Supplemental Appendix: Hierarchical Models for Spatial Data with Errors that are Correlated with the Latent Process

Jonathan R. Bradley[1], Christopher K. Wikle[2], and Scott H. Holan[2,3]

[1]*Florida State University,* [2]*University of Missouri,* [3]*U.S. Census Bureau*

Appendix A: Technical Results

We provide statements and proofs of five results claimed in the main text.

1. *Statement 1 (Positive Definite, Special Case 1): Let* $\boldsymbol{\Sigma}_1 \equiv \boldsymbol{\Sigma}_Y - \boldsymbol{\Sigma}_w$, $\sigma^2 > 0$, *and* $\boldsymbol{\Sigma}_Y$ *and* $\boldsymbol{\Sigma}_w$ *are symmetric positive-definite real valued matrices. Then the matrix,*

$$\begin{pmatrix} \boldsymbol{\Sigma}_Y & -\boldsymbol{\Sigma}_1 \\ -\boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_1 + \sigma^2 \mathbf{I}_n \end{pmatrix}, \tag{A.1}$$

is positive definite provided that $\boldsymbol{\Sigma}_w$ *and* $\boldsymbol{\Sigma}_1$ *are positive definite.*

Proof: The Schur complement of (A.1) is given by

$$\begin{pmatrix} \boldsymbol{\Sigma}_Y & \mathbf{0}_{n,n} \\ \mathbf{0}_{n,n} & \sigma^2 \mathbf{I}_n + \boldsymbol{\Sigma}_w - \boldsymbol{\Sigma}_w' \boldsymbol{\Sigma}_Y^{-1} \boldsymbol{\Sigma}_w \end{pmatrix}, \tag{A.2}$$

where $\mathbf{0}_{n,n}$ is a $n \times n$ matrix of zeros. If $\boldsymbol{\Sigma}_w - \boldsymbol{\Sigma}_w' \boldsymbol{\Sigma}_Y^{-1} \boldsymbol{\Sigma}_w$ is positive definite then we have that the Schur complement of (A.1) (and hence (A.1)) is positive definite.

Consider the matrix,

$$\begin{pmatrix} \boldsymbol{\Sigma}_w & -\boldsymbol{\Sigma}_w \\ -\boldsymbol{\Sigma}_w & \boldsymbol{\Sigma}_Y \end{pmatrix} = \begin{pmatrix} \mathbf{I}_n & \mathbf{0}_{n,n} \\ -\boldsymbol{I}_n & \mathbf{I}_n \end{pmatrix} \begin{pmatrix} \boldsymbol{\Sigma}_w & \mathbf{0}_{n,n} \\ \mathbf{0}_{n,n} & \boldsymbol{\Sigma}_Y - \boldsymbol{\Sigma}_w \end{pmatrix} \begin{pmatrix} \mathbf{I}_n & -\mathbf{I}_{n,n} \\ \mathbf{0}_{n,n} & \mathbf{I}_n \end{pmatrix}, \quad \text{(A.3)}$$

which is positive definite by the conditions of Statement 1. However, one can also rewrite (A.3) as,

$$\begin{pmatrix} \boldsymbol{\Sigma}_w & -\boldsymbol{\Sigma}_w \\ -\boldsymbol{\Sigma}_w & \boldsymbol{\Sigma}_Y \end{pmatrix} = \begin{pmatrix} \mathbf{I}_n & -\boldsymbol{\Sigma}_w \boldsymbol{\Sigma}_Y^{-1} \\ \mathbf{0}_{n,n} & \mathbf{I}_n \end{pmatrix} \begin{pmatrix} \boldsymbol{\Sigma}_w - \boldsymbol{\Sigma}_w' \boldsymbol{\Sigma}_Y^{-1} \boldsymbol{\Sigma}_w & \mathbf{0}_{n,n} \\ \mathbf{0}_{n,n} & \boldsymbol{\Sigma}_Y \end{pmatrix} \begin{pmatrix} \mathbf{I}_n & \mathbf{0}_{n,n} \\ -\boldsymbol{\Sigma}_Y^{-1} \boldsymbol{\Sigma}_w & \mathbf{I}_n \end{pmatrix}, \quad \text{(A.4)}$$

which is positive definite if and only if $\boldsymbol{\Sigma}_w - \boldsymbol{\Sigma}_w^\top \boldsymbol{\Sigma}_Y^{-1} \boldsymbol{\Sigma}_w$ is positive definite. Thus, it follows from (A.3) and (A.4) that $\boldsymbol{\Sigma}_w - \boldsymbol{\Sigma}_w^\top \boldsymbol{\Sigma}_Y^{-1} \boldsymbol{\Sigma}_w$ is positive definite.

2. *Statement 2 (Positive Definite, Special Case 3): Let* $\boldsymbol{\Sigma} \equiv \boldsymbol{\Sigma}_{Y,w} - \boldsymbol{\Sigma}_Y$, $\sigma^2 > 0$, *and* $\boldsymbol{\Sigma}_Y$ *and* $\boldsymbol{\Sigma}_{Y,w}$ *be real symmetric matrices. Then the matrix,*

$$\begin{pmatrix} \boldsymbol{\Sigma}_Y & \boldsymbol{\Sigma} \\ \boldsymbol{\Sigma}^\top & \sigma^2 \mathbf{I}_n - 2\boldsymbol{\Sigma} \end{pmatrix}, \quad \text{(A.5)}$$

provided $\boldsymbol{\Sigma}_Y$ *is positive semi-definite and* $\boldsymbol{\Sigma}_Y - \boldsymbol{\Sigma}_{Y,w}^\top \boldsymbol{\Sigma}_Y^{-1} \boldsymbol{\Sigma}_{Y,w}$ *is positive semi-definite.*

Proof: The Schur complement is given by

$$\begin{pmatrix} \boldsymbol{\Sigma}_Y & \mathbf{0}_{n,n} \\ \mathbf{0}_{n,n} & \sigma^2 \mathbf{I}_n + \boldsymbol{\Sigma}_Y - \boldsymbol{\Sigma}_{Y,w}^\top \boldsymbol{\Sigma}_Y^{-1} \boldsymbol{\Sigma}_{Y,w} \end{pmatrix}, \quad \text{(A.6)}$$

where $\mathbf{0}_{n,n}$ is a $n \times n$ matrix of zeros. Since $\boldsymbol{\Sigma}_Y - \boldsymbol{\Sigma}_{Y,w}^\top \boldsymbol{\Sigma}_Y^{-1} \boldsymbol{\Sigma}_{Y,w}$ and $\boldsymbol{\Sigma}_Y$ are positive

definite, the Schur complement, and the matrix in (A.5) is positive definite.

3. *Statement 3 (Positive Definite, Special Case 4): Let* $\mathbf{\Sigma}_Y$ *be a real symmetric positive definite matrix,* $\mathbf{\Psi} \in \mathbb{R}^n \times \mathbb{R}^r$, $\boldsymbol{P} = \mathbf{\Psi}(\mathbf{\Psi}^\top \mathbf{\Psi})^{-1}\mathbf{\Psi}^\top$, *and* $\sigma^2 > 0$. *Then the matrix,*

$$\text{cov}\left\{\begin{pmatrix} \mathbf{y} \\ \boldsymbol{\delta} \end{pmatrix}\right\} = \begin{pmatrix} \mathbf{\Sigma}_Y & -\mathbf{\Sigma}_Y(\mathbf{I}_n - \mathbf{P}) \\ -(\mathbf{I}_n - \mathbf{P})\mathbf{\Sigma}_Y & \mathbf{\Sigma}_Y + \mathbf{P}\mathbf{\Sigma}_Y\mathbf{P} + 2\mathbf{\Sigma}_Y(\mathbf{I}_n - \mathbf{P}) + \sigma^2\mathbf{I}_n \end{pmatrix}. \quad \text{(A.7)}$$

provided $\mathbf{\Sigma}_Y$ *is positive semi-definite.*

Proof: We can write

$$\begin{aligned} &\begin{pmatrix} \mathbf{\Sigma}_Y & -\mathbf{\Sigma}_Y(\mathbf{I}_n - \mathbf{P}) \\ -(\mathbf{I}_n - \mathbf{P})\mathbf{\Sigma}_Y & \mathbf{\Sigma}_Y + \mathbf{P}\mathbf{\Sigma}_Y\mathbf{P} + 2\mathbf{\Sigma}_Y(\mathbf{I}_n - \mathbf{P}) + \sigma^2\mathbf{I}_n \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{\Sigma}_Y & -\mathbf{\Sigma}_Y \\ -\mathbf{\Sigma}_Y & \mathbf{\Sigma}_Y + 2\mathbf{\Sigma}_Y(\mathbf{I}_n - \mathbf{P}) \end{pmatrix} + \begin{pmatrix} \mathbf{\Sigma}_Y & \mathbf{\Sigma}_Y\mathbf{P} \\ \mathbf{P}\mathbf{\Sigma}_Y & \mathbf{\Sigma}_Y + \sigma^2\mathbf{I}_n \end{pmatrix}. \end{aligned} \quad \text{(A.8)}$$

The sum of two positive semi-definite matrices is positive definite, and hence, we aim to show that both matrices on the right-hand-side of (A.8) are positive semi-definite. The Schur complement of the first matrix on the right-hand-side of (A.8) is given by

$$\begin{pmatrix} \mathbf{\Sigma}_Y & \mathbf{0}_{n,n} \\ \mathbf{0}_{n,n} & \mathbf{\Sigma}_Y + 2\mathbf{\Sigma}_Y(\mathbf{I}_n - \mathbf{P}) \end{pmatrix}, \quad \text{(A.9)}$$

where $\mathbf{0}_{n,n}$ is a $n \times n$ matrix of zeros. Since $\mathbf{\Sigma}_Y$ is positive definite, we only need to show that $2\mathbf{\Sigma}_Y(\mathbf{I}_n - \mathbf{P})$ is positive semi-definite. Recall the product of a stable and Hermitian square matrix $\mathbf{A}$ and a positive semi-definite matrix $\mathbf{B}$ satisfies the following

inequality (Zhang and Zhang, 2006),

$$\lambda_n(\mathbf{A})\lambda_1(\mathbf{B}) \leq \lambda_n(\mathbf{AB}) \leq \lambda_1(\mathbf{A})\lambda_1(\mathbf{B}),$$

where $\lambda_n(\mathbf{A})$ is the smallest eigenvalue of $\mathbf{A}$, $\lambda_1(\mathbf{B})$ is the largest eigenvalue of $\mathbf{B}$, $\lambda_n(\mathbf{AB})$ is the smallest eigenvalue of $\mathbf{AB}$, and $\lambda_1(\mathbf{A})$ is the largest eigenvalue of $\mathbf{A}$. Letting $\mathbf{A} = \mathbf{I}_n - \mathbf{P}$ and $\mathbf{B} = 2\boldsymbol{\Sigma}_Y$, we see that the smallest eigenvalue of $2\boldsymbol{\Sigma}_Y(\mathbf{I}_n - \mathbf{P})$ is in the set $[0, \lambda_1(2\boldsymbol{\Sigma}_Y)]$. Hence, the first matrix on the right-hand-side of (A.8) is positive semi-definite.

The Schur complement of the second matrix on the right-hand-side of (A.8) is given by

$$\begin{pmatrix} \boldsymbol{\Sigma}_Y & \mathbf{0}_{n,n} \\ \mathbf{0}_{n,n} & \sigma^2\mathbf{I}_n \end{pmatrix}, \tag{A.10}$$

which is positive definite.

4. *Statement 4: Assume the Regularity Conditions stated in Section D. Define* $\boldsymbol{\mu} \equiv (\mu(\mathbf{s}_1), \ldots, \mu(\mathbf{s}_n))^\top$, σ^2 *and* $\mathbf{s}_1, \ldots, \mathbf{s}_m \in D$. *Then, we have*

$$\widehat{f}(\mathbf{s}; \boldsymbol{\theta}, \mathbf{z}) = \mu(\mathbf{s}) + \operatorname{cov}\{Y(\mathbf{s}), \mathbf{z}|\boldsymbol{\theta}\}\left\{\operatorname{cov}(\boldsymbol{w}|\boldsymbol{\theta}) + \sigma^2\boldsymbol{I}_n\right\}^{-1}(\mathbf{z} - \boldsymbol{\mu}); \, \mathbf{s} \in D, \tag{A.11}$$

minimizes

$$\min_f \left(E\left[\{Y(\mathbf{s}) - f(\mathbf{s})\}^2 |\boldsymbol{\theta}\right]\right),$$

where f *falls in the space of linear unbiased estimators. Additionally, the variance of*

$\widehat{f}$ is given by

$$\begin{aligned}\sigma_f^2(\mathbf{s};\boldsymbol{\theta},\mathbf{z}) &\equiv E\left[\left\{Y(\mathbf{s})-\widehat{f}(\mathbf{s})\right\}^2|\boldsymbol{\theta}\right] \\ &= \operatorname{var}\left\{Y(\mathbf{s})|\boldsymbol{\theta}\right\} - \operatorname{cov}\left\{Y(\mathbf{s}),\mathbf{z}|\boldsymbol{\theta}\right\}\left\{\operatorname{cov}\left(\boldsymbol{w}|\boldsymbol{\theta}\right)+\sigma^2\boldsymbol{I}_n\right\}^{-1}\operatorname{cov}\left\{\mathbf{z},Y(\mathbf{s})|\boldsymbol{\theta}\right\},\end{aligned}$$

where $\mathbf{s}\in D$.

Proof: Write our linear estimator in $\mathbf{z}$ as,

$$q(\mathbf{s})+\boldsymbol{\lambda}(\mathbf{s})^\top\mathbf{z}, \tag{A.12}$$

where $q: D\to\mathbb{R}$ and $\boldsymbol{\lambda}: D\to\mathbb{R}^n$. For $\mathbf{s}\in D$ let $a(\mathbf{s})=q(\mathbf{s})+\boldsymbol{\lambda}(\mathbf{s})^\top\boldsymbol{\mu}$. Then the mean squared error for a given $\mathbf{s}\in D$ can be written as

$$\begin{aligned}&E\left[\left\{q(\mathbf{s})+\boldsymbol{\lambda}(\mathbf{s})^\top\mathbf{z}-Y(\mathbf{s})\right\}^2|\boldsymbol{\theta}\right] \\ &= E\left[\left\{a(\mathbf{s})+\boldsymbol{\lambda}(\mathbf{s})^\top(\mathbf{z}-\boldsymbol{\mu})-Y(\mathbf{s})\right\}^2|\boldsymbol{\theta}\right] \\ &= E\left[\left\{\boldsymbol{\lambda}(\mathbf{s})^\top(\mathbf{z}-\boldsymbol{\mu})-Y(\mathbf{s})\right\}^2|\boldsymbol{\theta}\right]+a(\mathbf{s})^2-2a(\mathbf{s})\mu(\mathbf{s}) \\ &= \boldsymbol{\lambda}(\mathbf{s})^\top\left\{\operatorname{cov}\left(\mathbf{w}|\boldsymbol{\mu},\boldsymbol{\theta}\right)+\sigma^2\mathbf{I}_n\right\}\boldsymbol{\lambda}(\mathbf{s})+E\left\{Y(\mathbf{s})^2|\boldsymbol{\theta}\right\} \\ &\qquad -2\boldsymbol{\lambda}(\mathbf{s})^\top\operatorname{cov}\left\{\mathbf{z},Y(\mathbf{s})|\boldsymbol{\theta}\right\}+a(\mathbf{s})^2-2a(\mathbf{s})\mu(\mathbf{s}).\end{aligned} \tag{A.13}$$

For a given $\mathbf{s}\in D$, take the derivative of (A.13) with respect to $\boldsymbol{\lambda}(\mathbf{s})$ and set it equal to zero to obtain,

$$\widehat{\boldsymbol{\lambda}}(\mathbf{s})=\left\{\operatorname{cov}\left(\mathbf{w}|\boldsymbol{\mu},\boldsymbol{\theta}\right)+\sigma^2\mathbf{I}_n\right\}^{-1}\operatorname{cov}\left\{\mathbf{z},Y(\mathbf{s})|\boldsymbol{\theta}\right\}.$$

Also, for a given $\mathbf{s}\in D$, take the derivative of (A.13) with respect to $a(\mathbf{s})$ and set it

equal to zero to obtain,

$$\widehat{a}(\mathbf{s}) = \mu(\mathbf{s}).$$

Substituting $\widehat{\boldsymbol{\lambda}}$ and $q(\mathbf{s}) = \widehat{a}(\mathbf{s}) - \widehat{\boldsymbol{\lambda}}(\mathbf{s})^\top \boldsymbol{\mu}$ into (A.12), we obtain $\widehat{f}$ in (A.11). Upon computing second derivatives of (A.13) it is clear that the Hessian matrix is positive definite indicating that $\widehat{f}$ in (A.11) is a minimum. Notice also that $\widehat{f}$ is unbiased for μ, so that it is the best linear unbiased predictor of Y.

The mean squared error of $\widehat{f}$ follows from straightforward algebra. For a given $\mathbf{s}$,

$$\begin{aligned}
&E\left[\left\{\widehat{f}(\mathbf{s};\,\boldsymbol{\theta},\mathbf{z}) - Y(\mathbf{s})\right\}^2 |\boldsymbol{\theta}\right] \\
&= E\left[\left\{\widehat{f}(\mathbf{s};\,\boldsymbol{\theta},\mathbf{z}) - \mu(\mathbf{s})\right\}^2 |\boldsymbol{\theta}\right] + E\left[\{Y(\mathbf{s}) - \mu(\mathbf{s})\}^2 |\boldsymbol{\theta}\right] \\
&\qquad\qquad - 2E\left[\left\{\widehat{f}(\mathbf{s};\,\boldsymbol{\theta},\mathbf{z}) - \mu(\mathbf{s})\right\}\{Y(\mathbf{s}) - \mu(\mathbf{s})\} |\boldsymbol{\theta}\right] \\
&= \mathrm{cov}\left\{Y(\mathbf{s}),\mathbf{z}|\boldsymbol{\theta}\right\}\left\{\mathrm{cov}\left(\mathbf{w}|\boldsymbol{\mu},\boldsymbol{\theta}\right) + \sigma^2\mathbf{I}_n\right\}^{-1}\mathrm{cov}\left\{\mathbf{z},Y(\mathbf{s})|\boldsymbol{\theta}\right\} + \mathrm{var}\left\{Y(\mathbf{s})|\boldsymbol{\theta}\right\} \\
&\quad - 2\mathrm{cov}\left\{Y(\mathbf{s}),\mathbf{z}|\boldsymbol{\theta}\right\}\left\{\mathrm{cov}\left(\mathbf{w}|\boldsymbol{\mu},\boldsymbol{\theta}\right) + \sigma^2\mathbf{I}_n\right\}^{-1}\mathrm{cov}\left\{\mathbf{z},Y(\mathbf{s})|\boldsymbol{\theta}\right\} \\
&= \mathrm{var}\left\{Y(\mathbf{s})|\boldsymbol{\theta}\right\} - \mathrm{cov}\left\{Y(\mathbf{s}),\mathbf{z}|\boldsymbol{\theta}\right\}\left\{\mathrm{cov}\left(\mathbf{w}|\boldsymbol{\mu},\boldsymbol{\theta}\right) + \sigma^2\mathbf{I}_n\right\}^{-1}\mathrm{cov}\left\{\mathbf{z},Y(\mathbf{s})|\boldsymbol{\theta}\right\},
\end{aligned}$$

which completes the proof.

Appendix B: Model Implementation

In Section 4.3 of the main text, we describe a two step method for implementation. Namely, the first step is to implement a Gibbs sampler to simulate from the posterior distribution of the augmented process $w(\cdot)$, and associated parameters. The second step is to simulate from the posterior predictive distribution of the latent process. We now outline these steps

in full to aid the reader in reproducing these results.

Step I:

The statistical model used to obtain $\boldsymbol{\beta}^{[b]}$, $\boldsymbol{\eta}^{[b]}$, $\boldsymbol{\xi}^{[b]}$, $\sigma_{\beta}^{2[b]}$, $\sigma_{\xi}^{2[b]}$, and $\boldsymbol{\theta}^{[b]}$ for $b = 1, ..., B$ is outlined in Algorithm 1 (below) using the hierarchical modeling notation of Berliner (1996). Let $\mathbf{K}$ be defined according to Equation (8), where

$$\text{cov}(\boldsymbol{\eta}|\boldsymbol{\theta}) \equiv (\boldsymbol{\Psi}^{\top}\boldsymbol{\Psi})^{-1}\boldsymbol{\Psi}^{\top}\boldsymbol{\Sigma}_Y(\boldsymbol{\theta})\boldsymbol{\Psi}(\boldsymbol{\Psi}^{\top}\boldsymbol{\Psi})^{-1}, \tag{B.1}$$

$\boldsymbol{\theta} = (\sigma_Y^2, \tau)'$, $\sigma_Y^2 > 0$, and $\tau > 0$. The basis functions $\boldsymbol{\Psi}$ are chosen using the algorithm outlined in Appendix F. We consider two choices for $\boldsymbol{\Sigma}_Y(\tau)$. In Section 5.1, we let $\boldsymbol{\Sigma}_Y(\tau)$ be formed by a Matérn covariogram; see Equation (E.4). Also, in Section 5.2, we set $\boldsymbol{\Sigma}_Y(\boldsymbol{\theta}) = \sigma_Y^2(\mathbf{I} - \tau\mathbf{A})^{-1}$, where $\mathbf{A}$ is a 72361×72361 first-order adjacency matrix associated with U.S. census tracts and $\boldsymbol{\theta} = (\sigma_Y^2, \tau)'$. The inverse $(\mathbf{I} - \tau\mathbf{A})^{-1}$ is straightforward to compute using sparse matrix inversion techniques.

It is important to emphasize that we never need to store every column of $\boldsymbol{\Sigma}_Y(\boldsymbol{\theta})$ at the same time. Write the columns of $\frac{1}{\sigma_Y^2}\boldsymbol{\Sigma}_Y(\boldsymbol{\theta})$ as $\frac{1}{\sigma_Y^2}\boldsymbol{\Sigma}_Y(\boldsymbol{\theta}) = (\mathbf{k}_1, \ldots, \mathbf{k}_r)$. A schematic of the code is given in Schematic 1. Storage, requires an N-dimensional vector $(\mathbf{k}_i)$, an r-dimensional vector $((\boldsymbol{\Psi}'\boldsymbol{\Psi})^{-1}\boldsymbol{\Psi}'\mathbf{k}_i)$, and an $r \times N$ matrix (LeftK), which requires $O(Nr)$ storage. If we place a continuous prior on τ, we have to repeat this loop in Schematic 1 every time we update τ, which is not computationally feasible. Thus, we place a discrete uniform prior on τ, where $\tau = \tau_1, \ldots, \tau_M$ with equal probability. Then, before implementing the Gibbs sampler, we use Schematic 1 to compute $\text{cov}(\boldsymbol{\eta}|\boldsymbol{\theta} = (1, \tau_1)'), \ldots, \text{cov}(\boldsymbol{\eta}|\boldsymbol{\theta} = (1, \tau_M)')$.

The statistical model used for inference is given in Algorithm 1. We now specify the full-conditional distributions for the process variables (i.e., $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$) and the parameters (i.e., $\boldsymbol{\beta}$, σ_Y^2, σ_{ξ}^2, σ_{β}^2, and τ). Using standard conjugacy results (Berger, 1985), the full conditional

# initialize a variable (called LeftCovEta) to store $\frac{1}{\sigma_Y^2}(\boldsymbol{\Psi}^\top\boldsymbol{\Psi})^{-1}\boldsymbol{\Psi}^\top\boldsymbol{\Sigma}_Y(\boldsymbol{\theta})$
LeftCovEta = empty;
for (i=1:n)

Step 1: compute $\mathbf{k}_i$;

Step 2: Let LeftCovEta be the column concatenation of LeftCovEta with $(\boldsymbol{\Psi}'\boldsymbol{\Psi})^{-1}\boldsymbol{\Psi}'\mathbf{k}_i$;

Step 3: remove $\mathbf{k}_i$ from memory;

end
$\text{cov}(\boldsymbol{\eta}|\boldsymbol{\theta} = (1,\tau)')$ = LeftCovEta times $\boldsymbol{\Psi}(\boldsymbol{\Psi}'\boldsymbol{\Psi})^{-1}$;
remove LeftCovEta from memory;

Schematic 1: A representation of the code used to compute and store (B.1).

Algorithm 1: Summary of statistical model to obtain $\boldsymbol{\beta}^{[b]}$, $\boldsymbol{\eta}^{[b]}$, $\boldsymbol{\xi}^{[b]}$, $\tau^{[b]}$, $\sigma_Y^{2[b]}$, $\sigma_\beta^{2[b]}$, and $\sigma_\xi^{2[b]}$ for $b = 1, ..., B$.

$$\text{Data Model}: Z(\mathbf{s})|\boldsymbol{\beta},\boldsymbol{\eta},\boldsymbol{\xi},\sigma_K^2,\theta \overset{\text{ind}}{\sim} \text{Normal}\left\{\mathbf{x}(\mathbf{s})^\top\boldsymbol{\beta} + \boldsymbol{\psi}(\mathbf{s})^\top\boldsymbol{\eta} + \xi(\mathbf{s}), \sigma_\epsilon^2(\mathbf{s})\right\};$$
$$\text{Process Model 1}: \boldsymbol{\eta}|\boldsymbol{\theta} \sim \text{Gaussian}\left\{\mathbf{0}, \text{cov}(\boldsymbol{\eta}|\boldsymbol{\theta})\right\};$$
$$\text{Process Model 2}: \boldsymbol{\xi}|\sigma_\xi^2 \sim \text{Gaussian}\left(\mathbf{0}, \sigma_\xi^2\mathbf{I}_N\right);$$
$$\text{Parameter Model 1}: \boldsymbol{\beta} \sim \text{Normal}\left(0, \sigma_\beta^2\mathbf{I}_p\right);$$
$$\text{Parameter Model 2}: \sigma_Y^2 \sim \text{IG}\left(1,1\right);$$
$$\text{Parameter Model 3}: \sigma_\xi^2 \sim \text{IG}\left(1,1\right);$$
$$\text{Parameter Model 4}: \sigma_\beta^2 \sim \text{IG}\left(1,1\right);$$
$$\text{Parameter Model 5}: \tau \sim I(\tau \in \{\tau_1, ..., \tau_M\});\ \mathbf{s} \in D_O.$$

distributions are as follows:

- The full conditional distribution for $\boldsymbol{\eta}$ is given by $\boldsymbol{\eta} \sim \text{Gaussian}\left(\boldsymbol{\mu}_{\eta}^{*}, \boldsymbol{\Sigma}_{\eta}^{*}\right)$, where $\boldsymbol{\Sigma}_{\eta}^{*} \equiv \left(\boldsymbol{\Psi}^{\top}\sigma^{2}\mathbf{I}_{n}^{-1}\boldsymbol{\Psi} + \mathbf{K}(\boldsymbol{\theta})\right)^{-1}$, $\boldsymbol{\mu}_{\eta}^{*} \equiv \boldsymbol{\Sigma}_{\eta}^{*} \times \boldsymbol{\Psi}^{\top}\sigma^{2}\mathbf{I}_{n}^{-1} \times (\mathbf{z} - \mathbf{X}\boldsymbol{\beta} - \boldsymbol{\xi})$, $\sigma^{2}\mathbf{I}_{n} = \text{diag}(\text{var}(\epsilon(\mathbf{s}_{i}) : i = 1, \ldots, n)$, $\boldsymbol{\Psi} \equiv (\boldsymbol{\psi}(\mathbf{s}_{1}), \ldots, \boldsymbol{\psi}(\mathbf{s}_{n}))^{\top}$, $\mathbf{X} \equiv (\mathbf{x}(\mathbf{s}_{1}), \ldots, \mathbf{x}(\mathbf{s}_{n}))^{\top}$, and $\mathbf{s}_{1}, \ldots, \mathbf{s}_{n}$ are the observed data locations.

- The full conditional distribution for $\boldsymbol{\xi}$ is given by $\boldsymbol{\xi} \sim \text{Gaussian}\left(\boldsymbol{\mu}_{\xi}^{*}, \boldsymbol{\Sigma}_{\xi}^{*}\right)$, where $\boldsymbol{\Sigma}_{\xi}^{*} \equiv \left(\sigma^{2}\mathbf{I}_{n}^{-1} + \frac{1}{\sigma_{\xi}^{2}}\mathbf{I}_{n}\right)^{-1}$, and $\boldsymbol{\mu}_{\eta}^{*} \equiv \boldsymbol{\Sigma}_{\eta}^{*} \times \sigma^{2}\mathbf{I}_{n}^{-1} \times (\mathbf{z} - \boldsymbol{\Psi}\boldsymbol{\eta} - \mathbf{X}\boldsymbol{\beta})$.

- The full-conditional distribution for $\boldsymbol{\beta}$ is given by $\boldsymbol{\beta} \sim \text{Normal}\left(\mu^{*}, \sigma_{\mu}^{*}\right)$, where $\sigma_{\mu}^{*} \equiv \left(\mathbf{X}^{\top}\sigma^{2}\mathbf{I}_{n}^{-1}\mathbf{X} + \sigma_{\mu}^{-2}\mathbf{I}_{p}\right)^{-1}$, and $\boldsymbol{\mu}^{*} \equiv \sigma_{\mu}^{*} \times \mathbf{X}^{\top}\sigma^{2}\mathbf{I}_{n}^{-1}(\mathbf{z} - \boldsymbol{\Psi}\boldsymbol{\eta})$.

- The full conditional distributions for σ_{β}^{2}, σ_{Y}^{2}, and σ_{ξ}^{2} are $\text{IG}(p/2 + 1,\ 1 + \boldsymbol{\beta}^{\top}\boldsymbol{\beta}/2)$, $\text{IG}(r/2 + 1,\ 1 + \boldsymbol{\eta}^{\top}\text{cov}(\boldsymbol{\eta}|\boldsymbol{\theta})^{-1}\boldsymbol{\eta}/2)$, and $\text{IG}(n/2 + 1,\ 1 + \boldsymbol{\xi}^{\top}\boldsymbol{\xi}/2)$ respectively.

- The full conditional distributions for τ is given by

$$p(\tau) = \frac{\frac{1}{\sigma_{Y}^{r}|\text{cov}(\boldsymbol{\eta}|\boldsymbol{\theta})|^{1/2}}\exp(-\boldsymbol{\eta}^{\top}\text{cov}(\boldsymbol{\eta}|\tau)^{-1}\boldsymbol{\eta})/2}{\sum_{\theta \in \{\theta_{1}, \ldots, \theta_{M}\}} \frac{1}{\sigma_{Y}^{r}|\text{cov}(\boldsymbol{\eta}|\boldsymbol{\theta})|^{1/2}}\exp(-\boldsymbol{\eta}^{\top}\text{cov}(\boldsymbol{\eta}|\tau)^{-1}\boldsymbol{\eta})/2},$$

  where $\tau = \tau_{1}, ..., \tau_{M}$.

The choices for $\tau_{1}, ..., \tau_{M}$ were chosen after a sensitivity analysis. For both the conditional autoregressive model and the Matérn model we let $\tau = 0.01, \cdots, 0.99$. A Gibbs sampler based on the full-conditional distributions in the bulleted list can be used to obtain $\boldsymbol{\beta}^{[b]}$, $\boldsymbol{\eta}^{[b]}$, $\boldsymbol{\xi}^{[b]}$, and $\boldsymbol{\theta}^{[b]}$ for $b = 1, ..., B$.

Step II:

We now discuss simulating from the posterior predictive distribution; that is, simulating

from $f(Y(\mathbf{s})|\boldsymbol{\eta}, \boldsymbol{\xi}, \boldsymbol{\theta}, \mathbf{z}) = \text{Gau}\left\{\mathbf{e}(\mathbf{s})^\top E(\mathbf{y}|\boldsymbol{\eta}, \boldsymbol{\xi}, \boldsymbol{\theta}), \mathbf{e}(\mathbf{s})^\top \mathbf{K}(\boldsymbol{\theta})\mathbf{e}(\mathbf{s})\right\}$. The steps involved for implementing the predictions in Section 5.1 are as follows.

1. Let $b = 1$.

2. Store a sparse $N \times N$ diagonal matrix, which we denote with $\mathbf{D}(\boldsymbol{\theta}^{[b]}) = \text{diag}\{\mathbf{e}(\mathbf{s}_i)^\top \mathbf{K}(\boldsymbol{\theta}^{[b]})\mathbf{e}(\mathbf{s}_i) : i = 1, \ldots, N\}$. Here, $\mathbf{e}(\mathbf{s}) = \{I(\mathbf{s} = \mathbf{s}_1), \ldots, I(\mathbf{s} = \mathbf{s}_n)\}$ and $\mathbf{K}(\boldsymbol{\theta}^{[b]}) = \text{cov}(\mathbf{y}|\boldsymbol{\theta}^{[b]}) - \text{cov}\left(\mathbf{y}, \boldsymbol{\eta}|\boldsymbol{\theta}^{[b]}\right)\mathbf{K}^{-1}(\boldsymbol{\theta}^{[b]})\text{cov}\left(\boldsymbol{\eta}, \mathbf{y}|\boldsymbol{\theta}^{[b]}\right)$. The specification in Section 5.1 is given by

$$\mathbf{K}(\boldsymbol{\theta}^{[b]}) = \sigma_Y^{2[b]}\boldsymbol{\Sigma}_Y(\tau^{[b]}) - \sigma_Y^{2[b]}\boldsymbol{\Sigma}_Y(\tau^{[b]})\boldsymbol{\Psi}(\boldsymbol{\Psi}^\top\boldsymbol{\Sigma}_Y(\tau^{[b]})\boldsymbol{\Psi})^{-1}\boldsymbol{\Psi}^\top\boldsymbol{\Sigma}_Y(\tau^{[b]}),$$

   where $\boldsymbol{\Sigma}_Y(\tau^{[b]})$ is formed from a Matérn covariogram with smoothing parameter 0.5 (i.e., an exponential covariagram), unit variance, and spatial range parameter $\tau^{[b]}$.

3. Store a sparse $N \times N$ diagonal matrix, which we denote with $\mathbf{D}(\boldsymbol{\theta}^{[b]}) = \text{diag}\{\mathbf{e}(\mathbf{s}_i)^\top \mathbf{K}(\boldsymbol{\theta})\mathbf{e}(\mathbf{s}_i) : i = 1, \ldots, N\}$.

4. Compute

$$\mathbf{y}^{[b]} = \mathbf{X}\boldsymbol{\beta}^{[b]} + \boldsymbol{\Sigma}_Y(\tau^{[b]})\boldsymbol{\Psi}(\boldsymbol{\Psi}^\top\boldsymbol{\Sigma}_Y(\tau^{[b]})\boldsymbol{\Psi})^{-1}(\boldsymbol{\Psi}^\top\boldsymbol{\Psi})\boldsymbol{\eta}^{[b]} + \boldsymbol{\xi}^{[b]} + \mathbf{D}(\boldsymbol{\theta}^{[b]})^{1/2}\boldsymbol{\phi},$$

   where $\boldsymbol{\phi}$ is a draw from a standard multivariate normal distribution. Store $\mathbf{y}^{[b]}$.

5. Let $b = b + 1$.

6. Repeat Steps 2 through 5 until $b = B$.

7. Compute and store,

$$\widehat{E}(\mathbf{y}|\mathbf{z}) = \frac{1}{B}\sum_{b=1}^{B}\mathbf{y}^{[b]}.$$

8. Let $resid_b(\mathbf{s}) = \mathbf{e}(\mathbf{s})^\top(\mathbf{y}^{[b]} - \widehat{E}(\mathbf{y}|\mathbf{z}))$ For each $\mathbf{s} \in \{\mathbf{s}_i : i = 1, \ldots, N\}$ compute and store,

$$\widehat{\text{var}}(Y(\mathbf{s})|\mathbf{z}) = \frac{1}{B}\sum_{b=1}^{B} resid_b(\mathbf{s})^2.$$

Steps $1 - 8$ above produce the predictions in Section 5.1 using Special Case 4. The predictions based on Special Case 1, illustrated in Section 5.2, are very similar. In fact there are only two differences. The first is that we have different specifications of $\text{cov}(\mathbf{y})$ and $\text{cov}(\mathbf{y}, \boldsymbol{\eta}|\boldsymbol{\theta})$, and the second difference is that we select σ^2_{Yw} and ρ using the criterion in Equation (14) of the main text. We outline the procedure used in Section 5.2 in Steps 1 through 17 listed below.

1. Let $\sigma^2_{Yw} = 0.1$.

2. Let $\rho = 0.01$.

3. Let $b = 1$.

4. Store a sparse $N \times N$ diagonal matrix, which we denote with $\mathbf{D}(\boldsymbol{\theta}^{[b]}) = \text{diag}\{\mathbf{e}(\mathbf{s}_i)^\top \mathbf{K}(\boldsymbol{\theta}^{[b]})\mathbf{e}(\mathbf{s}_i) : i = 1, \ldots, N\}$. Here, $\mathbf{e}(\mathbf{s}) = \{I(\mathbf{s} = \mathbf{s}_1), \ldots, I(\mathbf{s} = \mathbf{s}_n)\}$ and $\mathbf{K}(\boldsymbol{\theta}^{[b]}) = \sigma^2_{Yw}(\mathbf{I} - \rho\mathbf{A})^{-1}$.

5. Store a sparse $N \times N$ diagonal matrix, which we denote with $\mathbf{D}(\boldsymbol{\theta}^{[b]}) = \text{diag}\{\mathbf{e}(\mathbf{s}_i)^\top \mathbf{K}(\boldsymbol{\theta})\mathbf{e}(\mathbf{s}_i) : i = 1, \ldots, N\}$.

6. Compute

$$\mathbf{y}^{[b]} = \mathbf{X}\boldsymbol{\beta}^{[b]} + \boldsymbol{\Psi}\boldsymbol{\eta}^{[b]} + \boldsymbol{\xi}^{[b]} + \mathbf{D}(\boldsymbol{\theta}^{[b]})^{1/2}\boldsymbol{\phi},$$

where $\boldsymbol{\phi}$ is a draw from a standard multivariate normal distribution. Store $\mathbf{y}^{[b]}$.

7. Let $b = b + 1$.

8. Repeat Steps 4 through 7 until $b = B$.

9. Compute and store,
$$\widehat{E}(\mathbf{y}|\mathbf{z}, \rho, \sigma^2_{Yw}) = \frac{1}{B}\sum_{b=1}^{B} \mathbf{y}^{[b]}.$$

10. Let $resid_b(\mathbf{s}) = \mathbf{e}(\mathbf{s})^\top(\mathbf{y}^{[b]} - \widehat{E}(\mathbf{y}|\mathbf{z}))$ For each $\mathbf{s} \in \{\mathbf{s}_i : i = 1, \ldots, N\}$ compute and store,
$$\widehat{\text{var}}(Y(\mathbf{s})|\mathbf{z}, \rho, \sigma^2_{Yw}) = \frac{1}{B}\sum_{b=1}^{B} resid_b(\mathbf{s})^2.$$

11. Compute and store an estimate of Equation (17),
$$\begin{aligned}\widehat{criterion}(\rho, \sigma^2_{Yw}) = &\left\{\mathbf{z} - \widehat{E}(\mathbf{y}|\mathbf{z}, \rho, \sigma^2_{Yw})\right\}^\top \left\{\mathbf{z} - \widehat{E}(\mathbf{y}|\mathbf{z}, \rho, \sigma^2_{Yw})\right\} \\ &+ 2\frac{1}{B}\sum_{b=1}^{B} \left(\mathbf{z} - \mathbf{y}^{[b]}\right)^\top \widehat{E}(\mathbf{y}|\mathbf{z}, \tau, \sigma^2_{Yw}).\end{aligned}$$

12. Let $\rho = \rho + 0.01$.

13. Repeat Steps 3 through 11 until $\tau = 0.99$.

14. Let $\sigma^2_{Yw} = \sigma^2_{Yw} + 0.1$.

15. Repeat steps 2 through 14 until $\sigma^2_{Yw} = 10$.

16. Compute $(\widehat{\rho}, \widehat{\sigma}^2_{Yw}) = \arg\min \left\{\widehat{criterion}(\rho, \sigma^2_{Yw}) : \rho = 0.01, \ldots, 0.99, \sigma^2_{Yw} = 0.1, \ldots, 10\right\}$.

17. Use $\widehat{E}(\mathbf{y}|\mathbf{z}, \widehat{\rho}, \widehat{\sigma}^2_{Yw})$ and $\widehat{\text{var}}(Y(\mathbf{s})|\mathbf{z}, \widehat{\rho}, \widehat{\sigma}^2_{Yw})$ for prediction and uncertainty quantification, respectively.

Appendix C: Specification of Spatial Basis Functions

We consider two different choices for $\boldsymbol{\psi}(\mathbf{s}) \equiv \{\psi_1(\mathbf{s}), ..., \psi_r(\mathbf{s})\}^\top$; namely, the local bisquare radial basis function (Cressie and Johannesson, 2006, 2008), and the Moran's I basis function (Griffith, 2000, 2002, 2004). These two basis functions are chosen to represent a commonly used point-referenced basis function and areal-referenced basis function, respectively. However, many other choices are available and can be used within our framework. For other choices of point-referenced basis functions see Wikle (2010) and Bradley et al. (2015a), and see Bradley et al. (2017) for other choices of areal-referenced basis functions. The local bisquare radial basis functions are defined as follows:

$$\psi_j(\mathbf{s}) \equiv \begin{cases} \{1 - (||\mathbf{s} - \mathbf{c}_j||/w_r)^2\}^2 & \text{if } ||\mathbf{s} - \mathbf{c}_j|| \leq w_r \\ 0 & \text{otherwise; } \mathbf{s} \in D, \ j = 1, ..., r \end{cases}, \tag{C.1}$$

with pre-specified knots $\mathbf{c}_j$ and w_r is 1.5 times the smallest distance between two different knots in the set $\{\mathbf{c}_j\}$.

The Moran's I basis function is motivated by removing random effects that are confounded with covariates. This is done in an effort to facilitate inference on $\boldsymbol{\beta}$. Then, the $N \times N$ matrix $\boldsymbol{\Psi}$ is specified to be contained within the orthogonal complement of the column space of $\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$. That is, define the MI operator as

$$\mathbf{G}(\mathbf{X}, \mathbf{A}) \equiv \left(\mathbf{I}_N - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\right)\mathbf{A}\left(\mathbf{I}_N - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\right); \ t = 1, \ldots, T, \tag{C.2}$$

where $\mathbf{A}$ is a generic $N \times N$ weight matrix. We let $\mathbf{A}$ be the adjacency matrix corresponding to the edges formed by D. Notice that the MI operator in (C.2) defines a column space that is orthogonal to $\mathbf{X}$. Then, let the spectral representation $\mathbf{G}(\mathbf{X}, \mathbf{A}) = \boldsymbol{\Phi}\boldsymbol{\Lambda}\boldsymbol{\Phi}^\top$, and denote the $N \times r$ real matrix formed from the first N columns of $\boldsymbol{\Phi}$ as $\boldsymbol{\Psi}_N$. As done in Bradley

et al. (2015b), we set the row of $\boldsymbol{\Psi}$ that corresponds to areal unit A equal to $\boldsymbol{\psi}(A)$.

There are many algorithms that exist to select r knot locations: see Bradley et al. (2011), Nychka (2001), and Section 12.4.4 of Banerjee et al. (2008), among many others. However, we would like to use our unique modeling perspective to guide this choice. Thus, in what follows we use the relationship between $Y(\cdot)$ and $w(\cdot)$ to determine r and the knot locations. We outline the choice of knots when using local bisquare radial basis functions in the following enumerated list.

0. Let $t = 0$. Define $\{\mathbf{c}_j\}$ to be a $g_0 \times g_0$ equally spaced grid of locations in D, where g_0 is a positive integer. Define a threshold value for

$$\sum_{\mathbf{s} \in D_O} \left[E\left\{w(\mathbf{s})|\mathbf{z}\right\} - E\left\{Y(\mathbf{s})|\mathbf{z}\right\}\right]^2. \tag{C.3}$$

   We consider the average squared distance between $Y(\cdot)$ and $w(\cdot)$ to be reasonable if it is less than or equal to 0.1. This decision was made based on simulation.

1. Set $t = t + 1$. If (C.3) is less than or equal to the threshold (e.g., 0.1) stop, otherwise set $g_t = 2g_{t-1}$.

2. Let $\{\mathbf{c}_j\}$ be the knots formed by a $g_t \times g_t$ grid.

3. Repeat steps 1−2.

The step-by-step instructions on choosing the rank for the MI basis functions are very similar to choosing knots of spatial basis functions.

0. Let $t = 0$. Let r_0 equal roughly 10% of the available basis functions from the Moran's I operator in (C.2) (note that this is the rule of thumb used in Hughes and Haran (2013)). Define a threshold distance between $Y(\cdot)$ and $w(\cdot)$. In Section 5, we consider

(C.3) to be reasonable if it is less than or equal to 0.1. This decision was made based on simulation.

1. Set $t = t + 1$. If (C.3) is less than or equal to the threshold (e.g., 0.1) stop, otherwise set $r_t = 2r_{t-1}$.

2. Let $r = r_t$.

3. Repeat steps 1−2.

From our experience, the two algorithms given above tend to stop after three iterations. Additionally, when using Special Case 4, one should check to see if the eigenvectors of $\mathbf{\Sigma}_Y$ fall in the column space of $\mathbf{P}$.

Appendix D: Regularity Conditions

The regularity conditions for Statement 4 in Appendix A, Theorem 1, and Corollary 1 are listed below.

1. Let $(\Omega, \mathcal{A}, \mathcal{P})$ be a probability space, where Ω is a generic sample space, $\mathcal{A}$ is defined to be a sigma-algebra on Ω, and $\mathcal{P}$ is a generic probability measure.

2. For the mapping $Z : D \times \Omega \to \mathbb{R}$, it is assumed that $Z(\mathbf{s})$ is measurable for every $\mathbf{s} \in D \subset \mathbb{R}^d$.

3. For the mapping $Y : D \times \Omega \to \mathbb{R}$, it is assumed that $Y(\mathbf{s})$ is measurable for every $\mathbf{s} \in D \subset \mathbb{R}^d$.

4. For the mapping $\delta : D \times \Omega \to \mathbb{R}$, it is assumed that $\delta(\mathbf{s})$ is measurable for every $\mathbf{s} \in D \subset \mathbb{R}^d$.

5. Assume var $\{\delta(\mathbf{s})\} < \infty$ for every $\mathbf{s} \in D$.

6. For every $\omega \in \Omega$ and $\mathbf{s} \in D$ let $Z(\mathbf{s}) = Y(\mathbf{s}) + \delta(\mathbf{s})$.

7. For the mapping $\epsilon : D \times \Omega \to \mathbb{R}$, it is assumed that $\epsilon(\mathbf{s})$ is measurable for every $\mathbf{s} \in D \subset \mathbb{R}^d$.

8. Assume var $\{\epsilon(\mathbf{s})\} < \infty$ for every $\mathbf{s} \in D$.

9. Assume $Y(\cdot)$ and $\epsilon(\cdot)$ are mutually independent for any finite collection of locations in D.

10. Let $\boldsymbol{\theta}$ be a general real-valued parameter vector.

11. Let $w(\cdot)$ be a mapping $w : D \times \Omega \to \mathbb{R}$, where $w(\mathbf{s})$ is measurable for every $\mathbf{s} \in D \subset \mathbb{R}^d$.

12. Assume that the expected value of $\epsilon(\mathbf{s})$ is zero for every $\mathbf{s} \in D$.

13. Assume that var $\{\epsilon(\mathbf{s})\} < \infty$ for every $\mathbf{s} \in D$.

14. Assume that cov $\{\epsilon(\mathbf{s}_1), \epsilon(\mathbf{s}_2)\} = 0$ for every $\mathbf{s}_1, \mathbf{s}_2 \in D$.

15. Assume $w(\cdot)$ and $\epsilon(\cdot)$ are mutually independent for any finite collection of locations in D.

16. For every $\omega \in \Omega$ and $\mathbf{s} \in D$ the General Assumption in the main text is given by the following: $\delta(\mathbf{s}) = w(\mathbf{s}) - Y(\mathbf{s}) + \epsilon(\mathbf{s})$; $\mathbf{s} \in D$.

17. Let $f(Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_k)|\mathbf{w}, \mathbf{z}, \boldsymbol{\theta})$ be a valid probability density on $\mathbb{R}^k$, where recall $\mathcal{S} \subset D \subset \mathbb{R}^d$ is an open set for each $k \in \mathbb{N} = \{1, 2, 3, \ldots\}$ and finite collection of locations $\mathbf{s}_1, \ldots, \mathbf{s}_k \in \mathcal{S}$.

18. Define $f(Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_k)|\mathbf{w}, \mathbf{z}, \boldsymbol{\theta})$, $f(\mathbf{z}|\mathbf{w}, \boldsymbol{\theta})$, $f(\mathbf{z}|\mathbf{w}, \boldsymbol{\theta}, Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_k))$, $f(Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_k)|\mathbf{w}, \boldsymbol{\theta})$, $f(\mathbf{w}|\boldsymbol{\theta})$ to be valid probability densities on $\mathbb{R}$.

19. Let $f(Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_k)|\mathbf{w}, \boldsymbol{\theta})$ be Kolmogorov consistent.

Appendix E: Review of Current Methods for Spatial Prediction

In this section, we review three spatial predictors already in the literature, which arise from a spatial mixed model. This is done, in part, to show the primary difference in our proposed method and current methods for spatial prediction. For an in-depth review of many of the current spatial predictors see Bradley et al. (2016a).

The Spatial Mixed Effects Model: Let

$$Y(\mathbf{s}) = \mathbf{x}(\mathbf{s})^\top \boldsymbol{\beta} + \gamma(\mathbf{s}) + \xi(\mathbf{s}),$$

where $\mathbf{x}(\mathbf{s})$ is a known p-dimensional vector of spatial covariates, $\gamma(\cdot)$ is a mean-zero spatial process with a generic spatial covariance function $C : D \times D \to \mathbb{R}$, and the process $\xi(\cdot)$ has mean-zero, finite variance, and no spatial covariances. Let $D_P \subset D$ consists of N prespecified prediction locations and define the N-dimensional random vector $\boldsymbol{\gamma} = \{\gamma(\mathbf{s}) : \mathbf{s} \in D_P\}^\top$. Denote the $N \times N$ covariance matrix of $\boldsymbol{\gamma}$ with $\boldsymbol{\Sigma}_Y$.

The spatial mixed effects model can be written as,

$$w(\mathbf{s}) = \mathbf{x}(\mathbf{s})^\top \boldsymbol{\beta} + \boldsymbol{\psi}(\mathbf{s})^\top \boldsymbol{\eta} + \xi(\mathbf{s}); \ \mathbf{s} \in D, \tag{E.1}$$

where $\boldsymbol{\psi} : D \to \mathbb{R}^r$ is a generic set of spatial basis functions, $r \ll n$, $\boldsymbol{\eta}$ is an r-dimensional random vector of expansion coefficients that is mean zero and has $r \times r$ covariance matrix $\mathbf{K}$, and $\boldsymbol{\eta}$ is mutually independent of the process $\xi(\cdot)$. The covariance of $\boldsymbol{\eta}$ is chosen so that

$\text{cov}(\mathbf{w}) \approx \text{cov}(\mathbf{y})$, where the N-dimensional random vector $\mathbf{y} = \{Y(\mathbf{s}) : \mathbf{s} \in D_P\}^\top$, and the N-dimensional random vector $\mathbf{w} = \{w(\mathbf{s}) : \mathbf{s} \in D_P\}^\top$. Note that

$$\mathbf{w} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\Psi}\boldsymbol{\eta} + \boldsymbol{\xi},$$

where the $N \times p$ matrix $\mathbf{X} \equiv \left\{\mathbf{x}(\mathbf{s})^\top : \mathbf{s} \in D_P\right\}^\top$, the $N \times r$ matrix $\boldsymbol{\Psi} \equiv \left\{\boldsymbol{\psi}(\mathbf{s})^\top : \mathbf{s} \in D_P\right\}^\top$, and the N-dimensional random vector $\boldsymbol{\xi} \equiv \{\boldsymbol{\xi}(\mathbf{s}) : \mathbf{s} \in D_P\}^\top$.

The spatial mixed effects model has been used for areal data as well (i.e., when D_P consists of areal units); (Griffith, 2000; Hughes and Haran, 2013; Porter et al., 2014; Bradley et al., 2016b, 2015c,b, 2017). For this setting $\boldsymbol{\Sigma}_Y$ is assumed to be the covariance matrix of an intrinsically autoregressive model, and $\boldsymbol{\psi}$ is defined to be the Moran's I basis function (Griffith, 2000, 2002, 2004). In Section 5, we use the spatial mixed effects model for areal data.

Fixed Rank Kriging: Cressie and Johannesson (2008) aggregate $\mathbf{w}$ and $\mathbf{y}$ by postmultiplying by an $M \times N$ matrix $\mathbf{J}$, which is row-normalized version of a matrix of zeros and ones. Then, Cressie and Johannesson (2008) define

$$\mathbf{K} \equiv \arg\min_{\mathbf{K}} \left(||\text{cov}(\mathbf{Jw}) - \text{cov}(\mathbf{Jy})||_F^2\right) = \arg\min_{\mathbf{K}} \left(||\mathbf{J}\boldsymbol{\Psi}\mathbf{K}\boldsymbol{\Psi}^\top\mathbf{J}^\top - \mathbf{J}\boldsymbol{\Sigma}_Y\mathbf{J}^\top||_F^2\right), \quad \text{(E.2)}$$

where for any real-valued square matrix $\mathbf{G}$ we have that $||\mathbf{G}||_F^2 = \text{trace}(\mathbf{G}^\top\mathbf{G})$. (The matrix-valued distance function $||\cdot||_F$ is referred to as the Frobenius norm.) The closed form expression for (E.2) is well-known (Cressie and Johannesson, 2008) and given by:

$$\mathbf{K} = (\boldsymbol{\Psi}^\top\mathbf{J}^\top\mathbf{J}\boldsymbol{\Psi})^{-1}\boldsymbol{\Psi}^\top\mathbf{J}\boldsymbol{\Sigma}_Y\mathbf{J}^\top\boldsymbol{\Psi}(\boldsymbol{\Psi}^\top\mathbf{J}^\top\mathbf{J}\boldsymbol{\Psi})^{-1}. \quad \text{(E.3)}$$

Cressie and Johannesson (2008) substitute a method of moments estimator of $\mathbf{J}\boldsymbol{\Sigma}_Y\mathbf{J}^\top$ into

(E.3), and use the Sherman-Morrison-Woodbury formula (Searle, 1982) to efficiently calculate the best linear unbiased predictor, which they refer to as fixed rank Kriging (FRK).

Now, Cressie and Johannesson (2008) assume that $w(\cdot) \equiv Y(\cdot)$ (cf. Equations (2.6) and (2.12) of Cressie and Johannesson (2008)), which is a key difference with our approach. That is, Cressie and Johannesson (2008) assume that their reduced rank approximation w is exactly equal to Y, where we assume $w(\mathbf{s}) \neq Y(\mathbf{s})$ for at least one location $\mathbf{s} \in D$.

Modified Predictive Processes: Consider the modified predictive process (MPP) (Finley et al., 2009) approach, which is motivated by defining an approximation to a full rank model. Let $\{\gamma(\mathbf{s}) : \mathbf{s} \in D\}$ be assumed to have an isotropic covariagram denoted by $C(||\mathbf{h}||)$ and $\mathbf{h} \in \mathbb{R}^d$. Specifically, let

$$\boldsymbol{\Sigma}_Y \equiv \{C_{\mathrm{M}}(\mathbf{h}; \boldsymbol{\theta}) : \mathbf{h} = ||\mathbf{s} - \mathbf{s}||_E, \mathbf{s}, \mathbf{s} \in D_P\},$$

where $||\cdot||_E$ is the Euclidean distance, the Matérn covariance function (Matérn, 1960) is defined as,

$$C_{\mathrm{M}}(\mathbf{h}; \boldsymbol{\theta}) = \frac{\sigma_Y^2}{\Gamma(\alpha)2^{\alpha-1}}(\tau||\mathbf{h}||_E)^\alpha K_\alpha(\tau||\mathbf{h}||_E);\ \mathbf{h} = ||\mathbf{s}_1 - \mathbf{s}_2||_E, \mathbf{s}_1, \mathbf{s}_2 \in D, \tag{E.4}$$

$\boldsymbol{\theta} \equiv (\sigma_Y^2, \alpha, \tau)^\top$, K_α is the modified Bessel function of the second kind of order $\alpha > 0$, the correlation parameter $\tau > 0$, and the variance parameter $\sigma_Y^2 > 0$.

Consider the spatial mixed effects model with $\boldsymbol{\psi}(\mathbf{s})^\top = \{C(||\mathbf{s} - \mathbf{s}_1^*||), \ldots, C(||\mathbf{s} - \mathbf{s}_r^*||)\}\,\mathbf{K}^{-1}$, $\mathbf{s}_1^*, \ldots, \mathbf{s}_r^* \in D$ are prespecifed knot locations, $r \ll n$, and the r-dimesensional random vector $\boldsymbol{\eta}$ is assumed to be Gaussian with mean zero and $r \times r$ covariance matrix $\mathbf{K} = \left\{C(||\mathbf{s}_i^* - \mathbf{s}_j^*||) : i, j = 1, \ldots, r\right\}$. The motivation here is that the likelihood for the r-dimensional random vector $\boldsymbol{\eta}$ is easier to compute than the likelihood associated with

the n-dimensional vector $\{Y(\mathbf{s}) : \mathbf{s} \in D\}$, since $r \ll n$. Additionally, if $r = N$ and $\{\mathbf{s}_1^*, \ldots, \mathbf{s}_r^*\} = D_P$ then for Gaussian $\boldsymbol{\eta}$ we have that $\boldsymbol{\psi}(\mathbf{s})^\top \boldsymbol{\eta}$ is almost surely equal in distribution to $\gamma(\mathbf{s})$ for $\mathbf{s} \in D_P$. Finley et al. (2009) note that

$$\text{var}\left\{\boldsymbol{\psi}(\mathbf{s})^\top \boldsymbol{\eta}\right\} = \boldsymbol{\psi}(\mathbf{s})^\top \mathbf{K}^{-1} \boldsymbol{\psi}(\mathbf{s}) \neq \text{var}\left\{\gamma(\mathbf{s})\right\},$$

and suggest setting

$$w_1(\mathbf{s}) = \mathbf{x}(\mathbf{s})^\top \boldsymbol{\beta} + \boldsymbol{\psi}(\mathbf{s})^\top \boldsymbol{\eta} + \xi(\mathbf{s}); \ \mathbf{s} \in D, \tag{E.5}$$

where $\text{var}\{\xi(\mathbf{s})\} = \text{var}\{\gamma(\mathbf{s})\} - \boldsymbol{\psi}(\mathbf{s})^\top \mathbf{K}^{-1} \boldsymbol{\psi}(\mathbf{s})$, $\xi(\mathbf{s}_1)$ is independent of $\xi(\mathbf{s}_2)$ for $\mathbf{s}_1 \neq \mathbf{s}_2$, and $\xi(\cdot)$ is mutually independent of $\boldsymbol{\eta}$. This modification guarantees that $w_1(\cdot)$ has the same variance as $\gamma(\cdot)$.

Further, Banerjee et al. (2008) and Finley et al. (2009) assume that $w_1(\cdot) \equiv Y(\cdot)$ (cf. Equations (5) and (7) of Finley et al. (2009)), which is a key difference with our approach. That is, Finley et al. (2009) assume that their reduced rank approximation w_1 is exactly equal to Y, where we assume $w_1(\cdot) \neq Y(\cdot)$.

Full Scale Approximation: Sang and Huang (2012) is another predictive process-type approach that bares some similarity to the augmented process $w(\cdot)$. They approximate the process $\{Y(\mathbf{s}) : \mathbf{s} \in D\}$ with

$$w_2(\mathbf{s}) = Y_{taper}(\mathbf{s}) + w_1(\mathbf{s}), \tag{E.6}$$

where

$$\text{cov}\left\{Y_{taper}(\mathbf{s}_1), Y_{taper}(\mathbf{s}_2)\right\} = \left[\text{cov}\left\{Y(\mathbf{s}_1), Y(\mathbf{s}_2)\right\} - \mathbf{g}(\mathbf{s}_1)^\top \mathbf{K}^{-1} \mathbf{g}(\mathbf{s}_2)\right] K_{taper}(\mathbf{s}_1, \mathbf{s}_2); \ \mathbf{s}_1, \mathbf{s}_2 \in D,$$

K_{taper} is a taper function, and $Y_{taper}(\cdot)$ and $w_1(\cdot)$ are mutually independent. Sang and Huang

(2012) assume their full-scale approximation $w_2(\cdot)$ is *exactly* equal to $Y(\cdot)$ (cf. Equation (13) of Sang and Huang (2012)). Again, the key difference with our approach is that we assume $w_2(\cdot) \neq Y(\cdot)$.

Appendix F: Proof of Theorem 1

Let $\mathbf{y} = (Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_k))^\top$. We have that $[\mathbf{y}|\mathbf{w}, \boldsymbol{\theta}, \mathbf{z}] = [\mathbf{y}|\mathbf{w}, \boldsymbol{\theta}, \boldsymbol{\epsilon} = \mathbf{z} - \mathbf{X}\boldsymbol{\beta} - \boldsymbol{\Psi}\boldsymbol{\eta}]$, where $\boldsymbol{\epsilon} \equiv (\epsilon(\mathbf{s}_{k+1}), \ldots, \epsilon(\mathbf{s}_{k+n}))'$, the $n \times p$ matrix $\mathbf{X} \equiv (\mathbf{x}(\mathbf{s}_{k+1}), \ldots, \mathbf{x}(\mathbf{s}_{k+n}))^\top$, and the $n \times r$ matrix $\boldsymbol{\Psi} \equiv (\boldsymbol{\psi}(\mathbf{s}_{k+1}), \ldots, \boldsymbol{\psi}(\mathbf{s}_{k+n}))^\top$. Since Y is independent of ϵ, $[\mathbf{y}|\mathbf{w}, \boldsymbol{\theta}, \mathbf{z}] = [\mathbf{y}|\mathbf{w}, \boldsymbol{\theta}, \boldsymbol{\epsilon} = \mathbf{z} - \mathbf{X}\boldsymbol{\beta} - \boldsymbol{\Psi}\boldsymbol{\eta}] = [\mathbf{y}|\mathbf{w}, \boldsymbol{\theta}]$, which completes the proof.

To prove Equation (20) of the main text, we show the conditions of the Kolmogorov Extension theorem. First, we have that,

$$\mathbb{P}\left\{Y(\mathbf{s}_1) \in A_1, ..., Y(\mathbf{s}_k) \in A_k | \mathbf{w}, \boldsymbol{\theta}, \mathbf{z}\right\} =$$
$$\int_{A_1} ... \int_{A_k} \int_{\mathbb{R}} ... \int_{\mathbb{R}} f(Y(\mathbf{s}_1), ..., Y(\mathbf{s}_k), Y(\mathbf{s}_{k+1}), ..., Y(\mathbf{s}_{t+m}) | \mathbf{w}, \boldsymbol{\theta}, \mathbf{z}) dY(\mathbf{s}_1) ... dY(\mathbf{s}_k) dY(\mathbf{s}_{k+1}) \ldots dY(\mathbf{s}_{t+m})$$
$$= \mathbb{P}\left\{Y(\mathbf{s}_1) \in A_1, ..., Y(\mathbf{s}_k) \in A_k, Y(\mathbf{s}_{k+1}) \in \mathbb{R}, ..., Y(\mathbf{s}_{t+m}) \in \mathbb{R} | \mathbf{w}, \boldsymbol{\theta}, \mathbf{z}\right\}.$$

Additionally, we have that $f(\mathbf{y}|\mathbf{w}, \boldsymbol{\theta}, \mathbf{z}) = f(\mathbf{y}|\boldsymbol{\theta}, \mathbf{w})$ from the conditional independence result shown above. Since $f(\mathbf{y}|\boldsymbol{\theta}, \mathbf{w})$ is Kolmogorov consistent, it follows from the Kolomogorov Extension Theorem that there exists a probability space (with sample space Ω, sigma-algebra $\mathcal{F}$, and probability measure $\mathbb{P}$) and stochastic process $Y : \mathcal{S} \times \Omega \to \mathbb{R}$, such that

$$\mathbb{P}\left\{Y(\mathbf{s}_1) \in A_1, \ldots, Y(\mathbf{s}_k) \in A_k\right\} = \int_{A_1} \ldots \int_{A_k} f(Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_k) | \mathbf{w}, \boldsymbol{\theta}, \mathbf{z}) dY(\mathbf{s}_1) \ldots dY(\mathbf{s}_k).$$

A similar argument to the one presented in the first paragraph of the proof provides the last

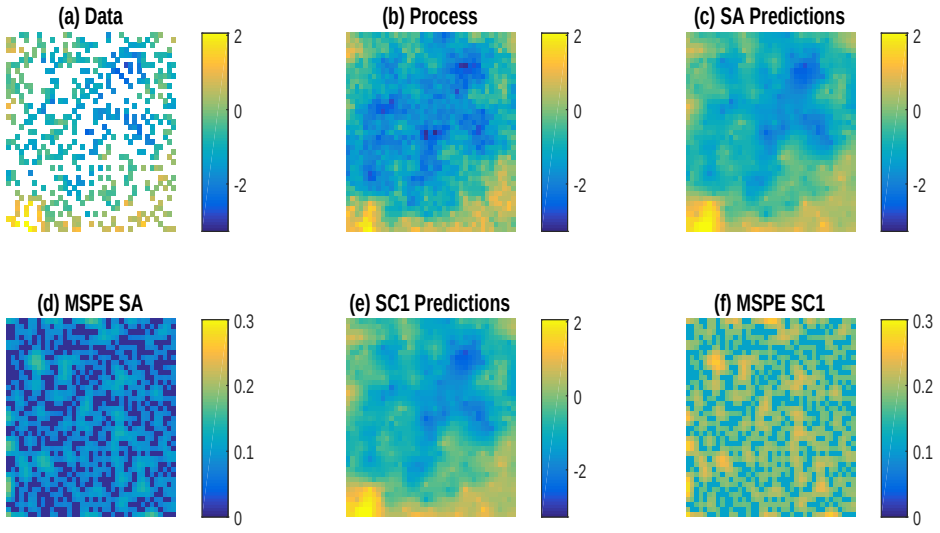

Figure 1: In Panel (a), we display the simulated data $\{Z(\cdot)\}$ over a collection of observed locations, which are generated by randomly selecting points outside of a rectangular region in D. Here, D is a 40×40 grid $D \equiv \{(s_1, s_2)^\top : s_1, s_2 = 0, 0.025, \ldots, 1\}$. White areas indicate a missing observation. Panel (b) represents a simulation of the latent process with a Matérn covariance function with unit variance and range parameter 1/12. In Panels (c), (d), (e), (f) we present the kriging predictor and variances under the Standard Assumption and Special Case 1, respectively. The data and the process were generated according to Special Case 1.

equation in (20) of the main text.

Appendix G: Simulation Study: Special Case 1

In this small simulation study we provide empirical results to investigate the performance of the kriging predictor assuming Special Case 1. The goal of this simulation study is to illustrate that the expression of the kriging predictor (with known covariances) is the same

for both the Standard Assumption and Special Case 1; however, the prediction variances become larger when Special Case 1 holds.

The spatial domain is set equal to a 40×40 grid $D \equiv \{(u_1, u_2)^\top : u_1, u_2 = 0, 0.025, \ldots, 1\}$. Then, let $\mathbf{\Sigma}_w = \mathbf{\Sigma}_{Y,w}$, where $\mathbf{\Sigma}_w$ consists of Matérn covariances (Matérn, 1960) with unit variance. The range parameter is set equal to $1/12$, so that the spatial range is moderate at $1/4$. Additionally, let $\mu(\cdot) \equiv 0$. Let $\mathbf{\Sigma}_Y = \mathbf{\Sigma}_w + \mathbf{\Sigma}_1$, where $\mathbf{\Sigma}_1$ consists of Matérn covariances (Matérn, 1960) with variance 0.001 and range parameter $1/12$. In Figure 2(a), we present simulated data, where $\{\mathbf{s}_1, \ldots, \mathbf{s}_m\}$ consists of locations randomly selected from D and $m = 800$. In Figure 2(c,d) we display the kriging predictor and kriging variances under the Standard Assumption. Figure 1(d,e) displays a map of the kriging predictor and kriging variances under Special Case 1. Both predictors were computed using the known parameters, and the true model was simulated from Special Case 1. Upon comparisons of Figures 2(c,d,e,f), we see that the two kriging predictors are identical, which is to be expected (see Proposition 2). However, the kriging variances are under estimated when using the Standard Assumption. This is true even though the variance of Σ_1 was very small.

Appendix H: Summary of Models Implemented in Section 5.2

In Table 1, we provide details on the models that were compared in Section 5.2 of the main text. Specifically, we outline the model, the assumptions made on the augmented process, the assumptions on the latent process, and notes on implementation.

Model Name (MD)	$w(\cdot)$ **Assumptions**	$Y(\cdot)$ **Assumptions**	**Implementation**
Special Case 1	Let $\mathbf{w} = \boldsymbol{\Psi}\boldsymbol{\eta} + \boldsymbol{\xi}$, where $\boldsymbol{\Psi}$ are the Moran's I basis functions, $\boldsymbol{\eta}$ is assumed to be Gaussian, and $\boldsymbol{\xi}$ is assumed to be Gaussian with independent components with mean zero and variance σ^2_ξ. The variances of $\epsilon(\cdot)$ are assumed known and constant.	Let $\mathbf{y}$ follow an intrinsically autoregressive model.	We perform a fully Bayesian implementation of this model. For details see Supplemental Appendix E.
Fixed Rank kriging (FRK)	Let $\mathbf{w} = \boldsymbol{\Psi}\boldsymbol{\eta} + \boldsymbol{\xi}$, where $\boldsymbol{\Psi}$ are the Moran's I basis functions, $\boldsymbol{\eta}$ is assumed to be Gaussian, and $\boldsymbol{\xi}$ is assumed to be Gaussian with independent components with mean zero and variance σ^2_ξ. The variances of $\epsilon(\cdot)$ are assumed known and constant.	Let $\mathbf{w} = \mathbf{y}$.	We perform a fully Bayesian implementation of this model. For details see Supplemental Appendix E.
Conditional Autoregressive (CAR)	Let $\mathbf{w}$ follow an intrinsically autoregressive model. The variances of $\epsilon(\cdot)$ are computed from the margins of errors made publicly available by ACS.	Let $\mathbf{w} = \mathbf{y}$.	We perform a fully Bayesian implementation of CAR. Predictions of $Y(\cdot)$ were implemented using R, and the Gibbs sampling details can be found in De Oliveira (2012).

Table 1: We list the models compared in Section 5.2. The leftmost column contains the model name (MD). We consider MD = SC1, FRK, and CAR. The middle two columns give the assumptions of the SC1 and the latent process imposed by MD. The rightmost column describes the implementation of MD.

Acknowledgments

We would like to express our gratitude to the editor, the editor, the associate editor, and the referees for their very helpful comments that improved this manuscript. This research was partially supported by the U.S. National Science Foundation (NSF) and the U.S. Census Bureau under NSF grant SES-1132031, funded through the NSF-Census Research Network (NCRN) program. This paper is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The views expressed are those of the authors and not necessarily those of the NSF or the U.S. Census Bureau.

References

Banerjee, S., Gelfand, A. E., Finley, A. O., and Sang, H. (2008). "Gaussian predictive process models for large spatial data sets." *Journal of the Royal Statistical Society Series B*, 70, 825–848.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. New York, NY: Springer-Verlag.

Berliner, L. M. (1996). *Hierarchical Bayesian Time-series Models*. Kluwer Academic Publishers, Dordrecht, NL.

Bradley, J. R., Cressie, N., and Shi, T. (2011). "Selection of rank and basis functions in the Spatial Random Effects model." In *Proceedings of the 2011 Joint Statistical Meetings*, 3393–3406. Alexandria, VA: American Statistical Association.

— (2015a). "Comparing and Selecting Spatial Predictors Using Local Criteria (with discussion)." *TEST*, 24, 1–28 (Rejoinder: pp. 54 – 60).

— (2016a). "A comparison of spatial predictors when datasets could be very large." *Statistics Surveys*, 10, 100–131.

Bradley, J. R., Holan, S. H., and Wikle, C. K. (2015b). "Multivariate spatio-temporal models for high-dimensional areal data with application to Longitudinal Employer-Household Dynamics." *The Annals of Applied Statistics*, 9, 1761–1791.

Bradley, J. R., Wikle, C. K., and Holan, S. H. (2015c). "Spatio-temporal change of support with application to American Community Survey multi-year period estimates." *Stat*, 4, 255 – 270.

— (2016b). "Bayesian spatial change of support for count-valued survey data." *Journal of the American Statistical Association*, 111, 472 – 487.

— (2017). "Regionalization of multiscale spatial processes using a criterion for spatial aggregation error." *Journal of the Royal Statistical Society: Series B*, forthcoming.

Cressie, N. and Johannesson, G. (2006). "Spatial prediction for massive data sets." In *Australian Academy of Science Elizabeth and Frederick White Conference*, 1–11. Canberra: Australian Academy of Science.

— (2008). "Fixed rank kriging for very large spatial data sets." *Journal of the Royal Statistical Society, Series B*, 70, 209–226.

De Oliveira, V. (2012). "Bayesian analysis of conditional autoregressive models." *Annals of the Institute of Statistical Mathematics*, 64, 107–133.

Finley, A. O., Sang, H., Banerjee, S., and Gelfand, A. E. (2009). "Improving the performance of predictive process modeling for large datasets." *Computational Statistics and Data Analysis*, 53, 2873–2884.

Griffith, D. (2000). "A linear regression solution to the spatial autocorrelation problem." *Journal of Geographical Systems*, 2, 141–156.

— (2002). "A spatial filtering specification for the auto-Poisson model." *Statistics and Probability Letters*, 58, 245–251.

— (2004). "A spatial filtering specification for the auto-logistic model." *Environment and Planning A*, 36, 1791–1811.

Hughes, J. and Haran, M. (2013). "Dimension reduction and alleviation of confounding for spatial generalized linear mixed model." *Journal of the Royal Statistical Society, Series B*, 75, 139–159.

Matérn, B. (1960). "Spatial Variation." *Meddelanden fran Statens Skogsforskningsinstitut*, 49, 1–144.

Nychka, D. (2001). "Spatial process estimates as smoothers." In *Smoothing and Regression: Approaches, Computation and Applications,* rev. edn, ed. M. G. Schmiek, 393–424. New York, NY: Wiley.

Porter, A. T., Holan, S. H., Wikle, C. K., and Cressie, N. (2014). "Spatial Fay-Herriot models for small area estimation with functional covariates." *Spatial Statistics*, 10, 27–42.

Sang, H. and Huang, J. (2012). "A full-scale approximation of covariance functions for large spatial data sets." *Journal of the Royal Statistical Society: Series B*, 74, 111–132.

Searle, S. R. (1982). *Matrix Algebra Useful for Statistics*. New York, NY: Wiley.

Wikle, C. K. (2010). "Low-rank representations for spatial processes." In *Handbook of Spatial Statistics*, eds. A. E. Gelfand, P. J. Diggle, M. Fuentes, and P. Guttorp, 107–118. Boca Raton, FL: Chapman & Hall/CRC Press.

Zhang, F. and Zhang, Q. (2006). "Eigenvalue Inequalities for Matrix Product." *IEEE Transactions on Automatic Control*, 51, 1506–1509.