(2), 169–185. https://doi.org/10.1111/2041-210X.13101

Graham, C. H., & Hijmans, R. J. (2006). A comparison of methods for mapping species ranges and species richness. *Global Ecology and Biogeography*, 15, 578–587. https://doi.org/10.1111/j.1466-8238.2006.00257.x

Grant, T. J., & Bradbury, S. P. (2019). The role of modeling in monarch butterfly research and conservation. *Frontiers in Ecology and Evolution*, 7, 197.

Guélat, J., & Kéry, M. (2018). Effects of spatial autocorrelation and imperfect detection on species distribution models. *Methods in Ecology and Evolution*, 9, 1614–1625. https://doi.org/10.1111/2041-210X.12983

Guisan, A., & Thuiller, W. (2005). Predicting species distribution: Offering more than simple habitat models. *Ecology Letters*, 8, 993–1009. https://doi.org/10.1111/j.1461-0248.2005.00792.x

Guisan, A., Tingley, R., Baumgartner, J. B., Naujokaitis-Lewis, I., Sutcliffe, P. R., Tulloch, A. I. T., Regan, T. J., Brotons, L., McDonald-Madden, E., Mantyka-Pringle, C., Martin, T. G., Rhodes, J. R., Maggini, R., Setterfield, S. A., Elith, J., Schwartz, M. W., Wintle, B. A., Broennimann, O., Austin, M., ... Buckley, Y. M. (2013). Predicting species distributions for conservation decisions. *Ecology Letters*, 16(12), 1424–1435. https://doi.org/10.1111/ele.12189

Homer, C. G., Dewitz, J. A., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N. D., Wickham, J. D., & Megown, K. (2015). Completion of the 2011 National Land Cover Database for the conterminous United States | representing a decade of land cover change information. *Photogrammetric Engineering and Remote Sensing*, 81(5), 345–354.

Johnson, D. S., Conn, P. B., Hooten, M. B., Ray, J. C., & Pond, B. A. (2013). Spatial occupancy models for large data sets. *Ecology*, 94(4), 801–808. https://doi.org/10.1890/12-0564.1

Ko, C.-Y., Schmitz, O. J., & Jetz, W. (2016). The limits of direct community modeling approaches for broad-scale predictions of ecological assemblage structure. *Biological Conservation*, 201, 396–404. https://doi.org/10.1016/j.biocon.2016.07.026

Kucera, T. E., & Barrett, R. H. (2011). A history of camera trapping. In A. F. O'Connell, J. D. Nichols, & K. U. Karanth (Eds.), *Camera traps in animal ecology: Methods and analyses* (pp. 9–26). Springer.

Larsen, D. P., Olsen, A. R., & Stevens, D. L. (2008). Using a master sample to integrate stream monitoring programs. *Journal of Agricultural, Biological, and Environmental Statistics*, 13(3), 243. https://doi.org/10.1198/108571108X336593

Legendre, P., & Fortin, M.-J. (1989). Spatial pattern and ecological analysis. *Vegetatio*, 40, 107–138. https://doi.org/10.1007/BF00048036

Levin, S. A. (1992). The problem of pattern and scale in ecology. *Ecology*, 73(6), 1943–1967.

Lewis, S. E. (1995). Roost fidelity of bats: A review. *Journal of Mammalogy*, 76(2), 481–496. https://doi.org/10.2307/1382357

Lichstein, J. W., Simons, T. R., Shriner, S. A., & Franzreb, K. E. (2002). Spatial autocorrelation and autoregressive models in ecology. *Ecological Monographs*, 72(3), 445–463. https://doi.org/10.1890/0012-9615(2002)072[0445:SAAAMI]2.0.CO;2

Lobo, J. M., Jiménez-Valverde, A., & Raimundo, R. (2008). AUC: A misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, 17, 145–151. https://doi.org/10.1111/j.1466-8238.2007.00358.x

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

MacKenzie, D. I., Nichols, J. D., Hines, J. E., Knutson, M. G., & Franklin, A. B. (2003). Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. *Ecology*, 84(8), 2200–2207. https://doi.org/10.1890/02-3090

MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Royle, J. A., & Langtimm, C. A. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, 83(8), 2248–2255. https://doi.org/10.1890/0012-9658(2002)083[2248:ESORWD]2.0.CO;2

Manley, P. N., Schlesinger, M. D., Roth, J. K., & Van Horne, B. (2005). A field-based evaluation of a presence-absence protocol for monitoring ecoregional-scale biodiversity. *Journal of Wildlife Management*, 63(3), 950–966. https://doi.org/10.2193/0022-541X(2005)069[0950:AFEOAP]2.0.CO;2

McCracken, G. F., Bernard, R. F., Gamba-Rios, M., Wolfe, R., Krauel, J. J., Jones, D. N., Russell, A. L., & Brown, V. A. (2018). Rapid range expansion of the Brazilian freetailed bat in the southeastern United States, 2008–2016. *Journal of Mammalogy*, 99(2), 312–320. https://doi.org/10.1093/jmammal/gyx188

Meyer, L., Diniz-Filho, J. A. F., & Lohmann, L. G. (2017). A comparison of hull methods for estimating species ranges and richness maps. *Plant Ecology and Diversity*, 10(5–6), 389–401. https://doi.org/10.1080/17550874.2018.1425505

Neece, B. D., Loeb, S. C., & Jachowski, D. S. (2019). Implementing and assessing the efficacy of the North American Bat Monitoring Program.

Journal of Fish and Wildlife Management, *10*(2), 391–409. https://doi.org/10.3996/092018-JFWM-087

Noon, B. R., Bailey, L. L., Sisk, T. D., & McKelvey, K. S. (2012). Efficient species-level monitoring at the landscape scale. *Conservation Biology*, *26*(3), 432–441. https://doi.org/10.1111/j.1523-1739.2012.01855.x

Ommundsen, P., Lausen, C., & Matthias, L. (2017). First acoustic records of the Brazilian free-tailed bat (*Tadarida brasiliensis*) in British Columbia. *Northwestern Naturalist*, *98*(2), 132–136.

Ovaskainen, O., Roy, D. B., Fox, R., & Anderson, B. J. (2016). Uncovering hidden spatial structure in species communities with spatially explicit joint species distribution models. *Methods in Ecology and Evolution*, *7*, 428–436. https://doi.org/10.1111/2041-210X.12502

Pebesma, E. (2018). Simple Features for R: Standardized support for spatial vector data. *The R Journal*, *10*(1), 439–446. https://doi.org/10.32614/RJ-2018-009

Peterson, A. T., Navarro-Sigüenza, A. G., & Gordillo, A. (2016). Assumption-versus databased approaches to summarizing species' ranges. *Conservation Biology*, *32*(3), 568–575.

R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rodhouse, T. J., Ormsbee, P. C., Irvine, K. M., Vierling, L. A., Szewczak, J. M., & Vierling, K. T. (2015). Establishing conservation baselines with dynamic distribution models for bat populations facing imminent decline. *Diversity and Distributions*, *21*, 1401–1413. https://doi.org/10.1111/ddi.12372

Rodriguez, R. M., Rodhouse, T. J., Barnett, J., Irvine, K., Banner, K. M., Lonneker, J., & Ormsbee, P. C. (2019). *North American Bat Monitoring Program regional protocol for surveying with stationary deployments for echolocation recording devices: Narrative version 1.0, Pacific Northwestern US (Natural Resources Report NPS/UCBN/NRR| 2019/1975)*. National Park Service.

Rushing, C. S., Royle, J. A., Ziolkowski, D. J., & Pardieck, K. L. (2019). Modeling spatially and temporally complex range dynamics when detection is imperfect. *Scientific Reports*, *9*, 12805. https://doi.org/10.1038/s41598-019-48851-5

Stan Development Team (2020). *RStan: The R interface to Stan [R package version 2.19.3]*. http://mc-stan.org/

Stevens, D. L. Jr, & Olsen, A. R. (2004). Spatially balanced sampling of natural resources. *Journal of the American Statistical Association*, *99*(465), 262–278.

Thomas, C. D., Cameron, A., Green, R. E., Bakkenes, M., Beaumont, L. J., Collingham, Y. C., Erasmus, B. F., De Siqueira, M. F., Grainger, A., Hannah, L. et al. (2004). Extinction risk from climate change. *Nature*, *427*(6970), 145–148.

Thornton, P. E., Thornton, M. M., Meyer, B. W., Wei, Y., Devarakonda, R., Vose, R. S., & Cook, R. B. (2018). Daymet: Daily surface weather data on a 1-km grid for North America. *Version*, *3*, https://doi.org/10.3334/ORNLDAAC/1328

Tikhonov, G., Duan, L., Abrego, N., Newell, G., White, M., Dunson, D., & Ovaskainen, O. (2020). Computationally efficient joint species distribution modeling with big spatial data. *Ecology*, *101*(2), e02929.

U.S. Geological Survey (USGS) Gap Analysis Project (GAP) (2018). *U.S. Geological Survey – Gap Analysis Project Species Range Maps CONUS 2001: U.S. Geological Survey data release*. https://doi.org/10.5066/F7Q81B3R

VanDerWal, J., Shoo, L. P., Graham, C., & Williams, S. E. (2009). Selecting pseudo-absence data for presence-only distribution modeling: How far should you stray from what you know? *Ecological Monitoring*, *220*, 589–594. https://doi.org/10.1016/j.ecolmodel.2008.11.010

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag. http://ggplot2.org

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., ... Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wiens, J. A., Stralberg, D., Jongsomjit, D., Howell, C. A., & Snyder, M. A. (2009). Niches, models, and climate change: Assessing the assumptions and uncertainties. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 19729–19736. https://doi.org/10.1073/pnas.0901639106

Zipkin, E. F., Grant, E. H. C., & Fagan, W. F. (2012). Evaluating the predictive abilities of community occupancy models using AUC while accounting for imperfect detection. *Ecological Applications*, *22*(7), 1962–1972. https://doi.org/10.1890/11-1936.1

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.