Received: 13 January 2023 | Revised: 8 December 2023 | Accepted: 1 February 2024

DOI: 10.1111/geb.13814

Global Ecology and Biogeography | A Journal of Macroecology

WILEY

METHOD

# Guidelines for the use of spatially varying coefficients in species distribution models

Jeffrey W. Doser[1,2] | Marc Kéry[3] | Sarah P. Saunders[4] | Andrew O. Finley[2,5,6] | Brooke L. Bateman[4] | Joanna Grand[4] | Shannon Reault[4] | Aaron S. Weed[7] | Elise F. Zipkin[1,2]

[1]Department of Integrative Biology, Michigan State University, East Lansing, Michigan, USA

[2]Ecology, Evolution, and Behavior Program, Michigan State University, East Lansing, Michigan, USA

[3]Swiss Ornithological Institute, Sempach, Switzerland

[4]Science Division, National Audubon Society, New York, New York, USA

[5]Department of Forestry, Michigan State University, East Lansing, Michigan, USA

[6]Department of Statistics and Probability, Michigan State University, East Lansing, Michigan, USA

[7]Northeast Temperate Inventory and Monitoring Network, National Park Service, Woodstock, Vermont, USA

**Correspondence**
Jeffrey W. Doser, Department of Integrative Biology, Natural Science Building, 288 Farm Lane, Rm. 203, Michigan State University, East Lansing, MI, 48824, USA.
Email: doserjef@msu.edu

**Funding information**
National Science Foundation, Grant/Award Number: DBI-1954406, DEB-2213565 and DMS-1916395

**Handling Editor:** Antoine Guisan

## Abstract

**Aim:** Species distribution models (SDMs) are increasingly applied across macroscales using detection-nondetection data. These models typically assume that a single set of regression coefficients can adequately describe species–environment relationships and/or population trends. However, such relationships often show nonlinear and/or spatially varying patterns that arise from complex interactions with abiotic and biotic processes that operate at different scales. Spatially varying coefficient (SVC) models can readily account for variability in the effects of environmental covariates. Yet, their use in ecology is relatively scarce due to gaps in understanding the inferential benefits that SVC models can provide compared to simpler frameworks.

**Innovation:** Here we demonstrate the inferential benefits of SVC SDMs, with a particular focus on how this approach can be used to generate and test ecological hypotheses regarding the drivers of spatial variability in population trends and species–environment relationships. We illustrate the inferential benefits of SVC SDMs with simulations and two case studies: one that assesses spatially varying trends of 51 forest bird species in the eastern United States over two decades and a second that evaluates spatial variability in the effects of five decades of land cover change on grasshopper sparrow (*Ammodramus savannarum*) occurrence across the continental United States.

**Main conclusions:** We found strong support for SVC SDMs compared to simpler alternatives in both empirical case studies. Factors operating at fine spatial scales, accounted for by the SVCs, were the primary divers of spatial variability in forest bird occurrence trends. Additionally, SVCs revealed complex species–habitat relationships with grassland and cropland area for grasshopper sparrow, providing nuanced insights into how future land use change may shape its distribution. These applications display the utility of SVC SDMs to help reveal the environmental factors that drive species distributions across both local and broad scales. We conclude by discussing the potential applications of SVC SDMs in ecology and conservation.

KEYWORDS
Bayesian, Breeding Bird Survey, context dependence, imperfect detection, nonstationarity, occupancy model, spatial autocorrelation

## 1 | INTRODUCTION

Elucidating the factors that drive species distributions is a fundamental objective of ecology. Species distribution models (SDMs) are the primary tool to study where and why species occur across space and time (Guisan & Zimmermann, 2000). While SDMs can leverage a variety of data types (e.g. presence-only, abundance), they are commonly used with presence–absence (or detection-nondetection) data in a parametric generalized-linear model (GLM)-based framework, allowing for assessments of species–environment relationships and probabilities of local-level occurrence. Parametric SDMs often describe species–environment relationships via a single set of regression coefficients (e.g. linear and quadratic terms) across the spatial extent of the data set, resulting in a set of linear or unimodal response curves along all environmental predictors included in the model (Guisan et al., 2002). When the spatial extent encompasses the entirety of the species range, the combination of such species–environment response curves characterizes the multivariate realized environmental niche of a species (Guisan et al., 2017). However, when fitting SDMs across large spatial extents, a single set of linear and quadratic regression coefficients may not adequately describe species–environment relationships, which can result from the true relationship taking a complex, nonlinear form or variability in the relationship across space (i.e. the relationship is context dependent, or nonstationary; Rollinson et al., 2021).

Complex, nonlinear and/or spatially varying species–environment relationships can arise from a variety of abiotic and biotic processes that operate at different scales (Miller, 2012; Osborne & Suárez-Seoane, 2002). Interactions with abiotic factors, such as historical disturbance regimes, fine-scale habitat characteristics (e.g. vegetation quality) and environmental conditions (e.g. soil content) can result in varying effects of environmental factors on species across their ranges (Rollinson et al., 2021). Given spatial heterogeneity in resource availability, effects of environmental factors on species occurrences may be stronger in areas with limited resources compared to areas with abundant resources (Pease, Pacifici, & Kays, 2022). If climate shapes species distributions, effects of climate change on species should be strongest in areas near their climatic limits (Amburgey et al., 2018). For example, Sultaire et al. (2022) found spatial variation in the effects of increasing temperature and snow cover duration on snowshoe hare (*Lepus americanus*) occurrence, suggesting that climate limits their distribution in different ways across the species range. Alternatively, spatially varying species–environment relationships may arise from biotic processes such as local genetic adaptations or spatial variation in species interactions. Pease, Pacifici, and Kays (2022) found that spatial variability in the effect of forest cover on white-tailed deer (*Odocoileus virginianus*) occurrence across North Carolina was partially driven by variation in predation pressure across the state. Failing to account for such spatially varying relationships can lead to misleading inferences on the abiotic factors that influence where species occur, which could have important implications for conservation recommendations (Rollinson et al., 2021).

In addition to characterizing species–environment relationships, monitoring data are often used in SDMs to quantify occurrence trends. Nonlinear and/or spatially varying occurrence trends primarily arise from spatio-temporal changes in abiotic and biotic factors that influence the species of interest (e.g. differences in land-use change across a species distribution). However, such patterns can also arise from complex relationships to such factors (e.g. the effect of land-use change is different in one part of a species range compared to another). Quantifying spatial variability in population change over time is a common objective of biodiversity monitoring programs (Babcock et al., 2016; Bled et al., 2013; Meehan et al., 2019), as such insights can help generate hypotheses about the drivers of population changes (e.g. Crossley et al., 2021) and identify priority areas for conservation or restoration (e.g. Ethier et al., 2017).

Numerous methods have been used to test hypotheses about spatial variability and nonlinearity in species–environment relationships or occurrence trends. GLMs with interactions between variables are simple, yet intuitive, ways to assess nonlinear and/or spatially varying relationships (Spake et al., 2023). In addition, estimating separate slopes (fixed or random) across pre-defined strata (e.g. ecoregions, management units) is another common alternative (e.g. Smith & Edwards, 2021). Furthermore, nonlinear functions (e.g. thresholds) can be readily incorporated into parametric SDMs (Hostetler & Chandler, 2015). However, these approaches require a priori knowledge of covariates that interact with the variables of interest, the functional forms of such relationships and/or the spatial resolution of variability in the relationship, most of which are unknown prior to analysis. Specifying interactions between multiple drivers operating at different spatial scales is particularly critical when working across macroscales, but difficult to accomplish using the aforementioned approaches. More flexible approaches that can readily accommodate complex species–environment relationships without a priori knowledge of all interacting variables and the nature of their interactions are thus needed.

Accordingly, there has been widespread use of MaxEnt (Phillips et al., 2006) and random forests (Liaw & Wiener, 2002) to model species distributions while accounting for complex species–environment relationships. MaxEnt uses combinations of different 'feature classes', or mathematical functions, on covariates to characterize nonlinear relationships in presence-only data, and random forests fit ensembles of classification or regression trees within partitions of the data based on covariate space (Valavi et al., 2021). Both approaches are widely used in ecology, but they are limited in their ability to provide uncertainty estimates of species–environment relationships and/or occurrence trends; they require all interacting variables to be known and incorporated into the model; and they do not account for imperfect detection (i.e. the failure to observe a species at a site when it is present; Kellner & Swihart, 2014; Kéry, 2011).

In this paper, we discuss the use of spatially varying coefficients (SVCs) in SDMs, a highly flexible approach for modelling nonlinear

and/or spatially varying species–environment relationships, regardless of how they arise. SVC models are intuitive extensions of GLMs that allow regression coefficients to vary smoothly across space. By fitting SVC models within a hierarchical Bayesian framework, we can generate predictions of species–environment effects and/or trends across a spatial region of interest with full uncertainty propagation, while simultaneously accounting for widespread observation errors, such as those due to imperfect detection. Recent studies suggest an increasing interest in this framework for a variety of ecological applications due to the prevalence of heterogeneity in species–environment relationships across macroscales (e.g. Meehan et al., 2019; Rollinson et al., 2021; Sultaire et al., 2022). However, a comprehensive understanding of the inferential benefits SVCs can provide compared to alternative approaches is lacking. Thorson et al. (2023) recently highlighted seven ecological questions that can be addressed with spatially varying coefficients. Here, we build on their work by explicitly demonstrating the inferential benefits of SVC SDMs compared to simpler approaches with simulations and two case studies on breeding birds in the United States. We conclude with a discussion of practical guidelines on when to use SVC SDMs instead of, or in addition to, simpler alternatives.

## 2 | What are spatially varying coefficients?

Here, we discuss SVCs in the context of occupancy models (MacKenzie et al., 2002; Tyre et al., 2003), a specific form of hierarchical GLM that models imperfect detection, although our findings are directly extensible to all parametric SDMs (i.e. GLMs). See Doser et al. (2024) for full statistical details.

Let $\boldsymbol{s}_j$ denote the spatial coordinates of site $j$, where $j = 1, \ldots, J$, which are each sampled across $t = 1, \ldots, T$ primary time periods (henceforth 'seasons'). Note that data may be obtained for only one season (i.e. $T = 1$), or for multi-season data sets in which sites do not need to be sampled every season (i.e. missing values are allowed). To account for imperfect detection, $k = 1, \ldots, K_t(\boldsymbol{s}_j)$ sampling replicates are obtained at site $j$ during season $t$ to estimate whether a nondetection of the target species is truly an absence (MacKenzie et al., 2002; Tyre et al., 2003). Note that the number of replicates can vary across site/season combinations. Such replicates typically come in the form of multiple visits to a site over a short period of time within a season, but other forms of replication such as spatial sub-sampling and multiple observers are possible (MacKenzie et al., 2017). Let $y_{t,k}(\boldsymbol{s}_j)$ denote the observed detection (1) or nondetection (0) of a study species at site $j$ during survey $k$ in season $t$, and let $z_t(\boldsymbol{s}_j)$ denote the true presence (1) or absence (0) of the species at site $j$ during season $t$. Note we assume $z_t(\boldsymbol{s}_j)$ does not change across replicate surveys within a given season (i.e. the 'closure' assumption). We model the observed data $y_{t,k}(\boldsymbol{s}_j)$ conditional on the true occurrence status of the species at site $j$ during season $t$ ($z_t(\boldsymbol{s}_j)$). Specifically, we have

$$y_{t,k}(\boldsymbol{s}_j) \sim \begin{cases} 0, & z_t(\boldsymbol{s}_j) = 0 \\ \text{Bernoulli}(p_{t,k}(\boldsymbol{s}_j)), & z_t(\boldsymbol{s}_j) = 1 \end{cases}, \quad (1)$$

where $p_{t,k}(\boldsymbol{s}_j)$ is the probability of detecting the species at site $j$ during replicate survey $k$ in season $t$. We model detection probability as a function of site, season and/or survey-level (i.e. observation-level) covariates according to

$$\text{logit}(p_{t,k}(\boldsymbol{s}_j)) = \boldsymbol{v}_{t,k}(\boldsymbol{s}_j)\boldsymbol{\alpha}, \quad (2)$$

where $\boldsymbol{\alpha}$ is a vector of regression coefficients (including an intercept) that describe the effect of covariates $\boldsymbol{v}_{t,k}(\boldsymbol{s}_j)$ on detection.

The true occurrence status $z_t(\boldsymbol{s}_j)$ is a partially observed variable, such that if $y_{t,k}(\boldsymbol{s}_j) = 1$, we know $z_t(\boldsymbol{s}_j) = 1$ (since we assume no false positives), but if $y_{t,k}(\boldsymbol{s}_j) = 0$ we do not know if the species is truly absent from the site, or if we failed to detect it. We model $z_t(\boldsymbol{s}_j)$ as

$$z_t(\boldsymbol{s}_j) \sim \text{Bernoulli}(\psi_t(\boldsymbol{s}_j)), \quad (3)$$

where $\psi_t(\boldsymbol{s}_j)$ is the occurrence probability of the species at site $j$ during season $t$. When fitting occupancy models, we can estimate species–environment relationships through the effect of covariates on occurrence probability, $\psi_t(\boldsymbol{s}_j)$, within a GLM framework. For simplicity, consider a single environmental variable, $x_t(\boldsymbol{s}_j)$, that varies across each spatial location $\boldsymbol{s}_j$ and season $t$ (e.g. temperature, precipitation). An SVC occupancy model has the form

$$\text{logit}(\psi_t(\boldsymbol{s}_j)) = \beta_0 + \eta_t + w_0(\boldsymbol{s}_j) + \beta_1 \cdot x_t(\boldsymbol{s}_j) + w_1(\boldsymbol{s}_j) \cdot x_t(\boldsymbol{s}_j), \quad (4)$$

where $\beta_0$ is an intercept, $\eta_t$ is a temporal random effect (i.e. season-specific intercept) to account for unmodelled temporal autocorrelation in occurrence probability (only applicable if $T > 1$), $w_0(\boldsymbol{s}_j)$ is a spatial random effect to account for unmodelled spatial variation in occurrence probability, $\beta_1$ is the non-spatial effect of the covariate $x_t(\boldsymbol{s}_j)$ and $w_1(\boldsymbol{s}_j)$ is the spatially varying effect of the covariate at each spatial location $\boldsymbol{s}_j$.

The spatially varying effect in the SVC model can be estimated in a variety of ways, including via generalized additive models (GAMs; Wood, 2006) and Gaussian processes (Banerjee et al., 2014). We focus on the latter due to their prevalence in spatial statistics, their comparatively higher predictive performance (Golding & Purse, 2016), and the potential for oversmoothing of relationships with GAMs (Stein, 2014). Specifically, we have

$$w_1(\boldsymbol{s}) \sim N(\boldsymbol{0}, \boldsymbol{C}(\boldsymbol{s}, \boldsymbol{s}', \theta)), \quad (5)$$

where $\boldsymbol{C}(\boldsymbol{s}, \boldsymbol{s}', \theta)$ is a $J \times J$ covariance matrix that is a function of the distances between any pair of site coordinates $\boldsymbol{s}$ and $\boldsymbol{s}'$ and a set of parameters ($\theta$) that govern the spatial process according to a spatial correlation function. Here we use Nearest Neighbour Gaussian Processes (Datta et al., 2016) as an efficient approximation to the full Gaussian process and an exponential correlation function, which models the correlation in the spatially varying effect of the covariate using two parameters, $\theta = \{\sigma^2, \phi\}$, where $\sigma^2$ is a spatial variance parameter and

$\phi$ is a spatial decay parameter. See Doser et al. (2024) for full details on SVC occupancy models.

## 3 | How do spatially varying coefficient models compare to simpler alternatives?

For comparison to alternative models, consider a more generalized model of occurrence probability $\psi_t(\boldsymbol{s}_j)$ with the form

$$\text{logit}(\psi_t(\boldsymbol{s}_j)) = \beta_0 + \eta_t + w_0(\boldsymbol{s}_j) + f(x_t(\boldsymbol{s}_j), \boldsymbol{\beta}), \quad (6)$$

where $\beta_0$, $\eta_t$ and $w_0(\boldsymbol{s}_j)$ are defined in (4), and $f(x_t(\boldsymbol{s}_j), \boldsymbol{\beta})$ is a generic function that relates the covariate $x_t(\boldsymbol{s}_j)$ to occurrence probability through a set of parameters $\boldsymbol{\beta}$. Note that $w_0(\boldsymbol{s}_j)$ and $\eta_t$ can be removed from (6) if not applicable for a target species/data set. We consider five functional relationships to describe the relationship between $\psi_t(\boldsymbol{s}_j)$ and $x_t(\boldsymbol{s}_j)$, which vary in their ability to estimate nonlinear and/or spatially varying species environment relationships:

1. **Linear**: $f(x_t(\boldsymbol{s}_j), \boldsymbol{\beta}) = \beta_1 \cdot x_t(\boldsymbol{s}_j)$
2. **Quadratic**: $f(x_t(\boldsymbol{s}_j), \boldsymbol{\beta}) = \beta_1 \cdot x_t(\boldsymbol{s}_j) + \beta_2 \cdot x_t^2(\boldsymbol{s}_j)$
3. **Stratum**: $f(x_t(\boldsymbol{s}_j), \boldsymbol{\beta}) = \beta_1 \cdot x_t(\boldsymbol{s}_j) + \beta_{2,\text{STRATUM}_j} \cdot x_t(\boldsymbol{s}_j)$
4. **Interaction**: $f(x_t(\boldsymbol{s}_j), \boldsymbol{\beta}) = \beta_1 \cdot x_t(\boldsymbol{s}_j) + \beta_2 \cdot x_t^*(\boldsymbol{s}_j) \cdot x_t(\boldsymbol{s}_j)$
5. **SVC**: $f(x_t(\boldsymbol{s}_j), \boldsymbol{\beta}) = \beta_1 \cdot x_t(\boldsymbol{s}_j) + w_1(\boldsymbol{s}_j) \cdot x_t(\boldsymbol{s}_j)$

The linear model simply assumes a linear species–environment relationship. The quadratic model extends the linear model by allowing the species–environment relationship to peak at some optimum level and subsequently decrease as one moves farther from the optimum or peak at the extremes of the environmental variable. The stratum model estimates an overall linear effect of the environmental predictor as well as stratum-specific adjustments (fixed or random) in the effect across a set of strata (e.g. management units). The interaction model similarly estimates a linear species–environment relationship, but allows for spatial variation in this relationship in the form of an interaction with a second covariate that varies across the $J$ sites ($x_t^*(\boldsymbol{s}_j)$). The SVC model estimates an overall linear species–environment relationship, but allows this linear relationship to vary across each site in the data set. The spatially varying adjustment in the SVC model ($w_1(\boldsymbol{s}_j)$) serves as a local adjustment of the species–environment relationship from the overall effect $\beta_1$. The SVC model can be viewed as an extension of the stratum model, where the strata are now individual sampling sites, or as an extension of the interaction model, where the interacting 'covariate' is unknown and estimated as part of the model fitting process.

We performed a simulation study to compare the five aforementioned models and assessed their ability to estimate species–environment relationships of different forms. Briefly, we simulated detection-nondetection data across a 20×20 grid, where occurrence probability was generated as a function of a single covariate that took negative values at the southern portion of the simulated area and positive values at the northern portion of the area. We varied the true species–environment relationship across six different functional forms (leftmost column, Figure 1): (1) linear; (2) quadratic; (3) a separate linear effect across nine strata; (4) an interaction with a second covariate that varied along the horizontal axis; (5) an interaction with an unknown ("missing") covariate; and (6) the sum of the five aforementioned components (i.e. "Full" effect). We simulated 50 data sets under each functional form of the species–environment relationship, and fit the five models to each data set using the `spOccupancy` R package (Doser et al., 2022), where the interaction model only incorporated the covariate that was assumed known, not the unknown covariate. Model performance was compared using the Widely Applicable Information Criterion (WAIC; Watanabe, 2010). See Supplemental Information S1 for complete details.

The SVC model was consistently able to capture the true relationship across all six forms of the species–environment relationship, while the simpler models only performed well in a subset of scenarios (Figure 1). When the true species–environment was linear, all models yielded virtually identical estimates of the species–environment relationship that closely resembled the truth. The simpler models with a pre-specified functional form (e.g. quadratic, stratum, known interaction) outperformed the SVC model according to WAIC when data were generated with the exact species–environment relationship, suggesting that when the true form of the species–environment relationship is known, simpler models should be used over the more complicated SVC model. However, when the true species–environment relationship was different from the form used to fit the model, the simpler models were limited in their ability to capture the underlying patterns (Figure 1) and performed substantially worse according to WAIC (Supplemental Information S1: Table S1). The SVC model was able to capture the true pattern in the relationship even when the data were not generated with an SVC, indicating the ability of SVC models to reveal simpler functional forms of species–environment relationships when such relationships are unknown prior to model fitting. Furthermore, the SVC model drastically outperformed all other models when the species–environment relationship interacted with an unknown variable (row 5, Figure 1), or when the species–environment relationship was determined by multiple components (row 6, Figure 1).

## 4 | Case study 1: Spatially varying occurrence trends in eastern US forest birds

Quantifying spatially explicit trends can help identify areas of conservation interest (e.g. climate change refugia) as well as provide insights on species range dynamics. In this case study, we assessed occurrence trends of forest bird species across a ~4.04 million km$^2$ region of the eastern United States (i.e. the continental United States east of the 100th meridian) from 2000 to 2019 ($T$ = 20 years) using detection-nondetection data from the North American Breeding


**FIGURE 1** Estimates of a simulated species–environment relationship from different models (columns) under different patterns of the species–environment relationship (rows). Red represents a negative effect, white no effect and blue a positive effect, with darker shades representing stronger effects. The simulated covariate varies from negative at the bottom of the grid to positive at the top of the grid. The true species–environment relationship is simulated as a linear effect (row 1), a quadratic effect (row 2), a separate linear effect across nine strata (row 3), an interaction with a second covariate that varies along the horizontal axis (row 4), an interaction with an unknown ("missing") covariate (row 5) and the sum of all the aforementioned components ("Full" effect, row 6). Estimates are shown from five candidate models relative to the truth (column 1), including a model with a: linear effect (column 2), quadratic effect (column 3), stratum-specific effect (column 4), interaction (column 5) and an SVC (column 6).

Bird Survey (BBS; Pardieck et al., 2020). We restricted our analysis to a community of 66 eastern forest bird species following the habitat classification of Bateman et al. (2020). We subsequently assessed trends for 51 of the 66 species whose breeding ranges (derived from BirdLife International, 2021) had at least 50% overlap with the study area (Supplemental Information S2: Table S3). Our objectives for this case study were to (1) develop spatially explicit maps of occurrence trends for each of the 51 species across the eastern United States, and (2) compare an SVC occupancy model to three alternative models that represent simple hypotheses regarding the drivers of distributional change. Specifically, our four hypotheses and associated models were:

1. The species has a constant, linear trend across the eastern United States (i.e. the linear model).
2. The species trend varies across broad ecologically distinct strata, Bird Conservation Regions (i.e. strata model), as a result of differences in bird communities and habitat types.
3. The species trend interacts with the 30-year (1981–2010) climate normal (i.e. temperature model). If climate shapes species distributions, we would expect differences in species trends near the climatic extremes as annual temperatures become increasingly warm (e.g. positive trends at northern range boundary and negative trends at southern range boundary indicating a northward shifting distribution).

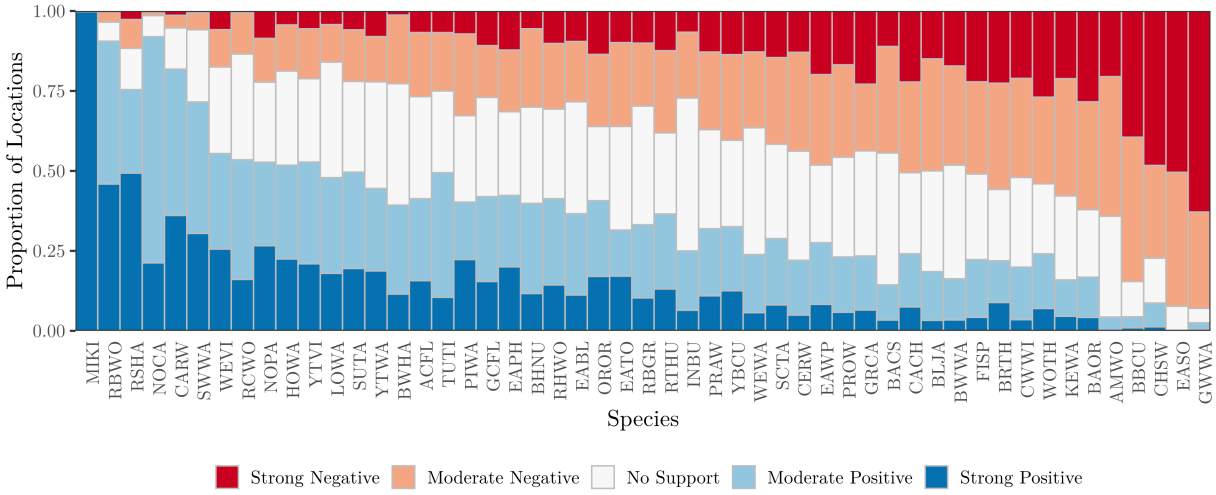


**FIGURE 2** Summary of the spatially varying trend of the 51 forest bird species. The height of each bar corresponds to the proportion of locations for the given species whose trend has the sign (i.e. positive, negative, no effect) and strength (i.e. strong, moderate) indicated by the colour. Blue indicates support for positive trends and red indicates support for negative trends. More specifically: (1) Dark blue = Strong Positive: P(trend > 0) > 0.8; (2) Light blue = Moderate Positive: 0.6 < P(trend > 0) ≤ 0.8; (3) White = No effect: 0.4 < P(trend > 0) ≤ 0.6; (4) Light red = Moderate Negative: 0.2 < P(trend > 0) ≤ 0.4; (5) Dark red = Strong Negative: P(trend > 0) < 0.2.

4. The species trend varies with a variety of interacting abiotic and biotic variables that result in a complex spatially varying trend across the species range (i.e. SVC model).

We used data from $J = 1846$ BBS routes (i.e. sites) sampled at least once between 2000 and 2019 (mean number of sampled years per route = 15). BBS observers performed a three-minute point count survey at each of 50 stops along each route, counting all birds seen or heard within a 0.4 km radius. We summarized the data for each species at each site into $K = 5$ spatial replicates (each comprising data from 10 of the 50 stops), where each replicate took value 1 if the species was detected at any of the 10 stops in that replicate, and value 0 if the species was not detected. While such an approach has been used in previous studies with BBS data (e.g. Rushing et al., 2020), this use of spatial replicates in an occupancy modelling framework likely leads to violation of the closure assumption (Kendall & White, 2009), and so we refer to our response as species-specific occurrence (or "use") rather than occupancy.

For each of the 51 species, we fit a multi-season occupancy model where occurrence probability at each site $j$ in each year $t$ was modelled as

$$\text{logit}(\psi_t(\boldsymbol{s}_j)) = \beta_0 + \text{w}_0(\boldsymbol{s}_j) + \beta_1 \cdot \text{TMAX}(\boldsymbol{s}_j) + f(\text{YEAR}_t, \boldsymbol{\beta}_{\text{TREND}}) + \eta_t, \quad (7)$$

where $\beta_0$ and $\text{w}_0(\boldsymbol{s}_j)$ together represent the spatially varying intercept, $\beta_1$ is the effect of the 30-year (1981–2010) maximum temperature climate normal on occurrence probability, $\eta_t$ is an AR(1) random year effect to accommodate residual temporal autocorrelation and $f(\text{YEAR}_t, \boldsymbol{\beta}_{\text{TREND}})$ is the estimated trend parameter(s) that varies in form across the four models. For all models, we expressed detection probability as a function of linear and quadratic effects of year, linear and quadratic effects of survey day (to account for seasonal variation in detection probability), and linear and quadratic effects of survey replicate (to account for variability in detection probability over the time of day). For each species, we used pre-existing published ranges from BirdLife International (2021) and only included routes that fell inside a 50-km buffer of the species range when fitting the occupancy model. Thirty-year average maximum temperature was calculated from TerraClimate (Abatzoglou et al., 2018).

We compared the four models using the WAIC as an assessment of model parsimony. Using realized observations from BBS in 2021, we additionally compared each model's ability to predict (i.e. forecast) future occurrence of the species in 2021 using the area under the receiver operating characteristic curve (AUC; Hosmer et al., 2013) following approaches outlined by Zipkin et al. (2012). See Supplemental Information S2 for full details.

We fit all models using Bayesian inference with Markov chain Monte Carlo (MCMC) algorithms implemented in the `spOccupancy` R package (Doser et al., 2022, 2024). Prior distributions were either vague or weakly informative (Supplemental Information S2). We ran three chains of each model for 100,000 MCMC iterations with a burn-in period of 50,000 iterations and a thinning rate of 50, yielding 3000 posterior samples. We assessed convergence using the potential scale reduction factor (i.e. $\hat{R}$; Brooks & Gelman, 1998). After fitting each of the four alternative models, we predicted across the range of each species in the study area to generate maps of occurrence trends across the 20-year period.

## 4.1 | RESULTS

There was strong support for spatial variability in 20-year occurrence trends across the eastern United States for the majority of species in our analysis (Figure 2). The SVC model substantially outperformed (i.e. ΔWAIC > 2) the linear trend model, strata model and

maximum temperature interaction model for almost all species (49, or 96%, out of 51) according to the WAIC. Among the three alternative models, the BCR model performed better than the temperature model and linear model for 78% and 90% of species, respectively, according to the WAIC, while the temperature model generally showed less support across the 51 species, outperforming the linear model for 57% of species. The two species with less support for the SVC model (American woodcock and Eastern screech owl) had low raw occurrence probabilities (i.e. <0.05). The SVC model requires more observations than the simpler strata model to yield estimates with reasonable uncertainty, which likely contributed to the higher performance of the strata model for these two species. We found more variable support for improvements in predictive performance for the SVC model relative to the three alternative models. The SVC model generally had the highest performance in predicting future species-specific occurrence, with AUC being highest for 69% (35 out of 51) of species (Supplemental Information S2: Table S3).

The SVC model revealed spatial heterogeneity in occurrence trends that was not evident in the three simpler alternative models. Figure 3 shows the trend estimates for three example species (grey catbird (GRCA), Eastern phoebe (EAPH), and wood thrush (WOTH)) from each of the four models. The SVC model revealed grey catbird had predominately negative trends in the southern portion of its range (except Louisiana and Mississippi) and positive or no directional trends in the northern states, indicating a potential northward shift in its range. Both the temperature model and the strata model were able to capture the general pattern of more negative trends in the southern portion of the eastern United States, but they both failed to capture more fine scale variability that was revealed by the SVC model (i.e. positive trends in Louisiana and Mississippi and

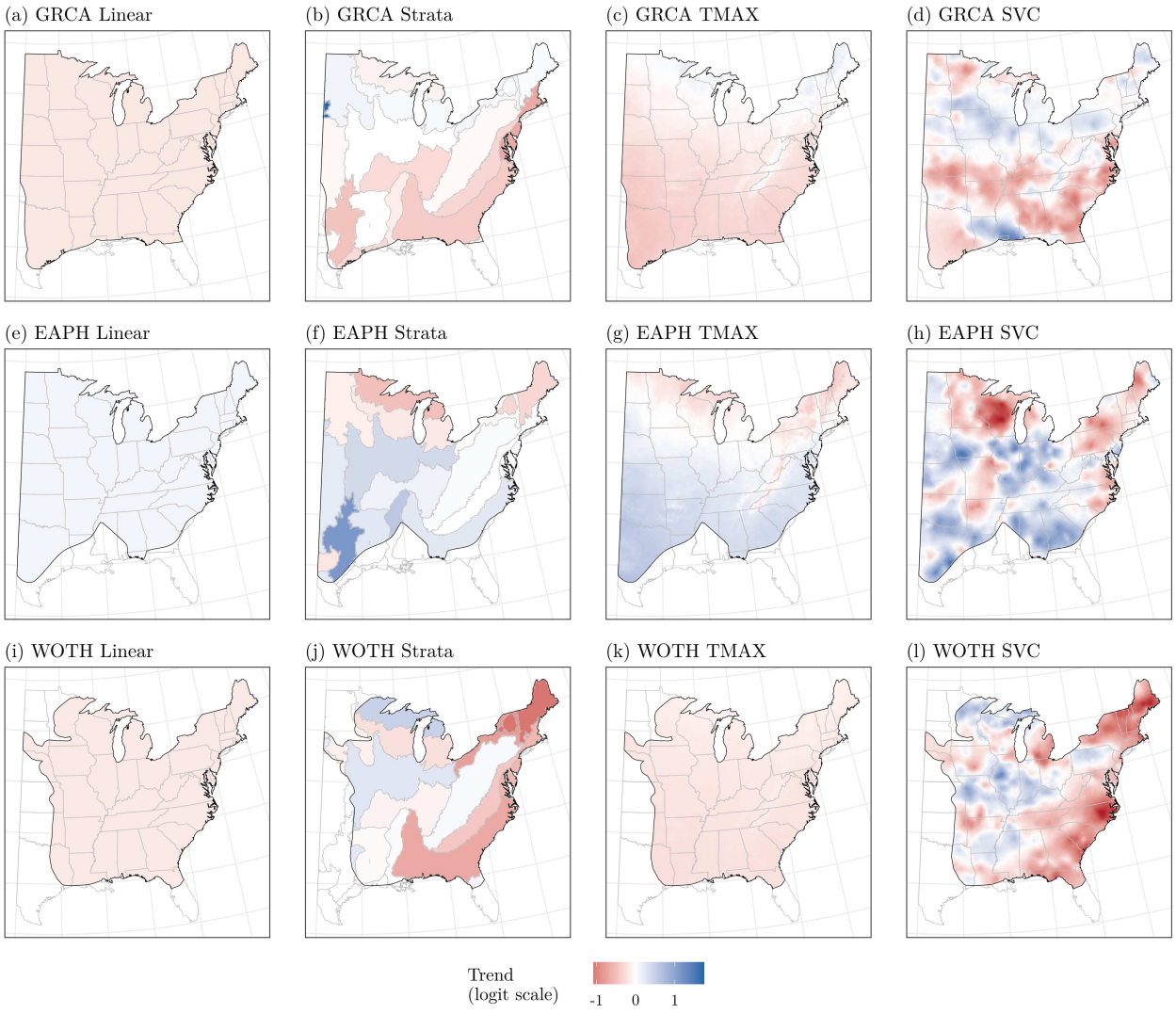


**FIGURE 3** Median predictions of an occurrence trend from 2000 to 2019 for three example species from the four models: a spatial occupancy model with a constant linear trend across the species range (Linear), a spatial occupancy model with a separate trend for each Bird Conservation Region (Strata), a spatial occupancy model with a trend that interacts with 30-year average maximum temperature (TMAX), and a spatially varying coefficient occupancy model estimating a spatially varying trend (SVC). Panels (a–d): grey catbird (3.79 million $km^2$); Panels (e–h): Eastern phoebe (3.55 million $km^2$); Panels (i–l): wood thrush (3.12 million $km^2$). Values represent the change in log odds of occurrence probability over one standard deviation of time (i.e. approximately 5.8 years).

negative trends in northern Minnesota). We found the opposite pattern for Eastern phoebe, where the strata and temperature models were able to adequately capture a generally positive trend in southern portions of the range and negative trend in northern portions of the range, but did not capture more fine-scale variability in trends revealed by the SVC model. Wood thrush trends were strongly negative along the eastern portion of its range and more positive in the northwestern part of its range, which were largely captured by both the SVC and strata model. In contrast, the lack of relationship between maximum temperature and trend for Wood Thrush resulted in essentially no spatial variation in the trend from the temperature model. This result illustrates the lack of flexibility in the interaction models if the interacting variable does not adequately explain spatial variability in the trend.

The temperature model revealed large heterogeneity in the sign and significance of the interaction between maximum temperature and the yearly trend (Supplemental Information S2: Figure S1), providing minimal support for climatic niche position being a consistent driver of spatial variation in eastern forest bird occurrence trends. Of the 51 total species, 18 species had a significant negative interaction (i.e. trends were less positive/more negative in areas with higher temperatures), while 8 species had a significant positive interaction (i.e. trends were more positive/less negative in areas with higher temperatures). While such significant interactions indicate some spatial variability in trends may be related to climatic position within a range, the large improvement in model fit (and to a lesser extent, prediction) of the SVC model compared to the temperature model suggests that this relationship is not the primary driver of spatial variability in occurrence trends for eastern forest birds. Similarly, the improved performance of the strata model compared to the constant linear trend model suggests spatial variability in trends may be partially attributed to broad-scale variation in habitat and climate conditions across Bird Conservation Regions. However, the improved performance of the SVC model compared to the strata model suggests that additional factors operating at finer spatial scales are important contributors to spatial variability in occurrence trends.

## 5 | Case study 2: Effects of land cover change on the grasshopper sparrow

Quantifying the effects of land cover change and resulting shifts in habitat availability on species distributions is crucial for understanding the primary drivers of large-scale avian population declines (Rosenberg et al., 2019). Grassland birds have experienced some of the steepest population declines across all bird groups in North America, largely due to habitat loss and agricultural intensification (Stanton et al., 2018). However, the effects of shifts in land cover on grassland bird species is unlikely to be constant across space as a result of complex interactive effects with local climate, farmland management practices or predation pressure. In this case study, we demonstrate the ability of SVC models to provide insight on the spatially varying effects of habitat change on the distribution of grasshopper sparrow (*Ammodramus savannarum*) across the continental United States from 1970 to 2019. We again use data from the North American BBS (Pardieck et al., 2020), collected over the 50-year period at $J = 2542$ routes within the range of the grasshopper sparrow (derived from BirdLife International [2021]) as described in Case Study 1. Our objectives for this case study were to (1) develop spatially explicit maps of the effect of change in grassland area and cropland area on grasshopper sparrow occurrence; and (2) compare a series of SVC and simpler alternative models to generate and test hypotheses regarding the drivers of spatial variability in the effects of habitat change.

We summarized the BBS data in the same manner as in Case Study 1, using $K = 5$ spatial replicates of detection-nondetection data at each BBS route to model route-level occurrence across the sparrow's range. We calculated annual amount of grassland area (natural grassland cover class) and cropland area (combined cropland and hay/pasture cover classes) within 1 km of each BBS route using data from the USGS EROS Center (Sohl et al., 2016). We calculated annual deviations in grassland and cropland area by subtracting the 50-year average value at each site from each yearly value to assess effects of temporal change in habitat separately from spatial variation in habitat availability (Clement et al., 2019; Saunders et al., 2022).

We hypothesized that grasshopper sparrow occurrence would be positively associated with grassland area change and that the effect would vary spatially as a result of: (1) a positive interaction with the average amount of grassland area given grasshopper sparrow's preference for landscapes comprised of large amounts of contiguous grassland (Shaffer et al., 2021); (2) an interaction with temperature such that effects of grassland area change are strongest near the climatic extremes of the species range (i.e. range boundaries); and (3) additional interactions with fine-scale habitat quality (e.g. grassland height, amount of litter; Shaffer et al., 2021) and management actions that were not available as covariates for the model. We expected grasshopper sparrow occurrence to be negatively related to cropland cover change across the Great Plains region, as an increase in cropland cover in this region would likely correspond to a decrease in grassland area (the dominant land cover type; Supplemental Information S3: Figure S1). However, in areas with minimal grassland, we predicted a positive association between occurrence and cropland area, as grasshopper sparrow occurrence has previously been associated with hay-fields and cultivated fields when native grassland is limited (Shaffer et al., 2021).

We fit five candidate models that varied in the functional forms of the effects of grassland area change and cropland area change to test our hypotheses (full details in Supplemental Information S3). Specifically, our five models consisted of: (1) a linear model with constant, linear effects of grassland and cropland change; (2) a habitat interaction model with linear effects of grassland and cropland change, an interaction of grassland change with 50-year average grassland area, and an interaction of cropland change with 50-year average cropland area; (3) a temperature interaction model with linear effects of grassland and cropland change that both also had an interaction with average temperature conditions (i.e. 30-year

maximum temperature calculated from TerraClimate as in Case Study 1); (4) an SVC model; (5) the "full" model that contained SVCs, interactive effects of temperature, and interactive effects of 50-year average land-cover.

For each of the five candidate models, we fit a Bayesian multi-season occupancy model using the `spOccupancy` R package (Doser et al., 2022). Detection probability was modelled consistently across the five models as a function of linear and quadratic ordinal date, a random effect of year and a separate intercept of survey replicate to account for variability in detection probability across the five spatial replicates within a BBS route. Given our focus on inference of the species–environment relationships, we compared candidate models using the WAIC. Prior distributions were either vague or weakly informative (Supplemental Information S3). For each model, we ran three chains for 100,000 MCMC iterations with a burn-in period of 50,000 iterations and a thinning rate of 50, yielding 3000 posterior samples. Convergence was assessed using the potential scale reduction factor and visual assessment of traceplots.

## 5.1 | RESULTS

We found strong support for spatial variability in the effects of grassland and cropland cover change, with all models that included an SVC and/or an interaction substantially outperforming (i.e. $\Delta$WAIC > 2) the model with constant effects (Supplemental Information S3: Table S1). The temperature interaction model outperformed the habitat interaction model ($\Delta$WAIC = 18.65), indicating maximum temperature was more important in explaining spatial variability in the effect of habitat change than the amount of habitat. Noticeably, including an SVC for the effect of grassland and cropland change reduced WAIC (i.e. $\Delta$WAIC = 774.90) substantially more than either of the interaction models compared to the constant model ($\Delta$WAIC = 20.69 for the habitat interaction model and $\Delta$WAIC = 39.34 for the temperature interaction model). The model including SVCs, a habitat interaction, and a temperature interaction slightly outperformed the model with only SVCs ($\Delta$WAIC = 3.32). Altogether, these results suggest that interactions with temperature and habitat explain some spatial variability in the effect of habitat change on occurrence, but most of the variation in these effects is the result of unexplained spatial variation that is accounted for by the SVCs. Maps of the predicted effects of grassland and cropland change from the candidate models reveal that models with the SVC capture far more spatial variation in the effects of habitat change than do models without SVCs (Figure 4).

The effects estimated from the candidate models revealed mixed support for our hypotheses. The best performing model revealed a range of positive and negative effects of habitat change across the breeding range of the grasshopper sparrow. As predicted, the effect of grassland change was strongly positive in the Northern Great Plains (where grassland availability is higher), suggesting that in heavily grassland-dominated landscapes, loss of grassland would result in declines in grasshopper sparrow occurrence probability. This is further supported by the habitat interaction model, which revealed a positive interaction between grassland change and average grassland area (Figure 4a). Surprisingly, we found near zero or negative

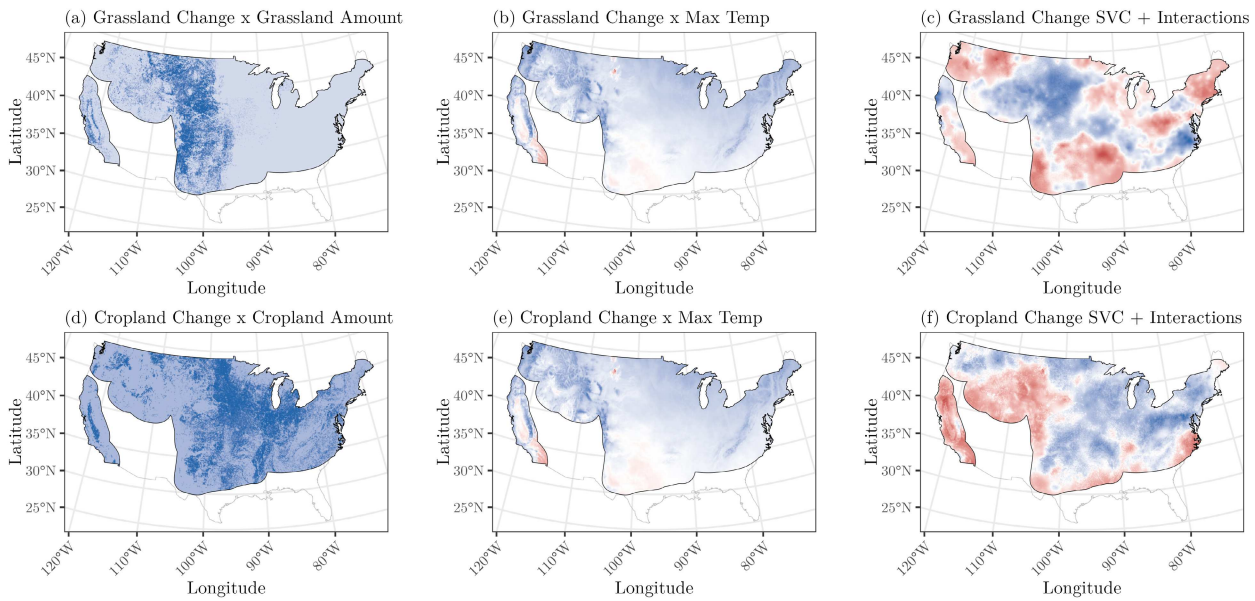


FIGURE 4 Median predictions of the effects of grassland change (top row) and cropland change (bottom row) on occurrence of grasshopper sparrow (*Ammodramus savannarum*) from three of the five candidate models. Panels (a) and (d) show estimates from a model with an interaction between land-cover change and average land-cover area over the 50-year period. Panels (b) and (e) show estimates from a model with an interaction between land-cover change and 30-year average maximum temperature. Panels (c) and (f) show estimates from a model with spatially varying coefficients for land-cover change and interactions with average land-cover area and maximum temperature. Blue indicates a positive effect, white indicates no effect and red indicates a negative effect.

effects of grassland area change in the Southern Great Plains, indicating increasing grassland in this area would result in no effect or even declines of occurrence probability. Given the relatively high amount of grassland and rangeland area in this region (Supplemental Information S3: Figure S1), this could indicate a regional peak in the optimal amount of grassland area for grasshopper sparrow occurrence probability, which is in line with previous work showing that the preferred grassland size of the species varies across ecoregions (Johnson & Igl, 2001). The temperature interaction model revealed this pattern was partially related to a negative interaction with maximum temperature (Figure 4b; Gorzo et al., 2016). Grasshopper sparrow occurrence probability was negatively related to cropland area along most of its southern range boundary, in part due to a negative interaction with maximum temperature (Figure 4). Alternatively, the effect of cropland change was generally positive throughout the Northeast and Midwest (Figure 4), which was related to a positive interaction between cropland change and cropland area (Figure 4d). In the northeastern and midwestern United States, where few native grasslands remain, the positive effect of cropland area change indicates that increases in cropland area would result in increases in occurrence probability. Furthermore, the northeast is largely dominated by forest, and thus increases in cropland cover are likely associated with declines in forest cover, which may partly explain the positive effect of cropland cover change, as grasshopper sparrows avoid forest (Grant et al., 2004). While such assessments are speculative, these insights are only possible because of the power of SVC models to reveal fine-scale, multifaceted species–environment relationships, which in turn can be used to inform local and regional management priorities.

## 6 | DISCUSSION

Accounting for complex species–environment relationships is increasingly important as the spatial and temporal extent of ecological research expands (Rollinson et al., 2021). Widely used statistical methods, such as interactions, stratification and nonlinear models, can partially account for such patterns, but they are limited in their ability to estimate spatially varying species–environment relationships that arise from multiple interacting factors that themselves vary spatially. Here we used spatially varying coefficients (SVCs) in SDMs to model nonlinear and/or spatially varying species environment relationships within a hierarchical GLM framework while simultaneously addressing observational biases common in both wildlife and plant datasets. Using simulations and two case studies, we highlighted the inferential benefits of SVC SDMs to generate and test ecological hypotheses regarding the factors driving spatial variability in estimated relationships and/or occurrence trends.

Our simulation study showed that SVC SDMs can accurately capture complex, spatially varying species–environment relationships under different forms, while they can also reveal more simple species–environment relationships (e.g. linear, quadratic; Figure 1) if such additional complexity is not supported by the data. Thus, when little is known regarding the form of the species–environment relationship prior to model fitting, SVC SDMs can be used to generate hypotheses on the true form of the relationship and associated abiotic and/or ecological drivers. When the true species–environment relationship is known a priori, simpler parametric GLMs will likely outperform SVC SDMs according to information criteria based on the principle of parsimony. In such situations, our simulation suggests that SVC SDMs will not generate erroneous conclusions, but rather will reveal the simpler, parametric form of the true relationship. When working across macroscales, it is unlikely that all interacting variables are known and/or available prior to model fitting, in which case SVC SDMs will outperform simpler alternatives (rows 5, 6; Figure 1) and help elucidate the ecological drivers of such patterns.

When assessing spatial variability in species–environment relationships and/or trends, we recommend comparing SVC SDMs with simpler parametric SDMs that represent explicit hypotheses, as such comparisons can reveal the amount of support for different drivers of spatially varying effects/trends (Pease, Pacifici, Kays, & Reich, 2022). For example, in the eastern forest bird case study, the temperature model revealed a significant negative interaction of trend and breeding season maximum temperature for 18 species and a significant positive interaction for 8 species (Supplemental Information S3 Figure S1). However, the SVC model was the best-performing model for all 26 species with significant temperature interactions, suggesting that while breeding season temperature often explains some variation in occurrence trends, there are additional factors that are important in explaining fine-scale variability in occurrence trends.

Whether SVC models improve predictive performance over models that only include a spatially varying intercept is an ongoing question. In our eastern forest bird case study, the SVC model provided relatively minor improvements in predictive performance compared to the simpler models when forecasting occurrence probability in 2021. All four models included a spatial random effect to account for spatial variability in occurrence probability, and given the likely small changes in the forest bird distributions from 2019 (the last year in the modelled data set) to 2021, they all had similar abilities to predict occurrence probabilities across the study region. SVC models in other ecological and natural resource applications have shown mixed results regarding their predictive benefits compared to models with only a spatially varying intercept; some studies found improved predictive performance of SVC models (May et al., 2023; Sultaire et al., 2022), while others showed improvements that vary depending on the species (Doser et al., 2024; Pease, Pacifici, & Kays, 2022) or region (Babcock et al., 2015). Regardless, we echo the statements of Thorson et al. (2023) that the primary benefits of SVC SDMs relate to their improved ability to test and generate hypotheses as well as answer relevant ecological questions regarding spatial variability in species–environment relationships and trends.

In addition to theoretical contributions, the results from SVC SDMs could be applied to multi-scale conservation and management decisions. For example, in the grasshopper sparrow case study, we found that loss of grassland area is most likely to have the largest

negative impact on the species in the Northern Great Plains, emphasizing the importance of providing large, contiguous patches of natural grassland to prevent further regional declines (Shaffer et al., 2021). By performing similar analyses for multiple grassland bird species, SVC SDM outputs could be used together as inputs for spatial prioritizations. This could offer major improvements in reserve design and help resource managers identify the exact locations where habitat restoration may be most beneficial to bird communities (Grand et al., 2019). Alternatively, estimates of species trends serve as the foundation for assigning conservation status to species of greatest conservation need. Spatially varying trends from SVC SDMs, like those generated in the eastern forest bird case study, could be used to tailor action plans at local (i.e. state-level), regional (i.e. ecoregion), and continental scales (Smith et al., 2024). Ultimately, this could improve our understanding of how and why conservation strategies in different regions lead to variable outcomes. Estimation of local trends with SVC SDMs can also improve predictions of species distribution changes (Barnett et al., 2021) in response to invasive species (Thorson et al., 2023) and future climate and/or land-use changes (Gonthier et al., 2014).

While other approaches (e.g. GAMs, random forests, MaxEnt) are commonly used in ecology to account for complex species–environment relationships, Bayesian spatially varying coefficient models are an attractive alternative as they (1) do not require a priori knowledge of interacting variables; (2) can readily provide uncertainty measures associated with all estimates; and (3) are easily embedded in hierarchical modelling frameworks (i.e. occupancy models) used to address observational biases prevalent in ecological data. Nevertheless, the flexibility provided by Bayesian SVC SDMs can lead to computational and practical difficulties in their implementations. While the Bayesian framework provides full uncertainty propagation into all estimates and predictions, models can take substantial time to run. For example, the full SVC model for the grasshopper sparrow case study with a data set comprised of nearly 400,000 observations took approximately 10 h to run a single MCMC chain of 100,000 samples using `spOccupancy` (Doser et al., 2022). Additionally, the ability of SVC SDMs to estimate complex spatially varying species–environment relationships can require large sample sizes to achieve reasonable levels of uncertainty compared to simpler alternatives. This is particularly true when working with detection-nondetection data, which provide relatively little information to estimate SVCs compared to count (e.g. abundance) or continuous (e.g. biomass) data sources used in many SDMs. In Supplemental Information S4, we provide additional simulation studies that give insights on how the reliability of SVC estimates scales with the number of spatial locations and number of seasons in the data set. When sample sizes are limited, simpler approaches like stratification or interactions may be more useful options to yield estimates of species–environment relationships without considerable uncertainty. Lastly, confounding can occur between the estimated spatially varying intercept and spatially varying coefficients, especially when working with modestly sized data sets (e.g. 500 data points), which could potentially lead to misleading conclusions. However, when estimating SVCs for covariates that vary across time (e.g. a temporal trend or habitat change as in our two case studies), confounding is minimized due to the added temporal component of multi-season data. We have found that recent guidelines for minimizing spatial confounding and understanding its effects in spatially explicit SDMs are applicable to SVC SDMs (Mäkinen et al., 2022), although further research is needed to understand when such confounding may occur and how to best mitigate it.

Spatial variability in species–environment relationships is prevalent throughout ecology (Rollinson et al., 2021) as a result of complex interactions with abiotic and biotic variables, which are rarely all known or available to be measured prior to statistical analysis. As we demonstrate in this study, the use of spatially varying coefficients in species distribution models can help elucidate the environmental factors that drive species distributional dynamics across both local and broad spatial scales. This provides an improved ability to test ecological hypotheses and inform multi-scale conservation and management initiatives. When fitting SDMs across macroscales, we encourage the comparison of SVC SDMs with simpler alternatives as a means of advancing our understanding of the drivers of species–environment relationships across space.

## ACKNOWLEDGEMENTS

We thank Antoine Guisan, William Godsoe and two anonymous reviewers for extremely insightful comments that substantially improved the manuscript.

## FUNDING INFORMATION

This work was supported by National Science Foundation (NSF) grants DBI-1954406, DMS 1916395, and DEB-2213565.

## CONFLICT OF INTEREST STATEMENT

We declare no conflicts of interest.

## DATA AVAILABILITY STATEMENT

All data and code used in the manuscript are available at https://doi.org/10.5281/zenodo.10294522 (Doser et al., 2023).

## ORCID

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

## REFERENCES

Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A., & Hegewisch, K. C. (2018). TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Scientific Data*, *5*(1), 1–12.

Amburgey, S. M., Miller, D. A., Campbell Grant, E. H., Rittenhouse, T. A., Benard, M. F., Richardson, J. L., Urban, M. C., Hughson, W., Brand, A. B., Davis, C. J., & Hardin, C. R. (2018). Range position and climate sensitivity: The structure of among-population demographic responses to climatic variation. *Global Change Biology*, *24*(1), 439–454.

Babcock, C., Finley, A. O., Bradford, J. B., Kolka, R., Birdsey, R., & Ryan, M. G. (2015). Lidar based prediction of forest biomass using hierarchical models with spatially varying coefficients. *Remote Sensing of Environment*, *169*, 113–127.

Babcock, C., Finley, A. O., Cook, B. D., Weiskittel, A., & Woodall, C. W. (2016). Modeling forest biomass and growth: Coupling long-term inventory and lidar data. *Remote Sensing of Environment*, *182*, 1–12.

Banerjee, S., Carlin, B. P., & Gelfand, A. E. (2014). *Hierarchical modeling and analysis for spatial data*. Chapman and Hall/CRC.

Barnett, L. A., Ward, E. J., & Anderson, S. C. (2021). Improving estimates of species distribution change by incorporating local trends. *Ecography*, *44*(3), 427–439.

Bateman, B. L., Wilsey, C., Taylor, L., Wu, J., LeBaron, G. S., & Langham, G. (2020). North American birds require mitigation and adaptation to reduce vulnerability to climate change. *Conservation Science and Practice*, *2*(8), e242.

BirdLife International. (2021). Handbook of the birds of the world 2021. Bird species distribution maps of the world, Ver. 2021. http://datazone.birdlife.org/species/requestdis

Bled, F., Sauer, J., Pardieck, K., Doherty, P., & Royle, J. A. (2013). Modeling trends from North American Breeding Bird Survey data: A spatially explicit approach. *PLoS ONE*, *8*(12), e81867.

Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, *7*(4), 434–455.

Clement, M. J., Nichols, J. D., Collazo, J. A., Terando, A. J., Hines, J. E., & Williams, S. G. (2019). Partitioning global change: Assessing the relative importance of changes in climate and land cover for changes in avian distribution. *Ecology and Evolution*, *9*(4), 1985–2003.

Crossley, M. S., Smith, O. M., Davis, T. S., Eigenbrode, S. D., Hartman, G. L., Lagos-Kutz, D., Halbert, S. E., Voegtlin, D. J., Moran, M. D., & Snyder, W. E. (2021). Complex life histories predispose aphids to recent abundance declines. *Global Change Biology*, *27*(18), 4283–4293.

Datta, A., Banerjee, S., Finley, A. O., & Gelfand, A. E. (2016). Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. *Journal of the American Statistical Association*, *111*(514), 800–812.

Doser, J. W., Finley, A. O., Kéry, M., & Zipkin, E. F. (2022). spOccupancy: An R package for single-species, multi-species, and integrated spatial occupancy models. *Methods in Ecology and Evolution*, *13*(8), 1670–1678.

Doser, J. W., Finley, A. O., Saunders, S. P., Kéry, M., Weed, A. S., & Zipkin, E. F. (2024). Modeling complex species-environment relationships through spatially-varying coefficient occupancy models. *Journal of Agricultural, Biological, and Environmental Statistics*, 1–26.

Doser, J. W., Kéry, M., Saunders, S. P., Finley, A. O., Bateman, B. L., Grand, J., Reault, S., Weed, A. S., & Zipkin, E. F. (2023). *Code and data for guidelines for the use of spatially-varying coefficients in species distribution models*. Zenodo. https://doi.org/10.5281/zenodo.10294522

Ethier, D. M., Koper, N., & Nudds, T. D. (2017). Spatiotemporal variation in mechanisms driving regional-scale population dynamics of a threatened grassland bird. *Ecology and Evolution*, *7*(12), 4152–4162.

Golding, N., & Purse, B. V. (2016). Fast and flexible Bayesian species distribution modelling using Gaussian processes. *Methods in Ecology and Evolution*, *7*(5), 598–608.

Gonthier, D. J., Ennis, K. K., Farinas, S., Hsieh, H.-Y., Iverson, A. L., Batáry, P., Rudolphi, J., Tscharntke, T., Cardinale, B. J., & Perfecto, I. (2014). Biodiversity conservation in agriculture requires a multi-scale approach. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1791), 20141358.

Gorzo, J. M., Pidgeon, A. M., Thogmartin, W. E., Allstadt, A. J., Radeloff, V. C., Heglund, P. J., & Vavrus, S. J. (2016). Using the north American Breeding Bird Survey to assess broad-scale response of the continent's most imperiled avian community, grassland birds, to weather variability. *The Condor: Ornithological Applications*, *118*(3), 502–512.

Grand, J., Wilsey, C., Wu, J. X., & Michel, N. L. (2019). The future of North American grassland birds: Incorporating persistent and emergent threats into full annual cycle conservation priorities. *Conservation Science and Practice*, *1*(4), e20.

Grant, T. A., Madden, E., & Berkey, G. B. (2004). Tree and shrub invasion in northern mixed-grass prairie: Implications for breeding grassland birds. *Wildlife Society Bulletin*, *32*(3), 807–818.

Guisan, A., Edwards, T. C., Jr., & Hastie, T. (2002). Generalized linear and generalized additive models in studies of species distributions: Setting the scene. *Ecological Modelling*, *157*(2–3), 89–100.

Guisan, A., Thuiller, W., & Zimmermann, N. E. (2017). *Habitat suitability and distribution models: With applications in R*. Cambridge University Press.

Guisan, A., & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, *135*(2–3), 147–186.

Hosmer Jr, D. W., Lemeshow, S., and Sturdivant, R. X. (2013). *Applied logistic regression*, volume 398. John Wiley & Sons.

Hostetler, J. A., & Chandler, R. B. (2015). Improved state-space models for inference about spatial and temporal variation in abundance from count data. *Ecology*, *96*(6), 1713–1723.

Johnson, D. H., & Igl, L. D. (2001). Area requirements of grassland birds: A regional perspective. *The Auk*, *118*(1), 24–34.

Kellner, K. F., & Swihart, R. K. (2014). Accounting for imperfect detection in ecology: A quantitative review. *PLoS ONE*, *9*(10), e111436.

Kendall, W. L., & White, G. C. (2009). A cautionary note on substituting spatial subunits for repeated temporal sampling in studies of site occupancy. *Journal of Applied Ecology*, *46*(6), 1182–1188.

Kéry, M. (2011). Towards the modelling of true species distributions. *Journal of Biogeography*, *38*(4), 617–618.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, *2*(3), 18–22.

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.

MacKenzie, D. I., Nichols, J. D., Royle, J. A., Pollock, K. H., Bailey, L. L., & Hines, J. E. (2017). *Occupancy estimation and modeling: Inferring patterns and dynamics of species occurrence*. Elsevier.

Mäkinen, J., Numminen, E., Niittynen, P., Luoto, M., & Vanhatalo, J. (2022). Spatial confounding in Bayesian species distribution modeling. *Ecography*, *2022*(11), e06183.

May, P., McConville, K. S., Moisen, G. G., Bruening, J., & Dubayah, R. (2023). A spatially varying model for small area estimates of biomass density across the contiguous United States. *Remote Sensing of Environment*, *286*, 113420.

Meehan, T. D., Michel, N. L., & Rue, H. (2019). Spatial modeling of Audubon Christmas bird counts reveals fine-scale patterns and drivers of relative abundance trends. *Ecosphere*, *10*(4), e02707.

Miller, J. A. (2012). Species distribution models: Spatial autocorrelation and non-stationarity. *Progress in Physical Geography*, *36*(5), 681–692.

Osborne, P. E., & Suárez-Seoane, S. (2002). Should data be partitioned spatially before building large-scale distribution models? *Ecological Modelling*, *157*(2–3), 249–259.

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

Pease, B. S., Pacifici, K., & Kays, R. (2022). Exploring spatial nonstationarity for four mammal species reveals regional variation in environmental relationships. *Ecosphere*, *13*(8), e4166.

Pease, B. S., Pacifici, K., Kays, R., & Reich, B. (2022). What drives spatially varying ecological relationships in a wide-ranging species? *Diversity and Distributions*, *28*, 1752–1768.

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, *190*(3–4), 231–259.

Rollinson, C. R., Finley, A. O., Alexander, M. R., Banerjee, S., Dixon Hamil, K.-A., Koenig, L. E., Locke, D. H., DeMarche, M. L., Tingley, M. W., Wheeler, K., Youngflesh, C., & Zipkin, E. F. (2021). Working across space and time: Nonstationarity in ecological research and application. *Frontiers in Ecology and the Environment*, *19*(1), 66–72.

Rosenberg, K. V., Dokter, A. M., Blancher, P. J., Sauer, J. R., Smith, A. C., Smith, P. A., Stanton, J. C., Panjabi, A., Helft, L., Parr, M., & Marra, P. P. (2019). Decline of the North American avifauna. *Science*, *366*(6461), 120–124.

Rushing, C. S., Royle, J. A., Ziolkowski, D. J., Jr., & Pardieck, K. L. (2020). Migratory behavior and winter geography drive differential range shifts of eastern birds in response to recent climate change. *Proceedings of the National Academy of Sciences*, *117*(23), 12897–12903.

Saunders, S. P., Meehan, T. D., Michel, N. L., Bateman, B. L., DeLuca, W., Deppe, J. L., Grand, J., LeBaron, G. S., Taylor, L., Westerkam, H., Wu, J. X., & Wilsey, C. B. (2022). Unraveling a century of global change impacts on winter bird distributions in the eastern United States. *Global Change Biology*, *28*, 2221–2235.

Shaffer, J. A., Igl, L. D., Johnson, D. H., Sondreal, M. L., Goldade, C. M., Nenneman, M. P., Wooten, T. L., & Euliss, B. R. (2021). *The effects of management practices on grassland birds—Grasshopper Sparrow (Ammodramus savannarum)*. U.S. Geological Survey Professional Paper.

Smith, A. C., Binley, D., A., Daly, L., Edwards, B. P. M., Ethier, D., Frei, B., Iles, D., Meehan, T. D., Michel, N. L., & Smith, P. A. (2024). Spatially explicit Bayesian hierarchical models improve estimates of avian population status and trends. *Ornithological Applications*, *126*(1), duad056.

Smith, A. C., & Edwards, B. P. (2021). North American Breeding Bird Survey status and trend estimates to inform a wide range of conservation needs, using a flexible Bayesian hierarchical generalized additive model. *The Condor*, *123*(1), duaa065.

Sohl, T., Reker, R., Bouchard, M., Sayler, K., Dornbierer, J., Wika, S., Quenzer, R., & Friesz, A. (2016). Modeled historical land use and land cover for the conterminous United States. *Journal of Land Use Science*, *11*(4), 476–499.

Spake, R., Bowler, D. E., Callaghan, C. T., Blowes, S. A., Doncaster, C. P., Antao, L. H., Nakagawa, S., McElreath, R., & Chase, J. M. (2023). Understanding 'it depends' in ecology: A guide to hypothesising, visualising and interpreting statistical interactions. *Biological Reviews*, *98*, 983–1002.

Stanton, R., Morrissey, C. A., & Clark, R. G. (2018). Analysis of trends and agricultural drivers of farmland bird declines in north america: A review. *Agriculture, Ecosystems & Environment*, *254*, 244–254.

Stein, M. L. (2014). Limitations on low rank approximations for covariance matrices of spatial data. *Spatial Statistics*, *8*, 1–19.

Sultaire, S. M., Humphreys, J. M., Zuckerberg, B., Pauli, J. N., & Roloff, G. J. (2022). Spatial variation in bioclimatic relationships for a snow-adapted species along a discontinuous southern range boundary. *Journal of Biogeography*, *49*(1), 66–78.

Thorson, J. T., Barnes, C. L., Friedman, S. T., Morano, J. L., & Siple, M. C. (2023). Spatially varying coefficients can improve parsimony and descriptive power for species distribution models. *Ecography*, *2023*, e06510.

Tyre, A. J., Tenhumberg, B., Field, S. A., Niejalke, D., Parris, K., & Possingham, H. P. (2003). Improving precision and reducing bias in biological surveys: Estimating false-negative error rates. *Ecological Applications*, *13*(6), 1790–1801.

Valavi, R., Elith, J., Lahoz-Monfort, J. J., & Guillera-Arroita, G. (2021). Modelling species presence-only data with random forests. *Ecography*, *44*(12), 1731–1742.

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, *11*(12), 3571–3594.

Wood, S. N. (2006). *Generalized additive models: An introduction with R*. Chapman and Hall/CRC.

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.

## BIOSKETCH

Jeffrey W. Doser is a statistical ecologist who develops and applies hierarchical Bayesian models to inform wildlife and natural resource conservation across spatial scales.

## SUPPORTING INFORMATION

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

**How to cite this article:** Doser, J. W., Kéry, M., Saunders, S. P., Finley, A. O., Bateman, B. L., Grand, J., Reault, S., Weed, A. S., & Zipkin, E. F. (2024). Guidelines for the use of spatially varying coefficients in species distribution models. *Global Ecology and Biogeography*, *33*, e13814. https://doi.org/10.1111/geb.13814