

Supplement to “Crop Yield Prediction Using Bayesian Spatially Varying Coefficient Models with Functional Predictors”

This supplement contains the full hierarchical models and MCMC sampling algorithm for the proposed model. We also provide MCMC diagnostics, additional Monte Carlo experiments, and additional corn yield data analysis results.

S1 Full Hierarchical Models

We present the full Bayesian Spatially Varying Functional Model (BSVFM) for data collected in K years, where data in different years are treated as independent replicates. As discussed in Section 5.1, the exponential correlation function is employed to regulate the spatial structure of varying coefficients $\beta_r(\mathbf{s}_i)$ and $\alpha_j(\mathbf{s}_i)$, FPC scores $\xi_{kr}(\mathbf{s}_i)$, and the latent process $v_r(\mathbf{s}_i)$ for variable selection. The common range parameters are assumed for each type of spatial variables or latent processes. Here, $IG(\cdot, \cdot)$ and $TN(\cdot, \cdot; a, b)$ denote the inverse gamma distribution and truncated Normal distribution in the range $[a, b]$, respectively.

(i) **Data stage:** for $k = 1, \dots, K$, and $i = 1, \dots, n$,

$$Y_k(\mathbf{s}_i) = \alpha_0(\mathbf{s}_i) + \sum_{j=1}^h Z_{kj}(\mathbf{s}_i) \alpha_j(\mathbf{s}_i) + \sum_{r=1}^p \xi_{kr}(\mathbf{s}_i) \{\beta_r(\mathbf{s}_i) \gamma_r(\mathbf{s}_i)\} + e_k(\mathbf{s}_i), \quad (\text{S1.1})$$

$$e_k(\mathbf{s}_i) \sim N\{0, \sigma_e^2 / \omega_k(\mathbf{s}_i)\},$$

$$\mathbf{W}_k(\mathbf{s}_i; t) = \sum_{r=1}^p \xi_{kr}(\mathbf{s}_i) \{\mathbf{I}_q \otimes \boldsymbol{\psi}^T(t)\} \mathbf{d}_r + \mathbf{u}_k(\mathbf{s}_i; t), \quad (\text{S1.2})$$

$$\mathbf{u}_k(\mathbf{s}_i; t) \stackrel{\text{i.i.d.}}{\sim} N\{0, \text{diag}(\sigma_{u1}^2, \dots, \sigma_{uq}^2)\}, \quad t \in \mathcal{T},$$

where $e_k(\mathbf{s}_i)$ and $\mathbf{u}_k(\mathbf{s}_i; t)$ are independent, and $\gamma_r(\mathbf{s}_i)$ is defined in (3) of the manuscript.

(ii) **Process stage:**

$$\begin{aligned}
\boldsymbol{\alpha}_0 &\sim N\{\mu_{\alpha_0} \mathbf{1}_n, \sigma_{\alpha_0}^2 \boldsymbol{\Sigma}(\phi_{\alpha_0})\}, \quad \boldsymbol{\alpha}_j \sim N\{\mu_{\alpha_j} \mathbf{1}_n, \sigma_{\alpha_j}^2 \boldsymbol{\Sigma}(\phi_{\alpha})\}, \quad j = 1, \dots, h, \\
\boldsymbol{\beta}_r &\sim N\{\mu_{\beta_r} \mathbf{1}_n, \sigma_{\beta_r}^2 \boldsymbol{\Sigma}(\phi_{\beta})\}, \quad r = 1, \dots, p, \\
\boldsymbol{\xi}_{kr} &\sim N\{\mathbf{0}_n, \sigma_{\xi_r}^2 \boldsymbol{\Sigma}(\phi_{\xi})\}, \quad k = 1, \dots, K, \quad r = 1, \dots, p, \\
\mathbf{v}_r &\sim N\{\mu_{v_r} \mathbf{1}_n, \sigma_{v_r}^2 \boldsymbol{\Sigma}(\phi_v)\}, \quad r = 1, \dots, p, \\
\mathbf{d}_r &\sim N(\mathbf{0}_{qL}, \mathbf{I}_q \otimes \mathbf{D}_r), \text{ where } \mathbf{D}_r = \text{diag}(c, c, \lambda_r^{-1}, \dots, \lambda_r^{-1}), \quad r = 1, \dots, p.
\end{aligned} \tag{S1.3}$$

(iii) **Prior stage:**

For smoothing parameters,

$$\begin{aligned}
\lambda_r^{-1/2} &\sim \text{Uniform}(a_r, b_r), \text{ where} \\
a_1 &= 0, \quad a_r = \lambda_{r-1}^{-1/2}, \quad r = 2, \dots, p, \quad b_r = \lambda_{r+1}^{-1/2}, \quad r = 1, \dots, p-1, \quad b_p = c_b;
\end{aligned} \tag{S1.4}$$

For mean and variance parameters,

$$\begin{aligned}
\sigma_e^2 &\sim IG(A_e, B_e), \quad \sigma_{ul}^2 \sim IG(A_u, B_u), \quad l = 1, \dots, q, \\
\mu_{\alpha_j} &\sim N(0, s_{\alpha}^2), \quad \sigma_{\alpha_j}^2 \sim IG(A_{\alpha}, B_{\alpha}), \quad j = 0, \dots, h, \\
\mu_{\beta_r} &\sim N(0, s_{\beta}^2), \quad \sigma_{\beta_r}^2 \sim IG(A_{\beta}, B_{\beta}), \quad r = 1, \dots, p, \\
\sigma_{\xi_r}^2 &\sim IG(A_{\xi}, B_{\xi}), \quad r = 1, \dots, p, \\
\mu_{v_r} &\sim N(0, s_v^2), \quad \sigma_{v_r}^2 \sim IG(A_v, B_v), \quad r = 1, \dots, p;
\end{aligned} \tag{S1.5}$$

For range parameters,

$$\begin{aligned}
\phi_{\alpha_0} &\sim TN(m_{\phi_{\alpha_0}}, s_{\phi_{\alpha_0}}^2; 0, \infty), \quad \phi_{\alpha} \sim TN(m_{\phi_{\alpha}}, s_{\phi_{\alpha}}^2; 0, \infty), \\
\phi_{\beta} &\sim TN(m_{\phi_{\beta}}, s_{\phi_{\beta}}^2; 0, \infty), \quad \phi_{\xi} \sim TN(m_{\phi_{\xi}}, s_{\phi_{\xi}}^2; 0, \infty), \quad \phi_v \sim TN(m_{\phi_v}, s_{\phi_v}^2; 0, \infty).
\end{aligned} \tag{S1.6}$$

The data may have an irregular structure with observations collected from different sets of spatial locations in each year. Let $\mathcal{S}_k = \{\mathbf{s}_{k1}, \dots, \mathbf{s}_{kn_k}\}$ be the locations with records in year k and $\mathcal{S} = \cup_k \mathcal{S}_k$ be the set of spatial locations with information from at least one year.

Set $n = |\mathcal{S}|$. Then the full model (S1.1) is specified under \mathbf{s}_i , $i = 1, \dots, n$, by admitting missing values and the practical strategy to handle missing values in MCMC implementation is described in the following section. In (S1.2), $\mathbf{W}_k(\mathbf{s}_i, t)$ are the centered temperature trajectories as described in Section 3.1. Since the meteorological data is generally available and complete for the past half a century, we use the available climate records to estimate the temperature mean function at each location \mathbf{s}_i and center the temperature trajectories by these location-specific mean functions.

In (S1.3), $\mathbf{d}_r = (\mathbf{d}_{r1}^T, \dots, \mathbf{d}_{rq}^T)^T$ is a $(Lq \times 1)$ vector, where $\mathbf{d}_{rl} = (d_{rl,1}, \dots, d_{rl,L})^T$ is a vector of spline coefficients for f_{rl} based on basis functions $\boldsymbol{\psi}(t) = \{\psi_1(t), \dots, \psi_L(t)\}^T$, for $l = 1, \dots, q$. They are subject to orthogonal constraints, $\mathbf{d}_{rl}^T \mathbf{J}_\psi \mathbf{d}_{r'l} = 1$, if $r = r'$, 0 otherwise, where $\mathbf{J}_\psi = \int \boldsymbol{\psi}(t) \boldsymbol{\psi}^T(t) dt$. And we set $c = 10^8$ to make the prior non-informative. Using uniform priors on the smoothing parameters λ_r and enforcing the ordering constraint, $\lambda_1 > \lambda_2 > \dots > \lambda_p > 0$, the priors are as in (S1.4). We set sufficiently large $c_b = 10^4$ for the upper bound of smoothing parameter λ_r . By incorporating orthogonality constraint on \mathbf{f}_r , the full conditional posterior distributions of λ_r and \mathbf{d}_{rl} can be derived as truncated Gamma and Gaussian distributions, respectively, as we shall see in the following section.

There are no closed-forms for the full conditional distributions of the range parameters on the exponential correlation functions, ϕ_{α_0} , ϕ_α , ϕ_β , ϕ_ξ , and ϕ_v , in (S1.3), and we employ the Metropolis-Hastings (M-H) algorithm. The priors of these scale parameters are set to be truncated Normal distributions truncated at zero to assure positive posterior samples. In our analysis, we set $m_{\phi_{\alpha_0}} = m_{\phi_\alpha} = m_{\phi_\xi} = m_{\phi_\beta} = m_{\phi_v} = 300$ and $s_{\phi_{\alpha_0}} = s_{\phi_\alpha} = s_{\phi_\xi} = s_{\phi_\beta} = s_{\phi_v} = 100$, considering the distances among the counties in the selected Midwest states ranging from 10 to 1,500 (km). We have also compared the posterior distributions with the priors and verified that these choices are not restrictive. The M-H algorithm is also applied to sample the latent process \mathbf{v}_r for the model selection indicator variable. The implementation processes are described in the following section. Lastly, hyper priors are chosen as follows: for IG priors, $A_e = A_u = A_{\alpha_0} = A_\alpha = A_\xi = A_\beta = A_v = 2.8$ and $B_e = B_u = B_{\alpha_0} = B_\alpha = B_\xi = B_\beta = B_v = 1/2.8$; for Gaussian priors, $s_{\alpha_0} = s_\alpha = s_\xi = s_\beta = s_v = 50$. All these hyper

priors are non-informative in the sense that the posterior distributions are not restricted by the corresponding priors.

S2 MCMC Sampling Algorithm

Our MCMC sampling scheme is based on Gibbs sampling where we iteratively sample the parameters and latent variables from their full conditional distributions. A M-H algorithm is used when a full conditional distribution does not have a closed form. For notation simplicity, we denote the observed scalar predictor from location \mathbf{s}_i in year k by $Z_k(\mathbf{s}_i)$ and the corresponding coefficient by $\alpha(\mathbf{s}_i)$. In practice, functional trajectories are collected over discretized points, $t_1, \dots, t_T \in \mathcal{T}$, and we use $\mathbf{W}_k(\mathbf{s}_i; t_\tau) = \{W_{k1}(\mathbf{s}_i, t_\tau), W_{k2}(\mathbf{s}_i, t_\tau)\}^T$, for $\tau = 1, \dots, T$, in the model estimation. Under the missing at random assumption as discussed in Section 6, we ignore the missing values in the model estimation by replacing missing information in $Y_k(\mathbf{s}_i)$ and $Z_k(\mathbf{s}_i)$ with zero in data preparation and performing the same replacement process for corresponding FPC scores $\xi_{kr}(\mathbf{s}_i)$. Now the sampling algorithm consists of the following steps.

1. **Basis functions and smoothing parameters:** for $r = 1, \dots, p$,

(a) $[\lambda_r | \dots] \propto \text{Gamma}\{\frac{1}{2}(L-3), \frac{1}{2} \sum_{l=1}^q \sum_{h=3}^L d_{rl,h}^2\}$ truncated to (a_r^{-2}, b_r^{-2}) , where $a_1 = 0$, $a_r = \lambda_{r-1}^{-1/2}$, $r = 2, \dots, p$, $b_r = \lambda_{r+1}^{-1/2}$, $r = 1, \dots, p-1$, $b_p = 10^4$, and L denotes the number of known splines for the estimation of \mathbf{f}_r .

(b) $[\mathbf{d}_r | \dots] \propto N\{\text{diag}(\tilde{\mathbf{H}}_{r1}, \dots, \tilde{\mathbf{H}}_{rq})\mathbf{h}_r, \text{diag}(\tilde{\mathbf{H}}_{r1}, \dots, \tilde{\mathbf{H}}_{rq})\}$, where $\text{diag}(\tilde{\mathbf{H}}_{r1}, \dots, \tilde{\mathbf{H}}_{rq})$ denotes the $(Lq \times Lq)$ block-diagonal matrix based on $(L \times L)$ matrices $\tilde{\mathbf{H}}_{rl}$, and $\mathbf{h}_r = (\mathbf{h}_{r1}^T, \dots, \mathbf{h}_{rq}^T)^T$ is a $(Lq \times 1)$ vector based on $(L \times 1)$ vectors $\mathbf{h}_{rl} = (h_{rl,1}, \dots, h_{rl,L})^T$, for $l = 1, \dots, q$. Let $\mathbf{Q}_{l[-r]} = (\mathbf{J}_\psi \mathbf{d}_{1l}, \dots, \mathbf{J}_\psi \mathbf{d}_{r-1,l}, \mathbf{J}_\psi \mathbf{d}_{r+1,l}, \dots, \mathbf{J}_\psi \mathbf{d}_{pl})$ with $\mathbf{J}_\psi = \int_{\mathcal{T}} \boldsymbol{\psi}(t) \boldsymbol{\psi}^T(t) dt$. Then conditioning on $\mathbf{d}_{rl}^T \mathbf{Q}_{l[-r]} = \mathbf{0}$, we have

$$\tilde{\mathbf{H}}_{rl} = \mathbf{H}_{rl} - \mathbf{H}_{rl} \mathbf{Q}_{l[-r]} (\mathbf{Q}_{l[-r]}^T \mathbf{H}_{rl} \mathbf{Q}_{l[-r]})^{-1} \mathbf{Q}_{l[-r]}^T \mathbf{H}_{rl}, \text{ where}$$

$$\mathbf{H}_{rl}^{-1} = \mathbf{D}_r^{-1} + \sigma_{ul}^2 \sum_{k,i} \xi_{kr}^2(\mathbf{s}_i) \int_{\mathcal{T}} \boldsymbol{\psi}(t) \boldsymbol{\psi}^T(t) dt \text{ and}$$

$$\mathbf{h}_{rl} = \sigma_{ul}^2 T^{-1} \sum_{k,i} \xi_{kr}(\mathbf{s}_i) \sum_{\tau} \{W_{kl}(\mathbf{s}_i; t_{\tau}) - \sum_{r' \neq r} \xi_{kr'}(\mathbf{s}_i) \boldsymbol{\psi}^T(t_{\tau}) \mathbf{d}_{r'l}\} \boldsymbol{\psi}(t_{\tau}).$$

After sampling the λ_r , we sample and normalize \mathbf{d}_r with a modified version of the efficient Cholesky decomposition approach. Sampling under conditional distributions (a) and (b) can be implemented by R function `mfdlmF()` written by Kowal et al. (2017).

2. Spatially structured variables:

(a) Spatially varying intercept:

$$[\boldsymbol{\alpha}_0 | \dots] \propto N(\mathbf{A}_0 \mathbf{a}_0, \mathbf{A}_0), \text{ where}$$

$$\mathbf{A}_0 = \left\{ \frac{\mathbf{I}_n}{\sigma_e^2} + \frac{\boldsymbol{\Sigma}(\phi_{\alpha_0})^{-1}}{\sigma_{\alpha_0}^2} \right\}^{-1},$$

$$\mathbf{a}_0 = \frac{\sum_{k,i} \mathbf{1}_n^{(i)} \{Y_k(\mathbf{s}_i) - Z_k(\mathbf{s}_i) \alpha(\mathbf{s}_i) - \sum_r \xi_{kr}(\mathbf{s}_i) \beta_r^*(\mathbf{s}_i)\}}{\sigma_e^2} + \frac{\boldsymbol{\Sigma}(\phi_{\alpha_0})^{-1} \mu_{\alpha_0} \mathbf{1}_n}{\sigma_{\alpha_0}^2},$$

where $\{\boldsymbol{\Sigma}(\phi_{\alpha_0})\}_{i,i'} = \exp\{-d(\mathbf{s}_i, \mathbf{s}_{i'})/\phi_{\alpha_0}\}$ based on the Euclidean distance $d(\mathbf{s}_i, \mathbf{s}_{i'})$ between \mathbf{s}_i and $\mathbf{s}_{i'}$. Here, $\mathbf{1}_n^{(i)}$ denotes the $(n \times 1)$ vector with 1 at i th element and 0 elsewhere, and $\beta_r^*(\mathbf{s}_i) = \beta_r(\mathbf{s}_i) \gamma_r(\mathbf{s}_i)$.

(b) Spatially varying coefficients for scalar predictor:

$$[\boldsymbol{\alpha} | \dots] \propto N(\mathbf{A} \mathbf{a}, \mathbf{A}), \text{ where}$$

$$\mathbf{A} = \left\{ \frac{\sum_{k,i} \mathbf{I}_n^{(i)} \mathbf{Z}_k \mathbf{Z}_k^T \mathbf{I}_n^{(i)}}{\sigma_e^2} + \frac{\boldsymbol{\Sigma}(\phi_{\alpha})^{-1}}{\sigma_{\alpha}^2} \right\}^{-1},$$

$$\mathbf{a} = \frac{\sum_{k,i} \mathbf{1}_n^{(i)} Z_k(\mathbf{s}_i) \{Y_i(\mathbf{s}_i) - \alpha_0(\mathbf{s}_i) - \sum_r \xi_{kr}(\mathbf{s}_i) \beta_r^*(\mathbf{s}_i)\}}{\sigma_e^2} + \frac{\boldsymbol{\Sigma}(\phi_{\alpha})^{-1} \mu_{\alpha} \mathbf{1}_n}{\sigma_{\alpha}^2},$$

where $\mathbf{Z}_k = \{Z_k(\mathbf{s}_1), \dots, Z_k(\mathbf{s}_n)\}^T$, $\{\boldsymbol{\Sigma}(\phi_{\alpha})\}_{i,i'} = \exp\{-d(\mathbf{s}_i, \mathbf{s}_{i'})/\phi_{\alpha}\}$, and $\mathbf{I}_n^{(i)}$ is the $(n \times n)$ matrix with 1 at (i, i) th element and 0 elsewhere.

(c) Spatially correlated FPC scores and corresponding regression coefficients: for $r = 1, \dots, p$,

i $[\boldsymbol{\xi}_{kr} | \dots] \propto N(\mathbf{A}_{kr} \mathbf{a}_{kr}, \mathbf{A}_{kr}), \text{ where}$

$$\mathbf{A}_{kr} = \left[\frac{\sum_i \mathbf{I}_n^{(i)} \beta_r^*(\beta_r^*)^T \mathbf{I}_n^{(i)}}{\sigma_e^2} + \sum_l \left\{ \frac{\mathbf{I}_n \sum_{\tau} \mathbf{d}_{rl}^T \boldsymbol{\psi}(t_{\tau}) \boldsymbol{\psi}^T(t_{\tau}) \mathbf{d}_{rl}}{T \sigma_{ul}^2} \right\} + \frac{\boldsymbol{\Sigma}(\phi_{\xi})^{-1}}{\sigma_{\xi_r}^2} \right]^{-1},$$

$$\mathbf{a}_{kr} = \frac{\sum_i \mathbf{1}_n^{(i)} \beta_r^*(\mathbf{s}_i) \{Y_k(\mathbf{s}_i) - \alpha_0(\mathbf{s}_i) - Z_k(\mathbf{s}_i) \alpha(\mathbf{s}_i) - \sum_{r' \neq r} \xi_{kr'}(\mathbf{s}_i) \beta_{r'}^*(\mathbf{s}_i)\}}{\sigma_e^2} +$$

$$\sum_l \left[\frac{\sum_i \mathbf{1}_n^{(i)} \sum_{\tau} \boldsymbol{\psi}^T(t_{\tau}) \mathbf{d}_{rl} \{W_{kl}(\mathbf{s}_i, t_{\tau}) - \sum_{r' \neq r} \xi_{kr'}(\mathbf{s}_i) \boldsymbol{\psi}^T(t_{\tau}) \mathbf{d}_{r'l}\}}{T \sigma_{ul}^2} \right],$$

where $\beta_r^* = \{\beta_r^*(\mathbf{s}_1), \dots, \beta_r^*(\mathbf{s}_n)\}^T$ and $\{\boldsymbol{\Sigma}(\phi_{\xi})\}_{i,i'} = \exp\{-d(\mathbf{s}_i, \mathbf{s}_{i'})/\phi_{\xi}\}$.

ii $[\boldsymbol{\beta}_r | \dots] \propto N(\mathbf{A}_r \mathbf{a}_r, \mathbf{A}_r), \text{ where}$

$$\mathbf{A}_r = \left\{ \frac{\sum_{k,i} \mathbf{1}_n^{(i)} \boldsymbol{\xi}_{kr}^* (\boldsymbol{\xi}_{kr}^*)^T \mathbf{I}_n^{(i)}}{\sigma_e^2} + \frac{\boldsymbol{\Sigma}(\phi_\beta)^{-1}}{\sigma_{\beta_r}^2} \right\}^{-1},$$

$$\mathbf{a}_r = \frac{\sum_{k,i} \mathbf{1}_n^{(i)} \boldsymbol{\xi}_{kr}^* (\mathbf{s}_i) \{Y_k(\mathbf{s}_i) - \alpha_0(\mathbf{s}_i) - Z_k(\mathbf{s}_i) \alpha(\mathbf{s}_i) - \sum_{r' \neq r} \boldsymbol{\xi}_{kr'}^* (\mathbf{s}_i) \beta_{r'}(\mathbf{s}_i)\}}{\sigma_e^2} + \frac{\boldsymbol{\Sigma}(\phi_\beta)^{-1} \mu_{\beta_r} \mathbf{1}_n}{\sigma_{\beta_r}^2},$$

where $\boldsymbol{\xi}_{kr}^* = \{\boldsymbol{\xi}_{kr}^*(\mathbf{s}_1), \dots, \boldsymbol{\xi}_{kr}^*(\mathbf{s}_n)\}^T$ with $\boldsymbol{\xi}_{kr}^*(\mathbf{s}_i) = \xi_{kr}(\mathbf{s}_i) \gamma_r(\mathbf{s}_i)$ and $\{\boldsymbol{\Sigma}(\phi_\beta)\}_{i,i'} = \exp\{-d(\mathbf{s}_i, \mathbf{s}_{i'})/\phi_\beta\}$.

3. Variance parameters:

- (a) Measurement error variance on functional predictors: for $l = 1, \dots, q$,

$$[\sigma_{ul}^2 | \dots] \propto IG\left(\frac{(\sum_k n_k)T}{2} + A_u, \frac{1}{2} \sum_{k,i,\tau} \{W_{kl}(\mathbf{s}_i; t_\tau) - \sum_r \xi_{kr}(\mathbf{s}_i) \boldsymbol{\psi}^T(t_\tau) \mathbf{d}_{rl}\}^2 + B_u\right).$$

- (b) Observation error variance on scalar response:

$$[\sigma_e^2 | \dots] \propto IG\left(\frac{\sum_k n_k}{2} + A_e, \frac{1}{2} \sum_{k,i} \{Y_k(\mathbf{s}_i) - \alpha_0(\mathbf{s}_i) - Z_k(\mathbf{s}_i) \alpha(\mathbf{s}_i) - \sum_r \xi_{kr}(\mathbf{s}_i) [\beta_r(\mathbf{s}_i) 1_{\{v_r(\mathbf{s}_i) \geq 0\}}]\}^2 + B_e\right).$$

- (c) Variance on spatial intercept:

$$[\sigma_{\alpha_0}^2 | \dots] \propto IG\left\{\frac{n}{2} + A_{\alpha_0}, \frac{1}{2} (\boldsymbol{\alpha}_0 - \mu_{\alpha_0} \mathbf{1}_n)^T \boldsymbol{\Sigma}(\phi_{\alpha_0})^{-1} (\boldsymbol{\alpha}_0 - \mu_{\alpha_0} \mathbf{1}_n) + B_{\alpha_0}\right\}.$$

- (d) Variance on spatial coefficients for scalar predictors:

$$[\sigma_\alpha^2 | \dots] \propto IG\left\{\frac{n}{2} + A_\alpha, \frac{1}{2} (\boldsymbol{\alpha} - \mu_\alpha \mathbf{1}_n)^T \boldsymbol{\Sigma}(\phi_\alpha)^{-1} (\boldsymbol{\alpha} - \mu_\alpha \mathbf{1}_n) + B_\alpha\right\}.$$

- (e) Variances on spatial FPC scores and corresponding regression coefficients: for

$$r = 1, \dots, p,$$

$$\text{i } [\sigma_{\xi_r}^2 | \dots] \propto IG\left\{\frac{\sum_k n_k}{2} + A_\xi, \frac{1}{2} \sum_k \boldsymbol{\xi}_{kr}^T \boldsymbol{\Sigma}(\phi_\xi)^{-1} \boldsymbol{\xi}_{kr} + B_\xi\right\}.$$

$$\text{ii } [\sigma_{\beta_r}^2 | \dots] \propto IG\left\{\frac{n}{2} + A_\beta, \frac{1}{2} (\boldsymbol{\beta}_r - \mu_{\beta_r} \mathbf{1}_n)^T \boldsymbol{\Sigma}(\phi_\beta)^{-1} (\boldsymbol{\beta}_r - \mu_{\beta_r} \mathbf{1}_n) + B_\beta\right\}.$$

4. Mean parameters:

- (a) Mean on spatial intercept:

$$[\mu_{\alpha_0} | \dots] \propto N(M_0 m_0, M_0), \text{ where } M_0 = \left\{ \frac{1}{s_{\alpha_0}^2} + \frac{\mathbf{1}_n^T \boldsymbol{\Sigma}(\phi_{\alpha_0})^{-1} \mathbf{1}_n}{\sigma_{\alpha_0}^2} \right\}^{-1} \text{ and } m_0 = \frac{\mathbf{1}_n^T \boldsymbol{\Sigma}(\phi_{\alpha_0})^{-1} \boldsymbol{\alpha}_0}{\sigma_{\alpha_0}^2}.$$

- (b) Mean on spatial coefficients for scalar predictor:

$$[\mu_\alpha | \dots] \propto N(M m, M), \text{ where } M = \left\{ \frac{1}{s_\alpha^2} + \frac{\mathbf{1}_n^T \boldsymbol{\Sigma}(\phi_\alpha)^{-1} \mathbf{1}_n}{\sigma_\alpha^2} \right\}^{-1} \text{ and } m = \frac{\mathbf{1}_n^T \boldsymbol{\Sigma}(\phi_\alpha)^{-1} \boldsymbol{\alpha}}{\sigma_\alpha^2}.$$

- (c) Mean on spatial coefficients for FPC scores: for $r = 1, \dots, p$,

$$[\mu_{\beta_r} | \dots] \propto N(M_r m_r, M_r), \text{ where } M_r = \left\{ \frac{1}{s_{\beta_r}^2} + \frac{\mathbf{1}_n^T \boldsymbol{\Sigma}(\phi_{\beta_r})^{-1} \mathbf{1}_n}{\sigma_{\beta_r}^2} \right\}^{-1} \text{ and } m_r = \frac{\mathbf{1}_n^T \boldsymbol{\Sigma}(\phi_{\beta_r})^{-1} \boldsymbol{\beta}_r}{\sigma_{\beta_r}^2}.$$

5. **Spatial range parameters on spatial priors:** sampling through M-H algorithms as follows.

(a) Sampling ϕ_ξ at each iteration j :

i Draw ϕ_ξ^* from proposal distribution $q\{\phi_\xi^*|\phi_\xi^{(j-1)}\} \sim TN\{\phi_\xi^{(j-1)}, s_{prop,\xi}^2; 0, \infty\}$, truncated Normal distribution, and $s_{prop,\xi}^2$ is set to achieve the acceptance rate 25-50%.

ii Calculate acceptance ratio $r_\xi = \min\left\{1, \frac{p(\phi_\xi^*|\cdot)q\{\phi_\xi^{(j-1)}|\phi_\xi^*\}}{p\{\phi_\xi^{(j-1)}|\cdot\}q\{\phi_\xi^*|\phi_\xi^{(j-1)}\}}\right\}$.

Here, $p(\phi_\xi^*|\cdot) = \sum_{k,r} f(\boldsymbol{\xi}_{kr}|\phi_\xi^*)\pi_{\phi_\xi}(\phi_\xi^*)$, where $f(\cdot|\phi_\xi^*)$ is a density from multivariate normal distribution for $\boldsymbol{\xi}_{kr}$ in (S1.3) conditional on range parameter ϕ_ξ^* , and $\pi_{\phi_\xi}(\cdot)$ denotes a density from the truncated Normal prior distribution of range parameter ϕ_ξ in (S1.6).

iii Generate $\omega \sim \text{Uniform}(0, 1)$. Then $\phi_\xi^{(j)} = \begin{cases} \phi_\xi^*, & \text{if } r_\xi > \omega. \\ \phi_\xi^{(j-1)}, & \text{otherwise.} \end{cases}$

(b) Sampling ϕ_β at each iteration j :

i Draw ϕ_β^* from proposal distribution $q\{\phi_\beta^*|\phi_\beta^{(j-1)}\} \sim TN\{\phi_\beta^{(j-1)}, s_{prop,\beta}^2; 0, \infty\}$.

ii Calculate acceptance ratio $r_\beta = \min\left\{1, \frac{p(\phi_\beta^*|\cdot)q\{\phi_\beta^{(j-1)}|\phi_\beta^*\}}{p\{\phi_\beta^{(j-1)}|\cdot\}q\{\phi_\beta^*|\phi_\beta^{(j-1)}\}}\right\}$.

Here, $p(\phi_\beta^*|\cdot) = \sum_r f(\boldsymbol{\beta}_r|\phi_\beta^*)\pi_{\phi_\beta}(\phi_\beta^*)$, where $f(\cdot|\phi_\beta^*)$ is a density from multivariate normal distribution for $\boldsymbol{\beta}_r$ in (S1.3) conditional on range parameter ϕ_β^* , and $\pi_{\phi_\beta}(\cdot)$ denotes a density from the prior distribution of range parameter in (S1.6).

iii Generate $\omega \sim \text{Uniform}(0, 1)$. Then $\phi_\beta^{(j)} = \begin{cases} \phi_\beta^*, & \text{if } r_\beta > \omega. \\ \phi_\beta^{(j-1)}, & \text{otherwise.} \end{cases}$

(c) Sampling ϕ_{α_0} and ϕ_α at each iteration j follows (b) by replacing β with α_0 and α , respectively.

6. **Model selection parameters:** for $r = 1, \dots, p$,

(a) $[\sigma_{v_r}^2 | \dots] \propto IG\{\frac{n}{2} + A_v, \frac{1}{2}(\mathbf{v}_r - \mu_{v_r}\mathbf{1}_n)^T \boldsymbol{\Sigma}(\phi_v)^{-1}(\mathbf{v}_r - \mu_{v_r}\mathbf{1}_n) + B_v\}$.

(b) $[\mu_{v_r} | \dots] \propto N(M_r m_r, M_r)$, where $M_r = \left\{\frac{1}{s_v^2} + \frac{\mathbf{1}_n^T \boldsymbol{\Sigma}(\phi_v)^{-1} \mathbf{1}_n}{\sigma_v^2}\right\}^{-1}$ and $m_r = \frac{\mathbf{1}_n^T \boldsymbol{\Sigma}(\phi_v)^{-1} \mathbf{v}_r}{\sigma_v^2}$.

(c) Sampling ϕ_v at each iteration j by M-H algorithm:

i Draw ϕ_v^* from proposal distribution $q\{\phi_v^*|\phi_v^{(j-1)}\} \sim TN\{\phi_v^{(j-1)}, s_{prop,v}^2; 0, \infty\}$.

ii Calculate acceptance ratio $r_v = \min\left\{1, \frac{p(\phi_v^*|\cdot)q\{\phi_v^{(j-1)}|\phi_v^*\}}{p\{\phi_v^{(j-1)}|\cdot\}q\{\phi_v^*|\phi_v^{(j-1)}\}}\right\}$.

Here, $p(\phi_v^*|\cdot) = \sum_r f(\mathbf{v}_r|\phi_v^*)\pi_{\phi_v}(\phi_v^*)$, where $f(\cdot|\phi_v^*)$ is a density from multivariate normal distribution for \mathbf{v}_r in (S1.3) conditional on range parameter ϕ_v^* , and $\pi_{\phi_v}(\cdot)$ denotes a density from the prior distribution of range parameter in (S1.6).

iii Generate $\omega \sim \text{Uniform}(0, 1)$. Then $\phi_v^{(j)} = \begin{cases} \phi_v^*, & \text{if } r_v > \omega. \\ \phi_v^{(j-1)}, & \text{otherwise.} \end{cases}$

(d) Sampling \mathbf{v}_r at each iteration j by M-H algorithm:

i Draw \mathbf{v}_r^* from proposal distribution $q\{\mathbf{v}_r^*|\mathbf{v}_r^{(j-1)}\} \sim N\{\mathbf{v}_r^{(j-1)}, s_{prop,\mathbf{v}_r}^2 \mathbf{I}_n\}$.

ii Calculate acceptance ratio $r_{\mathbf{v}_r} = \min\left\{1, \frac{p(\mathbf{v}_r^*|\cdot)}{p(\mathbf{v}_r^{(j-1)}|\cdot)}\right\}$.

Here, $p(\mathbf{v}_r^*|\cdot) = \sum_{k,i} f\{Y_k(\mathbf{s}_i)|\mathbf{v}_r^*\}\pi_{\mathbf{v}_r}(\mathbf{v}_r^*)$, where $f(\cdot|\mathbf{v}_r^*)$ is a density from normal distribution for $Y_k(\mathbf{s}_i)$ in (S1.1) conditional on \mathbf{v}_r^* , and $\pi_{\mathbf{v}_r}(\cdot)$ denotes a density from multivariate normal distribution in (S1.3).

iii Generate $\omega \sim \text{Uniform}(0, 1)$. Then $\mathbf{v}_r^{(j)} = \begin{cases} \mathbf{v}_r^*, & \text{if } r_{\mathbf{v}_r} > \omega. \\ \mathbf{v}_r^{(j-1)}, & \text{otherwise.} \end{cases}$

S3 MCMC Diagnostics

We examine the MCMC convergence on corn yield data analysis using two diagnostics tools; (i) trace plots and (ii) potential scale reduction factor (PSFR; Gelman and Rubin, 1992). Figure S1 presents trace plots for posterior samples obtained through MCMC implementation based on a total of 15,000 iterations, where a posterior sample of size 2500 is acquired by using the first 5,000 iterations as burn-in and thinning the remaining 10,000 by a factor of 4. Figure S1 shows posterior samples of σ_e^2 , $\alpha_0(\mathbf{s}_1)$, $\alpha(\mathbf{s}_1)$, $\beta_1(\mathbf{s}_1)$, $\xi_{11}(\mathbf{s}_1)$, and ϕ_α . We then calculate the PSFR to assess and summarize the convergence of MCMC algorithm, using the `coda` package (Plummer et al., 2006) in R. A PSRF close to 1 indicates convergence of a Markov chain, while a large PSRF means that the chain has not yet converged. The

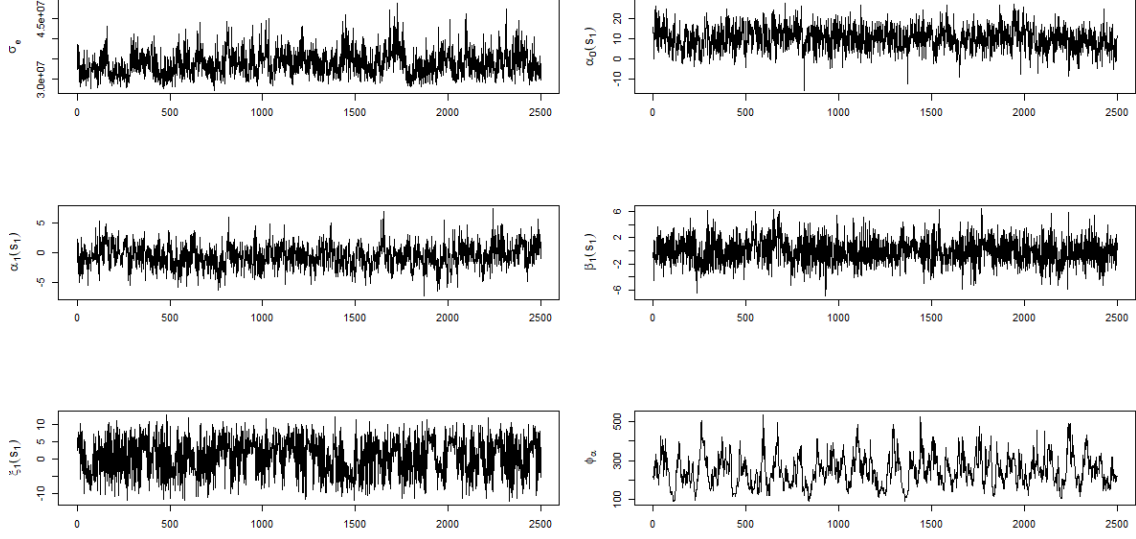


Figure S1: Trace plots for σ_e^2 , $\alpha_0(\mathbf{s}_1)$, $\alpha(\mathbf{s}_1)$, $\beta_1(\mathbf{s}_1)$, $\xi_{11}(\mathbf{s}_1)$, and ϕ_α from corn yield prediction analysis.

Table S1: Mean and median of potential scale reduction factor (PSRF)

	$\alpha(\mathbf{s}_i)$	$\beta_1(\mathbf{s}_i)$	$\beta_2(\mathbf{s}_i)$	$\beta_3(\mathbf{s}_i)$	$\beta_4(\mathbf{s}_i)$	$\beta_5(\mathbf{s}_i)$
Mean(RSPF)	1.13	1.05	1.06	1.00	1.00	1.02
Median(RSPF)	1.03	1.05	1.06	1.00	1.00	1.02

reference value of 1.1 was recommended by Gelman et al. (2014) and generally adopted by MCMC practitioners. To calculate the RSRF, we first run three sets of MCMC iterations for our corn yield data, following the same sampling and burn-in process. We then compute PSRFs for the regression coefficients, $\alpha(\mathbf{s}_i)$, $\beta_1(\mathbf{s}_i)$, \dots , $\beta_5(\mathbf{s}_i)$, for $i = 1, \dots, 403$. Table S1 shows the mean and median of PSRFs from the Markov chains of the regression coefficients. As we can see, most means and medians of PSRF are all very close to 1, indicating the chains have converged. The only exception is that the mean RSPF for $\alpha(\mathbf{s}_i)$ is slightly over 1.1, which might be due to the relatively weak effect of the precipitation on crop yield, as the regions with good irrigation systems may depend less on the precipitation. The weak signal leads to slower convergence in those locations, but the inflation of mean RSPF is tiny and the median RSPF is well below 1.1, so the convergence of $\alpha(\mathbf{s}_i)$ should not be a serious concern.

S4 Additional Monte Carlo Experiments and Results

S4.1 Performance Evaluation of Parameter Estimation

In the manuscript, we present the superior prediction performance of our proposed model estimated by MCMC implementation. Besides the prediction, in practice, the parameter identifiability is also important for the validity of the model interpretation based on the estimated coefficients. Since model interpretation is heavily centered on the regression coefficients α , β_1 , and β_2 , we examined the performance of their estimates in our simulations. Under the simulation setting described in Section 4.1, with a moderator spatial correlation range parameter $\phi = 200$, we calculated the Pearson correlation between the parameter estimates (posterior means) and the true parameters. The Pearson correlations are 0.73, 0.80, and 0.75 for α , β_1 , and β_2 , respectively. Figure S2 displays the scatter plots between the estimated and true parameters, pooling from five randomly selected simulation runs. We observe from Figure S2 that most of the points lie around the 45-degree reference line, which illustrates the desirable performance of our Bayesian estimators.

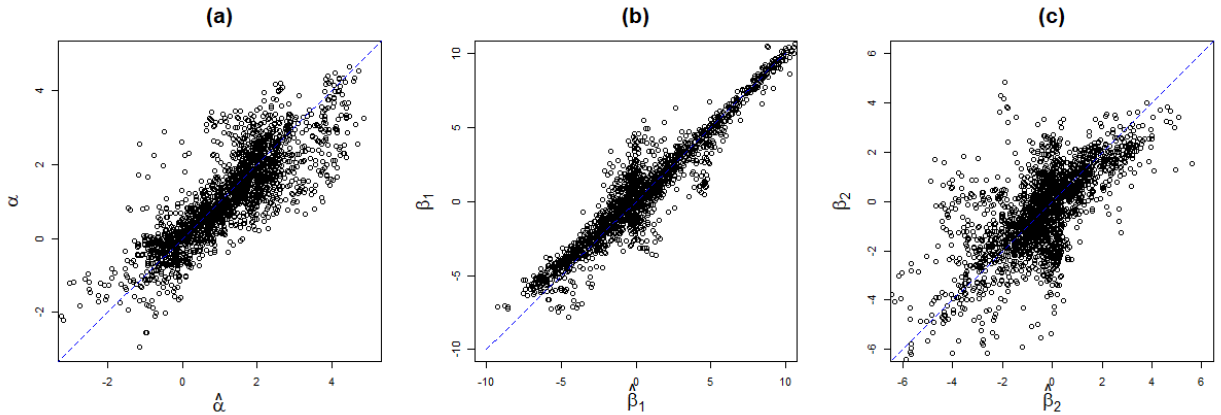


Figure S2: Scatter plots of estimated coefficients vs. true coefficients from five randomly selected simulation sets for (a) regression coefficient α for the scalar covariate, (b) regression coefficient β_1 for the first FPC scores, and (c) regression coefficient β_2 , for the second FPC scores.

S4.2 Prediction Performance in High Dimension

We conduct additional Monte Carlo experiments to examine the performance of our model under a higher dimension setting. Specifically, we double the number of scalar predictors and the number of functional covariates to $h = 2$ and $q = 4$, respectively. We also double the number of FPC's to $p = 4$. Following the same settings of the spatial location \mathbf{s}_i as in Section 4.1, we generate five replicates of the responses process based on

$$Y_k(\mathbf{s}_i) = \sum_{j=1}^2 Z_{kj}(\mathbf{s}_i) \alpha_j(\mathbf{s}_i) + \sum_{r=1}^4 \xi_{kr}(\mathbf{s}_i) \{\beta_r(\mathbf{s}_i) \gamma_r(\mathbf{s}_i)\} + e_k(\mathbf{s}_i),$$

where k and i are indices for years and counties, respectively, $\mathbf{Z}_{kj}(\mathbf{s}_i) \stackrel{\text{iid}}{\sim} \text{unif}(0, 3)$, for $j = 1, 2$. While the data generation settings for $\beta_1, \beta_2, \xi_{k1}, \xi_{k2}, \gamma_1$, and γ_2 , are kept the same as specified in Section 4.1, we generate $\alpha_1 \sim N\{2 \cdot \mathbf{1}_n, \Sigma(\phi)\}$, $\alpha_2 \sim N\{-2 \cdot \mathbf{1}_n, \Sigma(\phi)\}$, $\xi_{k3} \sim N\{\mathbf{0}_n, 2\Sigma(\phi)\}$, $\xi_{k4} \sim N\{\mathbf{0}_n, \Sigma(\phi)\}$, $\beta_3 \sim N\{1.5 \cdot \mathbf{1}_n, 9\Sigma(\phi)\}$, $\beta_4 \sim N\{-1.5 \cdot \mathbf{1}_n, 4\Sigma(\phi)\}$. The binary indicator variables γ_3 and γ_4 , defined based on (3) of the manuscript, are generated by $\mathbf{v}_3 \sim N\{\mathbf{0}_n, \Sigma(\phi)\}$ and $\mathbf{v}_4 \sim N\{\mathbf{0}_n, \Sigma(\phi)\}$, respectively. The correlation matrix $\Sigma(\phi)$ is governed by a Matérn correlation function with the spatial correlation at a moderate level $\phi = 200$ and smoothness parameter $\kappa = 1$. The errors $e_k(\mathbf{s}_i) \stackrel{\text{iid}}{\sim} N(0, 2^2)$. We then generate the 4-dimensional multivariate functional predictors as

$$\mathbf{W}_k(\mathbf{s}_i; t) = \sum_{r=1}^4 \xi_{kr}(\mathbf{s}_i) \mathbf{f}_r(t) + \mathbf{u}_k(\mathbf{s}_i; t),$$

where $\mathbf{f}_r(t) = \{f_{r1}(t), \dots, f_{r4}(t)\}^T$ with $f_{r1}(t) = \cos(2r\pi t)/\sqrt{2}$, $f_{r2}(t) = \sin(2r\pi t)/\sqrt{2}$, $f_{r3}(t) = \cos\{2(r+4)\pi t\}/\sqrt{2}$, $f_{r4}(t) = \sin\{2(r+4)\pi t\}/\sqrt{2}$, $t \in [0, 1]$, for $r = 1, \dots, 4$, satisfying $\int_0^1 \mathbf{f}_r^T(t) \mathbf{f}_{r'}(t) dt = I(r = r')$. The measurement errors $\mathbf{u}_k(\mathbf{s}_i; t) = \{u_{k1}(\mathbf{s}_i; t), \dots, u_{k4}(\mathbf{s}_i; t)\}^T$ are generated from $u_{kl}(\mathbf{s}_i; t) \stackrel{\text{iid}}{\sim} N(0, 1^2)$, for $l = 1, \dots, 4$. Functional trajectories are generated at a regular grid of 100 points in $[0, 1]$. By following the same MCMC implementation steps described in Section 4.1 for 100 simulation sets, we calculate the prediction error based on randomly selected 20% of observations as the testing data and the rest as training input.

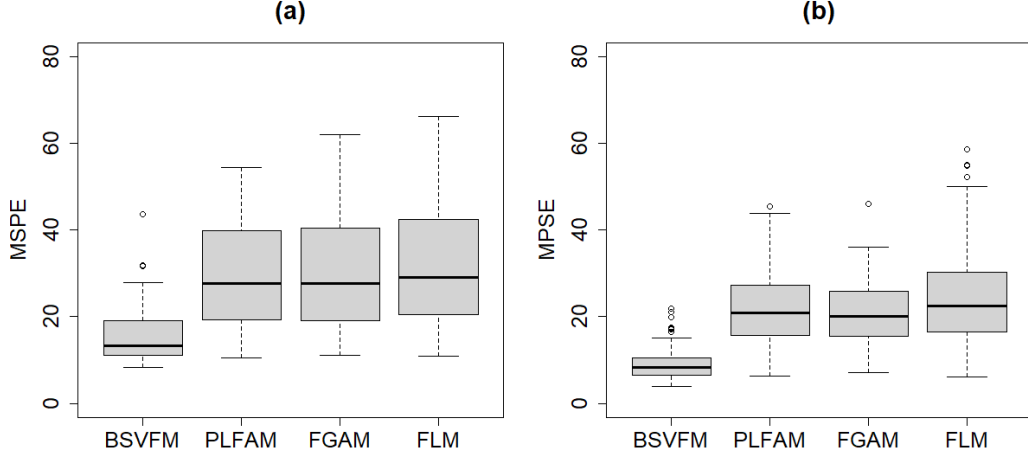


Figure S3: Box plots of Mean Squared Prediction Errors (MSPE) from the proposed BSVFM and the competing models, PLFAM, FGAM and FLM (a) when $h = 2$ and $q = 4$, and (b) when $h = 1$ and $q = 2$.

We fix $p = 4$ for our proposed BSVFM. Then we calculate prediction errors for competitors, PLFAM, FGAM, and FLM. For competing methods involving mFPC scores in the model, PLFAM and FLM, we determine the number of FPC that recovers at least 99.5% of the total variation.

Figure S3 (a) shows the MSPE under the new setting from our proposed BSVFM, and the competing methods including PLFAM, FGAM, and FLM. As we can see, our proposed method still outperforms the competing methods when the dimensions of the scalar and functional predictors are higher. For comparison, we present the results for $h = 1$, $q = 2$ and $\phi = 200$ in Figure S3 (b), which corresponds to Figure 3 (b) in the manuscript. When comparing the two panels in Figure S3, we observe overall larger MSPE's among all models under the high dimension setting. This is because the average marginal variance of Y is 62.1 under the new setting with more scalar and functional predictors, whereas the average marginal variance of Y is 51.3 under the old setting with $h = 1$ and $q = 2$.

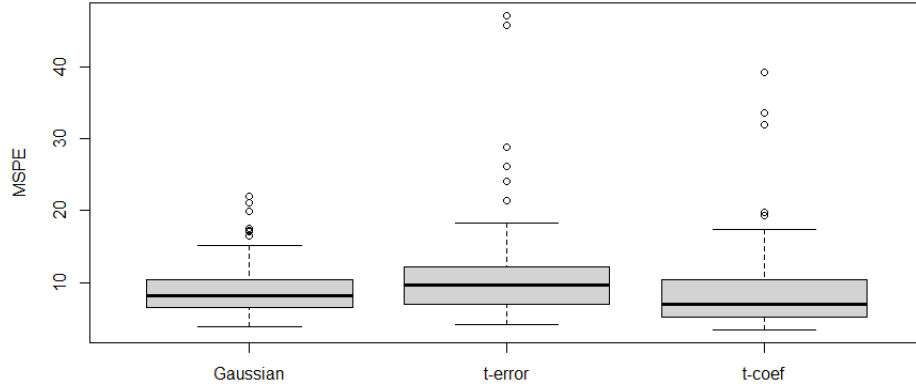


Figure S4: Boxplots of Mean Squared Prediction Errors (MSPE) from the proposed BSVFM under three data generation settings: ‘Gaussian’ where our model is correctly specified, ‘ t -error’ where the model and measurement errors follow t -distributions with 3 d.f., and ‘ t -coef’ where the regression coefficients and FPC scores follow multivariate t -distributions with 3 d.f.

S4.3 Robustness against Distribution Misspecifications

We have conducted additional simulation studies to examine the robustness of our model to distribution misspecification. Our proposed model, specified in (4)-(6) of the manuscript, assumes Gaussian distribution on error terms and the coefficients. We therefore consider two non-Gaussian distribution scenarios in the data generation:

- (i) ‘ t -error’ - both the model error $e_k(\mathbf{s}_i)$ and the measurement error $\mathbf{u}_k(\mathbf{s}_i; t)$ of the functional covariates follow t distributions with 3 degrees of freedom (d.f.)
- (ii) ‘ t -coef’ - the FPC scores, $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$, and the regression coefficients, $\boldsymbol{\alpha}$, $\boldsymbol{\beta}_1$, and $\boldsymbol{\beta}_2$, follow multivariate t distributions with 3 d.f.

All t distributed variables are re-scaled to have the same variance as their Gaussian counterparts as described in Section 4.1. The remaining settings are the same as Section 4.1, and we fix the range parameter of the spatial correlations at $\phi = 200$ which corresponds to the moderate spatial correlation considered in Section 4.1. The simulations are repeated 100 times.

Figure S4 shows the Mean Squared Prediction Errors (MSPE) of our method under three

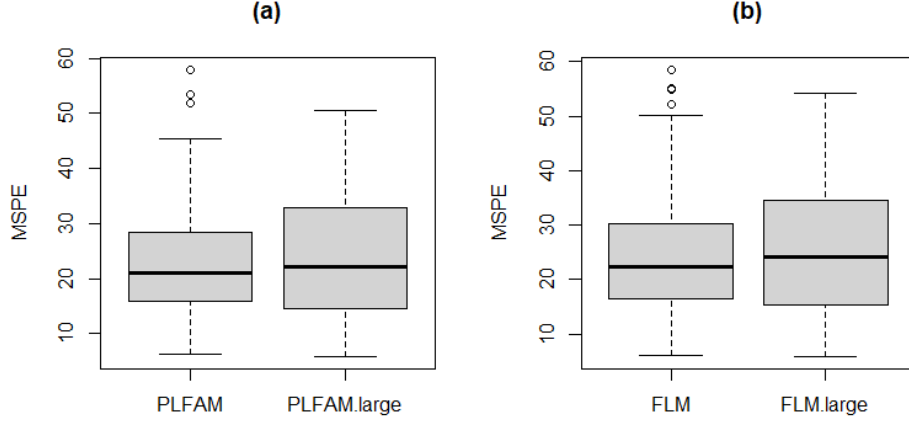


Figure S5: Boxplots of Mean Squared Prediction Errors (MSPE) from (a) PLFAM and (b) FLM by using different numbers p of mFPC scores. In each plot, the results are based on either a fixed $p = 2$ or a larger p chosen by 99.9% FVE.

data generation settings: the original ‘Gaussian’ data described in Section 4.1, the ‘ t -error’ and ‘ t -coef’ scenarios described above. Even under the two non-Gaussian settings where our model is misspecified, our method still achieves comparable MSPE as with the Gaussian data. Without surprise, we notice that the MSPE’s under the two non-Gaussian settings show more variability and outliers as Y now follows a heavy-tailed distribution. The median MSPE under the ‘ t -coef’ seems even slightly lower than the ‘Gaussian’ setting. We found this could be because the marginal variances of the response Y under the ‘ t -coef’ setting is slightly lower than under the ‘Gaussian’ after re-scaling the t -coefficients to match the variance of their Gaussian counterparts.

S4.4 Further Experiments on FLM and PLFAM

The true number of FPCs in our simulated data from Section 4.1 is $p = 2$, and on average the top 2 FPCs explain 99.3% of the total variation in the functional predictors. Among the three competing functional data methods, PLFAM and FLM are based on FPC scores. Following the referee’s suggestion, we have refitted the PLFAM and FLM with the number of FPCs chosen by 99.9% fraction of variation explained (FVE), which leads to $p = 11$ on

average. Figure S5 compares the MSPE of PLFAM and FLM with a fixed $p = 2$ vs a p chosen by 99.9% FVE. There seems to be no clear advantage of using a more extensive set of FPCs in our simulation study. Furthermore, we observe higher variation in the MSPE when using a larger p , possibly due to the overfitting of the training data.

S5 Additional Corn Yield Data Analysis Results

S5.1 Further Exploration of Machine Learning Methods

Existing machine learning literature on crop yield prediction uses temperature information in various ways, including (i) daily maximum and minimum temperatures (730 features each year) (Chu and Yu, 2020), (ii) weekly averages of the maximum of minimum temperatures (104 features) (Schwalbert et al., 2020; Khaki et al., 2020), or (iii) monthly averages of the maximum and minimum temperatures (24 features) (Khaki and Wang, 2019). While our paper followed option (i), we carry out analysis with reduced temperature information for the exploration. We therefore additionally ran the Neural Network (NN) and XGBoost (XGB) using weekly and monthly averages of max/min temperatures as input. We further consider directly using the 5 FPC scores identified in our model fitting as the input for machine learning methods.

Table S2 presents the MSPE by 10-fold cross-validation from Neural Network (NN) and XGBoost (XGB) using the daily, weekly, and monthly temperatures and 5 FPC scores as input. It appears that using reduced temperature information does not result in better prediction performance. Chu and Yu (2020) empirically illustrated through the real data

Table S2: MSPE and weighted MSPE based on the 10-fold cross validation calculated from BSVFM, Neural Network (NN), and XGBoost (XGB). The NN, NN(w), NN(m) and NN(fpc) represent the Neural Network model fitted using daily temperatures, weekly averaged temperatures, monthly averaged temperatures, and five mFPC scores, respectively. The XGB, XGB(w), XGB(m), and XGB(fpc) are named in the same way.

	BSVFM	NN	NN(w)	NN(m)	NN(fpc)	XGB	XGB(w)	XGB(m)	XGB(fpc)
MSPE	425.5	714.4	925.7	1005.1	817.2	448.3	507.4	448.5	710.8
Weighted MSPE	355.7	586.5	744.7	851.6	681.1	367.9	387.9	445.9	587.9

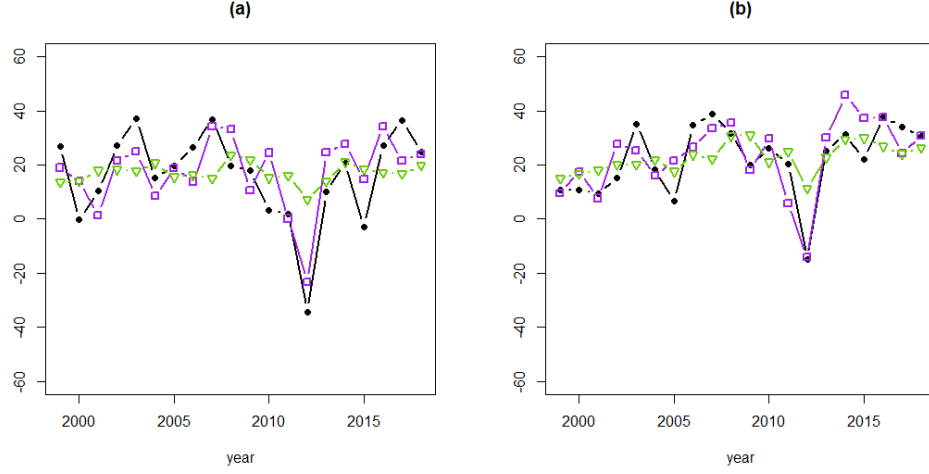


Figure S6: Observed corn yield after subtracting the year effect (\bullet) and the fitted corn yield using our BSVFM (\square) and using the SVCM (∇) over years from two randomly selected counties in (a) and (b).

analysis that the optimal choice for the number of hidden layers in NN could address the potential overfitting issue. We used cross-validation to choose the number of hidden layers in the NN, and the lowest MSPE of NN using the daily temperature in our results seems to corroborate their conclusion. The same phenomenon is observed for the XGBoost approach. The comparison of all methods using 5 FPCs indicates that our spatial functional regression approach outperforms the machine learning methods in corn yield prediction given the same information.

S5.2 Predictive Performance Comparison of BSVFM and SVCM

In Section 5.2, we considered a non-functional spatially varying coefficient Model (SVC) as one of competing methods, which simply replaces the functional temperature covariates in our model by the annual averages of maximum and minimum temperature. Figure S6 illustrates the comparison of prediction performance between our functional BSVFM and non-functional SVC for two randomly selected counties over the years, where the y-axis is the corn yield after removing the year effects. As discussed in Section 5.1, the year effects in crop yield can be largely explained by different genetic variants of corn being planted each year and

land fallows, and we remove these effects to better unveil the effects of temperature. We observe that our BSVFM outperforms the SVCM by tracking the true trajectory of Y more closely. This follows our intuition as the FPC scores capture the major modes of temperature trajectories and thus are likely able to catch the daily temperature influence on crop growth better than simple summary statistics of temperatures.

S5.3 Additional Figures from the Crop Yield Prediction Application

Figure S7 shows the spatially varying behaviors in posterior mean of coefficients $\beta_4(\mathbf{s})$ and $\beta_5(\mathbf{s})$ for the fourth and fifth FPC scores. As the model diagnostic, we calculate the autocorrelation function (ACF) for each county based on the residuals from the fitted yield prediction model. Figure S8 shows the ACF for randomly selected four counties with the complete corn yield data over 22 years. All ACF's fall within the confidence band based on the assumption of no temporal dependency, which supports the conditional independence assumption that we make.

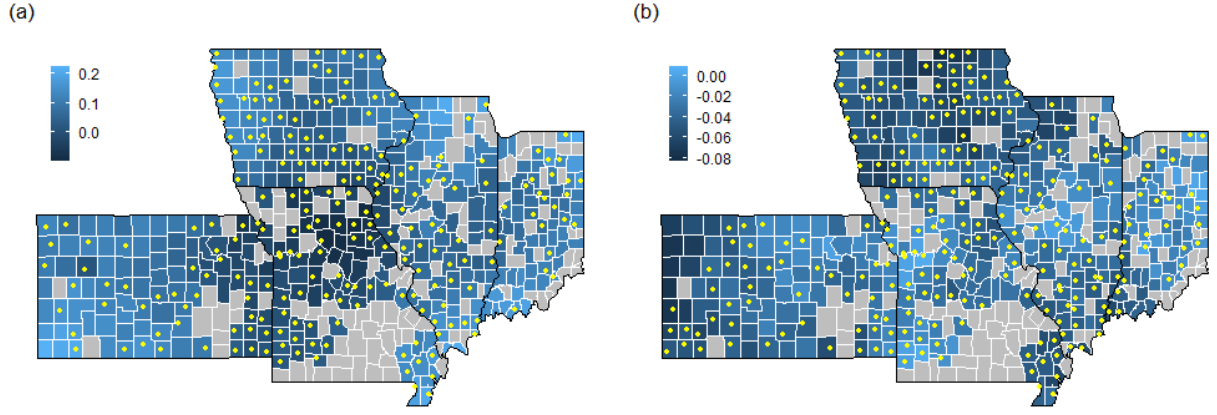


Figure S7: Posterior means of the spatially varying coefficients of (a) $\beta_4(\mathbf{s}_i)$ and (d) $\beta_5(\mathbf{s}_i)$ over Kansas, Iowa, Illinois, Indiana, and Missouri. Counties with non-null coefficients are marked by yellow circle. Counties with missing data are colored by grey.

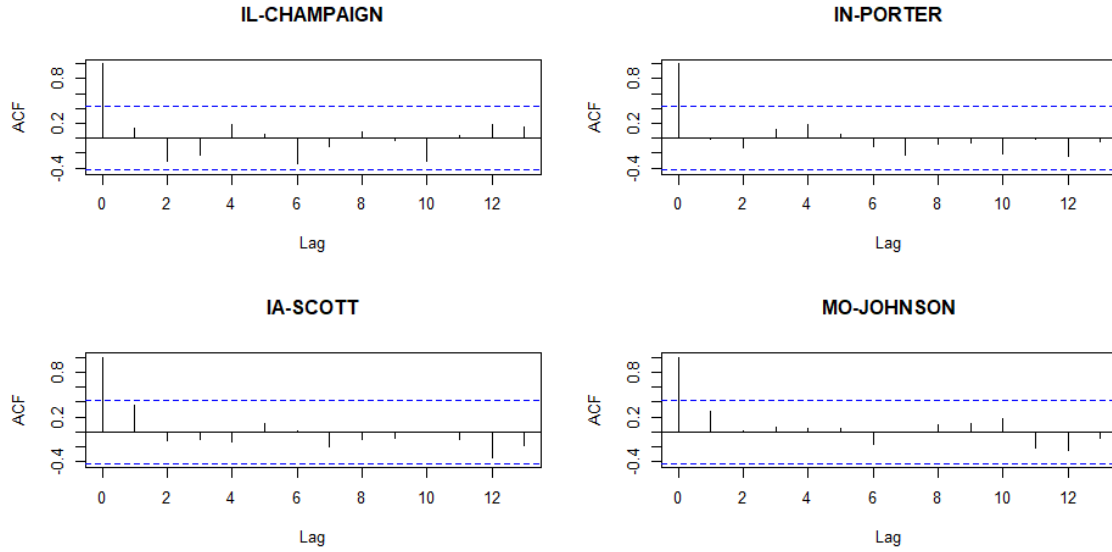


Figure S8: The ACF plot for randomly selected four counties with the complete corn yield information over 22 years, based on the residuals from the corn yield prediction model.

References

- Chu, Z. and Yu, J. (2020). An end-to-end model for rice yield prediction using deep learning fusion. *Computers and Electronics in Agriculture*, 174:105471.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2014). *Bayesian data analysis*. Chapman and Hall/CRC.
- Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):457–472.
- Khaki, S. and Wang, L. (2019). Crop yield prediction using deep neural networks. *Frontiers in Plant Science*, 10.
- Khaki, S., Wang, L., and Archontoulis, S. V. (2020). A cnn-rnn framework for crop yield prediction. *Frontiers in Plant Science*, 10.
- Kowal, D. R., Matteson, D. S., and Ruppert, D. (2017). A bayesian multivariate functional dynamic linear model. *Journal of the American Statistical Association*, 112:733–744.
- Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). Coda: convergence diagnosis and output analysis for mcmc. *R News*, 6(1):7–11.
- Schwalbert, R. A., Amado, T., Corassa, G., Pott, L. P., Prasad, P., and Ciampitti, I. A. (2020). Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in southern brazil. *Agricultural and Forest Meteorology*, 284:107886.