

NEW PARSIMONIOUS MULTIVARIATE SPATIAL MODEL: SPATIAL ENVELOPE

Hossein Moradi Rekabdarkolae, Qin Wang, Zahra Naji
and Montserrat Fuente

*South Dakota State University, The University of Alabama,
Shahid Chamran University of Ahvaz and The University of Iowa*

Abstract: Dimension reduction provides a useful tool for analyzing high-dimensional data. The recently developed *envelope* method is a parsimonious version of the classical multivariate regression model that identifies a minimal reducing subspace of the responses. However, existing envelope methods assume an independent error structure in the model. While the assumption of independence is convenient, it does not address the additional complications associated with spatial or temporal correlations in the data. Therefore, we propose a *Spatial Envelope* method for dimension reduction in the presence of dependencies across space. We study the asymptotic properties of the proposed estimators and show that the asymptotic variance of the estimated regression coefficients under the spatial envelope model is smaller than that of the traditional maximum likelihood estimation. Furthermore, we present a computationally efficient approach for inferences. The efficacy of the proposed method is investigated through simulation studies and an analysis of an Air Quality Standard data set provided by the US Environmental Protection Agency.

Key words and phrases: Dimension reduction, grassmanian manifold, matern covariance function, spatial dependency.

1. Introduction

In many research areas, including health science (Lave and Seskin (1973); Liang, Zeger and Qaqish (1992)), the environmental sciences (Guinness et al. (2014)), and business (Cooper, Schindler and Sun (2003)), it is common to observe multiple outcomes simultaneously. The traditional multivariate linear model has proved useful in such cases in terms of understanding the relationships between response variables and predictors. Mathematically, the model is typically presented as:

$$\mathbf{Y} = \boldsymbol{\alpha} + \boldsymbol{\beta}\mathbf{X} + \boldsymbol{\epsilon}, \quad (1.1)$$

where $\mathbf{Y} \in \mathbb{R}^r$ is the response vector, $\mathbf{X} \in \mathbb{R}^p$ is a predictor vector, $\boldsymbol{\alpha} \in \mathbb{R}^r$ is a vector of intercepts, $\boldsymbol{\beta} \in \mathbb{R}^{(r \times p)}$ is a matrix of regression coefficients, and $\boldsymbol{\epsilon} \sim N_r(\mathbf{0}, \boldsymbol{\Sigma})$ is an error vector, with $\boldsymbol{\Sigma} \geq 0$ indicating an unknown covariance matrix (Christensen (2001)). In order to completely specify a multivariate linear model, there are r unknown intercepts, $p \times r$ unknown parameters for the matrix of regression coefficients, and $r(r+1)/2$ unknown parameters needed to specify an unstructured covariance matrix. Therefore, one must estimate $r + pr + r(r+1)/2$ parameters, which becomes large as one or both of r or p increase.

Based on the observation that, in some cases, linear combinations of \mathbf{Y} may not depend on any of the predictors, Cook, Li and Chiaromonte (2010) proposed the *envelope* method as a parsimonious version of the classical multivariate linear model. This approach separates \mathbf{Y} into material and immaterial parts, which improves the estimation efficiency over that of the usual maximum likelihood estimation. The envelope approach constructs a link between the mean function and the covariance matrix using a minimal reducing subspace, such that the resulting number of parameters is maximally reduced. Cook, Li and Chiaromonte (2010) showed that the envelope estimator is at least as efficient as the standard maximum likelihood estimator (MLE). In related works, the concept of an envelope has been developed further from both theoretical and computational points of view. Such works include, but not restricted to, those of Su and Cook (2011, 2012, 2013), Cook and Zhang (2015), and Cook, Forzani and Su (2016). Furthermore, Li and Zhang (2017) and Zhang and Li (2017) extended the envelope model to tensor responses and tensor covariates, respectively.

The envelope methodology proposed by Cook, Li and Chiaromonte (2010) assumes observations are taken under identical conditions, where independence is assured. While models based on the independence assumption are extremely useful, their use is limited in applications in which the data have inherent dependency (Cressie (1993)). For example, in environment monitoring, each station collects data on several pollutants, such as ozone, carbon monoxide, and nitrogen dioxide. These data have a special type of dependency, called spatial correlation. Myers (1991) and Ver Hoef and Barry (1998) used a pseudo cross-variogram to model the multivariate spatial cross-correlation. In addition, Chiles and Delfiner (1999) and Wackernagel (2003) introduced several multivariate covariograms and cross-variograms that result in a nonnegative definite covariance matrix (also called a valid spatial covariance function). Linear coregionalization model (LCM) is popular in multivariate spatial data analyses. An LCM assumes that the observed variables are linear combinations of sets of independent under-

lying variables, and that they covary jointly over a region. Various methods have been proposed for fitting LCM including the least squares approach (Goulard and Voltz (1992)) and the expectation-maximization (EM) algorithm (Zhang (2007)), among others. Gneiting, Kleiber and Schlather (2010) introduced a flexible and interpretable Matern cross-covariance function for multivariate spatial random fields. Genton and Kleiber (2015) provide a comprehensive review on the approaches commonly used to build a valid spatial cross-covariance model. In this paper, we introduce a *spatial envelope* approach for spatially correlated data. This new approach addresses the impact of spatial correlation between observations in the model and, thus, provides more efficient estimators than those of the traditional multivariate linear model and linear coregionalization model. Accounting for the intrinsic spatial correlation facilitates appropriate inferences on the aforementioned data.

The rest of the paper is organized as follows. In Section 2, we briefly review the envelope methodology. The spatial envelope method is discussed in Section 3. Sections 4 and 5 provide the asymptotic variance and the prediction properties, respectively, of the proposed method. Sections 6 and 7 present a simulation study and an analysis of northeastern US air pollution data, respectively. We conclude the paper in Section 8. All technical details are provided in the online Supplementary Material.

2. Envelope Methodology

For model (1.1), suppose that we can find an orthogonal matrix $(\mathbf{\Gamma}_1, \mathbf{\Gamma}_0) \in \mathbb{R}^{r \times r}$ that satisfies the following two conditions: (i) $span(\boldsymbol{\beta}) \subseteq span(\mathbf{\Gamma}_1)$, and (ii) $\mathbf{\Gamma}_1^T \mathbf{Y}$ is conditionally independent of $\mathbf{\Gamma}_0^T \mathbf{Y}$, given \mathbf{X} . That is, $\mathbf{\Gamma}_0^T \mathbf{Y}$ is marginally independent of \mathbf{X} and conditionally independent of \mathbf{X} , given $\mathbf{\Gamma}_1^T \mathbf{Y}$. Then, we can rewrite $\boldsymbol{\Sigma}$ as

$$\boldsymbol{\Sigma} = \mathbf{P}_{\mathbf{\Gamma}_1} \boldsymbol{\Sigma} \mathbf{P}_{\mathbf{\Gamma}_1} + \mathbf{Q}_{\mathbf{\Gamma}_1} \boldsymbol{\Sigma} \mathbf{Q}_{\mathbf{\Gamma}_1}, \tag{2.1}$$

where $\mathbf{P}_{(\cdot)}$ represents an orthogonal projection operator with respect to the standard inner product, and $\mathbf{Q}_{(\cdot)} = \mathbf{I}_r - \mathbf{P}_{(\cdot)}$ is the projection onto its complement space. Cook, Li and Chiaromonte (2010) used this idea to construct the unique smallest subspace $span(\mathbf{\Gamma}_1)$ that satisfies (2.1) and contains $span(\boldsymbol{\beta})$. In summary, the goal is to find a subspace $span(\mathbf{\Gamma}_1) \subseteq \mathbb{R}^r$, such that

$$\mathbf{Q}_{\mathbf{\Gamma}_1} \mathbf{Y} | \mathbf{X} \sim \mathbf{Q}_{\mathbf{\Gamma}_1} \mathbf{Y}, \tag{2.2a}$$

$$\mathbf{Q}_{\mathbf{\Gamma}_1} \mathbf{Y} \perp\!\!\!\perp \mathbf{P}_{\mathbf{\Gamma}_1} \mathbf{Y} | \mathbf{X}. \tag{2.2b}$$

where \perp denotes statistical independence. This minimal subspace is called the Σ -envelope of $\text{span}(\beta)$ or, simply, the envelope. Here, $\mathbf{P}_{\Gamma_1}\mathbf{Y}$ and $\mathbf{Q}_{\Gamma_1}\mathbf{Y}$ are referred to as the material and immaterial parts of \mathbf{Y} , respectively, where $u \leq r$ is the dimension of the envelope subspace.

Following the envelope idea, model (1.1) can be rewritten as

$$\mathbf{Y} = \boldsymbol{\alpha} + \Gamma_1\boldsymbol{\eta}\mathbf{X} + \boldsymbol{\epsilon}, \quad (2.3)$$

where $\boldsymbol{\beta} = \Gamma_1\boldsymbol{\eta}$, $\boldsymbol{\eta} \in \mathbb{R}^{u \times p}$. In addition, $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0 + \boldsymbol{\Sigma}_1$, such that $\boldsymbol{\Sigma}_0 = \mathbf{Q}_{\Gamma_1}\boldsymbol{\Sigma}\mathbf{Q}_{\Gamma_1}^T$ is the variance of the immaterial part of the response, and $\boldsymbol{\Sigma}_1 = \mathbf{P}_{\Gamma_1}\boldsymbol{\Sigma}\mathbf{P}_{\Gamma_1}^T$ is the variance of the material part of the response. Cook, Li and Chiaromonte (2010) showed that $\boldsymbol{\Sigma} = \Gamma_1\boldsymbol{\Omega}_1\Gamma_1^T + \Gamma_0\boldsymbol{\Omega}_0\Gamma_0^T$, where $\boldsymbol{\Omega}_1 = \text{var}(\Gamma_1^T\mathbf{Y}) \in \mathbb{R}^{u \times u}$ and $\boldsymbol{\Omega}_0 = \text{var}(\Gamma_0^T\mathbf{Y}) \in \mathbb{R}^{(r-u) \times (r-u)}$ are unknown positive definite matrices, with $0 < u \leq r$. Here, we need only estimate $r + pu + r(r+1)/2$ parameters. The difference in the number of parameters between the envelope and the classical multivariate regression is $p(r-u)$. For further information, see Cook, Li and Chiaromonte (2010), and the references therein.

3. Spatial Envelope Method

In this section, we first review the spatial multivariate model. Then, we derive the likelihood function of the spatial envelope model and show the computational steps for the parameter estimation. Let $Y(s_i) = (y_1(s_i), \dots, y_r(s_i))^T$ be an r -variate stochastic spatial response vector, with p regressors $X(s_i) = (x_1(s_i), \dots, x_p(s_i))^T$ observed at locations $s = \{s_1, s_2, \dots, s_n; s_i \in \mathbb{R}^2; i = 1, 2, \dots, n\}$. The multivariate spatial regression model can be written as

$$Y(s_i) = \boldsymbol{\alpha} + \boldsymbol{\beta}X(s_i) + \boldsymbol{\epsilon}(s_i), \quad (3.1)$$

where $Y(s)$ denotes the $r \times 1$ response vector at location s_i , for $i = 1, \dots, n$, and $X(s)$ is a $p \times 1$ vector of fixed and nonstochastic covariates. Furthermore, $\boldsymbol{\alpha}$ denotes an $r \times 1$ vector of intercepts, $\boldsymbol{\beta}$ is an $r \times p$ matrix of regression coefficients, and $\boldsymbol{\epsilon}$ is a multivariate spatial process with mean zero. We assume that the data-generating process is second-order stationary, and that the covariance of the response vectors $Y(s_i)$ and $Y(s_j)$ at sites s_i and s_j is a function of the distance between the two sites. That is, the covariance can be written as

$$\text{Cov}(Y(s_i), Y(s_j)) = C_{ij}(\mathbf{h}), \quad \mathbf{h} = \|s_i - s_j\|, \quad (3.2)$$

where $\|\cdot\|$ denotes the Euclidean distance. The function $C(\mathbf{h}) = \{C_{ij}(\mathbf{h})\}$ is the multivariate covariogram, and $C_{ij}(\cdot)$ is the direct covariogram for $i = j$ and the cross-covariogram for $i \neq j$. By adopting the *proportional correlation model* (Chiles and Delfiner (1999)), the spatial covariance function can be written as

$$C_{ij}(\mathbf{h}) = \mathbf{V}\rho_{ij}(\mathbf{h}), \tag{3.3}$$

where \mathbf{V} is an $r \times r$ positive-definite matrix, and $\rho_{ij}(\mathbf{h})$ is the spatial correlation between sites s_i and s_j (Wackernagel (2003)). Estimating the correlation function solely from the data, without any structural assumptions, is difficult and sometimes infeasible. Usually, it is assumed that the form of the correlation function is a known function, but with unknown parameters $\boldsymbol{\theta}$ that control the range, smoothness, and other characteristics of the correlation function. Thus, instead of $\rho(\mathbf{h})$, we use $\rho(\mathbf{h}, \boldsymbol{\theta})$ to represent the unknown parameters $\boldsymbol{\theta}$ in the correlation function. For simplicity of notation, $\rho(\mathbf{h}, \boldsymbol{\theta})$ is denoted by $\rho(\boldsymbol{\theta})$ throughout the rest of the paper.

The matrix form of model (3.1) is

$$\mathbf{Y}(s) = \boldsymbol{\alpha}^T \otimes \mathbf{1}_n + \mathbf{X}(s)\boldsymbol{\beta}^T + \boldsymbol{\epsilon}(s), \tag{3.4}$$

where $\mathbf{Y}(s) = (Y^T(s_1), \dots, Y^T(s_n))'$ is an $n \times r$ response matrix, and $\mathbf{X}(s)$ is an $n \times p$ matrix of covariates. Furthermore, \otimes denotes the Kronecker product and $\mathbf{1}_n$ is an $n \times 1$ column vector with one at each entry. From the envelope idea, \mathbf{V} can be written as $\mathbf{V}_0 + \mathbf{V}_1$, where $\mathbf{V}_0 = \mathbf{Q}_{\Gamma_1}\mathbf{V}\mathbf{Q}_{\Gamma_1}$ denotes the covariance matrix associated with the immaterial part of the response, and $\mathbf{V}_1 = \mathbf{P}_{\Gamma_1}\mathbf{V}\mathbf{P}_{\Gamma_1}$ denotes the covariance matrix associated with the material part, where Γ_1 is the semi-orthogonal basis of $\text{span}(\mathbf{V}_1)$. Hence, the spatial covariance matrix of $C_{ij}(\mathbf{h})$ can be written as follows:

$$C_{ij}(\mathbf{h}) = \mathbf{V}\rho_{ij}(\boldsymbol{\theta}) = (\mathbf{V}_0 + \mathbf{V}_1)\rho_{ij}(\boldsymbol{\theta}). \tag{3.5}$$

Let $0 < u \leq r$ denote the structural dimension of the envelope. Here, u can be selected using a modified information criterion, such as the modified BIC (Li and Zhang (2017)), model-free dimension selection, such as the full Grassmanian (FG; Zhang and Mai (2017)), and the 1-D algorithm (Cook and Zhang (2016)), or cross-validation. For further information, see Zhang and Mai (2017) and Zhang, Wang and Wu (2018) and the references therein.

To illustrate the estimation, we use a *vec* operator on the response matrix.

That is, let $\mathbb{Y}(s) = \text{vec}(\mathbf{Y}(s))$ be an $nr \times 1$ vector for the vectorized response variable, and let $\mathbb{X}(s) = \mathbf{I}_r \otimes \mathbf{X}(s)$ be an $nr \times pr$ block diagonal matrix, with $\mathbf{X}_i(s)$ as blocks. Thus, the vectorized version of the multivariate spatial linear model can be written as

$$\mathbb{Y}(s) = \boldsymbol{\alpha} \otimes \mathbf{1}_n + \mathbb{X}(s)\boldsymbol{\beta}^* + \boldsymbol{\epsilon}^*(s), \quad (3.6)$$

where $\boldsymbol{\alpha}$ is an $r \times 1$ vector of intercepts, $\boldsymbol{\beta}^* = \text{vec}(\boldsymbol{\beta}^T)$ is a $pr \times 1$ vector of regression coefficients, and $\boldsymbol{\epsilon}^*(s)$ is an $nr \times 1$ vector of spatial errors with mean zero. Using the proportional covariance model and the vectorization of the response matrix, the $nr \times nr$ covariance matrix of the response variables, $\boldsymbol{\Sigma}_{\mathbb{Y}}$, can be written as $\mathbf{V} \otimes \boldsymbol{\rho}(\boldsymbol{\theta})$.

The likelihood function of model (3.6) is

$$\begin{aligned} L(\boldsymbol{\alpha}, \boldsymbol{\beta}^*, \mathbf{V}, \boldsymbol{\theta}) &= [\det(\mathbf{V} \otimes \boldsymbol{\rho}(\boldsymbol{\theta}))]^{-1/2} \\ &\times \exp \left\{ -\frac{1}{2} (\mathbb{Y}(s) - \boldsymbol{\alpha} \otimes \mathbf{1}_n - \mathbb{X}(s)\boldsymbol{\beta}^*)^T (\mathbf{V} \otimes \boldsymbol{\rho}(\boldsymbol{\theta}))^{-1} (\mathbb{Y}(s) - \boldsymbol{\alpha} \otimes \mathbf{1}_n - \mathbb{X}(s)\boldsymbol{\beta}^*) \right\}, \end{aligned} \quad (3.7)$$

where $\det(\cdot)$ denotes the determinant of the matrix. Suppose the response vector can be decomposed into material and immaterial parts, $\mathbb{Y}_1 = (\mathbf{I}_r \otimes \mathbf{P}_{\Gamma_1})\mathbb{Y}(s)$ and $\mathbb{Y}_0 = (\mathbf{I}_r \otimes \mathbf{Q}_{\Gamma_1})\mathbb{Y}(s)$, respectively. From (3.5), the covariance matrix of $\mathbb{Y}(s)$ can be written as follows:

$$\begin{aligned} \boldsymbol{\Sigma}_{\mathbb{Y}} &= \mathbf{V} \otimes \boldsymbol{\rho}(\boldsymbol{\theta}) \\ &= \mathbf{V}_0 \otimes \boldsymbol{\rho}(\boldsymbol{\theta}) + \mathbf{V}_1 \otimes \boldsymbol{\rho}(\boldsymbol{\theta}). \end{aligned} \quad (3.8)$$

Combining (3.7) and (3.8), we have

$$L^u(\boldsymbol{\alpha}, \boldsymbol{\beta}^*, \mathbf{V}_0, \mathbf{V}_1, \boldsymbol{\theta}) = L_1^u(\boldsymbol{\alpha}, \boldsymbol{\beta}^*, \mathbf{V}_1, \boldsymbol{\theta}) \times L_2^u(\boldsymbol{\alpha}, \mathbf{V}_0, \boldsymbol{\theta}), \quad (3.9)$$

with

$$\begin{aligned} L_1^u(\boldsymbol{\alpha}, \boldsymbol{\beta}^*, \mathbf{V}_1, \boldsymbol{\theta}) &= [\det_0(\mathbf{V}_1)]^{-n/2} [\det(\boldsymbol{\rho}(\boldsymbol{\theta}))]^{-r/2} \\ &\times \exp \left\{ -\frac{1}{2} (\mathbb{Y}(s) - \boldsymbol{\alpha} \otimes \mathbf{1}_n - \mathbb{X}(s)\boldsymbol{\beta}^*)^T \left(\mathbf{V}_1^\dagger \otimes \boldsymbol{\rho}^{-1}(\boldsymbol{\theta}) \right) (\mathbb{Y}(s) - \boldsymbol{\alpha} \otimes \mathbf{1}_n - \mathbb{X}(s)\boldsymbol{\beta}^*) \right\}, \\ L_2^u(\boldsymbol{\alpha}, \mathbf{V}_0, \boldsymbol{\theta}) &= [\det_0(\mathbf{V}_0)]^{-n/2} [\det(\boldsymbol{\rho}(\boldsymbol{\theta}))]^{-r/2} \\ &\times \exp \left\{ -\frac{1}{2} (\mathbb{Y}(s) - \boldsymbol{\alpha} \otimes \mathbf{1}_n)^T \left(\mathbf{V}_0^\dagger \otimes \boldsymbol{\rho}^{-1}(\boldsymbol{\theta}) \right) (\mathbb{Y}(s) - \boldsymbol{\alpha} \otimes \mathbf{1}_n) \right\}, \end{aligned} \quad (3.10)$$

where \dagger denotes the Moore–Penrose inverse, and $\det_0(\mathbf{A})$ denotes the product of the nonzero eigenvalues of a nonzero symmetric matrix \mathbf{A} . The likelihood in equation (3.7) can be factorized as equation (3.9), from $\text{span}(\boldsymbol{\beta}) \subseteq \text{span}(\mathbf{V}_1)$ and $(\mathbf{V}_0^\dagger \otimes \boldsymbol{\rho}^{-1}(\boldsymbol{\theta}))\mathbb{X}\boldsymbol{\beta}^* = \mathbf{0}$. This factorization is described in the Supplementary Material, Section S2.

The objective is to maximize the likelihood in (3.9) over $\boldsymbol{\beta}^*$, \mathbf{V}_0 , \mathbf{V}_1 , and $\boldsymbol{\theta}$, subject to the following constraints:

$$\begin{aligned} \text{span}(\boldsymbol{\beta}) &\subseteq \text{span}(\mathbf{V}_1), \\ \mathbf{V}_0\mathbf{V}_1 &= \mathbf{0}. \end{aligned} \tag{3.11}$$

Thus, the multivariate spatial model in (3.6) can be written as

$$\begin{aligned} \mathbb{Y}(s) &= \boldsymbol{\alpha} \otimes \mathbf{1}_n + \mathbb{X}(s)\text{vec}(\boldsymbol{\eta}^T \boldsymbol{\Gamma}_1^T) + \boldsymbol{\epsilon}^*(s), \\ \boldsymbol{\Sigma} &= (\boldsymbol{\Gamma}_1 \boldsymbol{\Omega}_1 \boldsymbol{\Gamma}_1^T + \boldsymbol{\Gamma}_0 \boldsymbol{\Omega}_0 \boldsymbol{\Gamma}_0^T) \otimes \boldsymbol{\rho}(\boldsymbol{\theta}), \end{aligned} \tag{3.12}$$

where $\boldsymbol{\Gamma}_1$ denotes the semi-orthogonal basis for $\text{span}(\mathbf{V}_1)$, $\boldsymbol{\Gamma}_0$ denotes the semi-orthogonal basis for the orthogonal complement space of $\text{span}(\mathbf{V}_1)$, $\boldsymbol{\Omega}_1$ denotes the covariance of the material part of the response, $\boldsymbol{\Omega}_2$ denotes the covariance of the immaterial part of the response, and $\boldsymbol{\eta} \in \mathbb{R}^{u \times r}$ is chosen such that $\boldsymbol{\beta}^* = \text{vec}(\boldsymbol{\eta}^T \boldsymbol{\Gamma}_1^T)$.

As mentioned by Cook, Li and Chiaromonte (2010), the gradient-based algorithms for Grassmann optimization (Edelman, Arias and Smith (1998)) require a coordinate version of the objective function, which must have continuous directional derivatives. The optimization depends on minimizing the logarithm of \mathbf{D} over the Grassmann manifold $\mathbb{G}^{r \times u}$, where

$$\mathbf{D} = \det(\mathbf{P}_{\mathbf{V}_1} \hat{\boldsymbol{\Sigma}}_{\text{res}} \mathbf{P}_{\mathbf{V}_1} + \mathbf{Q}_{\mathbf{V}_1} \hat{\boldsymbol{\Sigma}}_{\mathbb{Y}} \mathbf{Q}_{\mathbf{V}_1}),$$

and \mathbf{D} is the partially maximized likelihood function. The derivation of \mathbf{D} is discussed in the Supplementary Material, Section S3. Let $\hat{\boldsymbol{\Gamma}}_1$ be the semi-orthogonal basis for $\text{span}(\mathbf{V}_1)$, and let $\hat{\boldsymbol{\Gamma}}_0$ be the semi-orthogonal basis for $\text{span}(\mathbf{V}_0)$. Then, $\hat{\boldsymbol{\eta}} = \hat{\boldsymbol{\Gamma}}_1^T \hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\Omega}}_1 = \hat{\boldsymbol{\Gamma}}_1^T \hat{\boldsymbol{\Sigma}}_{\text{res}} \hat{\boldsymbol{\Gamma}}_1$, and $\hat{\boldsymbol{\Omega}}_0 = \hat{\boldsymbol{\Gamma}}_0^T \hat{\boldsymbol{\Sigma}}_{\mathbb{Y}} \hat{\boldsymbol{\Gamma}}_0$, where $\hat{\boldsymbol{\Sigma}}_{\mathbb{Y}}$ and $\hat{\boldsymbol{\Sigma}}_{\text{res}}$ are the marginal covariance matrix of \mathbb{Y} and the residual covariance matrix, respectively. Let $\log \det(\cdot)$ denote the composite function $\log \circ \det(\cdot)$. Then, the coordinate form of the $\log \mathbf{D}$ is as follows:

$$\log \mathbf{D} = \log \det \left(\Gamma_1^T \left(\mathbf{H}^T \hat{\rho}^{-1}(\boldsymbol{\theta}) \mathbf{H} - \mathbf{H}^T \hat{\rho}^{-1}(\boldsymbol{\theta}) \mathbf{G} \left(\mathbf{G}^T \hat{\rho}^{-1}(\boldsymbol{\theta}) \mathbf{G} \right)^{-1} \right. \right. \\ \left. \left. \mathbf{G}^T \hat{\rho}^{-1}(\boldsymbol{\theta}) \mathbf{H} \right) \Gamma_1 + \Gamma_0^T \left(\mathbf{H}^T \hat{\rho}^{-1}(\boldsymbol{\theta}) \mathbf{H} \right) \Gamma_0 \right), \quad (3.13)$$

where $\mathbf{H} = \mathbf{Y} - \bar{\mathbf{Y}} \otimes \mathbf{1}_n$, and $\mathbf{G} = \mathbf{X} - \bar{\mathbf{X}} \otimes \mathbf{1}_n$.

In order to obtain the parameters of the spatial envelope model, the objective function in (3.13) can be minimized using the gradient-based Grassmann optimization. To do this, first obtain initial values for $\hat{\Sigma}_{\mathbf{Y}}^0$, $\hat{\Sigma}_{\text{res}}^0$, and $\hat{\beta}_{MLE}$, the marginal covariance matrix of \mathbb{Y} , the residual covariance matrix, and the maximum likelihood estimate for $\boldsymbol{\beta}$ from the fit of the full model given in (3.6). Set $\Theta^1 = \Theta^0$, where $\Theta = \{\boldsymbol{\theta}, \mathbf{V}_0, \mathbf{V}_1\}$. Here, \mathbf{V}_0 and \mathbf{V}_1 can be obtained using the traditional envelope model, and $\boldsymbol{\theta}$ can be obtained using the linear coregionalization model. Then, we estimate $\mathbf{P}_{\mathbf{V}_1^m}$ by minimizing the objective function (3.13) over the Grassmann manifold $\mathbb{G}^{(r \times u)}$, and estimate $\mathbf{P}_{\mathbf{V}_0^m}$ by $\hat{\mathbf{P}}_{\mathbf{V}_0^m} = \mathbf{I} - \hat{\mathbf{P}}_{\mathbf{V}_1^m}$. In order to update the covariance function of the material and immaterial parts of the spatial envelope, fix $\boldsymbol{\theta}^m$ and estimate \mathbf{V}_0^m and \mathbf{V}_1^m by $\hat{\mathbf{V}}_0^m = \hat{\mathbf{P}}_{\mathbf{V}_0^m} \hat{\Sigma}_{\mathbf{Y}}^m \hat{\mathbf{P}}_{\mathbf{V}_0^m}$ and $\hat{\mathbf{V}}_1^m = \hat{\mathbf{P}}_{\mathbf{V}_1^m} \hat{\Sigma}_{\text{res}}^m \hat{\mathbf{P}}_{\mathbf{V}_1^m}$, respectively. Then, fix \mathbf{V}_0^m and \mathbf{V}_1^m and maximize $L^{(u)}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{V}_0^m, \mathbf{V}_1^m, \boldsymbol{\theta}^m)$ over $\boldsymbol{\theta}$ by solving the following minimization problem using a numerical algorithm, such as the Newton–Raphson method:

$$\hat{\boldsymbol{\theta}}^m = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \left\{ r \det(\boldsymbol{\rho}(\boldsymbol{\theta})) \right. \\ \left. + \frac{1}{2} \operatorname{tr} \left(\left(\mathbf{Q}_{(\boldsymbol{\rho}^{-1/2}(\boldsymbol{\theta}) \mathbf{G})} \boldsymbol{\rho}^{-1/2}(\boldsymbol{\theta}) \mathbf{H} \right) \mathbf{V}_1^{m\dagger} \left(\mathbf{Q}_{(\boldsymbol{\rho}^{-1/2}(\boldsymbol{\theta}) \mathbf{G})} \boldsymbol{\rho}(\boldsymbol{\theta})^{-1/2} \mathbf{H} \right)^T \right. \right. \\ \left. \left. + \boldsymbol{\rho}^{-1/2}(\boldsymbol{\theta}) \mathbf{H} \mathbf{V}_0^{m\dagger} \mathbf{H}^T \boldsymbol{\rho}^{-1/2}(\boldsymbol{\theta}) \right) \right\}. \quad (3.14)$$

Now, update $\hat{\Sigma}_{\mathbf{Y}}^m$ and $\hat{\Sigma}_{\text{res}}^m$ using the new estimates for $\mathbf{V}_0, \mathbf{V}_1$, and $\boldsymbol{\theta}$. Then, check the convergence. If $\|\Theta^{m+1} - \Theta^m\| < \delta$, where δ is a prespecified tolerance level, then stop the iteration, output the final spatial envelope estimators, and estimate $\boldsymbol{\beta}$ by $\hat{\boldsymbol{\beta}} = \hat{\mathbf{P}}_{\mathbf{V}_1} \hat{\beta}_{MLE}$; otherwise, set $m := m + 1$ and redo the procedure. Finally, estimate the intercept by $\hat{\boldsymbol{\alpha}} = \bar{\mathbf{Y}} - \bar{\mathbf{X}} \hat{\boldsymbol{\beta}}^T$. When the problem reduces to a standard envelope estimation problem, the fast algorithm for the envelope can be applied, such as that of Cook, Forzani and Su (2016).

4. Theoretical Properties

In what follows, we study the asymptotic properties of the spatial envelope parameter estimates. The regression coefficients can be written as $\beta = \Gamma_1 \eta$. Furthermore, $V_0 = \Gamma_0 \Omega_0 \Gamma_0^T$ and $V_1 = \Gamma_1 \Omega_1 \Gamma_1^T$ are the covariances of the immaterial part and the material part of the regression, respectively. Therefore, aside from the intercept, the parameters of the spatial envelope model in equation (3.6) can be combined into a vector, as follows:

$$\phi = \begin{bmatrix} \text{vec}(\eta) \\ \text{vec}(\Gamma_1) \\ \text{vech}(\Omega_1) \\ \text{vech}(\Omega_0) \end{bmatrix} \equiv \begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \phi_4 \end{bmatrix}, \tag{4.1}$$

where $\text{vec}(\cdot)$ denotes a vector operator, and $\text{vech}(\cdot)$ denotes a vector half operator. For background on these operators, see Seber (2008). Here, we focus on the following parameters under the spatial envelope model:

$$\psi(\phi) = \begin{bmatrix} \text{vec}(\beta^*) \\ \text{vech}(\mathbf{V}) \end{bmatrix} = \begin{bmatrix} \text{vec}(\eta^T \Gamma_1^T) \\ \text{vech}((\Gamma_1 \Omega_1 \Gamma_1^T + \Gamma_0 \Omega_0 \Gamma_0^T)) \end{bmatrix} \equiv \begin{bmatrix} \psi_1(\phi) \\ \psi_2(\phi) \end{bmatrix}. \tag{4.2}$$

Let

$$\Psi = \begin{bmatrix} \frac{\partial \psi_1}{\partial \phi_1^T} & \cdots & \frac{\partial \psi_1}{\partial \phi_4^T} \\ \frac{\partial \psi_2}{\partial \phi_1^T} & \cdots & \frac{\partial \psi_2}{\partial \phi_4^T} \end{bmatrix} \tag{4.3}$$

denote the gradient matrix. Using this matrix and following Cook, Li and Chiaromonte (2010), we present the following asymptotic properties of the proposed estimators.

Lemma 1. *Suppose $\bar{X} = 0$. Then, the Fisher information, J , for $\psi(\phi)$ in model (3.6) is as follows:*

$$\begin{aligned} J &= \begin{bmatrix} \frac{1}{n} \mathbb{X}^T (\mathbf{V}^{-1} \otimes \rho^{-1}(\theta)) \mathbb{X} & \mathbf{0} \\ \mathbf{0} & \frac{1}{2} \mathbf{E}_r^T (\mathbf{V}^{-1} \otimes \mathbf{V}^{-1}) \mathbf{E}_r \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{V}^{-1} \otimes \left(\frac{\mathbf{X}^T \rho^{-1}(\theta) \mathbf{X}}{n} \right) & \mathbf{0} \\ \mathbf{0} & \frac{1}{2} \mathbf{E}_r^T (\mathbf{V}^{-1} \otimes \mathbf{V}^{-1}) \mathbf{E}_r \end{bmatrix}, \end{aligned} \tag{4.4}$$

where $\mathbf{E}_r \in R^{r^2 \times r(r+1)/2}$ is an expansion matrix, such that, for a matrix \mathbf{A} , $\text{vec}(\mathbf{A}) = \mathbf{E}_r \text{vech}(\mathbf{A})$, and $\text{diag}(\mathbf{A})$ is a matrix of the diagonal elements of \mathbf{A} . The derivation of J is provided in the Supplementary Material, Section S4.

Theorem 1. Suppose $\bar{\mathbf{X}} = 0$ and \mathbf{J} is the Fisher information defined in Lemma 1. Let $\mathbf{\Lambda} = \mathbf{J}^{-1}$ be the asymptotic variance of the MLE under the full model. Then,

$$\sqrt{n}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}) \rightarrow N(\mathbf{0}, \mathbf{\Lambda}_0), \quad (4.5)$$

where $\mathbf{\Lambda}_0 = \Psi(\Psi^T \mathbf{\Lambda} \Psi)^\dagger \Psi$. Furthermore, $\mathbf{\Lambda}^{-1/2}(\mathbf{\Lambda} - \mathbf{\Lambda}_0)\mathbf{\Lambda}^{-1/2} \geq 0$, which means the asymptotic variance of the parameter estimation under the spatial envelope model is smaller than the estimate under the MLE. The proof for this theorem can be found in the Supplementary Material, Section S5.

Corollary 1. The asymptotic variance (avar) of $\sqrt{n}\boldsymbol{\beta}^*$ can be written as

$$\begin{aligned} & \text{avar}(\sqrt{n}\boldsymbol{\beta}^*) \\ &= K_{rp} \left\{ \left(\frac{\mathbf{X}^T \boldsymbol{\rho}(\boldsymbol{\theta})^{-1} \mathbf{X}}{n} \right)^{-1} \otimes \mathbf{\Gamma}_1 \mathbf{\Omega}_1 \mathbf{\Gamma}_1^T + (\boldsymbol{\eta}^T \otimes \mathbf{\Gamma}_0) (\Psi_2^T \mathbf{J} \Psi_2)^\dagger (\boldsymbol{\eta} \otimes \mathbf{\Gamma}_0^T) \right\} K_{rp}^T, \end{aligned} \quad (4.6)$$

where $\Psi_2 = (\partial\psi_1/\partial\phi_2^T, \partial\psi_2/\partial\phi_2^T)^T$, and $K_{rp} \in \mathbb{R}^{rp \times rp}$ is a unique matrix such that for a matrix \mathbf{A} , $\text{vec}(\mathbf{A}^T) = K_{rp} \text{vec}(\mathbf{A})$; that is, K_{rp} transforms the vec of a matrix into the vec of its transpose. The proof is available in the Supplementary Material, Section S6.

To gain further insight into the structure of the spatial envelope, we present a simplified version of the asymptotic variance of $\boldsymbol{\beta}^*$ for cases with one covariate, $\mathbf{\Omega}_1 = \sigma_1^2 \mathbf{I}_u$ and $\mathbf{\Omega}_0 = \sigma_0^2 \mathbf{I}_{r-u}$. Then, the asymptotic variance of $\boldsymbol{\beta}^*$ can be shown to be

$$\text{avar}(\sqrt{n}\boldsymbol{\beta}^*) = \frac{n\sigma_1^2}{\mathbf{X}^T \boldsymbol{\rho}^{-1}(\boldsymbol{\theta}) \mathbf{X}} \mathbf{\Gamma}_1 \mathbf{\Gamma}_1^T + \frac{n\sigma_0^2 \sigma_1^2 \|\boldsymbol{\beta}\|^2}{\mathbf{X}^T \boldsymbol{\rho}^{-1}(\boldsymbol{\theta}) \mathbf{X} \sigma_1^2 \|\boldsymbol{\beta}\|^2 + n(\sigma_0^2 - \sigma_1^2)^2} \mathbf{\Gamma}_0 \mathbf{\Gamma}_0^T. \quad (4.7)$$

For this simplified version, it can be shown that

$$\frac{\mathbf{V}_{SPEN}^{-1/2} \mathbf{V}_{EN} \mathbf{V}_{SPEN}^{-1/2}}{\mathbf{X}^T \boldsymbol{\rho}^{-1}(\boldsymbol{\theta}) \mathbf{X} / n \sigma_{\mathbf{X}}^2} = \mathbf{I}_r + \left(\frac{(\sigma_0^2 - \sigma_1^2)^2 (n\sigma_{\mathbf{X}}^2 / (\mathbf{X}^T \boldsymbol{\rho}^{-1}(\boldsymbol{\theta}) \mathbf{X}) - 1)}{(\sigma_0^2 - \sigma_1^2)^2 + \sigma_1^2 \sigma_{\mathbf{X}}^2 \|\boldsymbol{\beta}\|^2} \right) \mathbf{\Gamma}_0 \mathbf{\Gamma}_0^T, \quad (4.8)$$

where \mathbf{V}_{SPEN} is the asymptotic variance of the spatial envelope model, \mathbf{V}_{EN} is the asymptotic variance of the envelope model, and $\sigma_{\mathbf{X}}^2$ denotes the variance of \mathbf{X} , which is an $n \times 1$ vector. The proof for equation (4.8) can be found in the Supplementary Material, Section S7. This results indicates that when spatial

correlation does not exist, that is, $\rho(\boldsymbol{\theta}) = \mathbf{I}$, the two models have equal asymptotic variance. On the other hand, for cases in which spatial correlation exists, drawing an analytical conclusion that can be used to compare the asymptotic variances of the two models is very difficult. In this case, the comparison can be performed numerically.

5. Prediction

Prediction at an unsampled location is often a major objective of a spatial analysis. Let \mathbb{Y}_{new} be the $vec(\mathbf{Y}_{new})$ of the new multivariate response and \mathbb{X}_{new} be the predictor vector at an unsampled location. Then, the model can be written as follows:

$$\begin{pmatrix} \mathbb{Y}_{new} \\ \mathbb{Y} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\alpha} \otimes \mathbf{1}_{n_{new}} + \mathbb{X}_{new}\boldsymbol{\beta}^* \\ \boldsymbol{\alpha} \otimes \mathbf{1}_n + \mathbb{X}\boldsymbol{\beta}^* \end{pmatrix} + \begin{pmatrix} \boldsymbol{\epsilon}_{new} \\ \boldsymbol{\epsilon} \end{pmatrix} \sim N \left(\boldsymbol{\alpha} \otimes \mathbf{1}_N + \begin{pmatrix} \mathbb{X}_{new} \\ \mathbb{X} \end{pmatrix} \boldsymbol{\beta}^*, \boldsymbol{\Sigma} \right), \tag{5.1}$$

where $N = n + n_{new}$ and $\boldsymbol{\Sigma}$ is given as follows:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} = \begin{pmatrix} (\mathbf{V}_0 + \mathbf{V}_1) \otimes \boldsymbol{\rho}_{new,new}(\boldsymbol{\theta}) & (\mathbf{V}_0 + \mathbf{V}_1) \otimes \boldsymbol{\rho}_{new,\mathbf{Y}}(\boldsymbol{\theta}) \\ (\mathbf{V}_0 + \mathbf{V}_1) \otimes \boldsymbol{\rho}_{\mathbf{Y},new}(\boldsymbol{\theta}) & (\mathbf{V}_0 + \mathbf{V}_1) \otimes \boldsymbol{\rho}_{\mathbf{Y},\mathbf{Y}}(\boldsymbol{\theta}) \end{pmatrix}. \tag{5.2}$$

The conditional distribution $\mathbf{Y}_{new}|\mathbf{Y}$ is

$$\mathbf{Y}_{new}|\mathbf{Y}, \boldsymbol{\alpha}, \boldsymbol{\eta}, \mathbf{V}_0, \mathbf{V}_1, \boldsymbol{\theta} \sim N \left(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{Y} - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} \right), \tag{5.3}$$

where $\boldsymbol{\mu}_1 = \boldsymbol{\alpha} \otimes \mathbf{1}_{n_{new}} + \mathbb{X}_{new}\boldsymbol{\beta}^*$ and $\boldsymbol{\mu}_2 = \boldsymbol{\alpha} \otimes \mathbf{1}_n + \mathbb{X}\boldsymbol{\beta}^*$. Using the method described in Section 3, we can estimate the parameters of the model, and then estimate $E(\mathbf{Y}_{new}|\mathbf{Y})$ from the conditional distribution in (5.3).

6. Simulation

In this section, we carry out a simulation study to evaluate the finite-sample performance of the proposed spatial envelope model. Then, we compare the result with the performance of the traditional multivariate linear regression (MLR), LCM (Zhang (2007)), and envelope model (Cook, Li and Chiaromonte (2010)).

The data $\{(\mathbf{X}_1, \mathbf{Y}_1), \dots, (\mathbf{X}_n, \mathbf{Y}_n)\}$ are generated from the model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{6.1}$$

where $\mathbf{Y}_i \in \mathbb{R}^5$, $\mathbf{X}_i \in \mathbb{R}^6$, and the structural dimension $u = 2$. The matrix

$(\mathbf{\Gamma}_1; \mathbf{\Gamma}_0)$ is obtained by orthogonalizing a 5×5 matrix generated from uniform $(0, 1)$ variables. The elements of $\boldsymbol{\eta}$ follow a standard normal distribution, and $\boldsymbol{\beta} = \mathbf{\Gamma}_1 \boldsymbol{\eta}$. We generate $\boldsymbol{\Sigma}_Y = (\mathbf{\Gamma}_1 \boldsymbol{\Omega}_1 \mathbf{\Gamma}_1^T + 5 \mathbf{\Gamma}_0 \boldsymbol{\Omega}_0 \mathbf{\Gamma}_0^T) \otimes \boldsymbol{\rho}(\boldsymbol{\theta})$, where $\boldsymbol{\Omega}_1 = \{(-0.9)^{|i-j|}\}$ and $\boldsymbol{\Omega}_0 = \{(-0.5)^{|i-j|}\}$. For the spatial correlation function $\boldsymbol{\rho}(\boldsymbol{\theta})$, we use the following Matern covariance function:

$$\boldsymbol{\rho}(h; \boldsymbol{\theta}) = \frac{\sigma_m^2}{2^{\theta_2-1} \Gamma(\theta_2)} \left(\frac{\|h\|}{\theta_1} \right)^{\theta_1} \kappa_{\theta_2} \left(\frac{\|h\|}{\theta_1} \right),$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2)$, $\theta_1 > 0$ is the range parameter, θ_2 is the smoothness parameter, $\Gamma(\cdot)$ is the Gamma function, and κ_{θ_2} is the modified Bessel function of the second kind of order θ_2 (Abramowitz and Stegun (1964)). We assume $\boldsymbol{\epsilon}$ follows a normal distribution with mean zero and covariance $\boldsymbol{\Sigma}$. Three cases for the covariance $\boldsymbol{\Sigma}_Y$ are investigated. First, $\boldsymbol{\Sigma} = (\mathbf{\Gamma}_1 \boldsymbol{\Omega}_1 \mathbf{\Gamma}_1^T + 5 \mathbf{\Gamma}_0 \boldsymbol{\Omega}_0 \mathbf{\Gamma}_0^T)$, which serves as a benchmark in which the errors are independent of each other. For the second scenario, let $\boldsymbol{\rho}(\boldsymbol{\theta})$ be a Matern covariance function with $\sigma_m = 3$, $\theta_1 = 1$, and $\theta_2 = 0.5$. This case represents spatial correlation in the data with a short range of dependency. Finally, let $\boldsymbol{\rho}(\boldsymbol{\theta})$ be a Matern covariance function with $\sigma_m = 3$, $\theta_1 = 5$, and $\theta_2 = 0.5$. This case represents spatial correlation in the data with a long range of dependency.

The sample sizes are 100, 225, and 400. There are two ways in which to generate these samples. The first is based on 10×10 , 15×15 , and 20×20 , respectively, evenly spaced grids on $[0, 1]^2$. The second is to randomly choose 100, 225, and 400 locations from a 101×101 grid on $[0, 1]^2$. We use both sampling procedures to check whether the spatial distribution of the observations has any impact on the proposed estimation. All results reported here are based on 200 replications from the simulation model in each scenario. In order to compare the estimators, we use the Leave-One-Out Cross-Validation (LOCV) method, which provides a convenient approximation for the prediction error under a squared-error loss:

$$LOCV = \frac{1}{n} \sum_{i=1}^n (\hat{\mathbf{Y}}^{(-i)}(s_i) - \mathbf{Y}(s_{i,obs}))^T (\hat{\mathbf{Y}}^{(-i)}(s_i) - \mathbf{Y}(s_{i,obs})), \quad (6.2)$$

where $\mathbf{Y}(s_{i,obs})$ is the observed value for the response in location s_i , and $\hat{\mathbf{Y}}^{(-i)}(s_i)$ contains the predicted values of $\mathbf{Y}(s_i)$, computed after removing the i th row of the data. The Matlab package `Envlp` was used for all our simulation studies (Cook, Su and Yang (2014)). Tables 1 and 2 summarize the results of these simula-

Table 1. Prediction accuracy comparison based on the mean (standard deviation) of leave-one-out cross-validation (LOCV) for all 200 data sets from equally spaced samples. A smaller LOCV shows better performance.

| ϵ | n | MLR | LCM | Envelope | Spatial Envelope |
|------------|-----|-----------------|---------------|-----------------|------------------|
| 1 | 100 | 19.02 (1.537) | 20.01 (1.754) | 13.71 (1.547) | 14.28 (1.644) |
| | 225 | 18.49 (1.153) | 19.75 (1.659) | 11.49 (1.124) | 12.51 (1.234) |
| | 400 | 18.27 (0.828) | 19.02 (1.002) | 10.37 (0.812) | 10.87 (0.989) |
| 2 | 100 | 102.79 (35.570) | 22.54 (3.246) | 91.98 (36.379) | 20.21 (1.988) |
| | 225 | 101.57 (32.495) | 20.46 (2.897) | 89.24 (33.083) | 18.34 (1.450) |
| | 400 | 99.98 (32.185) | 18.89 (2.051) | 88.95 (31.855) | 17.68 (1.056) |
| 3 | 100 | 117.79 (48.834) | 24.19 (4.125) | 119.08 (47.852) | 21.36 (2.353) |
| | 225 | 103.22 (39.065) | 21.78 (3.278) | 104.73 (39.023) | 20.76 (2.012) |
| | 400 | 99.08 (37.718) | 19.45 (3.001) | 100.39 (36.896) | 18.10 (1.651) |

Table 2. Prediction accuracy comparison based on the mean (standard deviation) of leave-one-out cross-validation (LOCV) for all 200 data sets from random location samples. A smaller LOCV shows better performance.

| ϵ | n | MLR | LCM | Envelope | Spatial Envelope |
|------------|-----|-----------------|---------------|----------------|------------------|
| 1 | 100 | 20.12 (1.613) | 21.01 (1.863) | 14.32 (1.699) | 14.98 (1.722) |
| | 225 | 19.34 (1.231) | 19.68 (1.542) | 13.12 (1.234) | 13.19 (1.201) |
| | 400 | 17.83 (0.804) | 18.22 (1.101) | 11.73 (0.718) | 12.37 (0.819) |
| 2 | 100 | 104.02 (36.702) | 23.32 (4.111) | 93.02 (30.433) | 19.21 (2.004) |
| | 225 | 102.41 (34.521) | 21.41 (3.758) | 91.34 (27.211) | 17.34 (1.352) |
| | 400 | 100.39 (30.822) | 19.20 (3.201) | 89.21 (25.581) | 16.68 (1.110) |
| 3 | 100 | 116.34 (45.089) | 25.21 (4.821) | 97.01 (43.021) | 20.79 (2.115) |
| | 225 | 108.15 (34.211) | 22.35 (3.555) | 95.52 (31.774) | 18.92 (1.944) |
| | 400 | 101.54 (32.102) | 20.44 (2.998) | 90.94 (30.234) | 17.03 (1.234) |

tions. These tables provide the LOCV for different methods and different error distributions.

From the summary of all three error distributions, for the standard normal errors, where the observations are independent of each other, the spatial envelope provides comparable results to those of the envelope method, and both outperforms the MLR and LCM. In error distributions 2 and 3, where spatial dependency exists in the data, the spatial envelope method performs almost as well as it did in those cases without spatial dependency. In contrast, the original envelope method loses its efficiency. In addition, the spatial envelope outperforms

the LCM in both the independent and the dependent cases. Because the spatial envelope takes into account the spatial correlations between observations, its results are more accurate than those of the original envelope model. Furthermore, the spatial envelope uses only the material part of the data. As a result, the results are more efficient than those of the LCM, which uses both the material and the immaterial parts of the data. Therefore, we conclude that the proposed spatial envelope model provides consistent estimates with good prediction accuracy in all error distributions considered. This result is consistent for both sampling methods, which indicates that the spatial distribution of the observations has a minimal impact on the estimation.

As in Cook, Li and Chiaromonte (2010), it is possible for an objective function defined on Grassmann manifolds to have multiple local optimal points. One way to check this is to run the simulation with different starting values, and then to compare the results. In our numerical experiments, we did not find the local optima to be problematic for our method.

In order to investigate the accuracy of the asymptotic variance of $avar(\sqrt{n}\hat{\boldsymbol{\beta}}^*)$, presented in (4.7), we used the following simulation. The purpose of this simulation is to show that the variation of the spatial envelope estimator approaches its asymptotic variance derived in (4.7) when the sample size increases. The data are generated following model (6.1), with five responses and one covariate; that is, $\mathbf{Y}_i \in \mathbb{R}^5$, $\mathbf{X}_i \in \mathbb{R}$, and the structural dimension is $u = 1$. In addition, we let $\boldsymbol{\Omega}_1 = 5\mathbf{I}_u$, $\boldsymbol{\Omega}_0 = \mathbf{I}_{5-u}$, and $\boldsymbol{\eta} = \mathbf{1}$. The sample size n is 100, 225, 400, and 900, randomly chosen from a 101×101 grid on $[0, 1]^2$. For each sample size, 100 replications are performed to compute the estimation variance for the elements in $\hat{\boldsymbol{\beta}}$. For the spatial correlation, we used the Matern covariance function with $sigma_m = 3$, $\theta_1 = 2$, and $\theta_2 = 0.5$.

Figure 1 shows the simulation results of the asymptotic variance for a randomly selected element of $\hat{\boldsymbol{\beta}}$. The left panel of Figure 1 shows the asymptotic variance for the independent case, and the right panel shows the same results for the spatially correlated data for the envelope and the spatial envelope. The blue line shows the estimated standard deviation of the envelope estimator, and the black line denotes the estimated standard deviation of the spatial envelope estimator. Thus, for the standard normal errors, where the observations are independent of each other, the variances of the spatial envelope and the envelope method are very similar. On the other hand, when spatial dependency exists in the data, the spatial envelope method outperforms the envelope method.

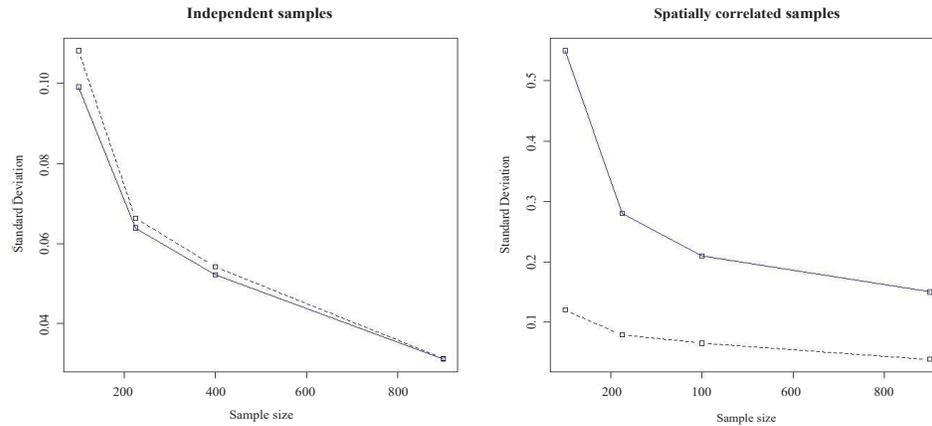


Figure 1. Simulation results of the asymptotic variance for a randomly selected element of $\hat{\beta}$ for the envelope and the spatial envelope for the independent case (**left panel**), and for spatially correlated data (**right panel**). The blue solid line shows the estimated standard deviation of the envelope estimator, and the black dashed line denotes the estimated standard deviation of the spatial envelope estimator.

7. Application

In this section, we apply the proposed methodology to air pollution data for the northeastern United States. Note that the main purpose of this data analysis is to determine the proposed approach can be used to find the reduced response space in a multivariate spatial data analysis. The data employed here have garnered attention from both statisticians and scientists in other areas. For example, researchers have used the data to examine climate change (Phelan et al. (2016)), health science (Kioumourtzoglou et al. (2016)), and air quality (Battye et al. (2016)). These studies showed that relationships exist between air pollution and meteorological factors, such as wind, temperature, and humidity. Most existing studies focus on one of these pollutants. However, because the pollutants are correlated, it is worth studying them simultaneously.

The data on pollutants and weather include the average levels of the following variables in January 2015. We choose a group of ambient air pollutants monitored by the EPA, because these present a high threat to human health. Specifically, we have eight response variables: ground level ozone, sulfur dioxide (SO_2), carbon monoxide (CO), nitrogen dioxide (NO_2), nitrogen monoxide (NO), lead, PM 2.5, and PM 10. PM 10 includes particles less than or equal to 10 micrometers in diameter. Similarly, PM 2.5 includes particles less than or equal to 2.5 micrometers,

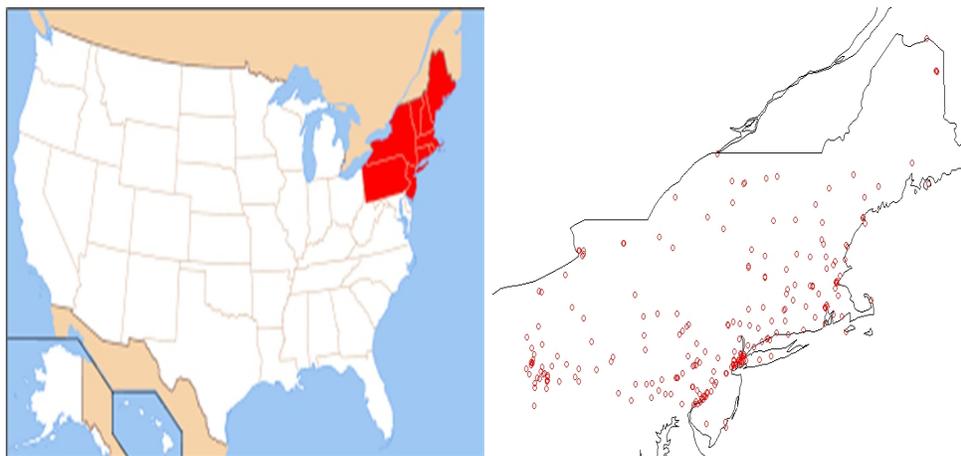


Figure 2. **Left:** Study area in the United States. States of interest are shaded in red. **Right:** Location of sites in the study area. Note that there are more sites in places with larger populations.

also called fine particle pollution. These data also include the following meteorological variables as predictors: wind, temperature, and relative humidity. Along with this information, the latitude and longitude of each monitoring location are used to model the spatial structure in the data. Our study area consists of nine states in the northeast of the United States: Connecticut, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, and Vermont. This data set is available at http://aqsd1.epa.gov/aqsweb/aqstmp/airdata/download_files.html#Daily. Figure 2 shows the study area and the location of 270 air monitoring sites.

The preliminary analysis using Moran's I and plots of the empirical variogram determined that spatial correlation does exist in these data. The results of the preliminary analysis can be found in the Supplementary Material, Section S8. Cross-validation shows that the best choice for the structural dimension is three. The Matern's covariance parameters, θ_1 and θ_2 , are estimated to be 0.51 and 0.91, respectively, indicating the existence of spatial dependency in the data. The corresponding direction estimates ($\hat{\Gamma}_1$) from the spatial envelope are given in Table 3. Note that $\hat{\Gamma}_1$ is not unique, and can be any orthonormal basis of the envelope subspace. The estimated regression coefficients and their standard deviations are available in the Supplementary Material, Section S9.

By checking the estimated basis coefficients of the minimal subspace (directions) and the regression coefficients, we find that sulfur dioxide, nitrogen dioxide, PM 10, and PM 2.5 dominate each of the three directions, respectively. Using

Table 3. Direction estimates using the spatial envelope for air pollution data on the northeastern United States.

| Variable | Direction 1 | Direction 2 | Direction 3 |
|-------------------|-------------|-------------|-------------|
| Ozone | -0.0464 | 0.0432 | -0.0080 |
| Carbon monoxide | 0.2840 | -0.3717 | -0.0179 |
| Lead | -0.0739 | 0.0872 | 0.0008 |
| Nitrogen dioxide | -0.5089 | 0.2612 | -0.4639 |
| Nitrogen monoxide | -0.3056 | -0.1137 | 0.2757 |
| Sulfur dioxide | -0.5335 | 0.0241 | -0.2981 |
| PM10 | -0.3257 | -0.8667 | -0.0506 |
| PM2.5 | -0.4106 | 0.1394 | 0.7855 |

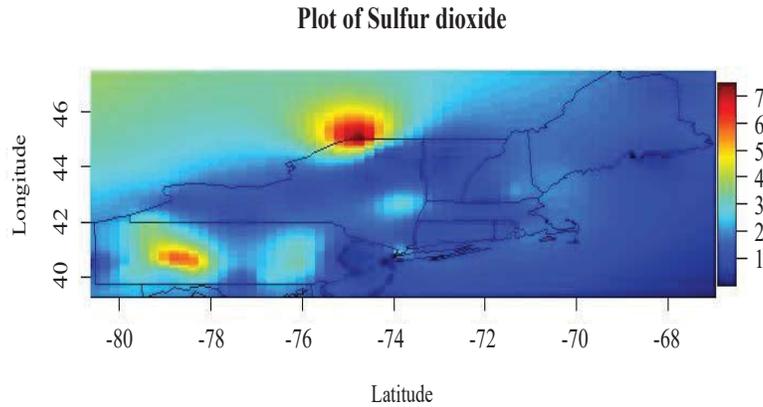


Figure 3. Prediction plot for sulfur dioxide for the study area. Sulfur dioxide is moderately high for most of the study area, and is extremely high in Johnstown, which is characterized by defense manufacturing.

fossil fuels creates sulfur dioxide, nitrogen monoxide, and nitrogen dioxide. The nitrogen monoxide will also become nitrogen dioxide in the atmosphere. The existence of particles in the air leads to a reduction in visibility and causes the air to become hazy when levels are elevated. Furthermore, because these particles can travel into human lungs, they can cause health problem such as lung cancer. The main source of these particles in the air is pollutants emitted from power plants, industries, and automobiles.

Figures 3 to 6 show the prediction plots for the three pollutants with the largest impact. Figure 3 shows the prediction plot for sulfur dioxide for the study area. Sulfur dioxide is moderately high for most of the study area. In addition,

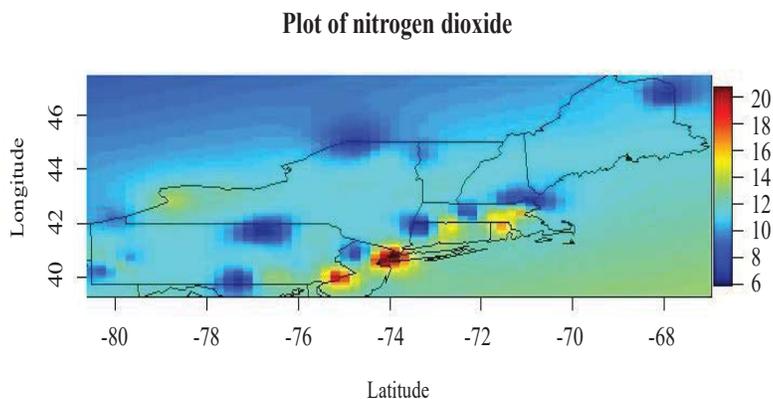


Figure 4. Prediction plot for nitrogen dioxide for the study area. Nitrogen dioxide is high in Newark, New York, Philadelphia, and Rhode Island, which are all highly populated areas.

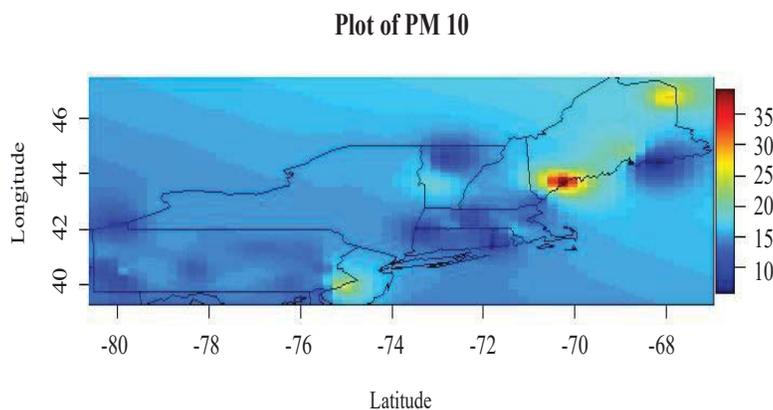


Figure 5. Prediction plot for PM 10 for the study area. PM 10 is high for most of the study area, especially in Philadelphia and Augusta.

levels are extremely high in Johnstown, which is characterized by defense manufacturing. Figure 4 shows the prediction plot for nitrogen dioxide for the study area. Nitrogen dioxide is high in Newark, New York, Philadelphia, and Rhode Island, which are all highly populated areas. Figure 5 shows the prediction plot for PM 10 for the study area. PM 10 is high in most of the study area, especially in Philadelphia and Augusta. Figure 6 shows the prediction plot for PM 2.5 for the study area. PM 2.5 is moderately high almost everywhere, especially in Pennsylvania state, Augusta, and the middle of Vermont state. Prediction plots

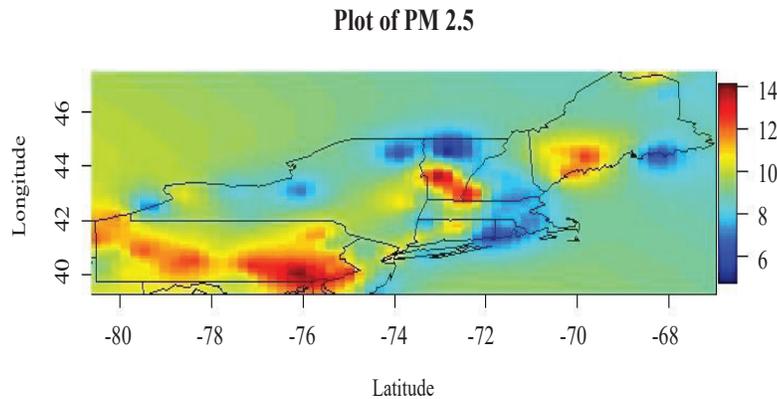


Figure 6. Prediction plot for PM 2.5 for the study area. PM 2.5 is moderately high almost everywhere in the study area, especially in Pennsylvania state, Augusta, and the middle of Vermont state.

of the other variables can be found in Supplementary Material, Section S10.

The square root of the leave-one-out cross-validation for the MLR, LCM, envelope, and spatial envelope are 7.537, 3.562, 4.876, and 1.978, respectively. This result shows that, the spatial envelope outperforms other methods and provides more accurate predictions. In summary, we find that the most important pollutants in January are particulates, sulfur, and nitrogen, and that other pollutants have minimal effect. These statistical conclusions support the environmental chemical claim that in cold weather, owing to the burning of fossil fuels and inversion, sulfur dioxide, nitrogen dioxide, and particulate matter are the most important pollutants (Byers (1959); Læg Reid, Bockman and Kaarstad (1999)).

8. Conclusion

Air pollution has a serious impact on human health. Research has greatly improved our understanding of pollutants and their relationship with weather conditions. However, relatively few studies examine the effects of meteorological variables on several pollutants together. Motivated by an analysis of air pollution in the northeastern United States, we have proposed a new parsimonious multivariate spatial model. Here, we focused on inferences and on constructing a method that can provide estimations for the parameters of interest more efficiently than traditional maximum likelihood estimators are able to do, by capturing the spatial structure in the data.

Our model is flexible enough to characterize complex dependency and cross-dependency structures of different pollutants. The results of a simulation study and a real-data analysis showed that the proposed spatial envelope model outperforms the multivariate linear regression, envelope, and linear coregionalization models. This new approach provides a more efficient estimation for regression coefficients than that of the traditional maximum likelihood approach.

The method presented in this paper is for a multivariate spatial response with a separable covariance matrix. This framework can be extended to cases in which the covariance matrix is nonseparable. Furthermore, the current work assumes normality in the derivations of the estimators. Confirming that the normality assumption is satisfied is more important for spatial random fields than when working with envelope models. The violation of the normality assumption brings computational and theoretical challenges Diggle, Tawn and Moyeed (1998); Liu et al. ((2017)). Incorporating the notation of an envelope with a multivariate nonGaussian spatial random field, which is beyond the scope of this paper, is a very interesting and challenging topic, as is the misspecification of the spatial structure. Investigating the potential cost of such a misspecifying is important, because it can affect the estimation of the coefficient and the prediction. Another possible extension of the proposed methodology is to include cases with spatiotemporal responses. These topics are left to future research.

Supplementary Material

The online Supplementary Material contains a brief description of the linear coregionalization model (LCM), as well the derivations and proof of the likelihood factorization, theorem, and corollaries presented in the main manuscript.

Acknowledgments

The authors would like to thank the Editor, an Associate Editor, and two referees for their careful reading and many constructive comments.

References

- Abramowitz, M. and Stegun, I. A. (1964). *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables*. Courier Corporation.
- Battye, W. H., Bray, C. D., Aneja, V. P., Tong, D., Lee, P. and Tang, Y. (2016). Evaluating ammonia (NH₃) predictions in the NOAA National Air Quality Forecast Capability (NAQFC) using in situ aircraft, ground-level, and satellite measurements from the DISCOVER-AQ Colorado campaign. *Atmospheric Environment* **140**, 342–351.

- Byers, H. R. (1959). *General Meteorology*. McGraw-Hill.
- Chiles, J. P. and Delfiner, P. (1999). *Geostatistics: Modeling Spatial Uncertainty*. John Wiley & Sons.
- Christensen, R. (2001). *Advanced Linear Modeling: Multivariate, Time Series, and Spatial Data; Nonparametric Regression and Response Surface Maximization*. Springer Science & Business Media.
- Cook, R. D., Li, B. and Chiaromonte, F. (2010). Envelope models for parsimonious and efficient multivariate linear regression. *Statistica Sinica* **20**, 927–960.
- Cook, R. D., Su, Z. and Yang, Y. (2014). Envlp: A MATLAB toolbox for computing envelope estimators in multivariate analysis. *Journal of Statistical Software* **62**, 1–20.
- Cook, R. D. and Zhang, X. (2015). Simultaneous envelopes for multivariate linear regression. *Technometrics* **57**, 11–25.
- Cook, R. D. and Zhang, X. (2016). Algorithms for envelope estimation. *Journal of Computational and Graphical Statistics* **25**, 284–300.
- Cook, R. D., Forzani, L. and Su, Z. (2016). A note on fast envelope estimation. *Journal of Multivariate Analysis* **150**, 42–54.
- Cooper, D. R., Schindler, P. S. and Sun, J. (2003). *Business Research Methods*. McGraw-Hill/Irwin New York, NY.
- Cressie, N. (1993). *Statistics for Spatial Data*. John Wiley & Sons.
- Diggle, P. J., Tawn, J. A. and Moyeed, R. A. (1998). Model based geostatistics (with discussion). *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **47**, 299–350.
- Edelman, A., Arias, T. A. and Smith, S T. (1998). The geometry of algorithms with orthogonality constraints. *SIAM journal on Matrix Analysis and Applications* **20**, 303–353.
- Genton, M. G. and Kleiber, W. (2015). Cross-covariance functions for multivariate geostatistics. *Statistical Science* **30**, 147–163.
- Gneiting, T., Kleiber, W. and Schlather, M. (2010). Matérn cross-covariance functions for multivariate random fields. *Journal of the American Statistical Association* **105**, 1167–1177.
- Goulard, M and Voltz, M. (1992). Linear coregionalization model: tools for estimation and choice of cross-variogram matrix. *Mathematical Geology* **24**, 269–286.
- Guinness, J., Fuentes, M., Hesterberg, D. and Polizzotto, M. (2014). Multivariate spatial modeling of conditional dependence in microscale soil elemental composition data. *Spatial Statistics* **9**, 93–108.
- Kioumourtzoglou, M. A., Schwartz, J. D., Weisskopf, M. G., Melly, S. J., Wang, Y., Dominici, F. and Zanobetti, A. (2016). Long-term PM2.5 exposure and neurological hospital admissions in the northeastern United States. *Environmental Health Perspectives (Online)* **124**, 23–29.
- Læg Reid, M and Bockman, O C and Kaarstad, O. (1999). *Agriculture, Fertilizers and the Environment*. CABI publishing.
- Lave, L. B. and Seskin, E. P. (1973). An analysis of the association between US mortality and air pollution. *Journal of the American Statistical Association* **68**, 284–290.
- Li, L. and Zhang, X. (2017). Parsimonious tensor response regression. *Journal of the American Statistical Association* **112**, 1–16.
- Liang, K. Y., Zeger, S. L. and Qaqish, B. (1992). Multivariate regression analyses for categorical data. *Journal of the Royal Statistical Society. Series B (Statistical Methodological)* **54**, 3–40.

- Liu, X., Chen, F., Lu, Y. C. and Lu, C. T. (2017). Prediction for Multivariate Non-Gaussian Data. *ACM Transactions on Knowledge Discovery from Data (TKDD)* **11**, 36:1–36:27.
- Myers, D. E. (1991). Pseudo-cross variograms, positive-definiteness, and cokriging. *Mathematical Geology* **23**, 805–816.
- Phelan, J., Belyazid, S., Jones, P., Cajka, J., Buckley, J. and Clark, C. (2016). Assessing the Effects of Climate Change and Air Pollution on Soil Properties and Plant Diversity in Sugar Maple–Beech–Yellow Birch Hardwood Forests in the Northeastern United States: Model Simulations from 1900 to 2100. *Water, Air, & Soil Pollution* **227**, 1–30.
- Seber, G. A. F. (2008) *A Matrix Handbook for Statisticians*. John Wiley & Sons.
- Su, Z. and Cook, R. D. (2011). Partial envelopes for efficient estimation in multivariate linear regression. *Biometrika* **98**, 133–146.
- Su, Z. and Cook, R. D. (2012). Inner envelopes: efficient estimation in multivariate linear regression. *Biometrika* **99**, 687–702.
- Su, Z. and Cook, R. D. (2013). Estimation of multivariate means with heteroscedastic errors using envelope models. *Statistica Sinica* **23**, 213–230.
- Ver Hoef, J. M. and Barry, R. P. (1998). Constructing and fitting models for cokriging and multivariable spatial prediction. *Journal of Statistical Planning and Inference* **69**, 275–294.
- Wackernagel, H. (2003). *Multivariate Geostatistics: an Introduction with Applications*. Springer Science & Business Media
- Zhang, H. (2007). Maximum-likelihood estimation for multivariate spatial linear coregionalization models. *Environmetrics* **18**, 125–139.
- Zhang, X. and Li, L. (2017). Tensor envelope partial least-squares regression. *Technometrics* **59**, 426–436.
- Zhang, X. and Mai, Q. (2017) Model-free envelope dimension selection. *arXiv preprint arXiv:1709.03945*.
- Zhang, X. Wang, C. and Wu, Y. (2018). Functional envelope for model-free sufficient dimension reduction. *Journal of Multivariate Analysis* **163**, 37–50.

Department of Mathematics and Statistics, South Dakota State University, Brookings, SD 57007, USA.

E-mail: hossein.moradirekabdarkolae@sdstate.edu

Department of Information Systems, Statistics, and Management Science, The University of Alabama, Tuscaloosa, AL 35487, USA.

E-mail: qwang57@culverhouse.ua.edu

Department of Statistics, Shahid Chamran University of Ahvaz, Ahvaz, Iran.

E-mail: z.naji64@gmail.com

Department of Statistics and Actuarial Science, The University of Iowa, Iowa City, IA 52242, USA.

E-mail: montserrat.fuentes@uiowa.edu

(Received October 2017; accepted September 2018)