Bayesian Analysis of Spatially-Dependent Functional Responses with Spatially-Dependent Multi-Dimensional Functional Predictors

Wen-Hsi Yang[1], Christopher K. Wikle[2], Scott H. Holan[2], D. Brenton Myers[2] and Kenneth A. Sudduth[3]

[1] *CSIRO,* [2] *University of Missouri and* [3] *USDA-ARS-Cropping Systems and Water Quality Unit*

Supplementary Material

This online supplement provides a brief overview of the model formulation, full-conditional distributions for the MCMC sampling algorithm, model choices, sensitivity analysis, and simulation study in support of Yang et al. (2014). As described in the paper, modeling choices and hyperparameter specification are application specific.

S1 Model Formulation

Recall that the data model for the functional response variables is given by

$$\widetilde{\mathbf{Y}} = \mathbf{H}_{(y)}\mathbf{W}\mathbf{A}\boldsymbol{\Psi} + \mathbf{E}_{\widetilde{y}},$$

where $\mathbf{E}_{\widetilde{y}} \sim N_{n_r,n_d}(\mathbf{0}, \boldsymbol{\Sigma}_{y,n_r}, \boldsymbol{\Sigma}_{y,n_d})$ with $\boldsymbol{\Sigma}_{y,n_r} = \text{diag}(\boldsymbol{\tau}_{y,n_r})$ and $\boldsymbol{\Sigma}_{y,n_d} = \text{diag}(\boldsymbol{\tau}_{y,n_d})$. Note, we also specify inverse gamma priors for $\tau_{y,r}$, $r = 1, \ldots, n_r$, and $\tau_{y,d}$, $d = 1, \ldots, n_d$, where $\{\kappa_{y,r}\}$, $\{\nu_{y,r}\}$, $\{\kappa_{y,d}\}$, and $\{\nu_{y,d}\}$ are hyperparameters. That is, $\tau_{y,r} \sim IG(\kappa_{y,r}, \nu_{y,r})$, $r = 1, \ldots, n_r$, and $\tau_{y,d} \sim IG(\kappa_{y,d}, \nu_{y,d})$, $d = 1, \ldots, n_d$, with associated hyperparameters $\{\kappa_{y,r}\}$, $\{\nu_{y,r}\}$, $\{\kappa_{y,d}\}$, and $\{\nu_{y,d}\}$.

Additionally, recall that the data model for the functional image covariates is specified as

$$\widetilde{\mathbf{X}}_j = \mathbf{H}_{(x)}\mathbf{W}\mathbf{F}_j\boldsymbol{\Phi} + \mathbf{E}_{\widetilde{x}_j}, \quad j = 1, \ldots, J,$$

where $\mathbf{E}_{\widetilde{x}_j} \sim N_{n_r,n_u}(\mathbf{0}, \boldsymbol{\Sigma}_{x_j,n_r}, \boldsymbol{\Sigma}_{x_j,n_u})$, with $\boldsymbol{\Sigma}_{x_j,n_r} = \text{diag}(\boldsymbol{\tau}_{x_j,n_r})$ and $\boldsymbol{\Sigma}_{x_j,n_u} = \text{diag}(\boldsymbol{\tau}_{x_j,n_u})$. Further, we specify inverse gamma priors for $\tau_{x_j,r}$, $r = 1, \ldots, n_r$, and $\tau_{x_j,u}$, $u = 1, \ldots, n_u$, where $\{\kappa_{x_j,r}\}$, $\{\nu_{x_j,r}\}$, $\{\kappa_{x_j,u}\}$, and $\{\nu_{x_j,u}\}$ are hyperparameters. That is, $\tau_{x_j,r} \sim IG(\kappa_{x_j,r}, \nu_{x_j,r})$, $r = 1, \ldots, n_r$, and $\tau_{x_j,u} \sim IG(\kappa_{x_j,u}, \nu_{x_j,d})$, $u = 1, \ldots, n_u$, with associated hyperparameters $\{\kappa_{x_j,r}\}$, $\{\nu_{x_j,r}\}$, $\{\kappa_{x_j,u}\}$, and $\{\nu_{x_j,u}\}$.

Next, recall the data model of covariate matrix is given by,

$$\widetilde{\mathbf{Z}} = \mathbf{H}_{(z)}\mathbf{W}\mathbf{Q} + \mathbf{E}_{\widetilde{z}},$$

where $\mathbf{E}_{\widetilde{z}} \sim N_{n_r,n_p}(\mathbf{0}, \mathbf{\Sigma}_{z,n_r}, \mathbf{\Sigma}_{z,n_p})$, with $\mathbf{\Sigma}_{z,n_r} = \text{diag}(\boldsymbol{\tau}_{z,n_r})$ and $\mathbf{\Sigma}_{z,n_p} = \text{diag}(\boldsymbol{\tau}_{z,n_p})$. Further, for $r = 1, \ldots, n_r$ and $p = 1, \ldots, n_p$, we specify $\tau_{z,r} \sim IG(\kappa_{z,r}, \nu_{z,r})$ and $\tau_{z,p} \sim IG(\kappa_{z,p}, \nu_{z,p})$, where $\{\kappa_{z,r}\}$, $\{\nu_{z,r}\}$, $\{\kappa_{z,p}\}$, and $\{\nu_{z,p}\}$ are associated hyperparameters.

Lastly, we recall the process model is given by

$$\mathbf{A} = \sum_{j=1}^{J} \mathbf{F}_j \mathbf{B}_j + \mathbf{Q}\mathbf{H} + \mathbf{G},$$

where $\mathbf{G} \sim N_{n_\ell,n_i}(\mathbf{0}, \mathbf{\Sigma}_{n_\ell}, \mathbf{\Sigma}_{n_i})$. We specify matrix normal priors for $\mathbf{F}_j$ and $\mathbf{Q}$, given by $\mathbf{F}_j \sim N_{n_\ell,n_{k_j}}(\boldsymbol{\mu}_{\mathbf{F}_j}, \mathbf{\Sigma}_{\mathbf{F}_j,n_\ell}, \mathbf{\Sigma}_{\mathbf{F}_j,n_{k_j}})$ and $\mathbf{Q} \sim N_{n_\ell,n_p}(\boldsymbol{\mu}_{\mathbf{Q}}, \mathbf{\Sigma}_{\mathbf{Q},n_\ell}, \mathbf{\Sigma}_{\mathbf{Q},n_p})$. Moreover, we also specify Wishart priors for various precision matrices. That is, $\mathbf{\Sigma}^{-1}_{\mathbf{F}_j,n_\ell} \sim W_{n_\ell}(\mathbf{V}_{\mathbf{F}_j,n_\ell}, v_{\mathbf{F}_j,n_\ell})$, $\mathbf{\Sigma}^{-1}_{\mathbf{F}_j,n_{k_j}} \sim W_{n_{k_j}}(\mathbf{V}_{\mathbf{F}_j,n_{k_j}}, v_{\mathbf{F}_j,n_{k_j}})$, $\mathbf{\Sigma}^{-1}_{\mathbf{Q},n_\ell} \sim W_{n_\ell}(\mathbf{V}_{\mathbf{Q},n_\ell}, v_{\mathbf{Q},n_\ell})$, and $\mathbf{\Sigma}^{-1}_{\mathbf{Q},n_p} \sim W_{n_p}(\mathbf{V}_{\mathbf{Q},n_p}, v_{\mathbf{Q},n_p})$, where $\mathbf{V}_{\mathbf{F}_j,n_\ell}$, $\mathbf{V}_{\mathbf{F}_j,n_{k_j}}$, $\mathbf{V}_{\mathbf{Q},n_\ell}$ and $\mathbf{V}_{\mathbf{Q},n_p}$ are given scale matrices and $v_{\mathbf{F}_j,n_\ell}$, $v_{\mathbf{F}_j,n_{k_j}}$, $v_{\mathbf{Q},n_\ell}$, and $v_{\mathbf{Q},n_p}$ are associated values for the degrees of freedom. For notational simplicity, we denote $\boldsymbol{\chi} = [\mathbf{F}_1, \ldots, \mathbf{F}_J, \mathbf{Q}]$ and $\mathbf{B} = [\mathbf{B}'_1, \ldots, \mathbf{B}'_J, \mathbf{H}']'$ and represent the above equation as follows

$$\mathbf{A} = \boldsymbol{\chi}\mathbf{B} + \mathbf{G}.$$

Typically, $\mathbf{B}$ is specified as a matrix normal distribution; i.e., $\mathbf{B} \sim N_{P,n_i}(\mathbf{0}, \mathbf{\Sigma}_{\mathbf{B}}, \mathbf{\Sigma}_{n_i})$, where $\mathbf{\Sigma}_{\mathbf{B}} = \tau_{\mathbf{B}}\mathbf{I}_P$ with $\tau_{\mathbf{B}}$ an associated hyperparameter and $\mathbf{I}_P$ a P-dimensional identity matrix. Again, we specify Wishart priors for various precision matrices; i.e., $\mathbf{\Sigma}^{-1}_{n_\ell} \sim W_{n_\ell}(\mathbf{V}_{G,n_\ell}, v_{G,n_\ell})$ and $\mathbf{\Sigma}^{-1}_{n_i} \sim W_{n_i}(\mathbf{V}_{G,n_i}, v_{G,n_i})$, where $\mathbf{V}_{G,n_\ell}$ and $\mathbf{V}_{G,n_i}$ are given scale matrices and v_{G,n_ℓ} and v_{G,n_i} are associated values for the degrees of freedom.

S2 MCMC: Full Conditional Distributions

The implementation of the Bayesian hierarchical model given here is fairly straightforward given that many of the modeling choices concerning rank reductions and prior distribution forms were selected to facilitate computation in a "big data" environment. The associated Gibbs sampling algorithm full conditional distributions is presented here. Implementation is application specific, with decisions having to be made concerning the measurement uncertainty of the responses and covariates, choice of basis functions for the various decompositions, choice of basis function truncation, and the specification of specific forms for the prior means and variance/covariance matrices, as well as the associated hyperparameters.

The full conditional distribution of $\mathbf{B}$ is a matrix variate normal distribution given by

$$\mathbf{B}|\cdot \sim N_{P,n_i}(\mathbf{K}\mathbf{M}, \mathbf{K}, \mathbf{\Sigma}_{n_i}),$$

with

$$\begin{aligned}\mathbf{K} &= \left(\boldsymbol{\chi}'\mathbf{\Sigma}^{-1}_{n_\ell}\boldsymbol{\chi} + \mathbf{\Sigma}^{-1}_{B}\right)^{-1}, \\ \mathbf{M} &= \boldsymbol{\chi}'\mathbf{\Sigma}^{-1}_{n_\ell}\mathbf{A},\end{aligned}$$

where $\mathbf{\Sigma}_B$ is a diagonal matrix with elements equal to $\tau_{\mathbf{B}}$.

The full conditional distribution of $\mathbf{\Sigma}_{n_\ell}^{-1}$ is

$$\mathbf{\Sigma}_{n_\ell}^{-1}|\cdot \sim W_{n_\ell}\left(\left\{(\mathbf{A}-\boldsymbol{\chi}\mathbf{B})\mathbf{\Sigma}_{n_i}^{-1}(\mathbf{A}-\boldsymbol{\chi}\mathbf{B})' + \mathbf{V}_{G,n_\ell}^{-1}\right\}^{-1}, n_i + v_{G,n_\ell}\right).$$

The full conditional distribution of $\mathbf{\Sigma}_{n_i}^{-1}$ is

$$\mathbf{\Sigma}_{n_i}^{-1}|\cdot \sim W_{n_i}\left(\left\{(\mathbf{A}-\boldsymbol{\chi}\mathbf{B})'\mathbf{\Sigma}_{n_\ell}^{-1}(\mathbf{A}-\boldsymbol{\chi}\mathbf{B}) + \mathbf{B}'\mathbf{\Sigma}_{\mathbf{B}}^{-1}\mathbf{B} + \mathbf{V}_{G,n_i}^{-1}\right\}^{-1}, n_\ell + P + v_{G,n_i}\right).$$

The full conditional distribution of $\text{vec}(\mathbf{A})$ is given by

$$\text{vec}(\mathbf{A})|\cdot \sim MVN(\mathbf{KM}, \mathbf{K}),$$

with

$$\begin{aligned}\mathbf{K} &= \left\{(\mathrm{I}_{n_i}\otimes\mathbf{H}_{(y)}\mathbf{W})'(\mathbf{\Sigma}_{y,n_i}\otimes\mathbf{\Sigma}_{y,n_r})^{-1}(\mathrm{I}_{n_i}\otimes\mathbf{H}_{(y)}\mathbf{W}) + (\mathbf{\Sigma}_{n_i}\otimes\mathbf{\Sigma}_{n_\ell})^{-1}\right\}^{-1},\\ \mathbf{M} &= (\mathrm{I}_{n_i}\otimes\mathbf{H}_{(y)}\mathbf{W})'(\mathbf{\Sigma}_{y,n_i}\otimes\mathbf{\Sigma}_{y,n_r})^{-1}\text{vec}(\widetilde{\mathbf{Y}}_\Psi) + (\mathbf{\Sigma}_{n_i}\otimes\mathbf{\Sigma}_{n_\ell})^{-1}\text{vec}(\boldsymbol{\chi}\mathbf{B}),\end{aligned}$$

where $\widetilde{\mathbf{Y}}_\Psi = \widetilde{\mathbf{Y}}\Psi'$ and $\mathbf{\Sigma}_{y,n_i} = \Psi\mathbf{\Sigma}_{y,n_d}\Psi'$.

For $r = 1,\ldots,n_r$, the full conditional distribution of the element $\tau_{y,r}$ of $\text{diag}(\boldsymbol{\tau}_{y,n_r})$ is

$$\tau_{y,r}|\cdot \sim IG\left(\frac{n_d + 2\kappa_{y,r}}{2}, \frac{\text{tr}(\mathbf{\Sigma}_{y,n_d}^{-1}(\widetilde{\mathbf{Y}} - \mathbf{H}_{(y)}\mathbf{WA}\mathbf{\Psi})'\mathbf{O}_r(\widetilde{\mathbf{Y}} - \mathbf{H}_{(y)}\mathbf{WA})\mathbf{\Psi}) + 2\nu_{y,r}}{2}\right),$$

where $\kappa_{y,r}$ and $\nu_{y,r}$ are hyperparameters and $\mathbf{O}_r$ is an $n_r \times n_r$ matrix with the rth diagonal element equal to one and the rest of the elements equal to zero.

For $d = 1,\ldots,n_d$, the full conditional distribution of the element $\tau_{y,d}$ of $\text{diag}(\boldsymbol{\tau}_{y,n_d})$ is

$$\tau_{y,d}|\cdot \sim IG\left(\frac{n_r + 2\kappa_{y,d}}{2}, \frac{\text{tr}(\mathbf{\Sigma}_{y,n_r}^{-1}(\widetilde{\mathbf{Y}} - \mathbf{H}_{(y)}\mathbf{WA}\mathbf{\Psi})\mathbf{O}_d(\widetilde{\mathbf{Y}} - \mathbf{H}_{(y)}\mathbf{WA})\mathbf{\Psi})' + 2\nu_{y,d}}{2}\right),$$

where $\kappa_{y,d}$ and $\nu_{y,d}$ are hyperparameters and $\mathbf{O}_d$ is an $n_d \times n_d$ matrix with the dth diagonal element equal to one and the rest of the elements equal to zero.

The full conditional distribution of $\text{vec}(\mathbf{F}_j)$ is

$$\text{vec}(\mathbf{F}_j)|\cdot \sim MVN(\mathbf{KM}, \mathbf{K}),$$

with

$$\begin{aligned}
\mathbf{K} &= \left\{\mathbf{L}'(\boldsymbol{\Sigma}_{x_j,n_{k_j}} \otimes \boldsymbol{\Sigma}_{x_j,n_r})^{-1}\mathbf{L} + (\mathbf{B}_j' \otimes \mathrm{I}_{n_\ell})'(\boldsymbol{\Sigma}_{n_i} \otimes \boldsymbol{\Sigma}_{n_\ell})^{-1}(\mathbf{B}_j' \otimes \mathrm{I}_{n_\ell}) + (\boldsymbol{\Sigma}_{\mathbf{F}_j,n_{k_j}} \otimes \boldsymbol{\Sigma}_{\mathbf{F}_j,n_\ell})^{-1}\right\}^{-1}, \\
\mathbf{M} &= \mathbf{L}'(\boldsymbol{\Sigma}_{x_j,n_{k_j}} \otimes \boldsymbol{\Sigma}_{x_j,n_r})^{-1}\mathrm{vec}(\widetilde{\mathbf{X}}_{j,\Phi}) + (\mathbf{B}_j' \otimes \mathrm{I}_{n_\ell})'(\boldsymbol{\Sigma}_{n_i} \otimes \boldsymbol{\Sigma}_{n_\ell})^{-1}\mathrm{vec}(\mathbf{A} - \boldsymbol{\chi}_{-j}\mathbf{B}_{-j}) \\
&\quad + (\boldsymbol{\Sigma}_{\mathbf{F}_j,n_{k_j}} \otimes \boldsymbol{\Sigma}_{\mathbf{F}_j,n_\ell})^{-1}\mathrm{vec}(\boldsymbol{\mu}_{\mathbf{F}_j}), \\
\mathbf{L} &= (\mathrm{I}_{n_{k_j}} \otimes \mathbf{H}_{(x)}\mathbf{W}),
\end{aligned}$$

where $\widetilde{\mathbf{X}}_{j,\Phi} = \widetilde{\mathbf{X}}_j\Phi'$, $\boldsymbol{\Sigma}_{x_j,n_{k_j}} = \Phi\boldsymbol{\Sigma}_{x_j,n_u}\Phi'$, and $\boldsymbol{\chi}_{-j}$ and $\mathbf{B}_{-j}$ exclude the components corresponding to $\widetilde{\mathbf{X}}_j$ and $\boldsymbol{\mu}_{\mathbf{F}_j}$, $\boldsymbol{\Sigma}_{\mathbf{F}_j,n_{k_j}}$, and $\boldsymbol{\Sigma}_{\mathbf{F}_j,n_\ell}$ are hyperparameters.

For $r = 1, \ldots, n_r$, the full conditional distribution of the element $\tau_{x_j,r}$ of $\mathrm{diag}(\boldsymbol{\tau}_{x_j,n_r})$ is

$$\tau_{x_j,r}|\cdot \sim IG\left(\frac{n_{k_j} + 2\kappa_{x_j,r}}{2}, \frac{\mathrm{tr}(\boldsymbol{\Sigma}_{x_j,n_{k_j}}^{-1}(\widetilde{\mathbf{X}}_{j,\Phi} - \mathbf{H}_{(x)}\mathbf{W}\mathbf{F}_j)'\mathbf{O}_r(\widetilde{\mathbf{X}}_{j,\Phi} - \mathbf{H}_{(x)}\mathbf{W}\mathbf{F}_j) + 2\nu_{x_j,r}}{2}\right),$$

where $\kappa_{x_j,r}$ and $\nu_{x_j,r}$ are hyperparameters and $\mathbf{O}_r$ is an $n_r \times n_r$ matrix with the rth diagonal element equal to one and the rest of the elements equal to zero.

For $u = 1, \ldots, n_u$, the full conditional distribution of the element $\tau_{x,u}$ of $\mathrm{diag}(\boldsymbol{\tau}_{y,n_u})$ is

$$\tau_{y,u}|\cdot \sim IG\left(\frac{n_r + 2\kappa_{x,u}}{2}, \frac{\mathrm{tr}(\boldsymbol{\Sigma}_{x,n_r}^{-1}(\widetilde{\mathbf{X}} - \mathbf{H}_{(x)}\mathbf{W}\mathbf{F}_j\boldsymbol{\Phi})\mathbf{O}_u(\widetilde{\mathbf{X}} - \mathbf{H}_{(x)}\mathbf{W}\mathbf{F}_j)\boldsymbol{\Phi})' + 2\nu_{x,u}}{2}\right),$$

where $\kappa_{x,u}$ and $\nu_{x,u}$ are hyperparameters and $\mathbf{O}_u$ is an $n_u \times n_u$ matrix with the uth diagonal element equal to one and the rest of the elements equal to zero.

The full conditional distribution of $\mathrm{vec}(\mathbf{Q})$ is given by

$$\mathrm{vec}(\mathbf{Q})|\cdot \sim MVN(\mathbf{KM}, \mathbf{K}),$$

with

$$\begin{aligned}
\mathbf{K} &= \left\{\mathbf{L}'(\boldsymbol{\Sigma}_{z,n_p} \otimes \boldsymbol{\Sigma}_{z,n_r})^{-1}\mathbf{L} + (\mathbf{H}' \otimes \mathrm{I}_{n_\ell})'(\boldsymbol{\Sigma}_{n_i} \otimes \boldsymbol{\Sigma}_{n_\ell})^{-1}(\mathbf{H}' \otimes \mathrm{I}_{n_\ell}) + (\boldsymbol{\Sigma}_{\mathbf{Q},n_p} \otimes \boldsymbol{\Sigma}_{\mathbf{Q},n_\ell})^{-1}\right\}^{-1}, \\
\mathbf{M} &= \mathbf{L}'(\boldsymbol{\Sigma}_{z,n_p} \otimes \boldsymbol{\Sigma}_{z,n_r})^{-1}\mathrm{vec}(\widetilde{\mathbf{Z}}) + (\mathbf{H}' \otimes \mathrm{I}_{n_\ell})'(\boldsymbol{\Sigma}_{n_i} \otimes \boldsymbol{\Sigma}_{n_\ell})^{-1}\mathrm{vec}(\mathbf{A} - \boldsymbol{\chi}_{-Q}\mathbf{B}_{-Q}) \\
&\quad + (\boldsymbol{\Sigma}_{\mathbf{Q},n_p} \otimes \boldsymbol{\Sigma}_{\mathbf{Q},n_\ell})^{-1}\mathrm{vec}(\boldsymbol{\mu}_{\mathbf{Q}}), \\
\mathbf{L} &= (\mathrm{I}_{n_p} \otimes \mathbf{H}_{(z)}\mathbf{W}),
\end{aligned}$$

where $\boldsymbol{\chi}_{-Q}$ and $\mathbf{B}_{-Q}$ exclude the components corresponding to $\widetilde{\mathbf{Q}}$ and $\boldsymbol{\mu}_{\mathbf{Q}}$, $\boldsymbol{\Sigma}_{\mathbf{Q},n_p}$, and $\boldsymbol{\Sigma}_{\mathbf{Q},n_\ell}$ are associated hyperparameters.

For $r = 1, \ldots, n_r$, the full conditional distribution of the element $\tau_{z,r}$ of $\text{diag}(\boldsymbol{\tau}_{z,n_r})$ is given by

$$\tau_{z,r}|\cdot \sim IG\left(\frac{n_p + 2\kappa_{z,r}}{2}, \frac{\text{tr}(\boldsymbol{\Sigma}_{z,n_r}^{-1}(\widetilde{\mathbf{Z}} - \mathbf{H}_{(z)}\mathbf{WQ})'\mathbf{O}_r(\widetilde{\mathbf{Z}} - \mathbf{H}_{(z)}\mathbf{WQ}) + 2\nu_{z,r}}{2}\right),$$

where $\kappa_{z,r}$ and $\nu_{z,r}$ are hyperparameters and $\mathbf{O}_r$ is an $n_r \times n_r$ matrix with the rth diagonal element equal to one and the rest of the elements equal to zero.

For $p = 1, \ldots, n_p$, the full conditional of the element $\tau_{z,p}$ of $\text{diag}(\boldsymbol{\tau}_{z,n_p})$ is

$$\tau_{z,p}|\cdot \sim IG\left(\frac{n_r + 2\kappa_{z,p}}{2}, \frac{\text{tr}(\boldsymbol{\Sigma}_{z,n_r}^{-1}(\widetilde{\mathbf{Z}} - \mathbf{H}_{(z)}\mathbf{WQ})\mathbf{O}_p(\widetilde{\mathbf{Z}} - \mathbf{H}_{(z)}\mathbf{WQ})' + 2\nu_{z,p}}{2}\right),$$

where $\kappa_{z,p}$ and $\nu_{z,p}$ are hyperparameters and $\mathbf{O}_p$ is an $n_p \times n_p$ matrix with the pth diagonal element equal to one and the rest of the elements equal to zero.

The full conditional of $\boldsymbol{\Sigma}_{\mathbf{F}_j,n_\ell}^{-1}$ is given by

$$\boldsymbol{\Sigma}_{\mathbf{F}_j,n_\ell}^{-1}|\cdot \sim W_{n_\ell}\left(\left\{(\mathbf{F}_j - \boldsymbol{\mu}_{\mathbf{F}_j})\boldsymbol{\Sigma}_{\mathbf{F}_j,n_{k_j}}^{-1}(\mathbf{F}_j - \boldsymbol{\mu}_{\mathbf{F}_j})' + \mathbf{V}_{\mathbf{F}_j,n_\ell}^{-1}\right\}^{-1}, n_{k_j} + v_{\mathbf{F}_j,n_\ell}\right).$$

The full conditional of $\boldsymbol{\Sigma}_{\mathbf{F}_j,n_{k_j}}^{-1}$ is given by

$$\boldsymbol{\Sigma}_{\mathbf{F}_j,n_{k_j}}^{-1}|\cdot \sim W_{n_{k_j}}\left(\left\{(\mathbf{F}_j - \boldsymbol{\mu}_{\mathbf{F}_j})'\boldsymbol{\Sigma}_{\mathbf{F}_j,n_\ell}^{-1}(\mathbf{F}_j - \boldsymbol{\mu}_{\mathbf{F}_j}) + \mathbf{V}_{\mathbf{F}_j,n_{k_j}}^{-1}\right\}^{-1}, n_\ell + v_{\mathbf{F}_j,n_{k_j}}\right).$$

The full conditional of $\boldsymbol{\Sigma}_{\mathbf{Q},n_\ell}^{-1}$ is given by

$$\boldsymbol{\Sigma}_{\mathbf{Q},n_\ell}^{-1}|\cdot \sim W_{n_\ell}\left(\left\{(\mathbf{Q} - \boldsymbol{\mu}_{\mathbf{Q}})\boldsymbol{\Sigma}_{\mathbf{Q},n_p}^{-1}(\mathbf{Q} - \boldsymbol{\mu}_{\mathbf{Q}})' + \mathbf{V}_{\mathbf{Q},n_\ell}^{-1}\right\}^{-1}, n_p + v_{\mathbf{Q},n_\ell}\right).$$

The full conditional of $\boldsymbol{\Sigma}_{\mathbf{Q},n_p}^{-1}$ is given by

$$\boldsymbol{\Sigma}_{\mathbf{Q},n_p}^{-1}|\cdot \sim W_{n_p}\left(\left\{(\mathbf{Q} - \boldsymbol{\mu}_{\mathbf{Q}})'\boldsymbol{\Sigma}_{\mathbf{Q},n_\ell}^{-1}(\mathbf{Q} - \boldsymbol{\mu}_{\mathbf{Q}}) + \mathbf{V}_{\mathbf{Q},n_p}^{-1}\right\}^{-1}, n_\ell + v_{\mathbf{Q},n_p}\right).$$

S3 Model Choices and Sensitivity Analysis

This section describes how the basis function and the specific hyperparameters used in our application were chosen. Many sensitivity analyses were run to facilitate these decisions as indicated below.

S3.1 Choice of Basis Functions

The major model choices in this framework are concerned with the basis functions – both the form and the number of functions. We chose the Karhunen-Loève decomposition for the image and ECp depth functions because of their efficiency in representing a large amount of variation using a relatively small number of functions. This is the basis for functional principal components analysis, which is common in functional data analysis (Ramsay and Silverman, 2005), as well as the basis for empirical orthogonal function (EOF) analysis in spatial and spatio-temporal statistics (see Cressie and Wikle (2011)). In practice we can calculate eigenvalues and eigenfunctions (i.e., EOFs) from the symmetric decomposition of the empirical covariance matrix of the ECp profiles as well as the depth/wavelength VNIR images. In our application, the first 5, 10 and 15 EOFs associated with the image covariates account for approximately 81.1%, 93.8% and 97.2% of the total variation, respectively. Additionally, the first 5, 10 and 15 EOFs of the ECp depth profiles account for approximately 98.7%, 98.8% and 99.9% of the total variation, respectively.

Although it is common to use EOFs from the singular value decomposition (SVD) in spatial statistics applications, it can be important in functional applications to consider the robustness of these estimates, since they are empirically derived (e.g., Bali et al. (2011)). Thus, we did a sensitivity analysis whereby we considered the SVD-based EOFs as well as those easily obtained from robust methods in the literature. Specifically, for ECp depth profiles, we used the "fast covariance estimation" (FACE) procedure and a two-step estimation procedure (SVDS) (Xiao et al. (2013b)). Both methods were implemented via bivariate P-splines, sometimes called a sandwich smoother (Xiao et al. (2013a)). In essence, FACE smooths both the observed curves and the corresponding sample covariance function to regularize the eigenvector estimates, whereas SVDS uses singular value decomposition to decompose the data matrix and then smooths the corresponding eigenvectors. These analyses (not shown) suggested very little difference in the first 10 eigenvectors obtained via basic SVD (EOFs), FACE, and SVDS.

For the image covariates, the FACE procedure is problematic since it is designed for one-dimensional functions, and the images are vectorized, leading to discontinuities in the vectorized "curve" that are over smoothed by the FACE procedure. Thus, we replaced this with population value decomposition (PVD) (Crainiceanu et al. (2011)), which can obtain the associated eigenvectors for the images directly (without vectorization). This procedure produces dominant eigenvectors for the image covariates that are nearly identical to the unregularized SVD (i.e., EOFs) and SVDS procedures. Although there is very little visual difference between the eigenvector basis functions obtained by these various procedures, we conducted an analysis to ensure that the spatial predictions were not sensitive to these choices. In particular, we considered five scenarios: (1) the basis functions for the image covariates and depth profiles are both EOFs; (2) the basis functions of image covariates and depth profile are obtained both obtained via SVDS; (3) the basis functions of image covariates and depth profiles are obtained via PVD and SVDS, respectively; (4) the basis functions of image covariates and depth profiles are obtained via SVDS and FACE, respectively; (5) the basis functions of image covariates and depth profiles are obtained via PVD and FACE, respectively. These results (not

shown) suggested no discernible difference in the spatial predictions obtained from our methodology. Consequently, we used the simpler EOF basis functions for the results presented herein. The number of EOF basis functions used in the analysis is selected based on sensitivity analysis as described in S3.3.

To predict the entire field for a given depth (including unobserved sites), we used reduced-rank spatial basis functions to facilitate prediction on large spatial domains. We considered a discrete kernel convolution approach (e.g., Berry and Ver Hoef (1996); Higdon (1998)) using bivariate Gaussian kernels. In general, we considered a bivariate Gaussian kernel with a diagonal covariance matrix $\sigma_g^2\mathbf{I}_2$, where $\mathbf{I}_2$ is a two-dimensional identity matrix giving an equal weight to all direction since we had no priori knowledge that there is anisotropy in the ECp profiles. However, if one has scientific knowledge supporting an anisotropic spatial dependence, a non-diagonal Gaussian kernel covariate matrix can be considered. Typically, it is more important to choose the number of support points (i.e., knots) n_ℓ in the discrete kernel convolution approach. We consider the criteria suggested in Ruppert et al. (2003) for splines,

$$n_\ell = \max\{20, \min(n_r/4, 150)\}.$$

In particular, since $n_r = 26$ in our application, we took $n_\ell = 20$. Further, we placed the knots in the domain by using a space filling design (Nychka and Saltzman (1998)) to select knots from a candidate set given by our 26 observed locations as shown in the paper. We note that other basis functions could be chosen, such as thin-plate splines (Ruppert et al. (2003, Chapter 13)) or a predictive process (Banerjee et al. (2008)). Typically, these various methods give similar results when used for spatial prediction. In the implementation, we standardized both easting and northing coordinates to facilitate numerical stability. The average distance between knot locations was then about 1.8 scaled units. The choice of the Gaussian kernel width (given by σ_g^2) was determined by sensitivity analysis (as described in S3.3 below).

S3.2 Covariance Matrix Hyperparameters

Data Model Parameters

Consider the specification of the hyperparameters associated with the data model measurement error and truncation covariance matrices. For empirical basis expansions, the truncation error can be fairly well-identified by prior data analysis given the specific choice of basis functions and the number of components in the expansion, which are ultimately chosen by sensitivity analysis as given below. Further, the measurement/nugget uncertainty is in some cases well known from knowledge of the instruments used to collect the data, or by preliminary variogram analysis to identify nugget effects. It is often reasonable in spatial analyses to assume these data model error processes to be uncorrelated, especially when factoring in the computational cost associated with the non-parsimonious alternative, so we assume $\mathbf{\Sigma}_{y,n_r} = \text{diag}(\boldsymbol{\tau}_{y,n_r})$, $\mathbf{\Sigma}_{y,n_d} = \text{diag}(\boldsymbol{\tau}_{y,n_d})$, $\mathbf{\Sigma}_{x,n_r} = \text{diag}(\boldsymbol{\tau}_{x,n_r})$, $\mathbf{\Sigma}_{x,n_u} = \text{diag}(\boldsymbol{\tau}_{x,n_u})$, $\mathbf{\Sigma}_{z,n_r} = \text{diag}(\boldsymbol{\tau}_{z,n_r})$ and $\mathbf{\Sigma}_{z,n_p} = \text{diag}(\boldsymbol{\tau}_{z,n_p})$. Critically, in each of these cases, the matrix normal distribution implies that the error

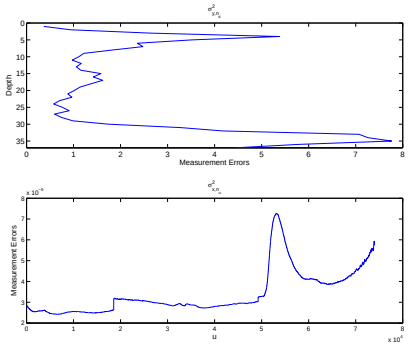

Figure 1: Measurement error variances used as prior information in the data analysis as discussed in the sensitivity analysis. Top panel: ECp measurement error variance (σ^2_{y,n_d}, measured in milliSiemens per meter squared) as a function of depth; Bottom panel: VNIR measurement error variance (σ^2_{x,n_u}, measured in nanometers squared) as a function of wavelength index, u.

covariance for each element of the random matrix arises from the Kronecker product of two diagonal matrices, so that each variance corresponds to the product of two variances (e.g., two τ values here). This makes identifiability problematic unless strong prior information is available.

In our application, we have strong prior information available about instrument measurement error corresponding to the ECp responses and the VNIR image covariates. Figure 1 shows the instrument measurement errors for the ECp responses as a function of depth (denote by σ^2_{y,n_d}), as well as the VNIR measurement errors as a function of wavelength (u) (denoted by σ^2_{x,n_u}). In both cases, expert opinion from subject matter experts suggested that these measurement errors are reasonable for the entire horizontal spatial domain considered in our application. Although this is useful information, we also have truncation error from the basis expansions so we included a common multiplicative variance parameter to scale these values; this parameter is assumed to be random and is estimated in the Bayesian framework. In addition, since the matrix normal diagonal covariance assumption prevents the estimation of two variance components, we specified one of the diagonal matrices to be the identity matrix. Specifically, $\mathbf{\Sigma}_{y,n_r} = \mathbf{I}_{n_r}$ and $\mathbf{\Sigma}_{y,n_d} = \tau_{y,n_d}\text{diag}(\boldsymbol{\sigma}^2_{y,n_d})$, where $\{\boldsymbol{\sigma}^2_{y,n_d}\}$ are the known measurement error variances discussed above. Then, τ_{y,n_d} was specified to have a conjugate inverse gamma prior; that is, $\tau_{y,n_d} \sim IG(\kappa_{y,n_d}, \nu_{y,n_d})$. Similarly, $\mathbf{\Sigma}_{x,n_r} = \mathbf{I}_{n_r}$ and $\mathbf{\Sigma}_{x,n_u} = \tau_{x,n_u}\text{diag}(\boldsymbol{\sigma}^2_{x,n_u})$, where $\{\boldsymbol{\sigma}^2_{x,n_u}\}$ are known measurement error variances described previously. Again, τ_{x,n_u} was specified to have an inverse gamma prior such that $\tau_{x,n_u} \sim IG(\kappa_{x,n_u}, \nu_{x,n_u})$. Finally, $\mathbf{\Sigma}_{z,p} = \mathbf{I}_p$ and $\mathbf{\Sigma}_{z,n_r} = \tau_{z,n_r}\mathbf{I}_{n_r}$, respectively where τ_{z,n_r} was specified to have an inverse gamma prior, $\tau_{z,n_r} \sim IG(\kappa_{z,n_r}, \nu_{x,n_r})$.

The choice of the hyperparameters associated with the $\kappa = \kappa_{y,n_d} = \kappa_{x,n_u} = \kappa_{z,n_r}$ and $\nu = \nu_{y,n_d} = \nu_{x,n_u} = \nu_{z,n_r}$ parameters were chosen based on a sensitivity analysis.

Specifically, we considered two combination of hyperparameters for κ and ν: (κ, ν)= (1,1) and (2.001, 1.001). An inverse-gamma prior with $(\kappa, \nu) = (1, 1)$ is vague and the prior with $(\kappa, \nu) = (2.001, 1.001)$ has mean one and variance 1000. The posterior distributions corresponding to these parameters showed very small differences for these two priors, and therefore we took the most non-informative (i.e., $(\kappa, \nu) = (1, 1)$) for the analyses presented here.

We also considered a sensitivity analysis in which we did not make use of the prior measurement error variance information. That is, we assumed $\mathbf{\Sigma}_{y,n_r} = \mathbf{I}_{n_r}$, $\mathbf{\Sigma}_{y,n_d} = \tau_{y,n_d}\mathbf{I}_{n_d}$, $\mathbf{\Sigma}_{x,n_r} = \mathbf{I}_{n_r}$, $\mathbf{\Sigma}_{x,n_u} = \tau_{x,n_u}\mathbf{I}_{n_u}$, $\mathbf{\Sigma}_{z,p} = \mathbf{I}_p$ and $\mathbf{\Sigma}_{z,n_r} = \tau_{z,n_r}\mathbf{I}_{n_r}$, with the same inverse gamma prior distributions on the τ parameters as described above. These results (not shown) gave very similar predictive mean fields as when including the known prior measurement error variances, but not surprisingly, the predictive variances were less. We feel that the prior information should be used if it is available, and that slightly larger posterior prediction variances are probably more realistic. Thus, we considered the model with the prior measurement error variances for the analyses presented here.

Process Model Parameters

Recall that we specified Wishart prior distributions for all of the precision matrices associated with the random effects in the process model. These require a specification of hyperparameters associated with the scale matrices and degrees of freedom (e.g., $\mathbf{V}$ and v in our notation, respectively). To increase model flexibility, we specified the degree of freedom values to equal the dimension of the associated covariance matrices so that the Wishart distribution is relatively non-informative. For scale matrices, if prior or subjective knowledge is available, we may indicate a specific structure and values. For example, if the columns of $\mathbf{F}_j$ are composed of EOFs, $\mathbf{V}_{\mathbf{F}_j,n_{k_j}}$ may be specified as a diagonal matrix with elements equal to the associated reciprocal eigenvalues. However, in most cases we may have no such information, and so we generally assume naive specifications such as $\mathbf{V}_{\mathbf{F}_j,n_\ell} = \tau_{\mathbf{F}_j,n_\ell}\mathbf{I}_{n_\ell}$, $\mathbf{V}_{\mathbf{F}_j,n_{k_j}} = \tau_{\mathbf{F}_j,n_{k_j}}\mathbf{I}_{n_{k_j}}$, $\mathbf{V}_{\mathbf{Q},n_\ell} = \tau_{\mathbf{Q},n_\ell}\mathbf{I}_{n_\ell}$, and $\mathbf{V}_{\mathbf{Q},n_p} = \tau_{\mathbf{Q},n_p}\mathbf{I}_{n_p}$, where the variance parameter can be chosen via sensitivity analysis. In our application, we conducted a sensitivity analysis for specification of diagonal structures for the hyperparameter scale matrices compared to this simpler constant formulation for precision matrices associated with coefficients from empirical (EOF) basis functions. These results (not shown) gave essentially indistinguishable predictions in the ECp fields, suggesting that the posterior predictions are not sensitive to this hyperparameter specification. Thus, we utilized the simpler constant structures for the results presented here.

S3.3 Additional Sensitivity Analysis

We conducted a sensitivity analysis to help select the number of image and depth profile EOFs, the Gaussian kernel width, and the variance hyperparameters. We consider sixteen scenarios corresponding to the combination of the number of image EOFs (n_k),

depth profile EOFs (n_i), and Gaussian kernel widths (σ_g^2): $(n_k, n_i, \sigma_g^2) = (10, 5, 1)$, $(10, 5, 1.5)$, $(10, 5, 1.8)$, $(10, 5, 2)$, $(10, 10, 1)$, $(10, 10, 1.5)$, $(10, 10, 1.8)$, $(10, 10, 2)$, $(15, 5, 1)$, $(15, 5, 1.5)$, $(15, 5, 1.8)$, $(15, 5, 2)$, $(15, 10, 1)$, $(15, 10, 1.5)$, $(15, 10, 1.8)$, and $(15, 10, 2)$. In addition, each scenario considered $\tau_{\mathbf{B}} = \{10^2, 10^4, 10^6\}$. Recall that when $\tau_{\mathbf{B}}$ becomes larger, the matrix normal prior of $\mathbf{B}$ becomes less informative. Since we have only one set of image covariates and one spatial covariate (elevation) in this data analysis application, we let $\mathbf{B} \equiv \{\mathbf{B}_1, \mathbf{H}\}$ and $\mathbf{F}_j = \mathbf{F}_1 \equiv \mathbf{F}$. For simplicity, we assumed $\tau = \tau_{\mathbf{F},n_\ell} = \tau_{\mathbf{F},n_k} = \tau_{\mathbf{Q},n_\ell} = \tau_{\mathbf{Q},n_p} = \tau_{\mathbf{G},n_\ell} = \tau_{\mathbf{G},n_i}$ and consider a range $\tau = \{10^{-5}, 10^{-6}, 10^{-7}, 10^{-8}\}$. The degree of freedom parameters for the Wishart priors were specified to be equal to the dimension of the associated covariance matrix, to make the priors as vague as possible. As we had no *a priori* scientific prior mean for the $\mathbf{F}$ and $\mathbf{Q}$ matrices, we specified $\boldsymbol{\mu}_{\mathbf{F}}$ and $\boldsymbol{\mu}_{\mathbf{Q}}$ to be zero matrices.

Model performance was evaluated by MSPE (as described in S4) averaged across depths and spatial locations, with posterior predictive distributions of fitted ECp curves obtained from an MCMC analysis based on 15,000 iterations and 5,000 burn-in. Convergence was evaluated through visual inspection of the trace plots of the sample chains for a wide-array of randomly selected parameters, with no evidence of lack of convergence detected.

Table 1 shows that, in general, the model is not sensitive to the covariance matrix hyperparameters ($\tau_{\mathbf{B}}$, τ) since the MSPEs show little variation within a given scenario. Regarding the EOF truncation and Gaussian kernel width parameters, scenarios $(n_k, n_i, \sigma_g^2) = (10, 10, 1)$ and $(n_k, n_i, \sigma_g^2) = (15, 10, 1)$ have the smallest MSPEs. Although the smallest MSE (54.8211) occurs when using scenario $(n_k, n_i, \sigma_g^2) = (15, 10, 1)$, including five extra image EOFs in the model provides only a minimal improvement in MSPE relative to the $(n_k, n_i, \sigma_g^2) = (10, 10, 1)$ case, at the cost of many more parameters. Therefore, we considered the more parsimonious model (i.e., scenario $(n_k, n_i, \sigma_g^2) = (10, 10, 1)$) for our ECp field predictions.

.

S4 Simulation Study

In order to evaluate the spatial predictive properties of our model, we design a simulation study to resemble the data analysis presented in the paper. Specifically, the simulation study mimics the soil science application and uses the bivariate Gaussian kernel matrix, $\mathbf{W}$, the first 10 empirical orthogonal functions (EOFs) of the image covariates, $\boldsymbol{\Phi}$, and the first 5 EOFs from the depth profiles, $\boldsymbol{\Psi}$. In the soil science application, 10 image EOFs and 5 depth EOFs correspond to 93.8% and 98.7% of the variation explained, respectively, whereas in the simulation these constitute the full basis set. Using these basis functions, we simulate a response matrix, $\widetilde{\mathbf{Y}}$, an image covariate matrix, $\widetilde{\mathbf{X}}$, and a spatial covariate vector, $\widetilde{\mathbf{Z}}$, on a grid of size 1680, where 1680 is the same grid size used to predict the ECp field in our data analysis. Additionally, at a given location $\mathbf{s}_i$ $(i = 1, \ldots, 1680)$, the dimensions of $\mathbf{Y}(\mathbf{s}_i)$, $\mathbf{X}(\mathbf{s}_i)$, and $\mathbf{Z}(\mathbf{s}_i)$ are 37×1, $74037 \times 1 =$

Table 1: The sensitivity analysis for MSPEs of the model predictions of ECp profiles, including image covariates and elevation. Note that n_k and n_i denote the number of image and depth EOFs, respectively, and σ_g^2 refers to the Gaussian kernel width.

				(n_k, n_i, σ_g^2)				
$(\tau_{\mathbf{B}}, \tau)$	(10,5,1)	(10,5,1.5)	(10,5,1.8)	(10,5,2)	(10,10,1)	(10,10,1.5)	(10,10,1.8)	(10,10,2)
$(10^2, 10^{-5})$	55.8850	57.4957	58.0602	58.2970	54.8231	56.4096	56.9565	57.2095
$(10^2, 10^{-6})$	55.8838	57.4901	58.0620	58.3113	54.8221	56.4046	56.9637	57.2102
$(10^2, 10^{-7})$	55.8817	57.4904	58.0588	58.3077	54.8228	56.4029	56.9657	57.2121
$(10^2, 10^{-8})$	55.8842	57.4885	58.0574	58.3083	54.8260	56.4039	56.9711	57.2158
$(10^4, 10^{-5})$	55.8880	57.4882	58.0571	58.2887	54.8217	56.3967	56.9771	57.1772
$(10^4, 10^{-6})$	55.8864	57.4908	58.0576	58.3080	54.8273	56.4041	56.9656	57.2164
$(10^4, 10^{-7})$	55.8832	57.4925	58.0590	58.3092	54.8234	56.4023	56.9679	57.2156
$(10^4, 10^{-8})$	55.8852	57.4918	58.0573	58.3096	54.8233	56.4042	56.9647	57.2174
$(10^6, 10^{-5})$	55.8865	57.4910	58.0577	58.3107	54.8238	56.3974	56.9716	57.2211
$(10^6, 10^{-6})$	55.8863	57.4908	58.0576	58.3085	54.8232	56.4029	56.9708	57.2155
$(10^6, 10^{-7})$	55.8847	57.4887	58.0604	58.3078	54.8241	56.4080	56.9673	57.2118
$(10^6, 10^{-8})$	55.8848	57.4894	58.0578	58.3091	54.8253	56.4067	56.9657	57.2106
				(n_k, n_i, σ_g^2)				
$(\tau_{\mathbf{B}}, \tau)$	(15,5,1)	(15,5,1.5)	(15,5,2)	(15,5,2)	(15,10,1)	(15,10,1.5)	(15,10,1.8)	(15,10,2)
$(10^2, 10^{-5})$	55.8835	57.4866	58.0602	58.3054	54.8229	56.3987	56.9606	57.2308
$(10^2, 10^{-6})$	55.8832	57.4893	58.0585	58.3108	54.8221	56.4018	56.9662	57.2135
$(10^2, 10^{-7})$	55.8835	57.4887	58.0585	58.3100	54.8240	56.4035	56.9660	57.2133
$(10^2, 10^{-8})$	55.8843	57.4900	58.0594	58.3091	54.8244	56.4063	56.9636	57.2115
$(10^4, 10^{-5})$	55.8823	57.4944	58.0659	58.3124	54.8196	56.4054	56.9674	57.1958
$(10^4, 10^{-6})$	55.8831	57.4883	58.0586	58.3082	54.8256	56.4057	56.9668	57.2125
$(10^4, 10^{-7})$	55.8852	57.4929	58.0586	58.3077	54.8217	56.4049	56.9642	57.2108
$(10^4, 10^{-8})$	55.8848	57.4903	58.0601	58.3071	54.8236	56.4057	56.9666	57.2152
$(10^6, 10^{-5})$	55.8854	57.4940	58.0537	58.3130	54.8211	56.4036	56.9723	57.2036
$(10^6, 10^{-6})$	55.8852	57.4880	58.0596	58.3085	54.8254	56.4041	56.9660	57.2130
$(10^6, 10^{-7})$	55.8833	57.4903	58.0617	58.3092	54.8234	56.4020	56.9673	57.2144
$(10^6, 10^{-8})$	55.8857	57.4899	58.0618	58.3041	54.8244	56.4042	56.9654	57.2108

$(37 * 2001) \times 1$, and 1×1, respectively, where 37 corresponds to the maximum number of usable depth segments and 2001 corresponds to the number of wavelengths.

The noisy response matrix $\widetilde{\mathbf{Y}}$ was generated as $\widetilde{\mathbf{Y}} = \mathbf{Y} + \mathbf{E}_{\widetilde{y}}$, where $\mathbf{Y}$ denotes the true response matrix and $\mathbf{E}_{\widetilde{y}} \sim N_{1680,37}(\mathbf{0}, \mathbf{I}_{1680}, 10\,\mathbf{I}_{37})$. Now, $\mathbf{Y}$ can be represented as follows $\mathbf{Y} = \mathbf{WA\Psi}$, where $\mathbf{W}$ is a 1680×20 spatial basis composed of bivariate Gaussian kernels with covariance equal to the identity matrix (i.e., variances equal to one). It is important to emphasize that the choice of identity covariance in this context reflects an assumption of isotropy and that the 20 knots are selected using a space filling design from the 26 observed locations of the ECp response data (see the online supplement for details). Finally, $\mathbf{A}$ corresponds to a low-rank coefficient matrix that is a combination of $\mathbf{F}_1$, $\mathbf{B}_1$, $\mathbf{Q}$, $\mathbf{H}$ and $\mathbf{G}$; i.e., $\mathbf{A} = \mathbf{F}_1\mathbf{B}_1 + \mathbf{QH} + \mathbf{G}$.

In this simulation, $\mathbf{F}_1$ and $\mathbf{Q}$ were taken as the estimated values of $\widehat{\mathbf{F}}_1$ and $\widehat{\mathbf{Q}}$ from the fitted model corresponding to an analysis of the ECp data using one image covariate and one spatial covariate. For $i = 1, 3, 5, 9$, we set the rows $\mathbf{B}_{1,(i,:)} = \widehat{\mathbf{B}}_{1,(i,:)}$, otherwise

row i is set equal to the zero vector. The specific values used were: $\mathbf{B}_{1,(1,:)}$ =(-8.2687, -6.1593, -8.7686, 4.8659, 0.6453); $\mathbf{B}_{1,(3,:)}$ = (7.0458, -1.6005, 30.6569, -57.2508, -24.5970); $\mathbf{B}_{1,(5,:)}$ = (19.3338, -34.9112, 62.6131, 9.5348, 17.8897); $\mathbf{B}_{1,(9,:)}$ = (-13.1880, -41.6825, 67.7879, 44.5034, 34.3638). Lastly, we specified $\mathbf{H}' = \widehat{\mathbf{H}}'$ =(-0.9343, 0.0144, 3.0850, -7.2958, -2.4153) and let $\mathbf{G} \sim N_{20,5}(\mathbf{0}, \mathbf{I}_{20}, \mathbf{I}_5)$.

Given $\mathbf{F}_1 = \widehat{\mathbf{F}}_1$, we simulated noisy $\widetilde{\mathbf{X}}$ as $\widetilde{\mathbf{X}} = \mathbf{W}\mathbf{F}_1\mathbf{\Phi} + \mathbf{E}_{\widetilde{x}}$, where we specify $\mathbf{E}_{\widetilde{x}} \sim N_{1680,74037}(\mathbf{0}, \mathbf{I}_{1680}, (4 * 10^{-5})\mathbf{I}_{74037})$. Similarly, given $\mathbf{Q} = \widehat{\mathbf{Q}}$, we simulated noisy $\widetilde{\mathbf{Z}}$ as $\widetilde{\mathbf{Z}} = \mathbf{W}\mathbf{Q} + \mathbf{E}_{\widetilde{z}}$, where $\mathbf{E}_{\widetilde{z}} \sim N_{1680,1}(\mathbf{0}, \mathbf{I}_{1680}, 0.04)$. Accordingly, we simulated 100 noisy datasets.

For each dataset, we considered eight scenarios consisting of various numbers of sample points (n_s) and knots (k_s): (n_s, k_s) = (26,13), (26,26), (52,13), (52,26), (52,39), (78,13), (78, 26), and (78, 39). For each number of sample points, n_s, five sets of observed locations were randomly generated from the 1680 grid points and, subsequently, knots were selected using a space filling design. This gave a total of 500 "observed" data sets for spatial prediction. For each, four combinations of the number of image EOFs and depth EOFs were considered: (n_k, n_i) = (5,3), (5,5), (10,3) and (10,5). To construct a spatial basis given κ_s, we used bivariate Gaussian kernels with the covariance matrix equal to $\sigma_g^2\mathbf{I}$, where the value of σ_g^2 varies according to k_s as follows: for $k_s = 13$, $\sigma_g^2 = \{1, 1.5, 2\}$; for $k_s = 26$, $\sigma_g^2 = \{0.5, 0.6, 0.7\}$; for $k_s = 39$, $\sigma_g^2 = \{0.3, 0.4, 0.5\}$. Importantly, σ_g^2 was chosen to be the largest value practicable, relative to the scale of the distance between locations, in order to ensure that the covariance matrix remains positive definite.

To evaluate model performance, we examined the mean squared prediction error (MSPE) given by MSPE $= \{1/(n_s \times 37)\} \sum_{j=1}^{37} \sum_{i=1}^{n_s} \left\{\widehat{Y}(s_i, d_j) - Y(s_i, d_j)\right\}^2$, where $\widehat{Y}(s_i, d_j)$ denotes posterior predicted values of $Y(s_i, d_j)$ at location s_i and depth d_j. Further, we fixed the hyperparameters $\boldsymbol{\mu}_{\mathbf{F}} = \mathbf{0}$, $\boldsymbol{\mu}_{\mathbf{Q}} = \mathbf{0}$, $\tau_{\mathbf{B}} = 10^4$ ($\mathbf{\Sigma}_{\mathbf{B}} = \tau_{\mathbf{B}}\mathbf{I}_P$), and specified the following Wishart distributions: $\mathbf{\Sigma}_{n_\ell}^{-1} \sim \mathrm{W}_{n_\ell}(10^{-6}\mathbf{I}, n_\ell)$, $\mathbf{\Sigma}_{n_i}^{-1} \sim \mathrm{W}_{n_i}(10^{-6}\mathbf{I}, n_i)$, $\mathbf{\Sigma}_{\mathbf{F},n_\ell}^{-1} \sim \mathrm{W}_{n_\ell}(10^{-6}\mathbf{I}, n_\ell)$, $\mathbf{\Sigma}_{\mathbf{F},n_k}^{-1} \sim \mathrm{W}_{n_k}(10^{-6}\mathbf{I}, n_k)$, $\mathbf{\Sigma}_{\mathbf{Q},n_\ell}^{-1} \sim \mathrm{W}_{n_\ell}(10^{-6}\mathbf{I}, n_\ell)$, and finally $\mathbf{\Sigma}_{\mathbf{Q},n_p}^{-1} \sim \mathrm{W}_{n_p}(10^{-6}\mathbf{I}, n_p)$. These hyperparameters were selected to be consistent with the sensitivity analysis detailed in Section S3.

The MCMC algorithm results were based on 15,000 iterations with a 5,000 iteration burn-in. Convergence was evaluated through visual inspection of the trace plots of the sample chains for a wide-array of randomly selected parameters, with no evidence of lack of convergence detected. For each scenario, Table 2 presents the optimal model performance based on MSPE from the pool of σ_g^2. From Table 2 it can be seen that there is a reduction in MSPE as the sample size increases, and, for a given sample size, the MSPE decreases as the number of knots increases. In addition, holding n_k fixed and increasing n_i results in reduced MSPE, as does holding n_i fixed and increasing n_k (note that ties in the table are resolved by additional significant digits). In summary, this simulation demonstrates good predictive properties for the models under consideration.

Table 2: Model performance based on MSPE for the simulation study presented in Section S4. Here n_s and k_s denote the number of sample locations and number of knots, respectively, while, n_k and n_i denote the number of image and depth EOFs considered, respectively. For each of the $n = 100$ generated datasets, the simulation consists of five sets of observed locations that are randomly generated from the 1680 grid points and, subsequently, knots are selected using a space filling design.

	(n_k, n_i)			
(n_s, k_s)	(5,3)	(5,5)	(10,3)	(10,5)
(26,13)	1837.12	1837.08	1837.14	1837.12
(26,26)	1305.30	1313.25	1349.11	1339.86
(52,13)	1197.26	1197.26	1197.28	1197.27
(52,26)	549.12	549.08	549.13	549.13
(52,39)	524.65	524.55	527.95	528.10
(78,13)	1004.19	1004.19	1004.21	1004.18
(78,26)	364.85	364.85	364.85	364.86
(78,39)	332.58	332.57	332.64	332.65

References

Bali, J. L., Boente, G., Tyler, D. E., and Wang, J.-L. (2011), "Robust Functional Principal Components: a projection-pursuit approach," *The Annals of Statistics*, 39, 2852–2882.

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 B*, 70, 825–848.

Berry, R. P. and Ver Hoef, J. (1996), "Blackbox kriging: Spatial prediction without specifying variogram models," *Journal of Agricultural, Biological and Ecological Statistics*, 1, 297–322.

Crainiceanu, C. M., Caffo, B. S., Luo, S., Zipunnikov, V., and Punjabi, N. M. (2011), "Population value decomposition, a framework for the analysis of image populations," *Journal of the American Statistical Association*, 106, 775–790.

Cressie, N. and Wikle, C. K. (2011), *Statistics for Spatio-Temporal Data*, John Wiley and Sons.

Higdon, D. (1998), "A process-convolution approach to modelling temperatures in the North Atlantic Ocean," *Environmental and Ecological Statistics*, 5, 173–190.

Nychka, D. and Saltzman, N. (1998), "Design of air-quality monitoring networks," in *Case Studies in Environmental Statistics*, (Edited by D. Nychka and W. W. Piegorsch and L. H. Cox), 51–76. Springer.

Ramsay, J. O. and Silverman, B. W. (2005), *Functional Data Analysis: 2nd ed*, Springer-Verlag.

Ruppert, D., Wand, M. P., and Carroll, R. J. (2003), *Semiparametric Regression*, Cambridge University Press.

Xiao, L., Li, Y., and Ruppert, D. (2013a), "Fast bivariate P-splines: the sandwich smoother," *Journal of the Royal Statistical Society B*, 75, 577–599.

Xiao, L., Ruppert, D., Zipunnikov, V., and Crainiceanu, C. (2013b), "Fast covariance estimation for high-dimensional functional data," *arXiv:1306.5718v1*.