Online Supplementary Materials

Calibrating MODIS Aerosol Optical Depth for Predicting Daily $PM_{2.5}$ Concentrations via Statistical Downscaling

Howard H Chang, Xuefei Hu, and Yang Liu

Contact: howard.chang@emory.edu

Section 1 Additional Figures

Figure S1. Estimates and standard error (SE) of the residual spatial random intercepts $\beta_0(s)$ and slopes $\beta_1(s)$. $PM_{2.5}$ monitoring locations are indicated by the red triangles.

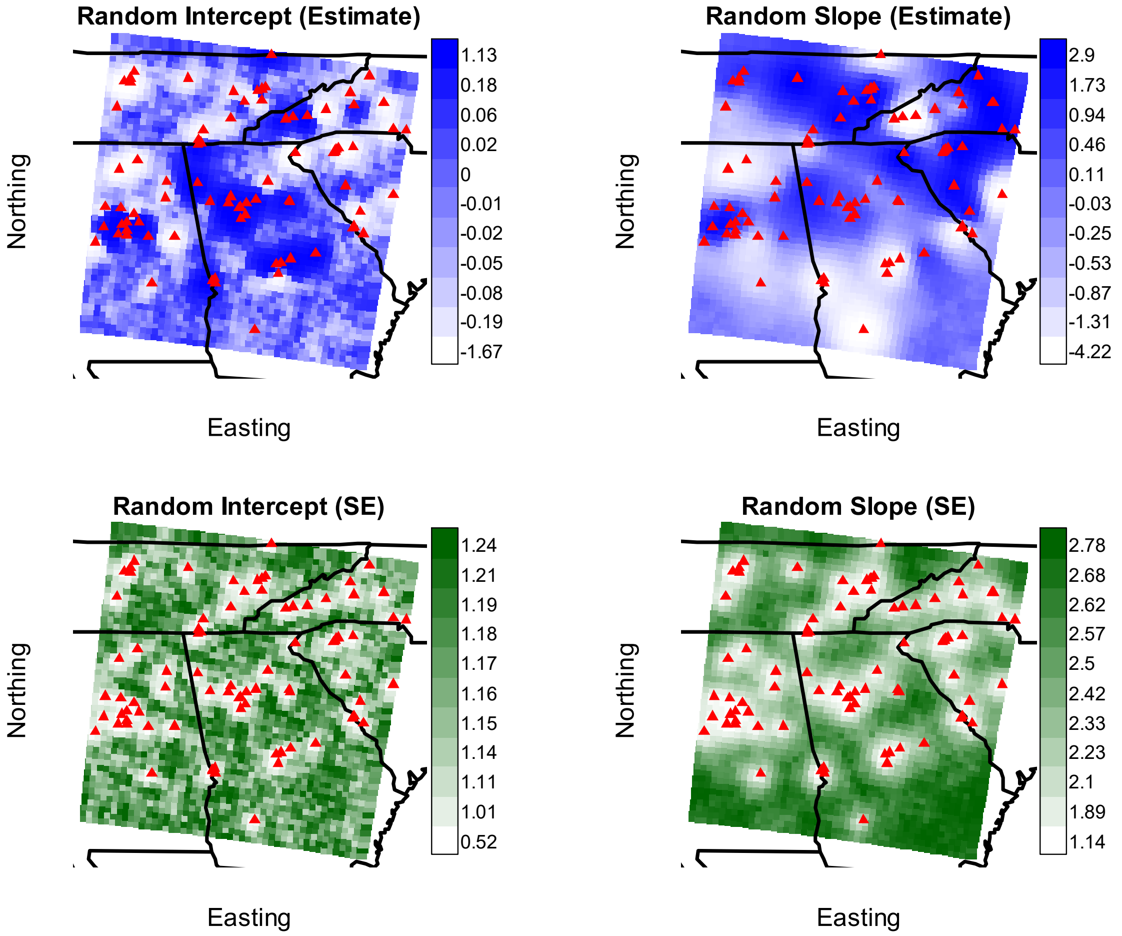


Figure S2. Predicted annual average $PM_{2.5}$ concentrations (left panel) at AOD grid cell centers and their prediction standard errors (right panel). Concentrations associated with missing AOD values were predicted with a Bayesian statistical downscaler without AOD as a covariate.

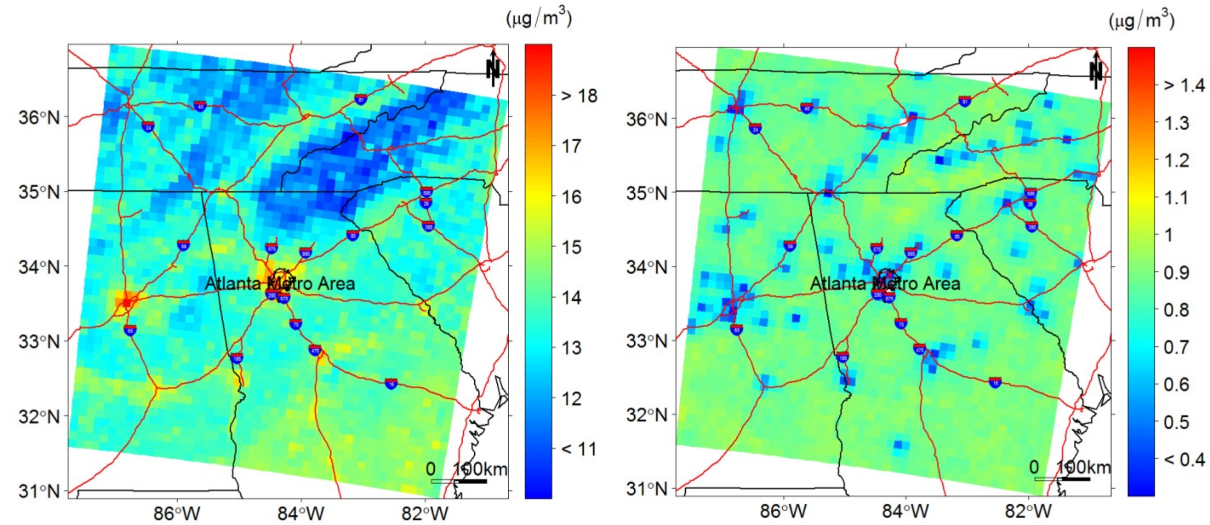
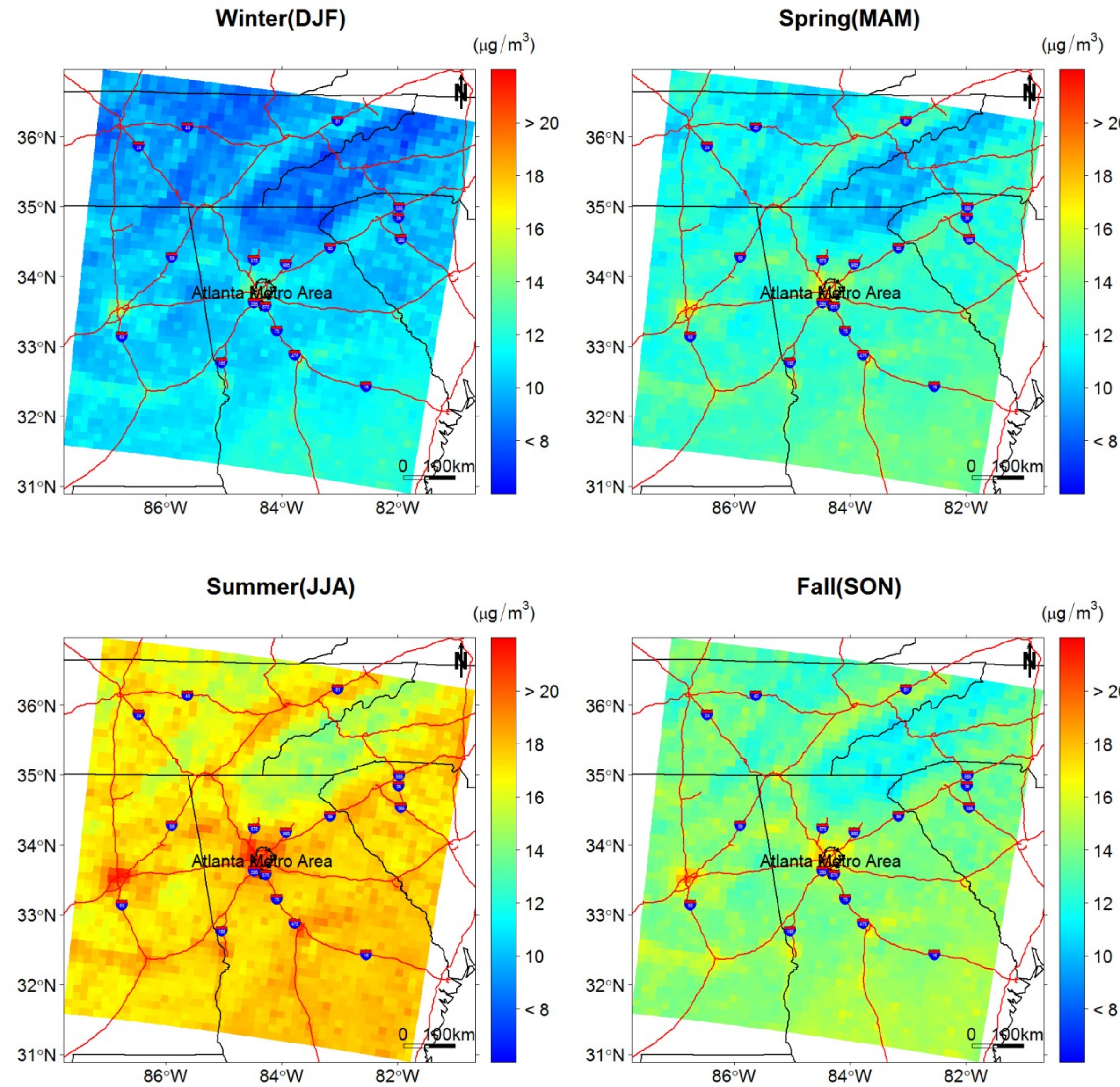


Figure S3. Predicted seasonal average $PM_{2.5}$ concentrations at AOD grid cell centers.



Section 2. Detailed Model Formulation and Estimation

Let $Y(\mathbf{s}, t)$ denote the observed $PM_{2.5}$ concentration on day t at monitoring location $\mathbf{s}$. Let $X(\mathbf{s}, t)$ denote the corresponding linked AOD value in the grid cell that includes $\mathbf{s}$. We also have a vector of p covariates $\mathbf{Z}(\mathbf{s}, t)$ that for land use and meteorology variables.

The statistical downscaler is given by

$$Y(\mathbf{s}, t) = \alpha_1(\mathbf{s}) + \beta_1(t) + \alpha_2(\mathbf{s})X(\mathbf{s}, t) + \beta_1(t)X(\mathbf{s}, t) + \mathbf{Z}(\mathbf{s}, t)\gamma + \epsilon(\mathbf{s}, t) \ ,$$

where γ is the vector of fixed-effect regression coefficients and $\epsilon(\mathbf{s}, t)$ is the residual white noise error, $\epsilon(\mathbf{s}, t) \overset{iid}{\sim} N(0, \sigma^2)$.

Spatial Random Effects

Coefficients $\alpha_1(\mathbf{s})$ and $\alpha_2(\mathbf{s})$ are spatially-varying intercept and slope. We assume

$$\begin{bmatrix} \alpha_1(\mathbf{s}) \\ \alpha_2(\mathbf{s}) \end{bmatrix} = \mathbf{A} \begin{bmatrix} W_1(\mathbf{s}) \\ W_2(\mathbf{s}) \end{bmatrix} = \begin{bmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} W_1(\mathbf{s}) \\ W_2(\mathbf{s}) \end{bmatrix} \tag{1}$$

where A_{11}, A_{21}, A_{22} are unknown constants. $W_1(\mathbf{s})$ and $W_2(\mathbf{s})$ are two *independent* zero-mean Gaussian spatial processes. For $i = 1, 2$, the spatial correlation function between location $\mathbf{s}$ and $\mathbf{s}'$ is

$$corr[\, W_i(\mathbf{s}), W_i(\mathbf{s}') \,] = \exp\left(\frac{1}{\rho_i}||\mathbf{s} - \mathbf{s}'||\right) \times R(||\mathbf{s} - \mathbf{s}'||; \delta_i) \ ,$$

where $||\mathbf{s} - \mathbf{s}'||$ is the Euclidean distance between $\mathbf{s}$ and $\mathbf{s}'$, ρ_i is the range parameter that controls the rate of decay, and $R(\,\cdot\,; \delta_i)$ is the Wendland tapering function that forces the spatial correlation to be zero when $||\mathbf{s} - \mathbf{s}'||$ exceeds the threshold δ_i.

Equation (1) entails that the covariance matrix between the spatial intercept and slope at the same location $\mathbf{s}$ is

$$Cov[\, \alpha_1(\mathbf{s}), \alpha_2(\mathbf{s}) \,] = \begin{bmatrix} A_{11}^2 & A_{11}A_{21} \\ A_{11}A_{21} & A_{21}^2 + A_{22}^2 \end{bmatrix} \ .$$

The covariance between intercepts or slopes at two locations are

$$Cov[\, \alpha_1(\mathbf{s}), \alpha_1(\mathbf{s}') \,] = A_{11}^2 \, Corr[\, W_1(\mathbf{s}), W_1(\mathbf{s}') \,]$$

$$Cov[\, \alpha_2(\mathbf{s}), \alpha_2(\mathbf{s}') \,] = A_{21}^2 \, Corr[\, W_1(\mathbf{s}), W_1(\mathbf{s}') \,] + A_{22}^2 \, Corr[\, W_2(\mathbf{s}), W_2(\mathbf{s}') \,] \ .$$

Temporal Random Effects

Coefficients $\beta_1(t)$ and $\beta_2(t)$ are modeled as two independent first-order random walk. Let T be the maximum number of days observed. The conditional distribution of $\beta_i(t)$ given the other days is assumed to be Normal with mean and variance given by

$$E[\beta_i(t')\,|\,t \neq t'] = \begin{cases} \psi_i \beta_i(t+1) & t' = 1 \\ \psi_i \dfrac{\beta_i(t-1)+\beta_i(t+1)}{2} & t' = 2,\ldots,T-1 \\ \psi_i \beta_i(t-1) & t' = T \end{cases},$$

$$Var[\beta_i(t')\,|\,t \neq t'] = \begin{cases} \tau_i^2 & t' = 1 \\ \tau_i^2/2 & t' = 2,\ldots,T-1 \\ \tau_i^2 & t' = T \end{cases}.$$

Let $\boldsymbol{\beta}_i(t)$ denote the $T \times 1$ vector of all temporal random intercepts ($i = 1$) or slopes ($i = 2$). The joint distribution of $\boldsymbol{\beta}_i(t)$ is multivariate Normal:

$$\boldsymbol{\beta}_i(t) \sim N\left(\mathbf{0},\ \tau_i^2[\mathbf{Q} - \psi_i \mathbf{W}]\right),$$

where $\mathbf{W}$ is a $T \times T$ adjacency matrix with the (i,j) element equal to 1 if $|i-j| = 1$ and 0 otherwise. $\mathbf{Q}$ is a diagonal matrix with the (i,i) element equal to the sum of row i of W.

Prior Distributions

The downscaler model includes the following unknown parameters: γ, A_{11}, A_{21}, A_{22}, ρ_1, ρ_2, τ_1, τ_2, ψ_1, ψ_2, and σ^2. The prior distributions are:

$$\gamma \propto 1$$

$$A_{11} \sim \text{Inv-Gamma}\,(a_{A1},\ b_{A1}) \qquad A_{22} \sim \text{Inv-Gamma}\,(a_{A2},\ b_{A2})$$

$$A_{21} \sim N(0, s_{A21}^2)$$

$$\rho_1 \sim \text{Uniform}(0,1) \qquad \rho_2 \sim \text{Uniform}(0,1)$$

$$\tau_1^2 \sim \text{Inv-Gamma}\,(a_{t1},\ b_{t1}) \qquad \tau_2^2 \sim \text{Inv-Gamma}\,(a_{t2},\ b_{t2})$$

$$\rho_1 \sim \text{Gamma}\,(a_{r1},\ b_{r1}) \qquad \rho_2 \sim \text{Gamma}\,(a_{r2},\ b_{r2})$$

$$\sigma^2 \sim \text{Inv-Gamma}\,(a_s,\ b_s)$$

Markov Chain Monte Carlo Algorithm

Let T denote the total number of days and S denote the total number of monitoring locations. For the i^{th} observation, $Y(\mathbf{s}_i, t_i)$, $X(\mathbf{s}_i, t_i)$, and $\mathbf{Z}(\mathbf{s}_i, t_i)$ represent, respectively, PM$_{2.5}$ level, AOD value, and other predictors at location $\mathbf{s}_i$ on day t_i, where $\mathbf{s}_i \in 1, 2, \ldots, S$, $t_i \in 1, 2, \ldots, T$, and $i = 1, 2, \ldots, N$.

Let $\mathbf{Y}$ denote the $N \times 1$ column vector of all observed monitoring data. Let $\mathbf{B}$ be the corresponding $N \times 1$ spatial effects $\mathbf{B}[i,] = [\alpha_1(\mathbf{s}_i) + \alpha_2(\mathbf{s}_i) X(\mathbf{s}_i, t_i)]$. Let $\mathbf{C}$ denote the $N \times 1$ vector of temporal intercept $\mathbf{C}[i,] = [\beta_1(\mathbf{s})]$. Let $\mathbf{D}$ denote the $N \times 1$ vector of temporal AOD effect $\mathbf{D}[i,] = [\beta_2(\mathbf{s}_i) X(\mathbf{s}_i, t_i)]$.

The MCMC algorithm, at iteration k, obtains posterior samples for all the unknown parameters through the following steps.

1. Update $\boldsymbol{\gamma}$.

   Let $\mathbf{Z}$ denote the design matrix including an overall intercept and the AOD values, and $\mathbf{R} = \mathbf{Y} - \mathbf{B} - \mathbf{C} - \mathbf{D}$. Sample $\boldsymbol{\gamma}^{(k)}$ from

$$\boldsymbol{\gamma}^{(k)} \,|\, \text{rest} \sim N\left((\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}\mathbf{R},\, \sigma^2 \mathbf{I} \right).$$

2. Update σ^2.

   Let $\mathbf{R} = \mathbf{Y} - \mathbf{B} - \mathbf{C} - \mathbf{D} - \mathbf{Z}\boldsymbol{\gamma}$. Sample $\sigma^{2,(k)}$ from

$$\sigma^{2,(k)} \,|\, \text{rest} \sim \text{Inv-Gamma}\left(N/2 + a_{t2},\ \mathbf{R}'\mathbf{R} + b_{t2}\right).$$

3. Update spatial random effects $\boldsymbol{\alpha}' = [\alpha_1(\mathbf{s}_1), \alpha_2(\mathbf{s}_1), \ldots, \alpha_1(\mathbf{s}_S), \alpha_2(\mathbf{s}_S)]$.

   Let $\mathbf{H}_1$ and $\mathbf{H}_2$ denote the covariance matrices for the latent processes $\mathbf{W}_1$ and $\mathbf{W}_2$ at monitoring locations calculated using parameters ρ_1 and ρ_2. Let $\mathbf{T}_j = \mathbf{a}_j \mathbf{a}_j'$, where $\mathbf{a}_j$ is the jth column of $\mathbf{A}$. Let $\mathbf{R}_s$ denote the residual vector $\mathbf{Y} - \mathbf{C} - \mathbf{D} - \mathbf{Z}\boldsymbol{\gamma}$ sorted by spatial location $\mathbf{s}$. Finally, denote $\tilde{\mathbf{X}}_s$ the $N \times 2S$ block diagonal matrix of the form

$$\tilde{\mathbf{X}}_s = \begin{bmatrix} \mathbf{X}_{\mathbf{s}_1} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_{\mathbf{s}_2} & \ldots & \mathbf{0} \\ \vdots & \vdots & \ldots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{X}_{\mathbf{s}_S} \end{bmatrix}.$$

   Sample $\boldsymbol{\alpha}^{(k)}$ from

$$\boldsymbol{\alpha}^{(k)} \,|\, \text{rest} \sim N\left(\mathbf{V}^{-1}\mathbf{M},\, \mathbf{V}^{-1} \right)$$

$$\mathbf{M} = \frac{1}{\sigma^2} \tilde{\mathbf{X}}' \mathbf{R}_s \qquad \mathbf{V} = \sigma^2 \tilde{\mathbf{X}}'\tilde{\mathbf{X}} + [\mathbf{H}_1 \otimes \mathbf{T}_1 + \mathbf{H}_2 \otimes \mathbf{T}_2]^{-1}.$$

4. Update temporal intercepts $\boldsymbol{\beta}_1' = [\beta_1(t_1), \beta_1(t_2), \ldots, \beta_1(t_T)]$.

Let $\mathbf{N}_s$ denote a diagonal matrix where the (i, i) element is the number of observations on day i. Define $\bar{\mathbf{R}}_t$ as the $T \times 1$ vector where the i^{th} element is the sum of the residuals of $\mathbf{Y} - \mathbf{B} - \mathbf{D} - \mathbf{Z}\gamma$ on day i.

Sample $\boldsymbol{\beta}_1^{(k)}$ from

$$\boldsymbol{\beta}_1^{(k)} \,|\, \text{rest} \sim N\left(\sigma^2 \left[\sigma^2 \mathbf{N}_s + \tau_1(\mathbf{D} - \psi_1 \mathbf{W})\right]^{-1} \bar{\mathbf{R}}_t,\ \left[\sigma^2 \mathbf{N}_s + \tau_1(\mathbf{D} - \psi_1 \mathbf{W})\right]^{-1}\right) .$$

5. Update temporal AOD slopes $\boldsymbol{\beta}_2' = [\beta_2(t_1), \beta_2(t_2), \ldots, \beta_2(t_T)]$.

   Denote $\tilde{\mathbf{X}}_t$ the vector of AOD values ordered by day t. Similarly, let $\mathbf{R}_t$ denote the residual $\mathbf{Y} - \mathbf{B} - \mathbf{C} - \mathbf{Z}\gamma$ ordered also by date. Sample $\boldsymbol{\beta}_2^{(k)}$ from

$$\boldsymbol{\beta}_2^{(k)} \,|\, \text{rest} \sim N\left(\sigma^2 \left[\sigma^2 \tilde{\mathbf{X}}_t' \tilde{\mathbf{X}}_t + \tau_2(\mathbf{D} - \psi_2 \mathbf{W})\right]^{-1} \tilde{\mathbf{X}}_t \mathbf{R}_t,\ \left[\sigma^2 \tilde{\mathbf{X}}_t' \tilde{\mathbf{X}}_t + \tau_2(\mathbf{D} - \psi_2 \mathbf{W})\right]^{-1}\right) .$$

6. Update temporal random effects parameters τ_1, τ_2, ψ_1, and ψ_2.

   For $j = 1, 2$, sample $\tau_j^{(k)}$ from

$$\tau_j^{(k)} \,|\, \text{rest} \sim \text{Inv-Gamma}\left(\frac{1}{2} + a_{tj}\,, \frac{1}{2}\boldsymbol{\beta}_j(\mathbf{Q} - \psi_j \mathbf{W})\boldsymbol{\beta}_j + b_{tj}\right) .$$

   We discretize the range of ψ_1 and ψ_2 into 1,000 equally-spaced values between [0,1]. The discrete prior results in discrete full conditional distributions. The full conditional probabilities are proportional to the likelihood

$$\begin{aligned} P(\psi_j^{(k)} = \psi^* \,|\, \text{rest}) &\propto |\mathbf{Q} - \psi^* \mathbf{W}|^{-1/2} \exp\left(-\frac{1}{2\tau_j^2}\boldsymbol{\beta}_j'(\mathbf{Q} - \psi^* \mathbf{W})\boldsymbol{\beta}_j\right) \\ &\propto |\mathbf{Q} - \psi^* \mathbf{W}|^{-1/2} \exp\left(\frac{\psi^*}{2\tau_j^2}\boldsymbol{\beta}_j' \mathbf{W} \boldsymbol{\beta}_j\right) . \end{aligned}$$

7. Update elements in matrix $\mathbf{A}$.

   Let $\tilde{\boldsymbol{\alpha}}_1 = \mathbf{H}_1^{-1/2}\boldsymbol{\alpha}_1$. Sample $A_{11}^{(k)}$ from

$$A_{11}^{(k)} \,|\, \text{rest} \sim \text{Inv-Gamma}\left(S/2 + a_{A1},\ \tilde{\boldsymbol{\alpha}}_1' \tilde{\boldsymbol{\alpha}}_1 + b_{A1}\right) .$$

   Let $\tilde{\boldsymbol{\alpha}}_2 = \mathbf{H}_1^{-1/2}[\boldsymbol{\alpha}_2 - A_{21}/A_{11}\boldsymbol{\alpha}_1]$. Sample $A_{22}^{(k)}$ from

$$A_{22}^{(k)} \,|\, \text{rest} \sim \text{Inv-Gamma}\left(S/2 + a_{A2},\ \tilde{\boldsymbol{\alpha}}_2' \tilde{\boldsymbol{\alpha}}_2 + b_{A2}\right) .$$

   Element A_{21} is updated using random-walk Metropolis-Hastings. First generate a proposal $A_{21}^{(k)}$ from proposal distribution $N(A_{21}^{(k-1)}, \kappa)$, where κ is the tuning parameter.

Accept $A_{21}^{(k)}$ with probability

$$\frac{|\mathbf{H}_2|^{-1/2}\exp\left([\boldsymbol{\alpha}_2 - A_{21}^{(k)}/A_{11}\boldsymbol{\alpha}_1]'\mathbf{H}_2^{-1}[\boldsymbol{\alpha}_2 - A_{21}^{(k)}/A_{11}\boldsymbol{\alpha}_1]\right)}{|\mathbf{H}_2|^{-1/2}\exp\left([\boldsymbol{\alpha}_2 - A_{21}^{(k-1)}/A_{11}\boldsymbol{\alpha}_1]'\mathbf{H}_2^{-1}[\boldsymbol{\alpha}_2 - A_{21}^{(k-1)}/A_{11}\boldsymbol{\alpha}_1]\right)}.$$

8. Update spatial range parameter ρ_1 and ρ_2.

   Generate proposal $\rho_1^{(k)}$ from a log-normal distribution with mean $\rho_1^{(k-1)}$ and variance κ_1. Accept $\rho_1^{(k)}$ with probability

$$\frac{f_1\left(\boldsymbol{\alpha}/A_{11}\,|\,\mathbf{0},\mathbf{H}_1(\rho_1^{(k)})\,|\right)\times f_2(\rho_1^{(k)}\,|\,a_{r1},b_{r1})\times\rho_1^{(k)}}{f_1\left(\boldsymbol{\alpha}/A_{11}\,|\,\mathbf{0},\mathbf{H}_1(\rho_1^{(k-1)})\right)\times f_2(\rho_1^{(k-1)}\,|\,a_{r1},b_{r1})\times\rho_1^{(k-1)}},$$

   where $f_1(\cdot)$ is the multivariate normal density for the data likelihood, and $f_2(\cdot)$ is the univariate Gamma density for the prior likelihood.

   Parameter $\rho_2^{(k)}$ is obtained similarly with log-normal proposal distribution with mean $\rho_2^{(k-1)}$ and variance κ_2. The acceptance probability is

$$\frac{f_1\left(\boldsymbol{\alpha}_2 - A_{21}^{(k)}/A_{11}\boldsymbol{\alpha}_1\,|\,\mathbf{0},\mathbf{H}_2(\rho_2^{(k)})\,|\right)\times f_2(\rho_2^{(k)}\,|\,a_{r2},b_{r2})\times\rho_2^{(k)}}{f_1\left(\boldsymbol{\alpha}_2 - A_{21}^{(k)}/A_{11}\boldsymbol{\alpha}_1\,|\,\mathbf{0},\mathbf{H}_2(\rho_2^{(k-1)})\right)\times f_2(\rho_2^{(k-1)}\,|\,a_{r2},b_{r2})\times\rho_2^{(k-1)}}.$$