



Crop Yield Prediction Using Bayesian Spatially Varying Coefficient Models with Functional Predictors

Yeonjoo Park, Bo Li & Yehua Li

To cite this article: Yeonjoo Park, Bo Li & Yehua Li (2022): Crop Yield Prediction Using Bayesian Spatially Varying Coefficient Models with Functional Predictors, Journal of the American Statistical Association, DOI: [10.1080/01621459.2022.2123333](https://doi.org/10.1080/01621459.2022.2123333)

To link to this article: <https://doi.org/10.1080/01621459.2022.2123333>



View supplementary material [↗](#)



Published online: 10 Oct 2022.



Submit your article to this journal [↗](#)



Article views: 217



View related articles [↗](#)



View Crossmark data [↗](#)



Crop Yield Prediction Using Bayesian Spatially Varying Coefficient Models with Functional Predictors

Yeonjoo Park^a, Bo Li^b, and Yehua Li^c 

^aManagement Science and Statistics, University of Texas at San Antonio, San Antonio, TX; ^bDepartment of Statistics, University of Illinois at Urbana-Champaign, Urbana, IL; ^cDepartment of Statistics, University of California at Riverside, Riverside, CA

ABSTRACT

Reliable prediction for crop yield is crucial for economic planning, food security monitoring, and agricultural risk management. This study aims to develop a crop yield forecasting model at large spatial scales using meteorological variables closely related to crop growth. The influence of climate patterns on agricultural productivity can be spatially inhomogeneous due to local soil and environmental conditions. We propose a Bayesian spatially varying functional model (BSVFM) to predict county-level corn yield for five Midwestern states, based on annual precipitation and daily maximum and minimum temperature trajectories modeled as multivariate functional predictors. The proposed model accommodates spatial correlation and measurement errors of functional predictors, and respects the spatially heterogeneous relationship between the response and associated predictors by allowing the functional coefficients to vary over space. The model also incorporates a Bayesian variable selection device to further expand its capacity to accommodate spatial heterogeneity. The proposed method is demonstrated to outperform other highly competitive methods in corn yield prediction, owing to the flexibility of allowing spatial heterogeneity with spatially varying coefficients in our model. Our study provides further insights into understanding the impact of climate change on crop yield. Supplementary materials for this article are available online.

ARTICLE HISTORY

Received August 2021
Accepted August 2022

KEYWORDS

Bayesian hierarchical model;
Scalar-on-function
regression; Spatial functional
data; Spatial model selection;
Spatially varying coefficient
model

1. Introduction

Crop yield prediction is a fundamental problem in Agriculture and Economics. In the United States, agriculture is a major industry by contributing 136.1 billion dollars to the national gross domestic product (GDP) and providing 11% of total employment in 2019 (USDA 2020b). The United States is also a vital contributor to global food security as a leading international producer of agricultural products. The American Midwest is especially one of the world's largest crop production areas with over 127 million acres of agricultural land and produces 85% of U.S. corn and soybean (USDA 2020a). Specifically, corn is a primary grain crop globally, widely used for human food, livestock feed, and biofuel (ethanol production). Timely and accurate crop yield prediction in the Midwest is crucial for policy-makers and planners to create appropriate strategies for the storage, distribution, and trade of products for agricultural risk management.

We aim to develop a crop yield forecasting model at large spatial scales to predict the corn yield for five Midwestern states. In this region, where the well-known “Corn Belt” lies, soils are deep, fertile, and rich in organic material and nitrogen, and the land is relatively level. The warm nights, hot days, and well-distributed rainfall of the region during the growing season are ideal conditions for raising corn. It is well known that climate variables, particularly temperature and precipitation, have a

significant impact on all stages of plant growth and hence affects agricultural productivity. Ray et al. (2015) estimated that 60% of corn yield variability and 36% of the soybean yield variability in the American Midwest can be explained by climate variability. Various predictive models have been developed by crop scientists for production monitoring, including process-based simulation models under comprehensive mechanism processes (Guan et al. 2017; Jones et al. 2017; Peng et al. 2018), empirical statistical crop models based on historical climate observations and yield measurements (Prasad et al. 2006; Gornott and Wechsung 2016; Kern et al. 2018), and machine learning methods in recent studies (Kang et al. 2020; van Klompenburg, Kassahun, and Catal 2020). These approaches typically use a summary of climate information, such as annual or monthly aggregated climate data, rather than the complete trajectories of meteorological variables in their prediction models.

Since measurements of meteorological variables, such as maximum and minimum temperatures, are available on a daily basis and their effects on crop yield vary at a different growing stage of the crop, it is natural to treat the trajectories of these variables as functional data in the predictive model for crop yield. Indeed, recent studies have shown that more accurate crop yield prediction can be obtained by using the complete temperature trajectories as functional predictors (Wong, Li, and Zhu 2019). On the other hand, estimating the potential impacts of weather patterns on crop productivity is of

widespread interest to agriculture and the economy as climate change becomes an imminent challenge for humanity (Lobell and Burke 2010; Holzkämper, Calanca, and Fuhrer 2012). Both aspects suggest investigating the relationship between crop yield and meteorology trajectories, for which functional regression models are a suitable tool.

The most widely used regression models for scalar responses against functional covariates are the functional linear models (Müller and Stadtmüller 2005; James, Wang, and Zhu 2009; Li, Wang, and Carroll 2010; Goldsmith et al. 2011; Zhao, Chen, and Ogden 2015). There has been some recent work on nonlinear regression models for independent functional data, including the functional additive models (Müller and Yao 2008; Zhu, Yao, and Zhang 2014; Fan, James, and Radchenko 2015), the functional generalized additive models (McLean et al. 2014), and the partially linear functional additive model (Wong, Li, and Zhu 2019; Cui, Lin, and Lian 2020). In addition, mixed models have been developed to accommodate the dependency of functional covariates collected in groups or repeatedly over time (Crainiceanu, Staicu, and Di 2009; Goldsmith et al. 2012). See Reiss et al. (2017) for a thorough review for scalar-on-function regression methods.

Although different approaches have been proposed to analyze functional data observed over a spatial domain and model the spatial correlation between curves (Zhou et al. 2010; Staicu, Crainiceanu, and Carroll 2010; Zhang et al. 2016; Kuenzer, Hörmann, and Kokoszka 2021), to our best knowledge, spatial correlation of functional covariates is rarely considered in functional regression models. This limits the applicability of functional regression models to prediction problems where the functional covariates are collected over a spatial domain. As detailed in Section 2, our functional covariates for corn yield prediction consist of county-level temperature trajectories which are known to be spatially correlated (North, Wang, and Genton 2011; Di Cecco and Gouhier 2018). Furthermore, it is no longer realistic to assume a homogeneous relationship between response variables and predictors when data are observed over a large spatial domain. In our study of corn yield prediction using meteorological variables for five Midwest states, the effect of temperatures on crop yield may vary across different regions depending on the local environmental conditions, such as soil moisture, soil PH, solar radiation, and wind velocity. Since continuously monitoring all confounding factors in a large spatial region is not feasible, it is essential for the crop yield model to allow for spatial heterogeneity.

We propose a spatially varying functional regression model to incorporate spatial correlation between functional covariates and also allow the functional coefficients to vary over space. Our model can be considered an extension of the spatially varying coefficient model of Gelfand et al. (2003) to functional regressions. There are two main statistical challenges for this extension: low-rank representation of spatially dependent functional predictors and regularization to avoid over-fitting while respecting the heterogeneous relationship between crop yield and meteorological variables. To address the first challenge, we develop a fully Bayesian functional principal component analysis (FPCA) framework for spatially dependent functional data. The within-curve dependence is captured by a low-rank FPC representation based on nonparametric basis functions.

Some existing FPCA methods such as James, Hastie, and Sugar (2000), Zhou, Huang, and Carroll (2008), Zhou et al. (2010), and Huang, Li, and Guan (2014) used reduced rank basis function representations fitted via Expectation-Maximization algorithms under frequentist paradigms. In contrast, we propose to embed FPCA estimation into a fully Bayesian hierarchical model to ensure uncertainties associated with FPCA estimation propagate into crop yield prediction. We accommodate the between-curve spatial dependence by allowing FPC scores to be spatially correlated. Over-parameterization of our model is likely to occur if insignificant FPC scores are used in the regression. To remedy this potential issue while respecting the possible spatial heterogeneity, we introduce a build-in Bayesian spatially varying variable selection device to ensure model parsimony. We also govern all spatially varying parameters by correlated processes to further reduce the risk of over-fitting.

The rest of the article is organized as follows. We introduce the corn yield and meteorology data in Section 2 which also includes a preliminary analysis to explore the heterogeneous relationship between the corn yield and the functional meteorology predictors. We describe the proposed spatially varying coefficient functional regression model with a fully Bayesian hierarchical structure in Section 3 and conduct simulations to examine its prediction performance in Section 4. We present a comprehensive study of the corn yield prediction in Section 5 and conclude the article with discussions in Section 6. Detailed MCMC algorithm, convergence diagnostics, additional simulation and data analysis results are relegated to the supplementary materials.

2. Crop and Weather Data

We obtain the county-level annual crop yield data (measured in bushels per acre) together with the size of harvested land (acre) of Illinois, Indiana, Iowa, Kansas, and Missouri between 1999 and 2020 from the National Agricultural Statistics Agency (<https://quickstats.nass.usda.gov/>). We will focus on the corn yield, which is the most important crop in this region. The corn yield data availability at each county varies over time. A total of 403 counties, including 79 in Illinois, 66 in Indiana, 93 in Iowa, 92 in Kansas, and 73 in Missouri, have data available for at least five years between 1999 and 2020, among the 102, 92, 99, 105, and 114 counties from their respective states. The median number of years of available corn yield record for those 403 counties is 17 out of 22, and our analysis will be performed on those counties. The harvest land size also varies over the years and is collected from the annual Agricultural Survey. In our data, the harvest land size ranges from 200 to 394,000 acres with a median size of 70,300 acres.

The meteorology measurements for each county, including annual precipitation, daily maximum and minimum temperature are obtained from the National Climatic Data Center (NCDC, <https://www.ncdc.noaa.gov/data-access>). Figure 1 shows temperature trajectories collected from counties in Iowa and Kansas as examples. Different temperature characteristics are observed in different regions. We model the annual precipitation as a scalar covariate and the daily maximum and minimum temperature trajectories as spatially dependent bivari-

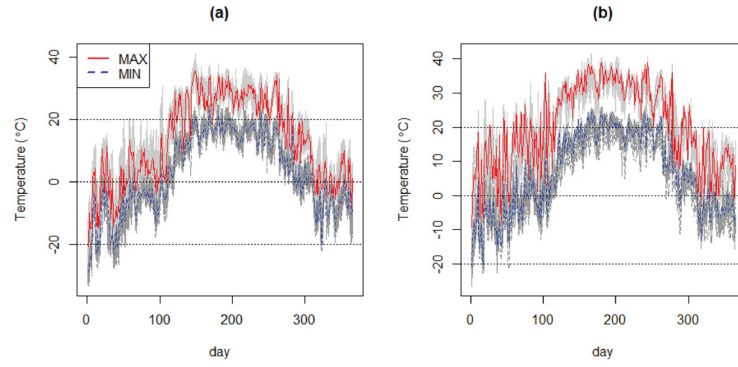


Figure 1. Daily maximum and minimum temperature trajectories collected from counties in (a) Iowa and (b) Kansas in the year 2018. Highlighted trajectories in each panel are from a randomly selected county from each state.

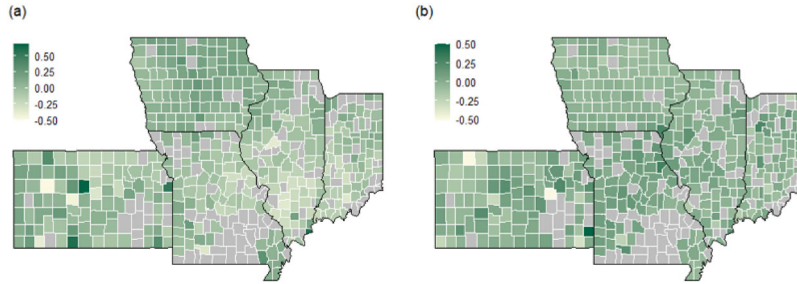


Figure 2. Spatial maps of estimated linear regression coefficients from the (a) first PC scores and the (b) second PC scores from counties in Illinois, Indiana, Iowa, Kansas, and Missouri. Counties with missing data are colored gray.

ate functional predictors. While the county-level agricultural data tend to have missing values on their historical records, the daily county-level temperature is available from the early 1900s for most areas of the United States from the NCDC. The long climate records allow us to reliably assess the long-term temperature pattern in each county.

A preliminary data exploration demonstrates heterogeneous relationships between corn yield and temperature trajectories. We treat the daily maximum and minimum temperature curves as multivariate functional data, conduct a multivariate FPCA (Wong, Li, and Zhu 2019), and fit a linear regression model on corn yield with the first and second FPC scores for each county independently, using different years as replicates. Figure 2 shows the estimated coefficients for all counties and coefficients exhibit a spatially clustered pattern. We then examine the spatial correlation of the coefficients by performing a Moran's I test to test the null hypothesis of no spatial correlation versus a one-sided alternative of positive spatial correlation. The resulting p -values < 0.001 for the regression coefficients on both the first and second FPC scores provide strong evidence of spatial correlation.

3. Bayesian Spatially Varying Functional Model

For ease of exposition, we present the model based on single-year data. Our analysis treats data from multiple years as conditionally independent replicates, with rationals detailed in Section 5.1. Let $Y(\mathbf{s})$ denote the scalar response at location $\mathbf{s} \in D$ for a spatial region $D \subset \mathbb{R}^2$, $\mathbf{X}(\mathbf{s}; t) = \{X_1(\mathbf{s}; t), \dots, X_q(\mathbf{s}; t)\}^T$ defined for $t \in \mathcal{T}$ denote q functional predictors, and $\mathbf{Z}(\mathbf{s}) =$

$\{Z_1(\mathbf{s}), \dots, Z_h(\mathbf{s})\}^T$ denote d scalar predictors associated with $Y(\mathbf{s})$. In our data, $Y(\mathbf{s})$ is the average corn yield per acre for the county located at $\mathbf{s}$, D is the spatial region of the five Midwestern states, and the time domain $\mathcal{T}$ is a year. We also have $q = 2$ and $h = 1$ with $X_1(\mathbf{s}; t)$ and $X_2(\mathbf{s}; t)$ being daily maximum and minimum temperature trajectories and $Z(\mathbf{s})$ the annual precipitation collected from county $\mathbf{s}$. Without loss of generality, we rescale the time domain to a unit interval and thus $\mathcal{T} = [0, 1]$. We consider spatially varying relationship between the scalar responses and associated predictors, and propose a general framework for spatially varying functional model:

$$Y(\mathbf{s}) = \eta_0(\mathbf{s}) + \mathbf{Z}^T(\mathbf{s})\boldsymbol{\alpha}(\mathbf{s}) + \int_{\mathcal{T}} \mathbf{X}^T(\mathbf{s}; t)\boldsymbol{\eta}(\mathbf{s}; t)dt + e(\mathbf{s}), \quad (1)$$

where $\eta_0(\mathbf{s})$ is a location-specific intercept, $\boldsymbol{\alpha}(\mathbf{s}) = \{\alpha_1(\mathbf{s}), \dots, \alpha_h(\mathbf{s})\}^T$ is a vector of coefficients for the scalar predictors, and $\boldsymbol{\eta}(\mathbf{s}; t) = \{\eta_1(\mathbf{s}; t), \dots, \eta_q(\mathbf{s}; t)\}^T$ denotes a vector of functional coefficients over $t \in \mathcal{T}$ modeling the effect of functional trajectories on scalar response at $\mathbf{s}$. As a special case, when scalar and functional coefficients are spatially invariant, that is, $\boldsymbol{\alpha}(\mathbf{s}) = \boldsymbol{\alpha}$ and $\boldsymbol{\eta}(\mathbf{s}; t) = \boldsymbol{\eta}(t)$, we obtain a partial functional linear model (Shin 2009). Because the spatially varying coefficients are flexible and expected to capture the spatial variability in $Y(\mathbf{s})$, imposing additional spatial correlation structure on the error $e(\mathbf{s})$ will cause identifiability issues to the model. Following Gelfand et al. (2003), we assume that $e(\mathbf{s})$ are zero-mean independent Gaussian errors. Since $Y(\mathbf{s})$ is the average crop yield per acre obtained from an agricultural survey, we assume $\text{var}\{e(\mathbf{s})\} = \sigma_e^2/\omega(\mathbf{s})$ following standard practice for survey data (USDA 2012), where σ_e^2 is an unknown variance parameter and $\omega(\mathbf{s})$ is the size of the harvest land.

The direct estimation of model (1) suffers from the curse of dimensionality, we thus need a low-dimensional representation of spatially dependent multivariate functional predictors. This dimension reduction in the functional model should fulfill the purposes of preserving the within-curve temporal correlation as well as the spatial correlation across $\mathbf{s}$. To accomplish this, we first define the mean function of the multivariate functional predictors as $\boldsymbol{\mu}(\mathbf{s}; t) = \mathbb{E}\{X(\mathbf{s}; t)\} = \{\mu_1(\mathbf{s}; t), \dots, \mu_q(\mathbf{s}; t)\}^T$, where $\mu_l(\mathbf{s}; t) = \mathbb{E}\{X_l(\mathbf{s}; t)\}$. Then, with a set of p principal component (PC) functions, $\mathbf{f}_r = (f_{r1}, \dots, f_{rq})^T$, satisfying $\int_{\mathcal{T}} \mathbf{f}_r^T(t) \mathbf{f}_{r'}(t) dt = I(r = r')$, we expand $X(\mathbf{s}; t)$ as

$$X(\mathbf{s}; t) = \boldsymbol{\mu}(\mathbf{s}; t) + \sum_{r=1}^p \xi_r(\mathbf{s}) \mathbf{f}_r(t),$$

where $\xi_r(\mathbf{s}) = \int_{\mathcal{T}} \{X(\mathbf{s}; t) - \boldsymbol{\mu}(\mathbf{s}; t)\}^T \mathbf{f}_r(t) dt$ represent FPC scores that are spatially correlated. Here, p determines the rank of FPC representation and will be determined by data-driven methods. Using the same PC functions, the spatially varying functional coefficient in (1) can be expanded as $\boldsymbol{\eta}(\mathbf{s}; t) = \sum_{r=1}^p \beta_r(\mathbf{s}) \mathbf{f}_r(t)$, where $\beta_r(\mathbf{s}) = \int_{\mathcal{T}} \boldsymbol{\eta}^T(\mathbf{s}; t) \mathbf{f}_r(t) dt$. Then model (1) can be written as

$$Y(\mathbf{s}) = \alpha_0(\mathbf{s}) + \sum_{j=1}^h Z_j(\mathbf{s}) \alpha_j(\mathbf{s}) + \sum_{r=1}^p \xi_r(\mathbf{s}) \beta_r(\mathbf{s}) + e(\mathbf{s}). \quad (2)$$

3.1. Spatial Functional Predictors with Measurement Errors

In practice, $X(\mathbf{s}; t)$ is often not directly observable, and instead, we only observe its surrogate containing measurement errors. It is well known that climate data products, especially the high-resolution data, are error prone for various reasons (e.g., Matthews, Mannshardt, and Gremaud 2013; Merchant et al. 2017). Since our functional covariates are daily maximum and minimum temperatures, they are likely contaminated with errors. The error prone functional predictors may lead to biased inference when directly used in the prediction model, we therefore should take into account measurement errors in the derivation of the low-dimensional representation of multivariate functional predictors. We then define $\mathbf{W}(\mathbf{s}; t) = \{W_1(\mathbf{s}; t), \dots, W_q(\mathbf{s}; t)\}^T$ to represent the functional covariates containing additive measure errors,

$$\mathbf{W}(\mathbf{s}; t) = X(\mathbf{s}; t) + \mathbf{u}(\mathbf{s}; t), \quad t \in \mathcal{T},$$

where $\mathbf{u}(\mathbf{s}; t) = \{u_1(\mathbf{s}; t), \dots, u_q(\mathbf{s}; t)\}^T$ is a vector of white noise with variance parameter $\sigma_{u_l}^2$ for the l th element.

Our meteorology records have a much longer history than the crop yield measurement. This allows us to estimate the local long-term temperature pattern $\boldsymbol{\mu}(\mathbf{s}; t)$ using historical records dated half a century back. The estimation errors in mean functions are negligible compared with the year-to-year variations. We therefore center each temperature curve by its long-term mean and expand the centered $\mathbf{W}(\mathbf{s}; t)$ as

$$\mathbf{W}(\mathbf{s}; t) = \sum_{r=1}^p \xi_r(\mathbf{s}) \mathbf{f}_r(t) + \mathbf{u}(\mathbf{s}; t); \quad \mathbf{u}(\mathbf{s}; t) \sim N(0, \mathbf{V}),$$

where $\mathbf{V} = \text{diag}(\sigma_{u1}^2, \dots, \sigma_{uq}^2)$. We assume that $\mathbf{u}(\mathbf{s}; t)$ and $e(\mathbf{s})$ in (1) are independent.

Let $\mathbf{s}_1, \dots, \mathbf{s}_n$ be the observed locations and $\boldsymbol{\xi}_r = \{\xi_r(\mathbf{s}_1), \dots, \xi_r(\mathbf{s}_n)\}^T$ be the n -dimensional vector of the r th order FPC scores. To model spatial dependency among functional predictors, we assume $\boldsymbol{\xi}_r \sim N(\mathbf{0}_n, \sigma_{\xi_r}^2 \boldsymbol{\Sigma}(\boldsymbol{\theta}_{\xi_r}))$, $r = 1, \dots, p$, where $\boldsymbol{\Sigma}(\boldsymbol{\theta}_{\xi_r})$ denotes an $n \times n$ correlation matrix parameterized by $\boldsymbol{\theta}_{\xi_r}$. This regularization on $\boldsymbol{\xi}_r$ helps preserve the spatial continuity of functional predictors, and more details on its correlation structure are elaborated in Section 3.3. We also assume that $\boldsymbol{\xi}_r$ are independent across r .

In the classic functional linear model literature, the PC functions, $\mathbf{f}_r(t)$, are often estimated by spectral decomposition of the empirical covariance function and held as if known in the regression analysis. Such an approach ignores the estimation error of the empirical principal components, and can lead to underestimated model variation and spurious statistical inference. We instead propose to combine FPC estimation and functional regression in a unified Bayesian framework. Specifically, we represent $\mathbf{f}_r$ by a penalized spline expansion similar to Zhou, Huang, and Carroll (2008), Zhou et al. (2010), and Huang, Li, and Guan (2014), and impose a Bayesian FPCA model similar to Kowal, Matteson, and Ruppert (2017). We first express each $\mathbf{f}_r$ as

$$\mathbf{f}_r(t) = \{\mathbf{I}_q \otimes \boldsymbol{\psi}_B^T(t)\} \mathbf{d}_r^B,$$

where $\mathbf{I}_q$ is a $q \times q$ identity matrix, $\otimes$ is the Kronecker product, $\boldsymbol{\psi}_B(t) = \{\psi_{B,1}(t), \dots, \psi_{B,L}(t)\}^T$ is a set of L spline basis functions, and $\mathbf{d}_r^B = \{(\mathbf{d}_{r1}^B)^T, \dots, (\mathbf{d}_{rq}^B)^T\}^T$ is a vector of length Lq with $\mathbf{d}_{rl}^B$ being the vector of L unknown coefficients for $f_{rl}(t)$. We choose cubic B-splines with equally spaced interior knots for $\boldsymbol{\psi}_B(t)$ in our application but other basis functions can also be employed. To penalize the roughness of $\mathbf{f}_r(t)$, a penalty function $\mathcal{P}$ as the L^2 -norm of the second derivative of $\mathbf{f}_r$ is constructed, that is, $\mathcal{P}(\mathbf{d}_r) = \sum_{l=1}^q \int \{\ddot{f}_{rl}(t)\}^2 dt = \sum_{l=1}^q (\mathbf{d}_{rl}^B)^T \boldsymbol{\Omega}_{\psi} \mathbf{d}_{rl}^B$, where $\ddot{f}_{rl}(t)$ is the second derivative of f_{rl} and $\boldsymbol{\Omega}_{\psi} = \int_{\mathcal{T}} \ddot{\boldsymbol{\psi}}_B(t) \ddot{\boldsymbol{\psi}}_B^T(t) dt$. Under the Bayesian spline approach, roughness penalty is imposed by constructing a prior on $\mathbf{d}_{rl}^B$ based on $\mathcal{P}$. Following Wand and Ormerod (2008), we use a reparameterized $\mathbf{d}_{rl}^B$, denoted by $\mathbf{d}_{rl}$, to make the penalty matrix diagonal for convenient computation. Specifically, let $\boldsymbol{\Omega}_{\psi} = \mathbf{U}_{\Omega} \mathbf{D}_{\Omega} \mathbf{U}_{\Omega}^T$ be the singular value decomposition of $\boldsymbol{\Omega}_{\psi}$. We reparameterize the basis into $\boldsymbol{\psi}(t) = \{1, t, \boldsymbol{\psi}_B^T(t) \mathbf{U}_{\Omega, P} \mathbf{D}_{\Omega, P}^{-1/2}\}^T$, where $\mathbf{D}_{\Omega, P}$ denotes the $(L-2)$ dimensional diagonal matrix of positive entries on $\mathbf{D}_{\Omega}$ and $\mathbf{U}_{\Omega, P}$ is the corresponding $L \times (L-2)$ submatrix of $\mathbf{U}_{\Omega}$. Now it satisfies $f_{rl}(t) = \boldsymbol{\psi}_B^T(t) \mathbf{d}_{rl}^B = \boldsymbol{\psi}^T(t) \mathbf{d}_{rl}$ and the roughness penalty for $\mathbf{d}_{rl}$ can be written as a diagonal matrix with smoothing parameter $\lambda_r > 0$. We constrain $\lambda_1 > \lambda_2 > \dots > \lambda_p > 0$ following Kowal, Matteson, and Ruppert (2017) to order $\mathbf{f}_1(t), \dots, \mathbf{f}_p(t)$ by decreasing smoothness. Then we construct a prior $\mathbf{d}_{rl} \sim N(\mathbf{0}_L, \text{diag}(c, c, \lambda_r^{-1}, \dots, \lambda_r^{-1}))$ with a sufficiently large constant $c > 0$, and also introduce orthogonal constraints, $\mathbf{d}_{rl}^T \mathbf{J}_{\psi} \mathbf{d}_{r'l} = 0$, for $r \neq r'$, where $\mathbf{J}_{\psi} = \int \boldsymbol{\psi}(t) \boldsymbol{\psi}^T(t) dt$. The expansion of $\mathbf{W}(\mathbf{s}; t)$ and the Bayesian spatial FPCA to estimate $\mathbf{f}_r(t)$ will be integrated into the full hierarchical models in Section 3.4.

3.2. Spatial Variable Selection

The model selection in the functional regression model is typically performed via truncation, which first determines a p based on how much total variation in X or W can be explained by its truncated expansion and then uses the same first p FPC scores in the regression model for all observations (Morris 2015). This selection procedure may not be ideal for data observed over a large spatial domain with a heterogeneous relationship between the response and functional covariates. For example, a specific r th FPC score may have nonzero coefficients at some locations while having zero coefficients elsewhere, depending on their local geological or ecological attributes. A more flexible model selection scheme that allows heterogeneous sets of FPC scores to be selected at different locations while preserving spatial continuity is desirable for our data. We, therefore, propose a Bayesian spatial variable selection method that can make location-specific decisions.

To facilitate the spatial variable selection, we introduce a binary indicator variable $\gamma_r(\mathbf{s})$ such that $\beta_r(\mathbf{s}) = 0$ if $\gamma_r(\mathbf{s}) = 0$, and $\beta_r(\mathbf{s}) \neq 0$ if $\gamma_r(\mathbf{s}) = 1$. Following Mendoza (2018), we then introduce a spatially correlated latent process $v_r(\mathbf{s})$ to ensure spatial continuity of $\gamma_r(\mathbf{s})$ by having

$$\gamma_r(\mathbf{s}) | v_r(\mathbf{s}) = \begin{cases} 1, & \text{if } v_r(\mathbf{s}) \geq 0, \\ 0, & \text{if } v_r(\mathbf{s}) < 0. \end{cases} \quad (3)$$

The choice of p associated with our Bayesian spatial variable selection will be discussed in Section 3.4. A similar variable selection scheme can be easily made for $Z(\mathbf{s})$ when h is high, but it is not considered in our analysis as $h = 1$ in our data.

3.3. Regularization of Spatially Varying Coefficients

Our proposed model employs spatially varying parameters to accommodate the spatially heterogeneous data structure. The varying coefficients $\beta_r(\mathbf{s})$ and $\alpha_j(\mathbf{s})$ model the location specific relationship between the crop yield and predictors, the FPC scores $\xi_r(\mathbf{s})$ represent the individual functional predictors for each location, and the binary coefficients $\gamma_r(\mathbf{s})$ further allows idiosyncratic response-predictor relationships for different locations. Allowing the coefficients to vary with no constraint could potentially lead to over-fitting or algorithm failure. Following Gelfand et al. (2003), we regularize all the spatially varying parameters by Gaussian random processes. The imposed spatial correlation in the Gaussian process priors can force neighboring coefficients to behave similarly, and hence mitigate over-fitting.

Specifically, we model the spatial coefficients $\beta_r(\mathbf{s})$ and $\alpha_j(\mathbf{s})$ as spatially correlated Gaussian processes, where the correlation decays as the locations become further apart. For example, we assume $\text{cov}\{\beta_r(\mathbf{s}_i), \beta_r(\mathbf{s}_{i'})\} = \sigma_{\beta_r}^2 \rho(\|\mathbf{s}_i - \mathbf{s}_{i'}\|, \boldsymbol{\theta}_{\beta_r})$, where $\rho(\cdot)$ is a correlation function defined on the distance between locations, $\|\mathbf{s}_i - \mathbf{s}_{i'}\|$, and a parameter vector $\boldsymbol{\theta}_{\beta_r}$. Different correlation functions can be employed for different spatial coefficients. Similarly, we assume the latent variable $\xi_r(\mathbf{s})$ and the latent process $v_r(\mathbf{s})$ for $\gamma_r(\mathbf{s})$ all governed by Gaussian processes with their own covariance structures.

Although our data are aggregated at the county level so the conditional autoregressive (CAR) model traditionally constructed for lattice data may apply, we argue a covariance func-

tion for continuous geostatistical data is a reasonable approximation for modeling the dependency structure for the crop yield and climate data. The CAR model usually treats regional observations sharing a border as neighbors and then assumes a certain correlation structure for neighbors. This aligns well with the inherent characteristics of certain datasets, such as the epidemiology data for which the spread of disease is often due to the interaction of humans or animals of adjacent regions sharing a border. However, the dependency of the crop yield data mainly comes from similar soil properties, weather, and environmental conditions when the two observations are close in geographic distance, rather than two counties sharing boundaries. Therefore, we choose a covariance function that depends on the distance to model the correlation. Our model aligns with arguments in Oyebamiji et al. (2015) and Park, Brorsen, and Harri (2018) who chose geostatistical covariance models to quantify the spatial dependency of areal crop yield data. The specific choice of covariance function for our data will be discussed in Section 5.1.

3.4. Bayesian Hierarchical Model

Due to the hierarchical nature of our model configuration, we resort to the Bayesian hierarchical model (BHM) for model fitting and statistical inference. The BHM typically consists of three stages. The data stage contains the likelihood of data given the unknown parameters and processes, while the process stage models the latent processes. The third level closes the hierarchy by specifying the priors of the unknown parameters. Below we summarize the first two stages of our model and defer the priors to the supplementary materials.

(i) Data stage:

$$Y(\mathbf{s}_i) = \alpha_0(\mathbf{s}_i) + \sum_{j=1}^h Z_j(\mathbf{s}_i) \alpha_j(\mathbf{s}_i) + \sum_{r=1}^p \xi_r(\mathbf{s}_i) \{\beta_r(\mathbf{s}_i) \gamma_r(\mathbf{s}_i)\} + e(\mathbf{s}_i),$$

$$e(\mathbf{s}_i) \sim N\{0, \sigma_e^2 / \omega(\mathbf{s}_i)\}, \quad i = 1, \dots, n, \quad (4)$$

$$W(\mathbf{s}_i; t) = \sum_{r=1}^p \xi_r(\mathbf{s}_i) \{\mathbf{I}_q \otimes \boldsymbol{\psi}^T(t)\} \mathbf{d}_r + \mathbf{u}(\mathbf{s}_i; t),$$

$$\mathbf{u}(\mathbf{s}_i; t) \stackrel{\text{iid}}{\sim} N\{0, \text{diag}(\sigma_{u1}^2, \dots, \sigma_{uq}^2)\}, \quad i = 1, \dots, n, \quad (5)$$

where $e(\mathbf{s}_i)$ and $\mathbf{u}(\mathbf{s}_i; t)$ are independent, and $\gamma_r(\mathbf{s}_i)$ is defined in (3). The positive weight $\omega(\mathbf{s}_i)$ is known at $\mathbf{s}_i$. The $\boldsymbol{\psi}(t)$ and $\mathbf{d}_r$ are reparameterized basis functions and their corresponding coefficients for the expansion of f_r , as described in Section 3.1.

(ii) Process stage:

$$\begin{aligned} \alpha_j &\sim N\{\mu_{\alpha_j} \mathbf{1}_n, \sigma_{\alpha_j}^2 \boldsymbol{\Sigma}(\boldsymbol{\theta}_{\alpha_j})\}, & j = 0, \dots, h, \\ \beta_r &\sim N\{\mu_{\beta_r} \mathbf{1}_n, \sigma_{\beta_r}^2 \boldsymbol{\Sigma}(\boldsymbol{\theta}_{\beta_r})\}, & r = 1, \dots, p, \\ \xi_r &\sim N\{\mathbf{0}_n, \sigma_{\xi_r}^2 \boldsymbol{\Sigma}(\boldsymbol{\theta}_{\xi_r})\}, & r = 1, \dots, p, \\ v_r &\sim N\{\mu_{v_r} \mathbf{1}_n, \sigma_{v_r}^2 \boldsymbol{\Sigma}(\boldsymbol{\theta}_{v_r})\}, & r = 1, \dots, p, \\ \mathbf{d}_r &\sim N(\mathbf{0}_{qL}, \mathbf{I}_q \otimes \mathbf{D}_r), \quad \text{where} \\ \mathbf{D}_r &= \text{diag}(c, c, \lambda_r^{-1}, \dots, \lambda_r^{-1}), \quad r = 1, \dots, p. \end{aligned} \quad (6)$$

The $L \times L$ roughness penalty matrix $\mathbf{D}_r$ in (6), placed on $\mathbf{d}_r$, controls the smoothness of basis functions with a sufficiently

large constant c and smoothing parameters $\lambda_1 > \dots > \lambda_p$. Ordered smoothing parameters are estimated through MCMC implementation and details can be found in the supplementary materials. The $n \times n$ correlation matrices $\Sigma(\theta_{\alpha_j})$, $\Sigma(\theta_{\beta_r})$, $\Sigma(\theta_{\xi_r})$, and $\Sigma(\theta_{\gamma_r})$ govern the dependency of each process, and are calculated for n spatial locations based on their corresponding correlation functions, as discussed in Section 3.3.

Most location and variance parameters are sampled through Gibbs due to their explicit posterior distributions, while the rest parameters, including parameters in correlation functions and latent indicator processes $\nu_r(\mathbf{s})$ are sampled through Metropolis-Hastings (M-H). The full conditional distributions of unknown parameters and the entire sampling algorithm are also deferred to the supplementary materials. Whenever the M-H algorithm is used, the acceptance rate is tuned to be between 25% and 50% to secure adequate mixing of posterior samples. Both simulations and data analysis results show that the results are robust against the choice of initial values.

The algorithm requires a specification of the truncation parameter p in the FPC expansion. We determine the optimal p via the Deviance Information Criterion (DIC) (Spiegelhalter et al. 2002), which is a generalization of the Akaike's Information Criterion (AIC) for hierarchical models. This requires to first measure the goodness of fit of the model, $D(\Theta)$, based on the deviance statistic given parameters Θ . For our model, we define $D(\Theta) = -2 \ln\{f(\mathbf{Y}|\xi_r, \Theta_1)f(\mathbf{W}|\xi_r, \Theta_2)\}$, where $\Theta = \{\Theta_1, \Theta_2\}$, $\mathbf{Y} = \{Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_n)\}^T$, $f(\cdot)$ denotes the likelihood function, and Θ_1 and Θ_2 denote parameters associated with the spatial functional regression and FPCA expansion, respectively. Next, the model complexity is measured to trade off model fit against the number of free parameters. Since the number of parameters is not clearly defined in complex hierarchical models, the penalty term measuring the model complexity is defined through P_D , known as the “effective number of parameters” in Spiegelhalter et al. (2002). It can be approximated as the difference between the posterior mean of the deviance and the deviance at the posterior estimates of the parameters. Based on this approximation, DIC is calculated as $\text{DIC} = 2\bar{D} - D(\bar{\Theta})$, where $\bar{D}$ denotes the posterior mean of $D(\Theta)$ and $\bar{\Theta}$ the posterior estimates of Θ . We then choose the optimal p as the value that minimizes the DIC.

3.5. Prediction

Given the hierarchical models, we simultaneously estimate model parameters, perform the spatial variable selection, and make predictions for new observations. To carry out such analysis, we incorporate new functional and scalar predictors, $\mathbf{W}_{\text{new}}$, $\mathbf{Z}_{\text{new}}$, and unknown response variables, $\mathbf{Y}_{\text{new}}$, into the BHM. The model treats $\mathbf{Y}_{\text{new}}$ as unknown variables and draws samples from a full conditional distribution at each iteration of the MCMC algorithm. By doing so, the new data participate in the estimation of PC functions and spatial variable selection. Finally, the prediction, $\hat{\mathbf{Y}}_{\text{new}}$, and their corresponding posterior $100(1 - \alpha)\%$ credible intervals are obtained from the posterior means, and the $100(\alpha/2)$ th and $100(1 - \alpha/2)$ th percentiles of the posterior samples, respectively.

4. Monte Carlo Experiments

4.1. Data Generation and Implementation

We conduct simulations to evaluate the numerical performance of our method in terms of model fitting and prediction and make comparisons with competing methods. We adopt the spatial locations of the 403 counties across five states in our real data as the spatial domain. The distances between those counties range from 16 to 1530 km, with an average of 516 km. We set the number of years as five, which represents a more challenging scenario in terms of estimation accuracy than the real data analysis due to the smaller sample size. We assume data at different years are conditionally independent given the predictors (see Section 5.1) and generate five replicates of spatial responses based on the following model:

$$Y_k(\mathbf{s}_i) = Z_k(\mathbf{s}_i)\alpha(\mathbf{s}_i) + \sum_{r=1}^2 \xi_{kr}(\mathbf{s}_i)\{\beta_r(\mathbf{s}_i)\gamma_r(\mathbf{s}_i)\} + e_k(\mathbf{s}_i), \quad (7)$$

where k and i are indices for years and counties, respectively, $Z_k(\mathbf{s}_i) \stackrel{\text{iid}}{\sim} \text{unif}(0, 3)$, and all the other vectors of spatially varying predictors and coefficients are generated from a Gaussian process with their own mean and covariance matrix parameterized by ϕ . Specifically, $\alpha \sim N(\mathbf{1}_n, \Sigma(\phi))$, $\xi_{k1} \sim N(\mathbf{0}_n, 9\Sigma(\phi))$, $\xi_{k2} \sim N(\mathbf{0}_n, 4\Sigma(\phi))$, $\beta_1 \sim N(\mathbf{1}_n, 9\Sigma(\phi))$, $\beta_2 \sim N(-\mathbf{1}_n, 4\Sigma(\phi))$. The binary indicator variable γ_r , defined in (3), is generated by $\nu_1 \sim N(\mathbf{0}_n, \Sigma(\phi))$, and $\nu_2 \sim N(\mathbf{0}_n, \Sigma(\phi))$. The errors $e_k(\mathbf{s}_i) \stackrel{\text{iid}}{\sim} N(0, 2^2)$. The correlation matrix $\Sigma(\phi)$ is governed by a Matérn correlation function $\rho(\Delta) = \{\Gamma(\kappa)2^{\kappa-1}\}^{-1}(\Delta/\phi)^\kappa K_\kappa(\Delta/\phi)$, where Δ is the distance between two observations, κ is the smoothness parameter, ϕ the range parameter, and $K_\kappa(\cdot)$ is the modified Bessel function of the second kind (Stein 1999). The parameter ϕ controls the rate of decay, with larger values of ϕ corresponding to a slower decay and thus stronger correlation. To investigate the effect of spatial correlation on prediction, we consider three values of ϕ : 50, 200, and 350, which correspond to low, moderate and high correlation of our simulated data. We set the smoothness parameter $\kappa = 1$. These settings are similar to Staicu, Crainiceanu, and Carroll (2010).

We further generate multivariate functional predictors $\mathbf{W}_k(\mathbf{s}_i; t)$ as follows,

$$\mathbf{W}_k(\mathbf{s}_i; t) = \sum_{r=1}^2 \xi_{kr}(\mathbf{s}_i)\mathbf{f}_r(t) + \mathbf{u}_k(\mathbf{s}_i; t),$$

where $\mathbf{f}_r(t) = \{f_{r1}(t), f_{r2}(t)\}^T$, with $f_{r1}(t) = \cos(2r\pi t)$, $f_{r2}(t) = \sin(2r\pi t)$, $t \in [0, 1]$, for $r = 1, 2$, satisfying $\int_0^1 \mathbf{f}_r^T(t)\mathbf{f}_{r'}(t)dt = I(r = r')$, and measurement errors $\mathbf{u}_k(\mathbf{s}_i; t) = \{u_{k1}(\mathbf{s}_i; t), u_{k2}(\mathbf{s}_i; t)\}^T$ are generated from $u_{kl}(\mathbf{s}_i; t) \stackrel{\text{iid}}{\sim} N(0, 1^2)$, for $l = 1, 2$. Functional trajectories are generated on a regular grid of 100 points in $[0, 1]$.

The simulation is repeated 100 times. Each time we randomly selected 20% of observations as the testing data and the rest as training data. Although our data are generated from a Matérn covariance structure, we choose to use an exponential covariance function in our model fitting for simplicity, which corresponds to a Matérn with $\kappa = 0.5$. This can also evaluate the robustness of prediction using a misspecified exponential

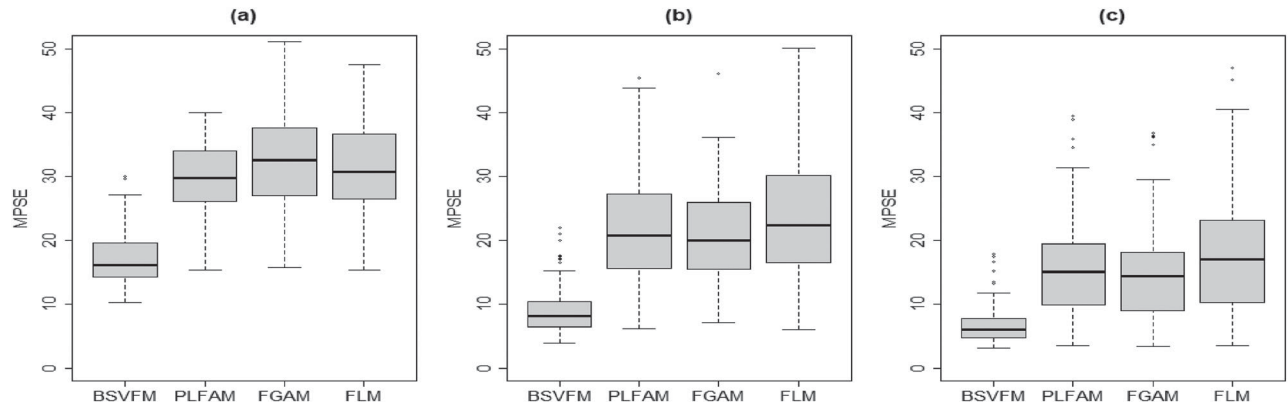


Figure 3. Boxplots of Mean Squared Prediction Errors (MSPE) from the proposed BSVFM and the competing models, PLFAM, FGAM, and FLM, under (a) $\phi = 50$, (b) $\phi = 200$, and (c) $\phi = 350$, based on 100 simulation runs.

covariance function, which has advantages in computation and estimation. More discussions on the choice of covariance function will be provided in Section 5.1. Each simulation samples 5000 times with its first 2500 as burn-in. In addition to making a prediction using our own Bayesian Spatially Varying Functional Model (BSVFM), we also try the following three models for comparison. All three comparison models include the parametric effect of a scalar covariate but employ different approaches to incorporate the effect of bivariate functional predictors.

1. *PLFAM*: Partially Linear Functional Additive Models (Wong, Li, and Zhu 2019) based on additive nonparametric effects of multivariate FPCA (mFPCA) scores.
2. *FGAM*: Functional Generalized Additive Models (McLean et al. 2014) based on the direct use of functional predictors via tensor-product B-splines with roughness penalties.
3. *FLM*: Functional Linear Model based on parametric effects of mFPCA scores.

To fit PLFAM and FLM, mFPCA scores are calculated based on the estimated cross-covariance function of bivariate functional trajectories by applying the algorithm in Wong, Li, and Zhu (2019). We fix $p = 2$, the actual value of p , for all FPCA-based methods including ours, FLM and PLFAM in the simulation. On average, the top two empirical FPCs explain 99.3% of the total variation in the functional predictors. We have also investigated the PLFAM and FLM using the FPCs that capture 99.9% of the total variation, which leads to $p = 11$ on average. We found no advantage of using a more extensive set of FPCs in those models, and thus only defer those results to the supplementary materials. For FGAM, the model is fitted using the R package “refund,” where the tuning parameter is determined by a built-in generalized cross-validation procedure. To fit our proposed model, we employ 20 cubic B-splines bases to represent the PC functions.

4.2. Evaluation of Prediction

We use Mean Squared Prediction Errors (MSPE) calculated from the testing set to measure the prediction performance of different methods. Figure 3(a)–(c) display the prediction results associated with the three values of the range parameter ϕ . Note that we observe much larger MSPE compared to the noise level of $\sigma_e^2 = 4$ in our data generation model (7). This is because

Table 1. Rejection rate (%) for no spatial correlation of residuals using Moran’s I test ($\alpha = 0.05$) under $\phi = 50, 200$, and 350 for the proposed BSVFM and the competing models of PLFAM, FGAM, and FLM, based on 100 simulation runs.

	BSVFM	PLFAM	FGAM	FLM
$\phi = 50$	9.7	11.4	11.3	11.7
$\phi = 200$	17.9	33.7	31.9	36.0
$\phi = 350$	13.6	23.3	21.4	26.1

the spatially varying coefficients add a substantial amount of variability to the responses so the total variation in Y is much larger than 4. In our simulation, the average marginal variances of the responses are 55.2, 51.3, and 46.8, for $\phi = 50, 200$, and 350, respectively. A stronger correlation associated with a larger ϕ leads to less variability in responses. The challenge of estimating spatially varying coefficients increases the prediction error. The measurement errors in the functional covariates additionally contribute to the uncertainty of the estimation and prediction.

For all three levels of spatial correlation, our proposed BSVFM consistently outperforms the competing models with a lower MSPE. When ϕ is small, the spatial correlation of the varying parameters is weak. This leaves the parameters more degrees of freedom to vary and thus more volatile to estimate. We therefore observe relatively higher MSPEs of all methods for a smaller ϕ . In all scenarios, our BSVFM demonstrates a significant advantage over other methods by allowing for spatially heterogeneous parameters. Furthermore, our BSVFM naturally provides prediction intervals which is challenging for the other three models. We then examine the accuracy of the prediction intervals derived from the BSVFM model. Based on the 100 simulation runs, the empirical coverage probabilities for the 95 % credible intervals under $\phi = 50, 200$, and 350 are 98.10%, 98.34%, and 98.38%, respectively, which are close to, albeit slightly larger than, the nominal level.

Ignoring the spatially varying feature of the parameters can also lead to clustered residuals, which is an indicator for regionally biased model fitting and prediction. To illustrate this, we performed Moran’s I test on residuals from different models, and Table 1 summarizes the percentages of rejecting the null hypothesis of no spatial correlation. A lower rejection rate implies a more random pattern of residuals and thus a more appropriate model with spatial variability properly incorporated. We find

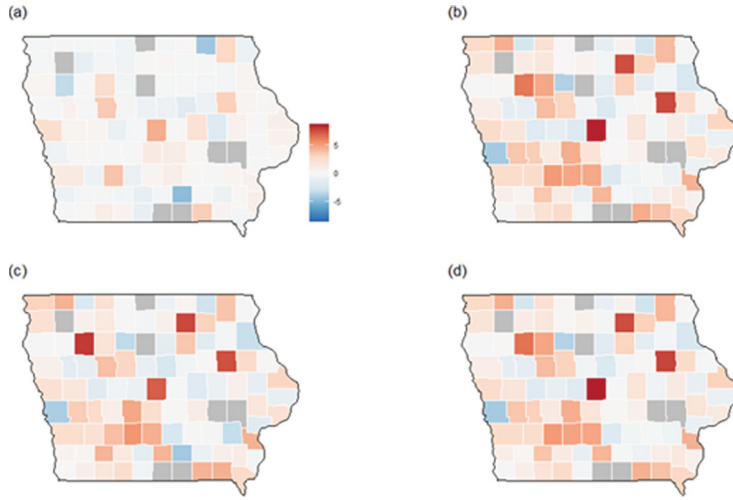


Figure 4. An example of residuals under $\phi = 200$ over the Iowa State calculated from the proposed (a) BSVFM, and the competing models: (b) PLFAM, (c) FGAM, and (d) FLM. Counties with missing data are colored gray.

a much lower reject rate for our model residuals when data has moderate to high spatial correlation, that is, $\phi = 200$ and 350. The data with $\phi = 50$ is very weakly correlated, so all model residuals show relatively small rejection rates. A uniformly decreasing rejection rate for $\phi = 350$ compared to $\phi = 200$ is due to the very strong correlation making the spatial data less varying. Figure 4 illustrates an example of residuals from four models over Iowa State under $\phi = 200$. Residuals from our BSVFM show milder clusters compared to other competing models.

5. Prediction of Corn Yield

5.1. Model Specifics

We apply the proposed BSVFM to predict the county-level averaged corn yield in the Midwest region using the annual precipitation and daily temperature trajectories as introduced in Section 2. Let $Y_k(\mathbf{s}_i)$ be the average corn yield per acre and $Z_k(\mathbf{s}_i)$ be the annual precipitation at county $\mathbf{s}_i$ and year k , for $i = 1, \dots, 403$, and $k = 1999, \dots, 2020$. Let $\mathbf{W}_k(\mathbf{s}_i; t) = \{W_{k1}(\mathbf{s}_i; t), W_{k2}(\mathbf{s}_i; t)\}^T$, where $W_{k1}(\mathbf{s}_i; t)$ and $W_{k2}(\mathbf{s}_i; t)$ are the centered daily maximum and minimum temperature trajectories in year k , respectively. The sustained improvements in seed genetics and biotechnologies have drastically affected the year-to-year independent variation of crop yield in the past several decades. The periodical land fallow or land rotation additionally reduces the temporal correlation on corn production. The land used to grow corn this year is usually used to grow other crops or left without sowing to rest the next year. We thus remove the year effect from the corn yield by constructing $Y_k^*(\mathbf{s}_i) = Y_k(\mathbf{s}_i) - \bar{Y}_k$, where $\bar{Y}_k$ is the annual average corn yields at year k over all regions. We then treat $Y_k^*(\mathbf{s}_i)$ over different years as conditionally independent replicates given meteorological variables. To verify this assumption, we evaluated the autocorrelation function of the residuals from the fitted prediction model. The results support the conditional independence assumption and are deferred to the Supplement.

We fit the proposed Bayesian hierarchical model described in (4)–(6) to the centered response $Y_k^*(\mathbf{s}_i)$ and the centered

meteorology trajectories $\mathbf{W}_k(\mathbf{s}_i; t)$. We choose an exponential correlation function in (6) to model the spatial correlation in α and β_r , and the latent random fields ξ_r and ν_r . To reduce the model complexity, we assume a common range parameter for each type of parameters, that is, a common ϕ_β across β_r , a common ϕ_ξ for ξ_r , and ϕ_ν for ν_r , for $r = 1, \dots, p$. Although this restriction can be relaxed, it largely improves the model parsimony and helps to stabilize the estimation. Additionally, it is reasonable to assume a common decaying rate because correlations in all model parameters trace back to the correlation structure of the original data. All range parameters will be estimated along with other parameters in our model fitting. Assuming a more general Matérn correlation function does not seem to have clear advantages for our data analysis. This is because the smoothness parameter in the Matérn covariance function is difficult to estimate unless there is a reasonable number of observations in proximity (Stein 1999). Our county-level crop data, however, does not align with such data structure. Moreover, it is well known that the three parameters in the Matérn covariance function cannot be all consistently estimated (Zhang 2004). As seen in simulation studies in Section 4, the exponential correlation assumption achieves satisfactory prediction accuracy even with the functional data generated from Matérn spatial correlation functions.

We employ cubic B-splines with 40 equal spaced interior knots to construct $\boldsymbol{\psi}(t) = \{\psi_1(t), \dots, \psi_L(t)\}^T$ for the estimation of f_r , which captures the local variations in daily temperature trajectories. We tried different numbers of knots, and the experiments showed that the prediction performance is insensitive to the choice of L , given it is sufficiently large to capture the features in meteorology trajectories. Using the DIC measure described in Section 3.4, we choose $p = 5$ for the FPC expansion of the temperature trajectories. Based on a total of 15,000 MCMC samples, a posterior sample of size 2500 was obtained by using the first 5000 iterations as burn-in and thinning the remaining 10,000 by a factor of 4. Similar burn-in and thinning practices were adopted by Peruzzi, Banerjee, and Finley (2020). Each MCMC iteration takes, approximately, 25 sec using a PC with a quad-core Intel(R) Core(TM) i-7 processor. Trace plots of the posterior samples and convergence

diagnostics using the potential scale reduction factor (PSFR; Gelman and Rubin 1992) are provided in the supplementary materials to show the convergence of our MCMC algorithm.

5.2. Prediction Assessment

We examine the prediction performance of our model using a 10-fold cross-validation and compare it with competing statistical and machine learning methods. The statistical methods include FLM, PLFAM, and FGAM, described in Section 4.1, and a nonfunctional spatially varying coefficient Model (SVCMM) which simply replaces the functional temperature covariates in our model by the annual averages of maximum and minimum temperatures. We also make comparisons with two machine learning methods that have gained popularity in crop science studies (van Klompenburg, Kassahun, and Catal 2020), the Neural Network (NN) and the extreme gradient boosting (XGB). To fit PLFAM and FLM, the number of FPC is chosen to recover at least 99.9% of the total variation in the functional predictors. Both NN and XGB are trained based on 731 variables including the annual precipitation and the daily minimum and maximum temperatures for the whole year, using R packages “nnet” and “XGB,” respectively. The tuning parameters are determined by cross-validation in each training set. Further exploration of machine learning methods based on various summary statistics of temperatures, such as weekly or monthly averaged temperatures, or FPCs of temperatures, are presented in the supplementary materials. There seems to be no advantage of using the reduced temperature information in the machine learning algorithms. This aligns with the conclusion of Chu and Yu (2020), which demonstrated that machine learning methods for crop yield prediction based on daily temperatures could avoid overfitting through the optimal choice of the number of hidden layers.

Two types of MSPE are calculated for model comparison: (i) Regular MSPE and (ii) Weighted MSPE using the size of the harvest land as the weight. The latter one is calculated by averaging

$$\sum_{(k,i) \in \mathcal{A}_m} \omega_k(\mathbf{s}_i) \{Y_k(\mathbf{s}_i) - \hat{Y}_k(\mathbf{s}_i)\}^2 / \sum_{(k,i) \in \mathcal{A}_m} \omega_k(\mathbf{s}_i),$$

over $m = 1, \dots, 10$, where $\mathcal{A}_m = \{(k, i); Y_k(\mathbf{s}_i) \text{ belongs to the } m\text{th validation set}\}$.

Table 2 summarizes the prediction results. Our BSVFM achieves both the lowest MSPE and the lowest weighted MSPE compared to all six competitors. It is worth mentioning that the SVCMM turns out to be a very competitive approach by allowing spatially varying regression coefficients while employing scalar rather than functional temperature covariates. This further demonstrates the essence of allowing spatially heterogeneous relationships in the corn yield prediction model. On the other hand, the better performance of BSVFM than SVCMM shows treating daily temperature trajectories as functional covariates enables gleaning more information from temperatures. More details of the comparison between SVCMM and BSVFM are illustrated in the supplementary materials. Both NN and XGB are highly sensitive to the choice of tuning parameters and the results in Table 2 represent their best performance we

Table 2. MSPE and weighted MSPE based on the 10-fold cross-validation calculated from BSVFM, SVCMM, PLFAM, FGAM, FLM, NN, and XGB.

	BSVFM	SVCMM	PLFAM	FGAM	FLM	NN	XGB
MSPE	425.5	487.9	481.9	473.7	599.4	714.4	448.3
Weighted MSPE	355.7	395.8	414.3	420.1	464.8	586.5	367.9

have obtained based on our attempts. In addition, results of machine learning methods are typically challenging to interpret. In contrast, our proposed BSVFM not only yields the best predictions, but also delivers insightful and interpretable messages, as detailed in Section 5.3, and provides prediction uncertainty through credible intervals.

Figure 5 illustrates predicted yields together with their corresponding 95% credible intervals using the BSVFM, based on a randomly selected subset of the data. Credible interval widths vary over counties due to different harvest land size. The displayed credible intervals in Figure 5 are from counties with the harvest land sizes ranging from 47,000 to 297,600 acres. The narrow interval widths are obtained from counties with large corn harvest lands. This numerical experiment demonstrates how to use our model for crop yield prediction when the meteorology trajectories are observed. When the meteorology data are not available for future yield prediction, we will use weather forecasting techniques to construct the unobserved temperature trajectories and precipitation to make the prediction.

5.3. Further Nuances

Following the convention of PCA, we sort the initially estimated f_r in the descending order of the posterior means of $\sigma_{\xi_r}^2$ (the variance of the corresponding FPC scores) so that the leading FPC functions $\hat{f}_r$ represent the dominant modes of variation in the temperature trajectories. The spatially varying effect of precipitation and temperatures on corn yield is illustrated in Figure 6 which shows the posterior mean of the coefficients, $\alpha(\mathbf{s})$, for annual precipitation, and $\beta_r(\mathbf{s})$, for the top three FPC scores. Counties marked by yellow dots in (b), (c), and (d) indicate that their corresponding FPC scores were chosen by our Bayesian variable selection device. FPC scores are selected into the final model if their posterior inclusion probability is greater than 0.5. The estimated coefficients are evidently county-specific while spatially clustered. Figure 6(a) shows a cluster of large positive $\hat{\alpha}(\mathbf{s})$ in central Missouri and negative $\hat{\alpha}(\mathbf{s})$ concentrated in northern Illinois and Indiana. In Iowa, a gradual change of annual precipitation effect from the west to east is observed with the sign of coefficients changing from positive to negative. The spatially varying pattern also appears in $\hat{\beta}_1(\mathbf{s})$, $\hat{\beta}_2(\mathbf{s})$ and $\hat{\beta}_3(\mathbf{s})$ in Figure 6(b)–(d). Indeed, the most striking pattern of the three estimated β processes is their strong spatial heterogeneity. Large clusters of negative and positive coefficients in $\hat{\beta}_1(\mathbf{s})$ and $\hat{\beta}_2(\mathbf{s})$ are observed, while $\hat{\beta}_3(\mathbf{s})$ tends to have multiple smaller clusters of coefficients. The nonnull sets of coefficients are captured via the spatial variable selection, and we observe spatial clusters of counties with nonzero coefficients. Estimates of other location-specific coefficients also show spatially varying and correlated patterns, and are presented in the supplementary materials.

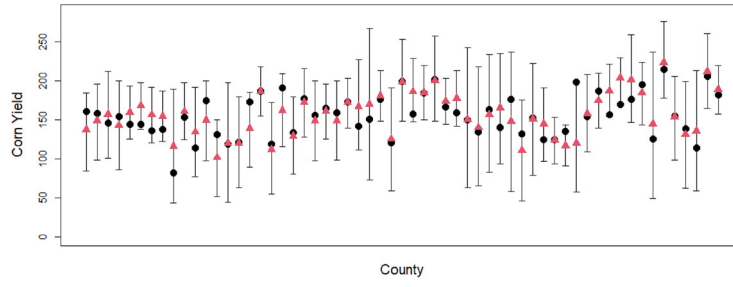


Figure 5. True (●) and predicted corn yields (▲) with their corresponding 95% credible intervals using the BSVFM from a randomly selected subset of the validation data.

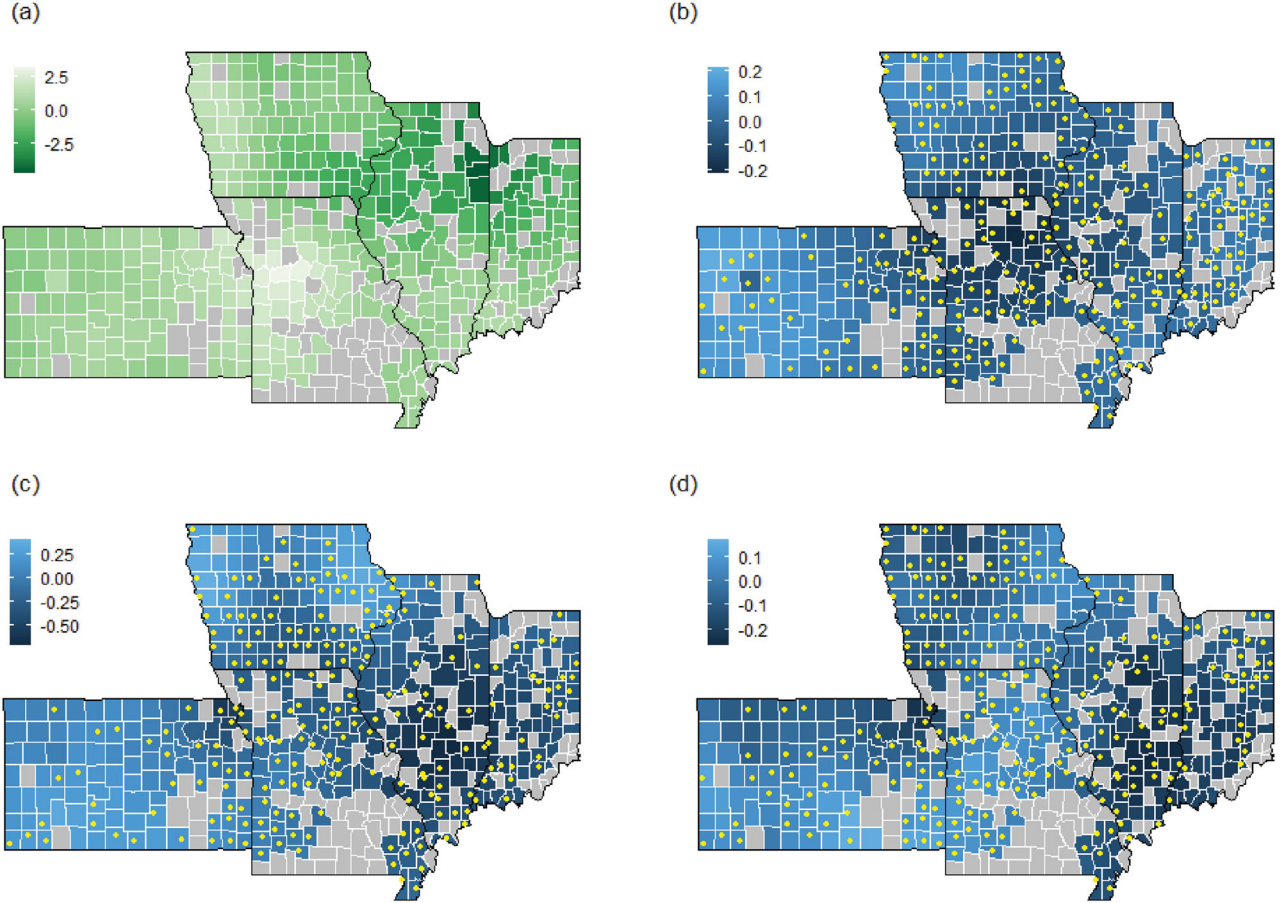


Figure 6. Posterior means of the spatially varying coefficients: (a) $\hat{\alpha}(s)$, (b) $\hat{\beta}_1(s)$, (c) $\hat{\beta}_2(s)$, and (d) $\hat{\beta}_3(s)$ over Kansas, Iowa, Illinois, Indiana, and Missouri. Counties with nonnull coefficients are marked by yellow circles. Counties with missing data are colored gray.

The estimated PC functions $\hat{f}_r = (\hat{f}_{r1}, \hat{f}_{r2})^T$, $r = 1, \dots, 5$, are presented in Figure 7. The two components of the FPC, $\hat{f}_{r1}(t)$ and $\hat{f}_{r2}(t)$, generally show similar patterns over time, which is not surprising since the trajectories of maximum and minimum temperatures are strongly correlated. Overall, $\hat{f}_1$ represents a temperature pattern featuring a warmer than average Winter, a cooler than average Summer and a hot Fall. Counties with a high loading on $\hat{f}_1$ may also experience a smaller than usual diurnal range of temperature in January and August. Then $\hat{f}_2$ appears to represent a colder than average Winter followed by monthly temperature oscillations that are agreed by both the maximum and minimum temperatures, such as higher than average temperature in July and September but lower in August and October. The third component $\hat{f}_3$ depicts a cooler July and October but hotter August and September. The two warmer fall

months also observe a larger than usual diurnal range of temperature through the large gap between $\hat{f}_{31}(t)$ and $\hat{f}_{32}(t)$. Then $\hat{f}_4$ features a warmer-than-average temperature pattern in February, June and November; and $\hat{f}_5$ appears to represent a cooler early summer followed by a hot September. The spatially varying coefficients $\hat{\beta}_r(s)$ in Figure 6 reveal how the effects of their corresponding climate patterns vary over different geographical regions. Similar weather patterns, represented by similar PC scores, can have different, sometimes even opposite effects on corn yield in different regions.

To illustrate the benefit of allowing spatially varying effects of functional temperature covariates on corn yield, Figure 8 compares the fitted yield using the BSVFM and the FLM in Illinois and Iowa in 2018. Both the BSVFM and the FLM treat temperatures as functional covariates yet the latter is a spatially

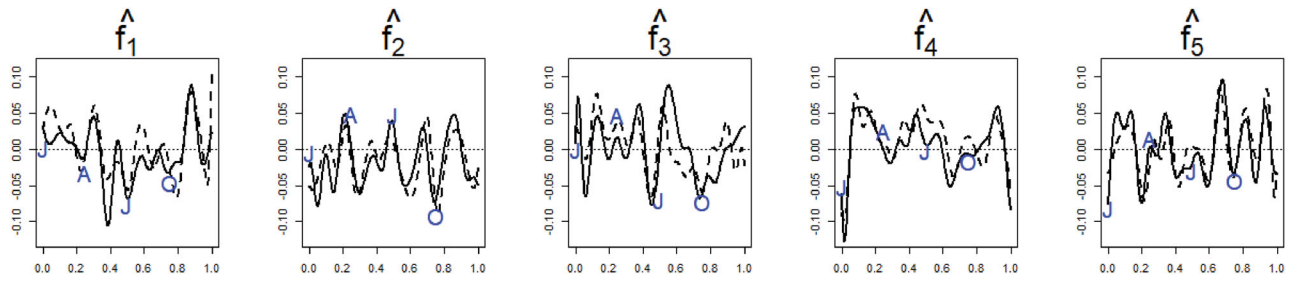


Figure 7. Estimated five PC functions, $\hat{f}_r$, $r = 1, \dots, 5$, from the multivariate Bayesian spatial FPCA. Each PC function is a vector $\hat{f}_r = (\hat{f}_{r1}, \hat{f}_{r2})^T$. The solid and dashed curves in each panel are the estimated $\hat{f}_{r1}$ and $\hat{f}_{r2}$, representing the PC functions for maximum and minimum temperatures, respectively. Alphabet symbols “J”, “A”, “J”, and “O” along with trajectories in each panel indicate the first day of January, April, July, and October, respectively, in the rescaled unit interval $\mathcal{T} = [0, 1]$.

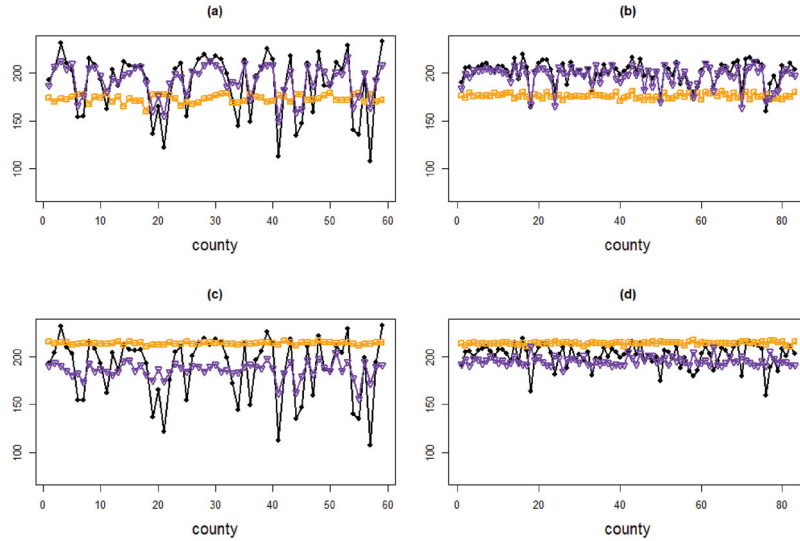


Figure 8. Observed corn yield ($\bullet$) and the fitted corn yield from our BSVFM using annual precipitation only ($\square$) and using both the annual precipitation and functional temperature covariates (∇) in (a) Illinois and (b) Iowa in the year 2018. Corresponding plots from the spatially invariant FLM in (c) Illinois and (d) Iowa are also shown for comparison.

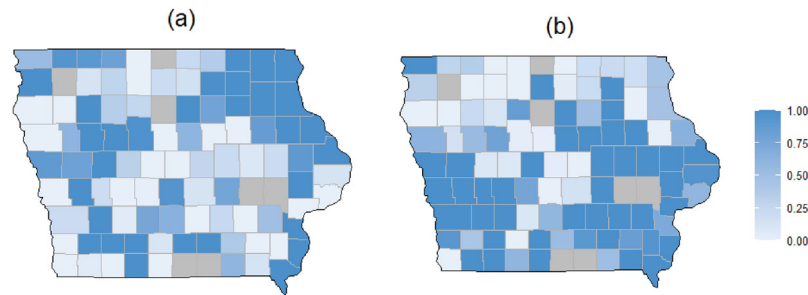


Figure 9. Posterior inclusion probabilities (a) $\hat{\gamma}_1(s)$ of the first and (b) $\hat{\gamma}_2(s)$ of the second FPC scores for Iowa. Counties with missing data are colored gray.

invariant model. Apparently, the BSVFM method using both temperature and precipitation captures significantly more variability of corn yield than the FLM. Figure 8 also shows the variations in corn yield explained merely by the annual precipitation and the added value of incorporating the functional temperature predictors. For both states, the temperature trajectories seem to be the main driver for corn yield variability compared to annual precipitation. All these results indicate the necessity of allowing spatially varying coefficients and employing functional temperature covariates for corn yield prediction. Similar patterns are observed in other states.

Taking Iowa as an example, we examine the variable selection results using the posterior inclusion probabilities for the first

and second FPC scores in Figure 9. Iowa is chosen for a better visualization due to it having the smallest number of missing counties among the five states. The probabilities are calculated as the posterior means of $\gamma_1(s)$ and $\gamma_2(s)$. Higher (lower) posterior probabilities imply a higher (lower) chance of the corresponding FPC scores being selected in the final model. Again, the inclusion probabilities display a heterogeneous feature, strongly supporting our conjecture of spatially varying influence of different PCs on crop yield. As a summary of the variable selection result for the five states, the average number of selected FPC scores in the final model is 3. The second FPC score shows the highest chance of being included with a 64.5% rate over all counties. The lowest selection percentage is observed on the fourth FPC

score with a 54.1% rate. The selection rates of the first, third, and fifth FPC are 59.5%, 57.1%, and 62.3%, respectively. Overall, 72 counties select all five FPCs, 85 select four FPCs, and the rest counties mostly have 2–3 FPCs in the final model. We perform Moran's I test on spatial residuals obtained from the BSVFM fit over the five states at each of the 22 years. Fifteen out of 22 years show a p -value larger than 0.05. This demonstrates that the spatially varying coefficients of the BSVFM properly address the spatial variability in our data, and the assumption of independent errors in the functional regression model appears reasonable.

6. Conclusions

We developed a fully Bayesian functional regression model for corn yield prediction over a large-scale spatial domain. Our model treats county-level maximum and minimum daily temperature trajectories as spatially dependent functional predictors and annual precipitation as a scale predictor. Our model allows for idiosyncratic response-predictor relationships by incorporating spatially varying coefficients as well as spatially heterogeneous variable selection. We estimate the low-rank representation of the multivariate functional covariates via a penalized spline expansion approach while considering additive measurement errors and spatial dependency in the functional data. The introduction of spatially varying coefficients enables us to achieve more accurate predictions than other methods, demonstrated by both simulation and real data studies. The Bayesian approach also has the advantage of providing uncertainty quantification for predictions. Although our model is motivated by the corn yield prediction, it can be broadly applied to scalar-on-function regression models with spatially indexed functional predictors.

Crop yield is often boosted by the advancement of seed genetics and biotechnologies. South et al. (2019) reported that crops engineered with a photorespiratory shortcut for photosynthetic glitch are 40% more productive in real-world agro-economic conditions. Those types of discontinuous year-to-year variation are beyond the scope of our current model, though our model can be naturally expanded to gain such power given appropriate predictors are available. To focus on modeling the variation arising mainly from meteorological variables, our 10-fold cross-validation randomly chose 10% from all observations rather than data of an individual year as the validation set. That allows model fitting to learn the year-specific effect. In practice, an estimated shift should be added to the prediction if a breakthrough technology is expected to significantly improve the crop yield in the prediction horizon. Other factors such as soil conditions and soil moisture, if available, should also be included in the prediction model as their fluctuation can directly impact crop yield.

As mentioned in Section 2, our analysis focuses on counties with corn yield data available for at least five years between 1999 and 2020. Ignoring counties with no observations for at least 15 years seems reasonable in our study since very sparse corn data is likely to indicate insignificant corn production in that county, possibly due to its geographic features or attributes. For instance, certain counties may include large urban areas that do not or only intermittently grow corn. Additionally, we could not

identify a systematic pattern on ignored counties regarding their geographic and time locations, and the missing mechanism is unknown. Thus, we assume missing at random in our analysis and only use the observed values with at least five repetitions for the model fitting. However, if we consider the extension of the model to other data or applications, the missing mechanism should be carefully investigated. For example, the public health data often suppress the small number of cases in a county for confidentiality issues (Shand et al. 2018). In such a case, the missing values are not at random and could place prediction at the risk of carrying bias if left unattended.

In addition to improving prediction accuracy, our BSVFM provides further insights in understanding the relationship between crop yield and temperature trends and variation as discussed in Section 5.3. Our model enables capturing spatially heterogeneous effects of temperature characteristics on crop yields, such as different effects of hot summer or warm fall on corn yield over different regions. With the accelerating climate change and more frequent extreme weather events, our study will shed new light on evaluating the impact of climate change on crop yield and food security.

Supplemental Materials

The online Supplement contains the full hierarchical model and MCMC sampling algorithm. It also contains MCMC diagnostics, additional Monte Carlo experiments, and additional corn yield data analysis results. R codes and data are also provided.

Acknowledgments

The authors thank the Editor, the Associate Editor and two anonymous referees for their very constructive comments and suggestions, which lead to significant improvements of the article.

Disclosure Statement

The authors report there are no competing interests to declare.

Funding

B. Li's work was partially supported by the National Science Foundation Grants DMS-1830312 and -2124576.

ORCID

Yehua Li  <http://orcid.org/0000-0002-5945-378X>

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. [9]
- Crainiceanu, C. M., Staicu, A.-M., and Di, C.-Z. (2009), "Generalized Multilevel Functional Regression," *Journal of the American Statistical Association*, 104, 155–1561. [2]
- Cui, X., Lin, H., and Lian, H. (2020), "Partially Functional Linear Regression in Reproducing Kernel Hilbert Spaces," *Computational Statistics & Data Analysis*, 150, 106978. [2]
- Di Cecco, G., and Gouhier, T. (2018), "Increased Spatial and Temporal Autocorrelation of Temperature under Climate Change," *Scientific Reports*, 8, 14850. [2]
- Fan, Y., James, G. M., and Radchenko, P. (2015), "Functional Additive Regression," *The Annals of Statistics*, 43, 2296–2325. [2]

- Gelfand, A. E., Kim, H.-J., Sirmans, C. F., and Banerjee, S. (2003), "Spatial Modeling with Spatially Varying Coefficient Processes," *Journal of the American Statistical Association*, 98, 387–396. [2,3,5]
- Gelman, A., and Rubin, D. B. (1992), "Inference from Iterative Simulation using Multiple Sequences," *Statistical Science*, 7, 457–472. [9]
- Goldsmith, J., Bobb, J., Crainiceanu, C. M., Caffo, B., and Reich, D. (2011), "Penalized Functional Regression," *Journal of Computational and Graphical Statistics*, 20, 830–851. [2]
- Goldsmith, J., Crainiceanu, C. M., Caffo, B., and Reich, D. (2012), "Longitudinal Penalized Functional Regression for Cognitive Outcomes on Neuronal Tract Measurements," *Journal of the Royal Statistical Society, Series C*, 61, 435–469. [2]
- Gornott, C., and Wechsung, F. (2016), "Statistical Regression Models for Assessing Climate Impacts on Crop Yields: A Validation Study for Winter Wheat and Silage Maize in Germany," *Agricultural and Forest Meteorology*, 217, 89–100. [1]
- Guan, K., Sultan, B., Biasutti, M., Baron, C., and Lobell, D. (2017), "Assessing Climate Adaptation Options and Uncertainties for Cereal Systems in West Africa," *Agricultural and Forest Meteorology*, 232, 291–305. [1]
- Holzkämper, A., Calanca, P., and Fuhrer, J. (2012), "Statistical Crop Models: Predicting the Effects of Temperature and Precipitation Changes," *Climate Research*, 51, 11–21. [2]
- Huang, H., Li, Y., and Guan, Y. (2014), "Joint Modeling and Clustering Paired Generalized Longitudinal Trajectories with Application to Cocaine Abuse Treatment Data," *Journal of the American Statistical Association*, 109, 1412–1424. [2,4]
- James, G., Hastie, T., and Sugar, C. (2000), "Principal Component Models for Sparse Functional Data," *Biometrika*, 87, 587–602. [2]
- James, G. M., Wang, J., and Zhu, J. (2009), "Functional Linear Regression that's Interpretable," *The Annals of Statistics*, 37, 2083–2108. [2]
- Jones, J., Antle, J., Basso, B., Boote, K., Conant, R., Foster, I., Godfray, H., Herrero, M., Howitt, R., Janssen, S., Keating, B., Munoz-Carpena, R., Porter, C., Rosenzweig, C., and Wheeler, T. (2017), "Brief History of Agricultural Systems Modeling," *Agricultural Systems*, 155, 240–254. [1]
- Kang, Y., Ozdogan, M., Zhu, X., Ye, Z., Hain, C., and Anderson, M. (2020), "Comparative Assessment of Environmental Variables and Machine Learning Algorithms for Maize Yield Prediction in the US Midwest," *Environmental Research Letters*, 15, 064005. [1]
- Kern, A., Barcza, Z., Marjanović, H., Árendás, T., Fodor, N., Bónis, P., Bognár, P., and Lichtenberger, J. (2018), "Statistical Modelling of Crop Yield in Central Europe using Climate Data and Remote Sensing Vegetation Indices," *Agricultural and Forest Meteorology*, 260, 300–320. [1]
- 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. [4]
- Kuenzer, T., Hörmann, S., and Kokoszka, P. (2021), "Principal Component Analysis of Spatially Indexed Functions," *Journal of the American Statistical Association* (to appear), DOI: 10.1080/01621459.2020.1732395. [2]
- Li, Y., Wang, N., and Carroll, R. J. (2010), "Generalized Functional Linear Models with Semiparametric Single-Index Interactions," *Journal of the American Statistical Association*, 105, 621–633. [2]
- Lobell, D., and Burke, M. (2010), "On the Use of Statistical Models to Predict Crop Yield Responses to Climate Change," *Agricultural and Forest Meteorology*, 150, 1443–1452. [2]
- Matthews, J. L., Mannshardt, E., and Gremaud, P. (2013), "Uncertainty Quantification for Climate Observations," *American Meteorological Society*, 94, ES21–ES25. [4]
- McLean, M. W., Hooker, G., Staicu, A. M., Scheipl, F., and Ruppert, D. (2014), "Functional Generalized Additive Models," *Journal of Computational and Graphical Statistics*, 23, 249–269. [2,7]
- Mendoza, N. A. (2018), "Variable Selection and Spatial Models for Identifying Chemical Plumes in Hyperspectral Imagery," PhD thesis, University of California Santa Cruz. Available at <https://escholarship.org/uc/item/5cc5d13d>. [5]
- Merchant, C. J., Paul, F., Popp, T., Ablain, M., Bontemps, S., Defourny, P., Hollmann, R., Lavergne, T., Laeng, A., de Leeuw, G., Mittaz, J., Poulsen, C., Povey, A. C., Reuter, M., Sathyendranath, S., Sandven, S., Sofieva, V. F., and Wagner, W. (2017), "Uncertainty Information in Climate Data Records from Earth Observation," *Earth System Science Data*, 9, 511–527. [4]
- Morris, J. S. (2015), "Functional Regression," *Annual Review of Statistics and Its Application*, 2, 321–359. [5]
- Müller, H.-G., and Stadtmüller, U. (2005), "Generalized Functional Linear Models," *Annals of Statistics*, 33, 774–805. [2]
- Müller, H.-G., and Yao, F. (2008), "Functional Additive Models," *Journal of the American Statistical Association*, 103, 1534–1544. [2]
- North, G. R., Wang, J., and Genton, M. G. (2011), "Correlation Models for Temperature Fields," *Journal of Climate*, 24, 5850–5862. [2]
- Oyebamiji, O. K., Edwards, N. R., Holden, P. B., Garthwaite, P. H., Schaphoff, S., and Gerten, D. (2015), "Emulating Global Climate Change Impacts on Crop Yields," *Statistical Modelling*, 6, 499–525. [5]
- Park, E., Brorsen, B. W., and Harri, A. (2018), "Using Bayesian Kriging for Spatial Smoothing in Crop Insurance Rating," *American Journal of Agricultural Economics*, 101, 330–351. [5]
- Peng, B., Guan, K., Chen, M., Lawrence, D., Pokhrel, Y., Suyker, A., Arkebauer, T., and Lu, Y. (2018), "Improving Maize Growth Processes in the Community Land Model: Implementation and Evaluation," *Agricultural and Forest Meteorology*, 250, 64–89. [1]
- Peruzzi, M., Banerjee, S., and Finley, A. O. (2020), "Highly Scalable Bayesian Geostatistical Modeling via Meshed Gaussian Processes on Partitioned Domains," *Journal of the American Statistical Association*, 117, 969–982. [8]
- Prasad, A. K., Chai, L., Singh, R. P., and Kafatos, M. (2006), "Crop Yield Estimation Model for Iowa using Remote Sensing and Surface Parameters," *International Journal of Applied Earth Observation and Geoinformation*, 8, 26–33. [1]
- Ray, D. K., Gerber, J. S., MacDonald, G. K., and West, P. C. (2015), "Climate Variation Explains a Third of Global Crop Yield Variability," *Nature Communications*, 6, 5989. [1]
- Reiss, P. T., Goldsmith, J., Shang, H. L., and Ogden, R. T. (2017), "Methods for Scalar-on-Function Regression," *International Statistical Review*, 85, 228–249. [2]
- Shand, L., Li, B., Park, T., and Albarracín, D. (2018), "Spatially Varying Auto-Regressive Models for Prediction of New Human Immunodeficiency Virus Diagnoses," *Journal of the Royal Statistical Society, Series C*, 67, 1003–1022. [12]
- Shin, H. (2009), "Partial Functional Linear Regression," *Journal of Statistical Planning and Inference*, 139, 3405–3418. [3]
- South, P. F., Cavanagh, A. P., Liu, H. W., and Ort, D. R. (2019), "Synthetic Glycolate Metabolism Pathways Stimulate Crop Growth and Productivity in the Field," *Science*, 363, eaat9077. [12]
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002), "Bayesian Measures of Complexity and Fit," *Journal of the Royal Statistical Society, Series B*, 64:583–639. [6]
- Staicu, A.-M., Crainiceanu, C. M., and Carroll, R. J. (2010), "Fast Methods for Spatially Correlated Multilevel Functional Data," *Biostatistics*, 11, 177–194. [2,6]
- Stein, M. L. (1999), *Interpolation of Spatial Data*, New York: Springer. [6,8]
- USDA (2012), "The Yield Forecasting Program of NASS," available at https://www.nass.usda.gov/Publications/Methodology_and_Data_Quality/Advanced_Topics/Yield%20Forecasting%20Program%20of%20NASS.pdf. [3]
- (2020a), "Crop Production 2019 Summary," Available at https://www.nass.usda.gov/Publications/Todays_Reports/reports/cropan20.pdf. [1]
- (2020b), "What is Agriculture's Share of the Overall U.S. Economy?" Available at <https://www.ers.usda.gov/data-products/chart-gallery/chart-detail?chartId=58270>. [1]
- van Klompenburg, T., Kassahun, A., and Catal, C. (2020), "Crop Yield Prediction using Machine Learning: A Systematic Literature Review," *Computers and Electronics in Agriculture*, 177, 105709. [1,9]
- Wand, M. P., and Ormerod, J. T. (2008), "On semiparametric Regression with 'osullivan Penalized Splines,'" *Australian and New Zealand Journal of Statistics*, 50, 179–198. [4]
- Wong, R. K., Li, Y., and Zhu, Z. (2019), "Partially Linear Functional Additive Models for Multivariate Functional Data," *Journal of the American Statistical Association*, 114, 406–418. [1,2,3,7]
- Zhang, H. (2004), "Inconsistent Estimation and Asymptotically Equal Interpolations in Model-based Geostatistics," *Journal of the American Statistical Association*, 99, 250–261. [8]
- Zhang, L., Baladandayuthapani, V., Zhu, H., Baggerly, K. A., Majewski, T., Czerniak, B. A., and Morris, J. S. (2016), "Functional Car Models for

- Large Spatially Correlated Functional Dataset,” *Journal of the American Statistical Association*, 111, 772–786. [2]
- Zhao, Y., Chen, H., and Ogden, R. T. (2015), “Wavelet-based Weighted Lasso and Screening Approaches in Functional Linear Regression,” *Journal of Computational and Graphical Statistics*, 24, 655–675. [2]
- Zhou, L., Huang, J. Z., and Carroll, R. J. (2008), “Joint Modelling of Paired Sparse Functional Data using Principal Components,” *Biometrika*, 95, 601–619. [2,4]
- Zhou, L., Huang, J. Z., Martinez, J. G., Maity, A., Baladandayuthapani, V., and Carroll, R. J. (2010), “Reduced Rank Mixed Effects Models for Spatially Correlated Hierarchical Functional Data,” *Journal of the American Statistical Association*, 105, 390–400. [2,4]
- Zhu, H., Yao, F., and Zhang, H. H. (2014), “Structured Functional Additive Regression in Reproducing Kernel Hilbert Spaces,” *Journal of the Royal Statistical Society, Series B*, 76, 581–603. [2]