

# Application of Spatially-Varying Coefficient Models

---

Jeffrey Doser<sup>1</sup> & Andrew Finley<sup>2</sup>

May 15, 2023

<sup>1</sup>Department of Integrative Biology, Michigan State University.

<sup>2</sup>Department of Forestry, Michigan State University.

# Spatially-Varying Coefficient Models

- Extension of spatial regression approaches that allow regression coefficients to vary across space, and not just the intercept
- SVC models are random slopes models, with spatial structure given to the random slopes

$$y(\mathbf{s}_i) \sim f(\mu(\mathbf{s}_i), \tau)$$

$$\mu(\mathbf{s}_i) = g^{-1}(\eta(\mathbf{s}_i)) = \mathbf{x}(\mathbf{s}_i)^\top \tilde{\boldsymbol{\beta}}(\mathbf{s}_i)$$

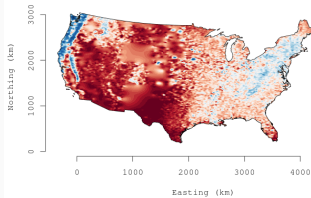
$$\tilde{\beta}_r(\mathbf{s}_i) = \beta_r + \mathbf{w}_r(\mathbf{s}_i) \text{ for each } r = 1, \dots, p$$

- We can model  $\mathbf{w}(\mathbf{s}_i)$  using a GP, predictive process, or **NNGP**
- We can envision modeling  $\mathbf{w}(\mathbf{s}_i)$  in two ways:
  1. Multivariate NNGP (see previous forest biomass example)
  2. **Independent NNGPs**
- Here we focus on the latter
- Pros and cons to both approaches, similar to correlations between random slopes and intercepts in mixed models

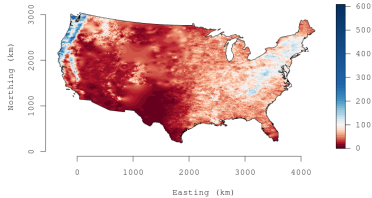
## Potential benefits of SVC models

- Improved predictive performance
- Tremendous flexibility to accommodate spatial variability in effects
- Hypothesis testing and generation
- Accommodate highly non-linear relationships
- Model spatial variability in trends over time

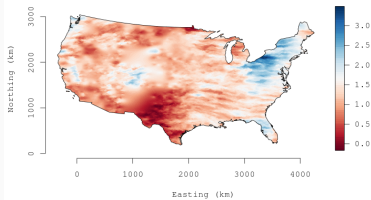
# Improved predictive performance



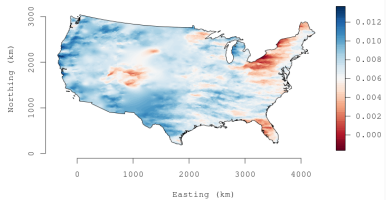
Observed biomass



Fitted biomass



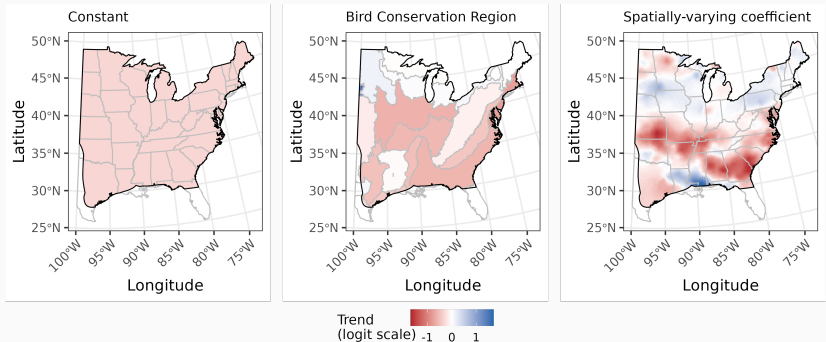
$\beta_0(s)$



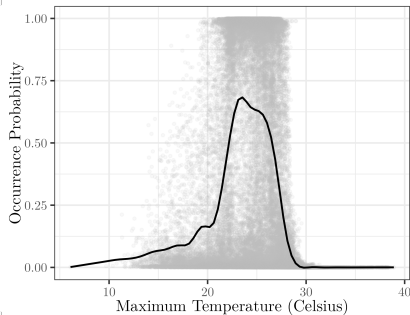
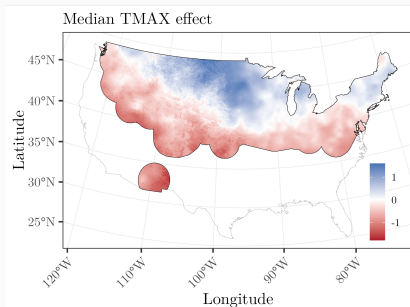
$\beta_{NDVI}(s)$

# More flexibility to accommodate spatial variability in effects

Gray Catbird occurrence trend across the eastern US from 2000-2019



# Example: Effect of max temperature on Bobolink occurrence

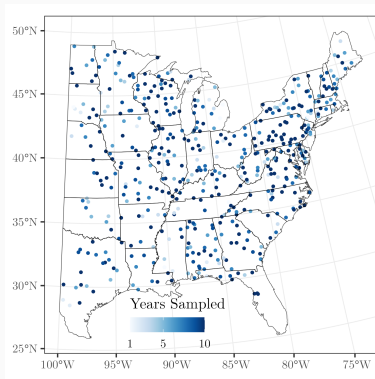


- spBayes: univariate Gaussian SVC with full GPs
- spOccupancy: univariate Binomial SVC with NNGPs (multivariate on its way)
- varycoef: maximum likelihood Gaussian SVCs (Dambon et al. 2021 *Spatial Stats.*)
- sdmTMB: penalized likelihood and Bayesian SVC GLMMs (Anderson et al. 2022 *bioRxiv*)



## Exercise: 10-year occurrence trend of Wood Thrush

- Data come from USGS North American Breeding Bird Survey
- We desire to account for observational biases in detection of the birds (i.e., false negatives).
- Add on an additional observational layer to our hierarchical model



## Exercise: Process model

- Let  $z_t(\mathbf{s}_i)$  denote the true presence (1) or absence (0) of the species at site  $i = 1, \dots, 500$  during year  $t = 1, \dots, 10$ .

## Exercise: Process model

- Let  $z_t(\mathbf{s}_i)$  denote the true presence (1) or absence (0) of the species at site  $i = 1, \dots, 500$  during year  $t = 1, \dots, 10$ .
- If the bird was detected at a site and year, we know  $z_t(\mathbf{s}_i) = 1$ . If not, it might be there and we just missed it during the surveys.

## Exercise: Process model

- Let  $z_t(\mathbf{s}_i)$  denote the true presence (1) or absence (0) of the species at site  $i = 1, \dots, 500$  during year  $t = 1, \dots, 10$ .
- If the bird was detected at a site and year, we know  $z_t(\mathbf{s}_i) = 1$ . If not, it might be there and we just missed it during the surveys.
- We model  $z_t(\mathbf{s}_i)$  just as before with a Bernoulli GLM, with a SVC for trend

$$z_t(\mathbf{s}_i) \sim \text{Bernoulli}(\psi_t(\mathbf{s}_i))$$
$$\text{logit}(\psi_t(\mathbf{s}_i)) = \tilde{\beta}_0(\mathbf{s}_i) + \tilde{\beta}_1(\mathbf{s}_i) \cdot \text{YEAR}_t$$

- $\tilde{\beta}_0(\mathbf{s}_i)$  and  $\tilde{\beta}_1(\mathbf{s}_i)$  are modeled as independent SVCs with NNGPs

## Exercise: Observation model

- Let  $y_{t,k}(\mathbf{s}_i)$  denote the observed detection (1) or nondetection (0) of the bird at site  $i$  during year  $t$  and survey  $k = 1, \dots, 5$ .
- We model  $y_{t,k}(\mathbf{s}_i)$  conditional on the true presence/absence of the species  $z_t(\mathbf{s}_i)$

$$y_{t,k}(\mathbf{s}_i) \mid z_t(\mathbf{s}_i) \sim \text{Bernoulli}(p_{i,t,k} \cdot z_t(\mathbf{s}_i))$$
$$\text{logit}(p_{i,t,k}) = \alpha_{0,t} + \alpha_1 \cdot \text{DAY}_{i,t,k} + \alpha_2 \cdot \text{DAY}_{i,t,k}^2$$

- A key assumption for identifiability is that  $z_t(\mathbf{s}_i)$  does not change across the 5 replicate surveys at site  $i$  during year  $t$ .