Hierarchical spatial modelling for applied population and community ecology

Jeffrey W. Doser, Marc Kéry, Gesa von Hirschheydt

24-27 June 2024

Spatially-varying coefficient occupancy models

Jeffrey W. Doser

24-27 June 2024

Spatial modelling

- So far, our discussion of spatial models has focused on accounting for residual spatial autocorrelation.
- As we've seen, this has a lot of benefits, e.g., more accurate predictions, better uncertainty estimates
- All of the models we have talked about have done so via some form of **spatially-varying intercept**

Recall our spatial SSOM

Occupancy (ecological) sub-model

$$
\begin{gathered}
z_j \sim \mathrm{Bernoulli}(\psi_j) \\
\mathrm{logit}(\psi_j) = \beta_1 + \beta_2 \cdot X_{2,j} + \cdots + \beta_r \cdot X_{r,j} + \mathrm{w}_j \\
\mathbf{w} \sim \mathrm{Normal}(\mathbf{0}, \tilde{\boldsymbol{C}}(d, \phi, \sigma^2))
\end{gathered}
$$

Detection (observation) sub-model

$$
\begin{gathered}
y_{j,k} \sim \mathrm{Bernoulli}(p_{j,k} \cdot z_j) \\
\mathrm{logit}(p_{j,k}) = \alpha_1 + \alpha_2 \cdot V_{2,j,k} + \cdots + \alpha_r \cdot V_{r,j,k}
\end{gathered}
$$

$j = 1, \ldots, J$ (site)
$k = 1, \ldots, K_j$ (replicate)

Recall our spatial SSOM

Occupancy (ecological) sub-model

$$\boxed{\begin{aligned} z_j &\sim \text{Bernoulli}(\psi_j) \\ \text{logit}(\psi_j) &= \beta_1 + \beta_2 \cdot X_{2,j} + \cdots + \beta_r \cdot X_{r,j} + \mathrm{w}_j \\ \mathbf{w} &\sim \text{Normal}(\mathbf{0}, \tilde{\boldsymbol{C}}(d, \phi, \sigma^2)) \end{aligned}}$$

Detection (observation) sub-model

$$\begin{aligned} y_{j,k} &\sim \text{Bernoulli}(p_{j,k} \cdot z_j) \\ \text{logit}(p_{j,k}) &= \alpha_1 + \alpha_2 \cdot V_{2,j,k} + \cdots + \alpha_r \cdot V_{r,j,k} \end{aligned}$$

$j = 1, \ldots, J$ (site)
$k = 1, \ldots, K_j$ (replicate)

Reorganization and notation

$$\mathrm{logit}(\psi_j) = \beta_1 + \beta_2 \cdot X_{2,j} + \cdots + \beta_r \cdot X_{r,j} + \mathrm{w}_j$$

Reorganization and notation

$$\mathrm{logit}(\psi_j) = \beta_1 + \beta_2 \cdot X_{2,j} + \cdots + \beta_r \cdot X_{r,j} + \mathrm{w}_j$$

$$\mathrm{logit}(\psi_j) = (\beta_1 + \mathrm{w}_j) + \beta_2 \cdot X_{2,j} + \cdots + \beta_r \cdot X_{r,j}$$

Reorganization and notation

$$\mathrm{logit}(\psi_j) = \beta_1 + \beta_2 \cdot X_{2,j} + \cdots + \beta_r \cdot X_{r,j} + \mathrm{w}_j$$

$$\mathrm{logit}(\psi_j) = (\beta_1 + \mathrm{w}_j) + \beta_2 \cdot X_{2,j} + \cdots + \beta_r \cdot X_{r,j}$$

$$\mathrm{logit}(\psi_j) = \tilde{\beta}_{1,j} + \beta_2 \cdot X_{2,j} + \cdots + \beta_r \cdot X_{r,j}$$

$$\tilde{\beta}_{1,j} = \beta_1 + \mathrm{w}_j$$

Reorganization and notation

$$\mathrm{logit}(\psi_j) = \beta_1 + \beta_2 \cdot X_{2,j} + \cdots + \beta_r \cdot X_{r,j} + \mathrm{w}_j$$

$$\mathrm{logit}(\psi_j) = (\beta_1 + \mathrm{w}_j) + \beta_2 \cdot X_{2,j} + \cdots + \beta_r \cdot X_{r,j}$$

$$\mathrm{logit}(\psi_j) = \tilde{\beta}_{1,j} + \beta_2 \cdot X_{2,j} + \cdots + \beta_r \cdot X_{r,j}$$

$$\tilde{\beta}_{1,j} = \beta_1 + \mathrm{w}_j$$

Spatially-varying intercept (SVI)

Non-spatial component

Spatial component

Reorganization and notation

$$\text{logit}(\psi_j) = \beta_1 + \beta_2 \cdot X_{2,j} + \cdots + \beta_r \cdot X_{r,j} + \text{w}_j$$

$$\text{logit}(\psi_j) = (\beta_1 + \text{w}_j) + \beta_2 \cdot X_{2,j} + \cdots + \beta_r \cdot X_{r,j}$$

$$\text{logit}(\psi_j) = \tilde{\beta}_{1,j} + \beta_2 \cdot X_{2,j} + \cdots + \beta_r \cdot X_{r,j}$$

$$\tilde{\beta}_{1,j} = \beta_1 + \text{w}_j$$

Spatially-varying intercept (SVI)

Non-spatial component

Spatial component

There is nothing restricting us from doing this only with the intercept. We can make any coefficient spatial!

Why might species-environment relationships vary spatially?

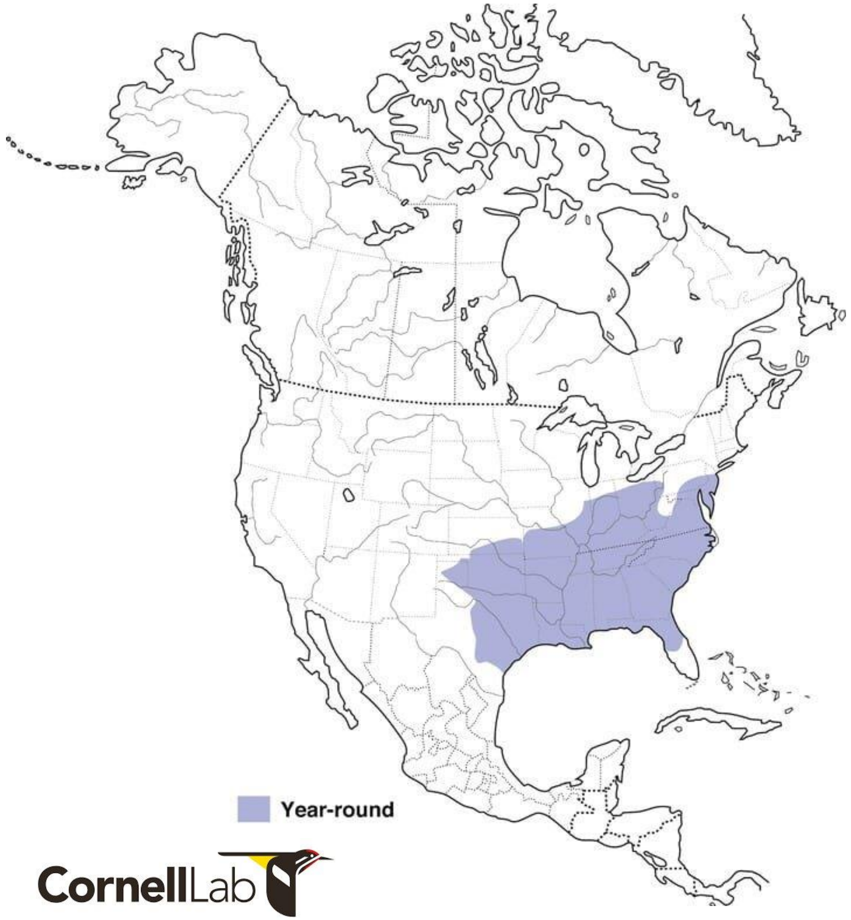
Why might species-environment relationships vary spatially?

- Interactions with abiotic factors such as:
  - Historical disturbance regimes
  - Fine-scale habitat characteristics (e.g., vegetation quality)
  - Local site conditions (e.g., soil content)
  - Spatial variation in resource availability

Why might species-environment relationships vary spatially?

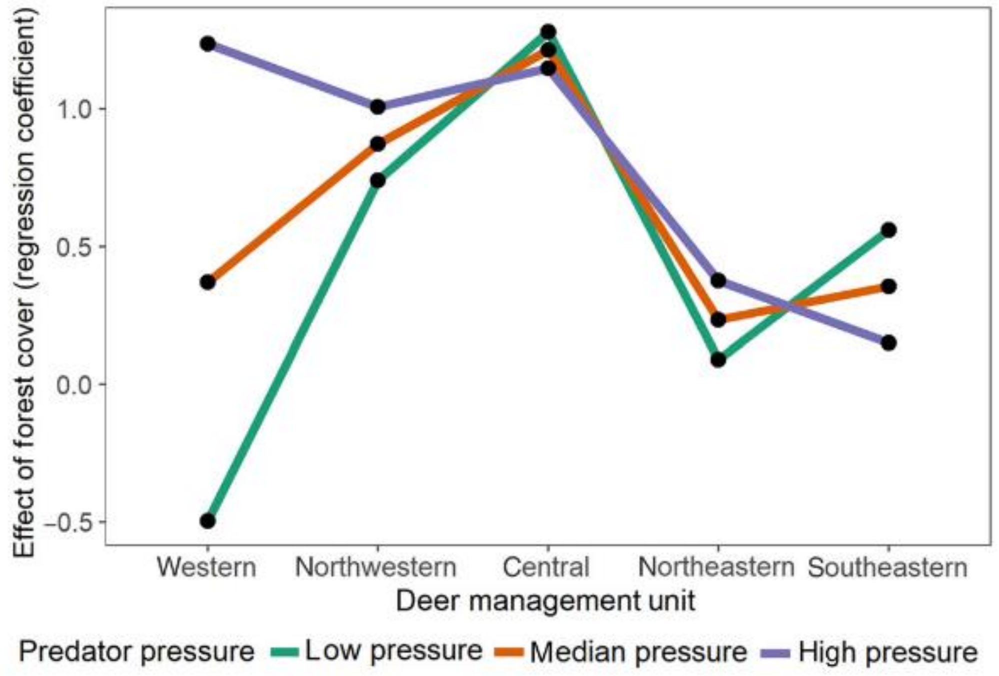
- Impacts of climate may differ between range core and range boundaries

Carolina chickadee



Why might species-environment relationships vary spatially?

- Biotic processes
  - Local genetic adaptations
  - Spatial variation in species interactions



Pease, Pacifici, Kays (2022) Ecosphere

Spatially-varying species-environment relationships

- Often called **nonstationary** relationships (but some statisticians might yell at you).

Spatially-varying species-environment relationships

- Often called **nonstationary** relationships (but some statisticians might yell at you).
- By estimating spatial variability in a species-environment relationship, we can:
    1. Test and generate hypotheses regarding what is driving the relationship.

Spatially-varying species-environment relationships

- Often called **nonstationary** relationships (but some statisticians might yell at you).
- By estimating spatial variability in a species-environment relationship, we can:
  1. Test and generate hypotheses regarding what is driving the relationship.
  2. Better understand the relative importance of different drivers across a species range.

Spatially-varying species-environment relationships

- Often called **nonstationary** relationships (but some statisticians might yell at you).
- By estimating spatial variability in a species-environment relationship, we can:
  1. Test and generate hypotheses regarding what is driving the relationship.
  2. Better understand the relative importance of different drivers across a species range.
  3. Inform conservation and management at both local and broad scales.

Estimating spatial variability in species-environment relationships

- Consider a general model for a spatial occupancy model with one covariate

$$\text{logit}(\psi_j) = (\beta_1 + \mathrm{w}_{1,j}) + f(X_{2,j}, \boldsymbol{\beta})$$

Estimating spatial variability in species-environment relationships

- Consider a general model for a spatial occupancy model with one covariate

$$\text{logit}(\psi_j) = (\beta_1 + \text{w}_{1,j}) + f(X_{2,j}, \boldsymbol{\beta})$$

A generic function that relates the covariate to occurrence probability through a set of parameters

Estimating spatial variability in species-environment relationships

- Consider a general model for a spatial occupancy model with one covariate

$$\text{logit}(\psi_j) = (\beta_1 + \text{w}_{1,j}) + f(X_{2,j}, \boldsymbol{\beta})$$

A generic function that relates the covariate to occurrence probability through a set of parameters

Let's consider 5 forms of this functional relationship

1. Linear

$$f(X_{2,j}, \boldsymbol{\beta}) = \beta_2 \cdot X_{2,j}$$

- The simplest form.
- Does not account for spatial variation in species-environment relationships.

2. Quadratic

$$f(X_{2,j}, \boldsymbol{\beta}) = \beta_2 \cdot X_{2,j} + \beta_3 \cdot X_{2,j}^2$$

- Allows occupancy probability to peak at some optimum level or peak at the extremes
- Useful if the species-environment relationship is non-linear

3. Stratum

$$f(X_{2,j}, \boldsymbol{\beta}) = \beta_2 \cdot X_{2,j} + \beta_{3,\mathrm{STRATUM}_j} \cdot X_{2,j}$$

3. Stratum

$$f(X_{2,j}, \boldsymbol{\beta}) = \beta_2 \cdot X_{2,j} + \beta_{3,\mathrm{STRATUM}_j} \cdot X_{2,j}$$

- Overall linear effect of the environmental predictor as well as stratum-specific adjustments in the effect across a set of strata (e.g., management units)

3. Stratum

$$f(X_{2,j}, \boldsymbol{\beta}) = \beta_2 \cdot X_{2,j} + \beta_{3,\mathrm{STRATUM}_j} \cdot X_{2,j}$$

- Overall linear effect of the environmental predictor as well as stratum-specific adjustments in the effect across a set of strata (e.g., management units)
- Stratum-specific effects could be fixed or random

3. Stratum

$$f(X_{2,j}, \boldsymbol{\beta}) = \beta_2 \cdot X_{2,j} + \beta_{3,\mathrm{STRATUM}_j} \cdot X_{2,j}$$

- Overall linear effect of the environmental predictor as well as stratum-specific adjustments in the effect across a set of strata (e.g., management units)
- Stratum-specific effects could be fixed or random
- Accounts for nonstationarity, but...
  - Spatial variation is limited to pre-defined strata.
  - Lots of uncertainty if some strata have small sample sizes.

4. Interaction

$$f(X_{2,j}, \boldsymbol{\beta}) = \beta_2 \cdot X_{2,j} + \beta_3 \cdot X_{3,j} \cdot X_{2,j}$$

4. Interaction

$$f(X_{2,j}, \boldsymbol{\beta}) = \beta_2 \cdot X_{2,j} + \beta_3 \cdot X_{3,j} \cdot X_{2,j}$$

- Allows for spatial variation in species-environment relationship through an interaction with another variable.
- Very useful for testing explicit hypotheses.
- Spatial variation is limited to the interaction with the additional variable.
- May not always have the interacting variable available (or may not know what the variable is).

5. Spatially-varying coefficient (SVC)

$$f(X_{2,j}, \boldsymbol{\beta}) = \beta_2 \cdot X_{2,j} + \mathrm{w}_{2,j} \cdot X_{2,j}$$

5. Spatially-varying coefficient (SVC)

$$f(X_{2,j}, \boldsymbol{\beta}) = \beta_2 \cdot X_{2,j} + \mathrm{w}_{2,j} \cdot X_{2,j}$$

- Our most flexible option. Estimates an overall linear effect of the covariate along with a site-specific adjustment that varies smoothly across space.

5. Spatially-varying coefficient (SVC)

$$f(X_{2,j}, \boldsymbol{\beta}) = \beta_2 \cdot X_{2,j} + \mathrm{w}_{2,j} \cdot X_{2,j}$$

- Our most flexible option. Estimates an overall linear effect of the covariate along with a site-specific adjustment that varies smoothly across space.
- The spatial adjustment is modelled with an NNGP just like everything we've done with spatial occupancy models.

5. Spatially-varying coefficient (SVC)

$$f(X_{2,j}, \boldsymbol{\beta}) = \beta_2 \cdot X_{2,j} + \mathrm{w}_{2,j} \cdot X_{2,j}$$

- Our most flexible option. Estimates an overall linear effect of the covariate along with a site-specific adjustment that varies smoothly across space.
- The spatial adjustment is modelled with an NNGP just like everything we've done with spatial occupancy models.
- Can be viewed as an extension of the strata model where the strata are now individual sites.

5. Spatially-varying coefficient (SVC)

$$f(X_{2,j}, \boldsymbol{\beta}) = \beta_2 \cdot X_{2,j} + \mathrm{w}_{2,j} \cdot X_{2,j}$$

- Our most flexible option. Estimates an overall linear effect of the covariate along with a site-specific adjustment that varies smoothly across space.
- The spatial adjustment is modelled with an NNGP just like everything we've done with spatial occupancy models.
- Can be viewed as an extension of the strata model where the strata are now individual sites.
- Can be viewed as an extension of the interaction model where we now estimate the interacting variable.

Simulation experiment

- Simulated data under six different species environment relationships:
    1. Linear
    2. Quadratic
    3. Stratum
    4. Interaction
    5. Interaction with an unknown "missing" covariate
    6. The sum of all the above
- Compared the performance of the 5 models

Linear

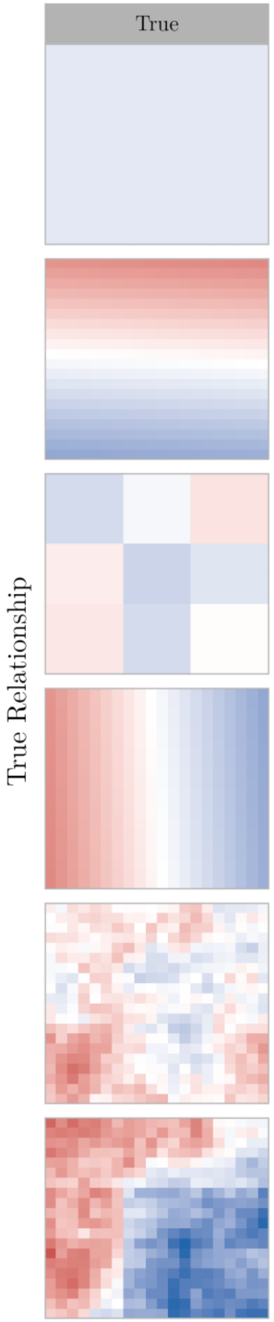
Quadratic

Stratum

Interaction

Missing
interaction

Full



Linear

Quadratic

Stratum

Interaction

Missing interaction

Full

SVC occupancy models

- A simple extension of spatial occupancy models, but we now allow the regression coefficients to vary spatially as well!
- Example SVC ocupancy model with an SVI and 1 SVC

$$\text{logit}(\psi_j) = \tilde{\beta}_{1,j} + \tilde{\beta}_{2,j} \cdot X_{2,j}$$
$$\tilde{\beta}_{1,j} = \beta_1 + \text{w}_{1,j}$$
$$\tilde{\beta}_{2,j} = \beta_2 + \text{w}_{2,j}$$

SVC occupancy models

- A simple extension of spatial occupancy models, but we now allow the regression coefficients to vary spatially as well!
- Example SVC ocupancy model with an SVI and 1 SVC

$$
\begin{aligned}
\text{logit}(\psi_j) &= \tilde{\beta}_{1,j} + \tilde{\beta}_{2,j} \cdot X_{2,j} \\
\tilde{\beta}_{1,j} &= \beta_1 + \mathrm{w}_{1,j} \\
\tilde{\beta}_{2,j} &= \beta_2 + \mathrm{w}_{2,j}
\end{aligned}
$$

Each w is modelled with an NNGP with its own set of spatial parameters

Fitting SVC Occupancy models

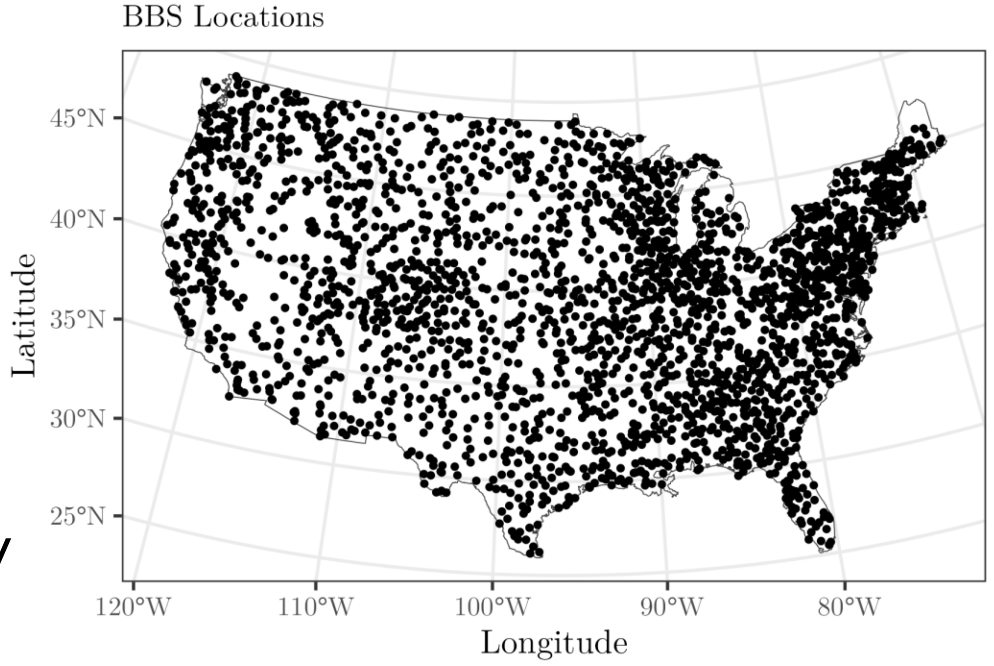
- `svcPGOcc()` function (spatially-varying coefficient Pólya-Gamma occupancy model)
- Each SVC (and SVI) have a different set of spatial parameters (spatial decay and spatial variance) that control their form. Can specify different priors on the different SVCs
- Often useful to set a more informative prior for the spatial decay parameter on SVCs to prevent extremely small effective spatial ranges
- Using `predict()` we can generate maps of the SVCs!!

Extensions to multiple seasons and multiple species

- `svcMsPGOcc()` fits single-season multi-species SVC occupancy models using a spatial factor approach.
- `svcTPGOcc()` fits multi-season single-species SVC occupancy models
- `svcTMsPGOcc()` fits multi-season, multi-species SVC occupancy models.

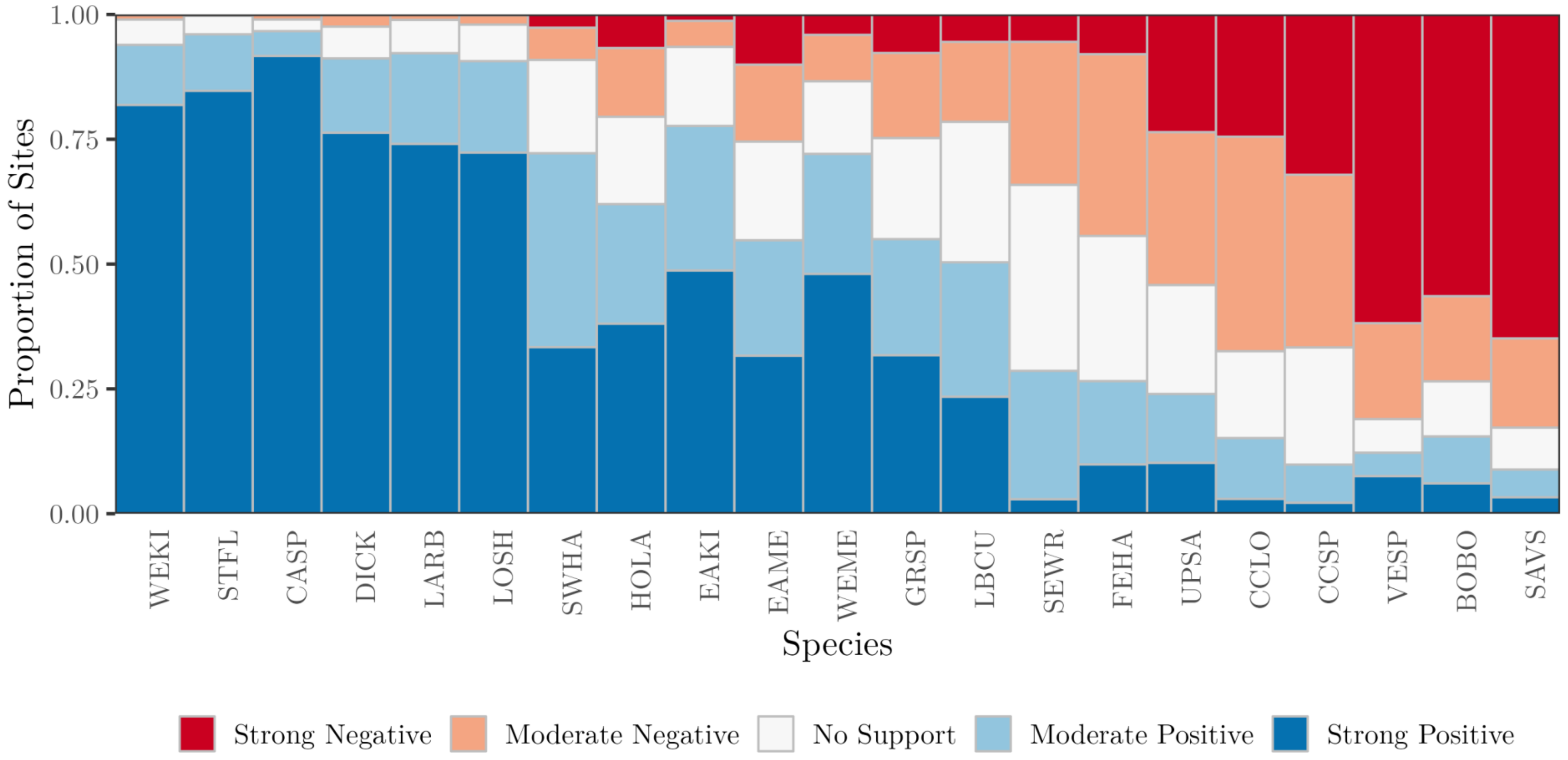
Example: Grassland birds in the US

- North American Breeding Bird Survey Data in 2019
- 21 grassland bird species across 2,486 BBS routes
- Does the effect of maximum temperature vary spatially?
- Fit a multi-species SVC occupancy model



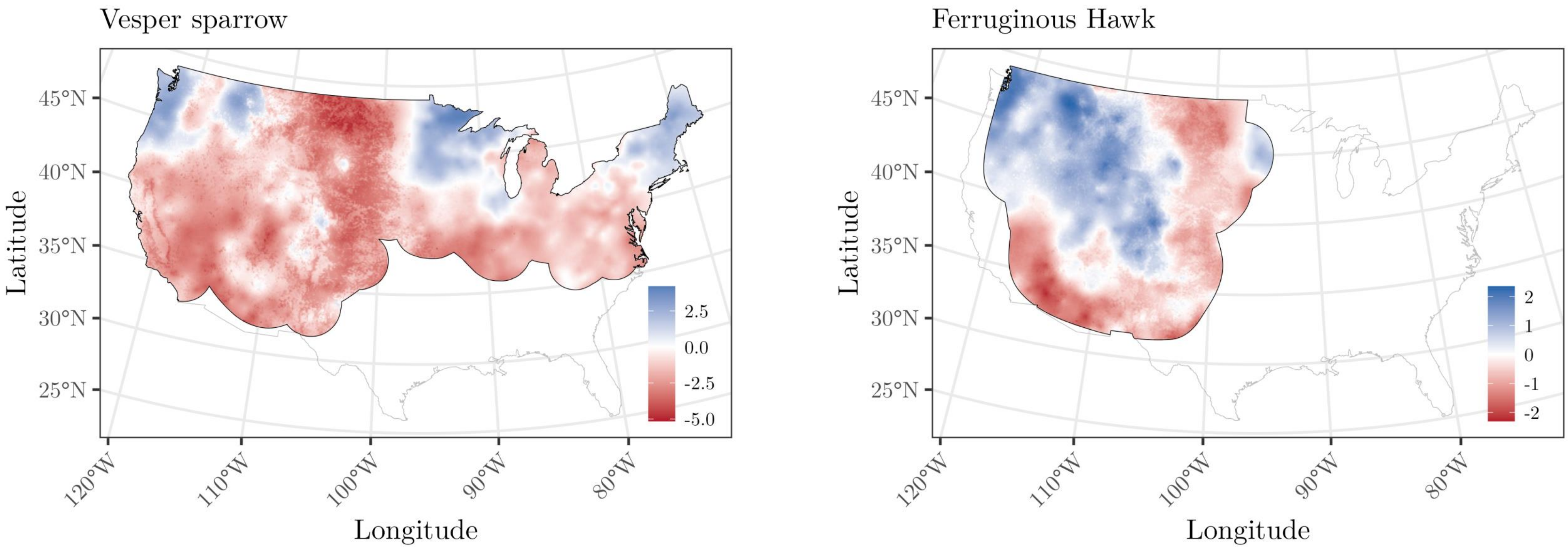
Doser et al. (2024) *Journal of Agricultural, Biological, and Environmental Statistics*

Large spatial variation in maximum temperature effect



Doser et al. (2024) *Journal of Agricultural, Biological, and Environmental Statistics*

Effects of maximum temperature



Doser et al. (2024) *Journal of Agricultural, Biological, and Environmental Statistics*

Estimating spatially-varying trends

- One important application of SVC models is estimating spatially-varying trends
- Can provide fine-scale insight on population/distribution changes

$$z_{j,t} \sim \text{Bernoulli}(\psi_{j,t})$$
$$\text{logit}(\psi_{j,t}) = (\beta_1 + \text{w}_{1,j}) + (\beta_2 + \text{w}_{2,j}) \cdot \text{YEAR}_t$$

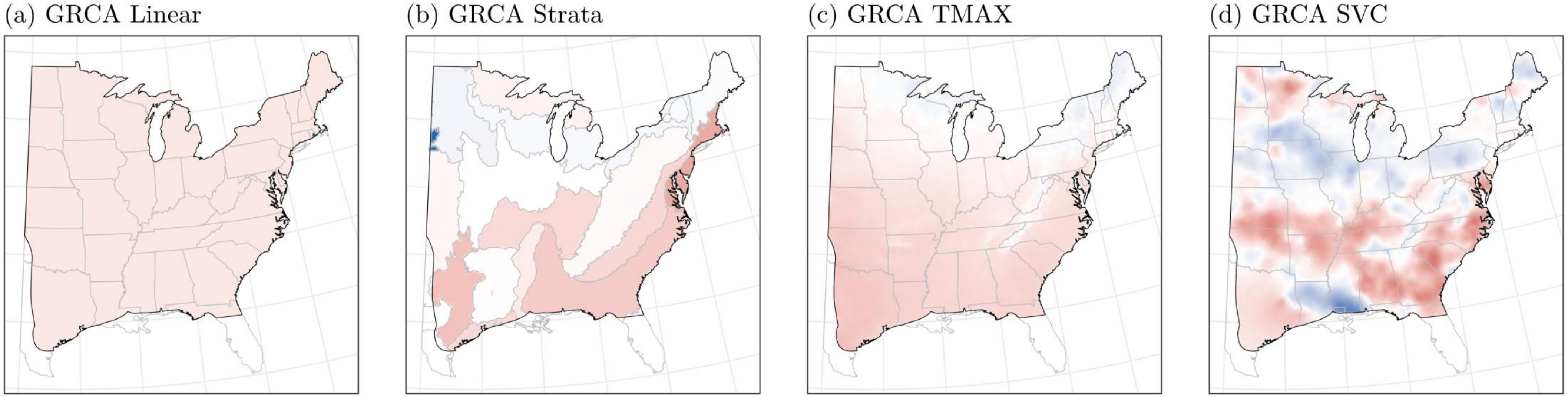
Estimating spatially-varying trends

- One important application of SVC models is estimating spatially-varying trends
- Can provide fine-scale insight on population/distribution changes

$$z_{j,t} \sim \text{Bernoulli}(\psi_{j,t})$$
$$\text{logit}(\psi_{j,t}) = (\beta_1 + \text{w}_{1,j}) + (\beta_2 + \text{w}_{2,j}) \cdot \text{YEAR}_t$$

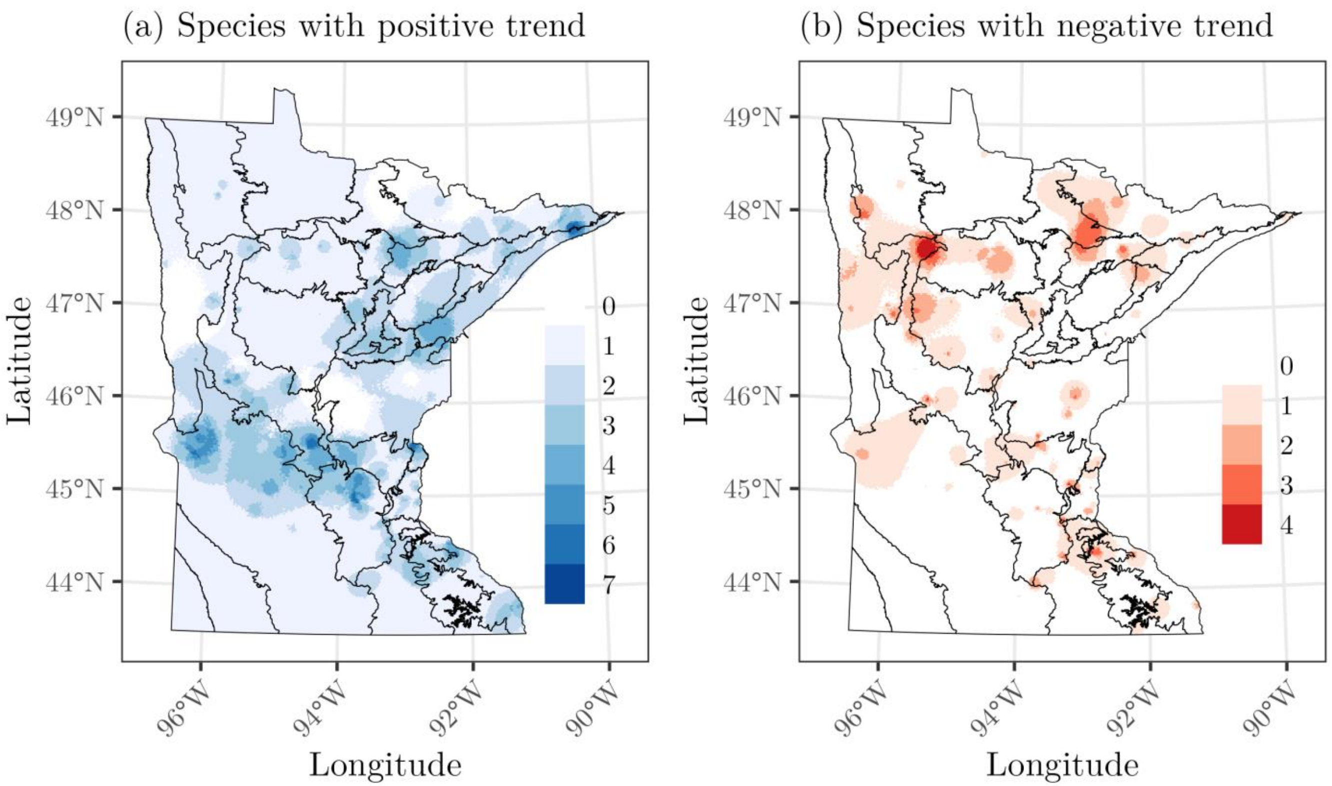
Could also model spatially-varying nonlinear trends by using SVCs with temporal spline covariates

Example: Spatial heterogeneity in forest bird trends in eastern USA



Doser et al. (2024) *Global Ecology and Biogeography*

Example: Identify hotspots of decline for targeted monitoring/management



Doser, Edwards, Zipkin, Jennelle (*In prep*)

Recommendations for fitting and interpreting SVC occupancy models

Recommendations for fitting and interpreting SVC occupancy models

- "With great power comes great responsibility" - Stan Lee

Recommendations for fitting and interpreting SVC occupancy models

- "With great power comes great responsibility" - Stan Lee
- Do not arbitrarily assign SVCs to covariates. Use SVCs for covariates when you hypothesize the relationship may vary spatially.

Recommendations for fitting and interpreting SVC occupancy models

- "With great power comes great responsibility" - Stan Lee
- Do not arbitrarily assign SVCs to covariates. Use SVCs for covariates when you hypothesize the relationship may vary spatially.
- Compare SVC models to simpler models that represent explicit hypotheses

Recommendations for fitting and interpreting SVC occupancy models

- "With great power comes great responsibility" - Stan Lee
- Do not arbitrarily assign SVCs to covariates. Use SVCs for covariates when you hypothesize the relationship may vary spatially.
- Compare SVC models to simpler models that represent explicit hypotheses
- Always include a spatially-varying intercept in SVC models

Recommendations for fitting and interpreting SVC occupancy models

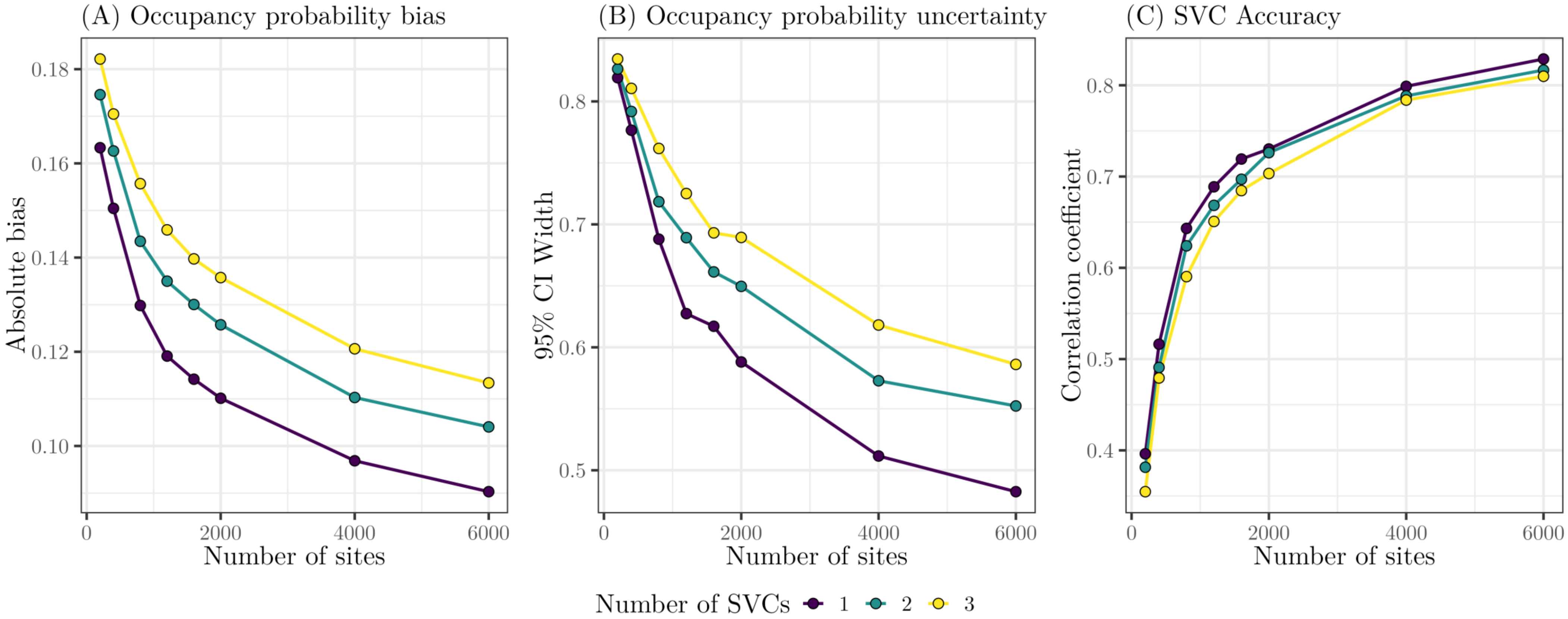
- "With great power comes great responsibility" - Stan Lee
- Do not arbitrarily assign SVCs to covariates. Use SVCs for covariates when you hypothesize the relationship may vary spatially.
- Compare SVC models to simpler models that represent explicit hypotheses
- Always include a spatially-varying intercept in SVC models
- Compare SVC models to a spatial occupancy model
- More here

Priors for spatial decay on SVCs

- Default `spOccupancy` prior for spatial decay parameter ϕ is Uniform(3/max, 3/min).
- Recall the effective spatial range is $3 / \phi$
- This can lead to overfitting in SVC models.
- Can often be useful to specify a more informative prior to restrict the effective spatial range.
- Best option is to set the bounds based on ecology of the species or the spatial scale of interest.
- Setting upper bound of ϕ to $3 / q_{0.25}$ can help minimize the potential for misleading inferences (where $q_{0.25}$ is the 25% quantile of the intersite distance matrix)

SVC occupancy models are data hungry

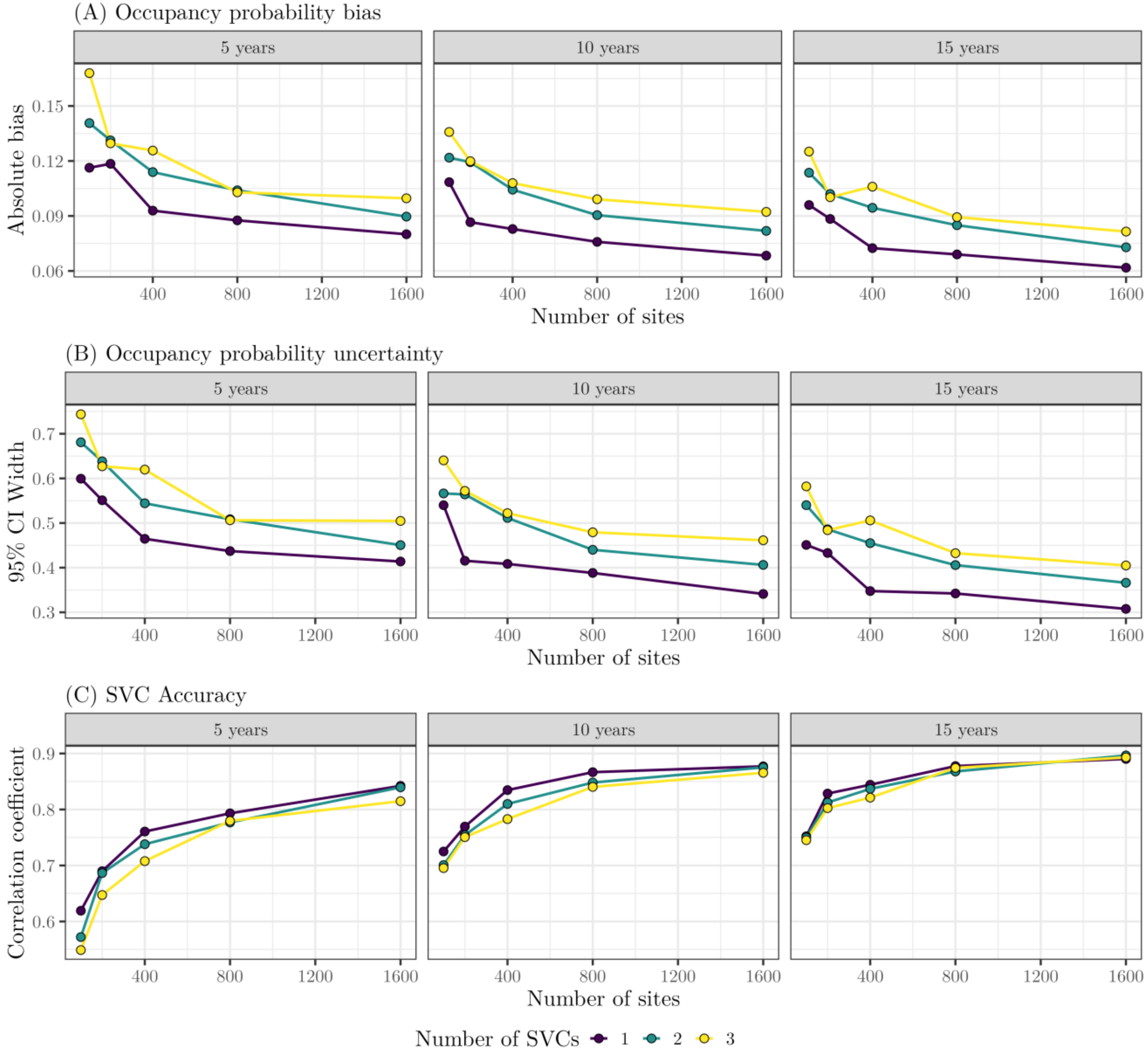
Single-season occupancy models



Doser et al. (2024) *Global Ecology and Biogeography*

SVC occupancy models are data hungry

Multi-season SVC occupancy models are somewhat less data hungry



Doser et al. (2024) *Global Ecology and Biogeography*

Exercise: Spatially-varying trend in wood thrush occupancy in the eastern US

09-wood-thrush-spatial-trend-occ.R