

## NONSTATIONARY SPATIAL MODELING USING PENALIZED LIKELIHOOD

Yin-Ping Chen and Hsin-Cheng Huang

*National Central University and Academia Sinica*

*Abstract:* We consider nonstationary spatial modeling using empirical orthogonal functions (EOFs) under the consideration that data may be observed only at some sparse, irregularly spaced locations with repeated measurements. Instead of obtaining EOFs by principal component analysis based on a class of pre-specified basis functions or a pre-smoothing step with data imputed on a regular grid, we propose a regularization approach using penalized likelihood, which provides a new EOF-type expansion in terms of a small number of functions with their degrees of smoothness controlled by a tuning parameter. An expectation-conditional-maximization algorithm is applied to obtain the penalized maximum likelihood estimates of the mean and the covariance parameters simultaneously. Some simulation results show that the proposed method performs well in both spatial prediction and covariance function estimation, regardless of whether the underlying spatial process is stationary or nonstationary. In addition, the method is applied to analyze precipitation in Colorado.

*Key words and phrases:* Empirical orthogonal function, Karhunen Loève expansion, smoothing splines, spatial prediction, thin-plate splines.

### 1. Introduction

The empirical orthogonal function (EOF) technique, termed by Lorenz (1956) in meteorology, is an effective tool for nonstationary spatial modeling. It can be regarded as an extension of the conventional principal component analysis (PCA) to a continuous spatial domain in terms of some basis functions (Obukhov (1960); Holmström (1963); Obled and Creutin (1986)). An early review of EOF analysis can be found in Alishouse et al. (1967). Some more recent reviews on EOF analysis are provided by Wilks (1995), Preisendorfer (1988), Jolliffe (2002), Hannachi, Jolliffe, and Stephenson (2007), and Cressie and Wikle (2011).

The aim of EOF analysis is to achieve a low dimensional representation of a continuous spatial (or spatial-temporal) process and its covariance function. Consider a sequence of zero-mean nonstationary Gaussian spatial processes  $\{Y_t(\mathbf{s}) : \mathbf{s} \in D\}$  defined on a compact region  $D \subset \mathbb{R}^d$ , which have a common covariance function  $C_Y(\mathbf{s}, \mathbf{s}') \equiv \text{cov}(Y_t(\mathbf{s}), Y_t(\mathbf{s}'))$ ,  $\mathbf{s}, \mathbf{s}' \in D$ , for  $t \in \mathbb{N} \equiv \{1, 2, \dots\}$ .

If  $C_Y(\cdot, \cdot)$  is continuous, then  $Y_t(\mathbf{s})$  is known to possess the following biorthogonal decomposition:

$$Y_t(\mathbf{s}) = \sum_{k=1}^{\infty} w_{t,k} \phi_k(\mathbf{s}), \quad (1.1)$$

where  $w_{t,k} \sim N(0, \lambda_k)$ , for  $k, t \in \mathbb{N}$ , are uncorrelated variables with  $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$ ,  $\phi_k(\cdot)$ 's are orthonormal functions in the sense that  $\int_D \phi_k(\mathbf{s}) \phi_{k'}(\mathbf{s}) d\mathbf{s} = \delta_{k,k'}$ , and  $\delta_{k,k'}$  denotes the Kronecker delta. The decomposition (1.1) is called the Karhunen-Loève expansion (Karhunen (1947); Loève (1978)). It follows that  $w_{t,k} = \int_D Y_t(\mathbf{s}) \phi_k(\mathbf{s}) d\mathbf{s}$ ,  $k, t \in \mathbb{N}$ , and the covariance function,  $C_Y(\mathbf{s}, \mathbf{s}')$ , satisfies  $C_Y(\mathbf{s}, \mathbf{s}') = \sum_{k=1}^{\infty} \lambda_k \phi_k(\mathbf{s}) \phi_k(\mathbf{s}')$ . Thus  $\phi_k(\cdot)$ 's are the solutions of the Fredholm integral equation:

$$\int_D C_Y(\mathbf{s}, \mathbf{s}') \phi_k(\mathbf{s}') d\mathbf{s}' = \lambda_k \phi_k(\mathbf{s}), \quad \mathbf{s} \in D, \quad k \in \mathbb{N}, \quad (1.2)$$

where  $\phi_k(\cdot)$  is the  $k$ th eigenfunction of  $C_Y(\cdot, \cdot)$  and  $\lambda_k$  is the corresponding eigenvalue.

In practice,  $C_Y(\cdot, \cdot)$  is unknown and the process  $Y_t(\cdot)$  is only observed at some finite number of locations. Hence the  $\phi_k(\cdot)$ 's and  $\lambda_k$ 's have to be estimated from data. There are basically two approaches when the data locations are irregularly located. One approach is to restrict the  $\phi_k(\cdot)$ 's to a function space generated by some pre-specified basis functions (e.g., Cohen and Jones (1969); Deville (1974); Creutin and Obled (1982); Obled and Creutin (1986); Braud, Obled, and Phamdihtuan (1993)). Then the conventional or a spatially weighted version of PCA can be directly applied. Nevertheless, the resulting eigenfunction estimates may be sensitive to the choice of the basis functions. The other approach utilizes a two-step procedure by first imputing data at each  $t$  on a fine grid, and then applying the conventional PCA to the processed data (e.g., Karl, Koscielny, and Diaz (1982); Wikle and Cressie (1999)). However, this approach may be sensitive to the choice of a smoothing method and is likely to underestimate the variances at locations where there are few data nearby.

In this paper, we derive a likelihood-based approach for EOF analysis incorporating the possibility that the data locations may be sparse and irregularly spaced. Instead of applying a two-step procedure, we propose a regularization approach under the penalized likelihood framework; this provides a new estimation method of EOFs with degrees of smoothness controlled by a tuning parameter. In addition, we incorporate a stationary process in the model similar to that considered by Wikle and Cressie (1999), which allows us to reduce the model dimension further.

The article is outlined as follows. Section 2 introduces our model and the proposed penalized likelihood method, as well as some formulae for spatial prediction. Section 3 develops an expectation-conditional-maximization (ECM; Meng and Rubin (1993)) algorithm for simultaneous estimation of the mean and the covariance parameters. Some simulation results on one- and two-dimensional spaces are given in Section 4, and an application of our method to precipitation in Colorado is given in Section 5. Finally, Section 6 contains a brief discussion.

## 2. The Proposed Method

Consider a sequence of Gaussian (nonstationary) spatial processes  $\{Y_t(\mathbf{s}) : \mathbf{s} \in D\}$  of interest defined on a compact region  $D \subset \mathbb{R}^d$ , which are independent at time  $t \in \mathbb{N}$ . Suppose that data  $\mathbf{Z}_t = (Z_t(\mathbf{s}_1), \dots, Z_t(\mathbf{s}_n))'$  are observed at locations  $\mathbf{s}_1, \dots, \mathbf{s}_n$  and times  $t = 1, \dots, T$ , with added Gaussian white noise according to

$$Z_t(\mathbf{s}_i) = Y_t(\mathbf{s}_i) + \varepsilon_t(\mathbf{s}_i), \text{ with } \varepsilon_t(\mathbf{s}_i) \sim N(0, \sigma_\varepsilon^2), \quad i = 1, \dots, n. \quad (2.1)$$

Motivated by (1.1), we propose a model for  $\{Y_t(\mathbf{s}) : \mathbf{s} \in D\}$  as

$$Y_t(\mathbf{s}) = \mathbf{x}_t(\mathbf{s})' \boldsymbol{\beta} + \sum_{k=1}^K w_{t,k} \phi_k(\mathbf{s}) + \xi_t(\mathbf{s}), \quad \mathbf{s} \in D, \quad t \in \mathbb{N}, \quad (2.2)$$

where  $\mathbf{x}_t(\mathbf{s})$  is the  $p$ -vector regressor at location  $\mathbf{s}$ ,  $\boldsymbol{\beta}$  is the corresponding regression parameter vector,  $w_{t,k} \sim N(0, \lambda_k)$ ,  $k = 1, \dots, K$ , are uncorrelated,  $\phi_k(\cdot)$ 's are unknown functions, and  $\xi_t(\cdot)$  is a zero-mean stationary spatial process with its covariance function parameterized by  $\boldsymbol{\theta}_\xi$ , and independent of the  $w_{t,k}$ 's. We can rewrite  $\mathbf{Z}_t$  in matrix form as

$$\mathbf{Z}_t = \mathbf{X}_t \boldsymbol{\beta} + \boldsymbol{\Phi}_K \mathbf{w}_t + \boldsymbol{\xi}_t + \boldsymbol{\varepsilon}_t, \quad t = 1, \dots, T, \quad (2.3)$$

where  $\mathbf{X}_t \equiv (\mathbf{x}_t(\mathbf{s}_1), \dots, \mathbf{x}_t(\mathbf{s}_n))'$ ,  $\boldsymbol{\phi}_k \equiv (\phi_k(\mathbf{s}_1), \dots, \phi_k(\mathbf{s}_n))'$ ,  $k = 1, \dots, K$ ,  $\boldsymbol{\Phi}_K \equiv (\boldsymbol{\phi}_1, \dots, \boldsymbol{\phi}_K)$ ,  $\mathbf{w}_t \equiv (w_{t,1}, \dots, w_{t,K})'$ ,  $\boldsymbol{\xi}_t \equiv (\xi_t(\mathbf{s}_1), \dots, \xi_t(\mathbf{s}_n))'$ , and  $\boldsymbol{\varepsilon}_t \equiv (\varepsilon_t(\mathbf{s}_1), \dots, \varepsilon_t(\mathbf{s}_n))'$ . As in Wikle and Cressie (1999), we consider a stationary component  $\boldsymbol{\xi}_t$  for global behavior which, as to be shown in simulations, is effective in further reducing the model dimension. Let  $\mathbf{V}_\xi(\boldsymbol{\theta}_\xi) \equiv \text{var}(\boldsymbol{\xi}_t)/\sigma_\varepsilon^2$ . Then the covariance matrix of  $\mathbf{Z}_t$  can be written as

$$\boldsymbol{\Sigma}_Z(\boldsymbol{\theta}) \equiv \text{var}(\mathbf{Z}_t) = \sum_{k=1}^K \lambda_k \boldsymbol{\phi}_k \boldsymbol{\phi}_k' + \sigma_\varepsilon^2 (\mathbf{V}_\xi(\boldsymbol{\theta}_\xi) + \mathbf{I}_n),$$

where  $\boldsymbol{\theta} \equiv (\lambda_1, \dots, \lambda_K, \boldsymbol{\phi}_1', \dots, \boldsymbol{\phi}_K', \sigma_\varepsilon^2, \boldsymbol{\theta}_\xi')$  consists of all the covariance parameters.

**2.1. Penalized maximum likelihood**

For simplicity, we take  $d = 1, 2$ . We consider a penalized maximum likelihood (PML) method for estimating  $\beta$  and  $\theta$  with the roughness penalties  $J(\phi_1), \dots, J(\phi_K)$  imposed on the functions,  $\phi_1(\cdot), \dots, \phi_K(\cdot)$ , leading to the objective function

$$T \log |\Sigma_Z(\theta)| + \sum_{t=1}^T (\mathbf{Z}_t - \mathbf{X}_t \beta)' \Sigma_Z^{-1}(\theta) (\mathbf{Z}_t - \mathbf{X}_t \beta) + \alpha (J(\phi_1) + \dots + J(\phi_K)), \tag{2.4}$$

subject to an identifiability constraint,  $\phi'_k(\mathbf{V}_\xi(\theta_\xi) + \mathbf{I}_n)^{-1} \phi_k = 1, k = 1, \dots, K$ , where  $\alpha > 0$  is a smoothing parameter and the roughness penalty has the form

$$J(\phi_k) = \begin{cases} \int_D (\phi''_k(\mathbf{s}))^2 d\mathbf{s}, & \text{for } d = 1, \\ \int_{\mathbf{s}=(x,y) \in D} \left\{ \left( \frac{\partial^2 \phi_k(x,y)}{\partial x^2} \right)^2 + 2 \left( \frac{\partial^2 \phi_k(x,y)}{\partial x \partial y} \right)^2 + \left( \frac{\partial^2 \phi_k(x,y)}{\partial y^2} \right)^2 \right\} d\mathbf{s}, & \text{for } d = 2. \end{cases}$$

Note that the minimizers for  $\phi_1(\cdot), \dots, \phi_K(\cdot)$  in (2.4) are natural cubic splines and thin-plate splines with knots at  $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$  for  $d = 1$  and  $d = 2$ , respectively (see Green and Silverman (1994)). Also note that when  $\mathbf{V}_\xi(\theta_\xi)$  is assumed to be  $\mathbf{0}$ , the cost function (2.4) is similar to those considered by Yao, Müller, and Wang (2005) and Huang, Shen, and Buja (2008) in a least squares setting for analyzing sparse longitudinal data.

For  $d = 1$ , Green and Silverman (1994) showed that  $J(\phi_k)$  can be written in the form of  $\phi'_k \mathbf{\Omega} \phi_k, k = 1, \dots, K$ . Specifically,  $\mathbf{\Omega}$  can be expressed in terms of  $h_i \equiv s_{i+1} - s_i, i = 1, \dots, n - 1$ , as  $\mathbf{\Omega} = \mathbf{Q} \mathbf{P}^{-1} \mathbf{Q}'$ , where  $\mathbf{Q}$  is an  $n \times (n - 2)$  banded matrix with entries

$$q_{j,k} = \begin{cases} h_k^{-1}, & \text{if } j = k = 1, \dots, n - 2, \\ -h_k^{-1} - h_{k+1}^{-1}, & \text{if } j = k + 1 = 2, \dots, n - 1, \\ h_{k+1}^{-1}, & \text{if } j = k + 2 = 3, \dots, n, \\ 0, & \text{otherwise,} \end{cases}$$

and  $\mathbf{P}$  is an  $(n - 2) \times (n - 2)$  banded symmetric matrix with entries

$$p_{j,k} = \begin{cases} \frac{h_k + h_{k+1}}{3}, & \text{if } j = k = 1, \dots, n - 2, \\ \frac{h_{k+1}}{6}, & \text{if } |j - k| = 1, k = 1, \dots, n - 3, \\ 0, & \text{otherwise.} \end{cases}$$

Similarly for  $d = 2$ , it follows from Green and Silverman (1994) that  $J(\phi_k) = \phi_k' \boldsymbol{\Omega} \phi_k$ ,  $k = 1, \dots, K$ , where

$$\boldsymbol{\Omega} = \mathbf{E}^{-1} - \mathbf{E}^{-1} \mathbf{T}' (\mathbf{T} \mathbf{E}^{-1} \mathbf{T}')^{-1} \mathbf{T} \mathbf{E}^{-1},$$

$\mathbf{T}$  is a  $3 \times n$  matrix with the  $i$ th column  $(1, \mathbf{s}'_i)'$ ,  $\mathbf{E}$  is an  $n \times n$  symmetric matrix with the  $(i, j)$ th entry  $E_{ij} = \eta(\|\mathbf{s}_i - \mathbf{s}_j\|)$ , and  $\eta(r) = (1/8\pi)r^2 \log r$ . Consequently, the penalized log-likelihood of (2.4) can be rewritten in matrix form as

$$\ell_Z(\boldsymbol{\beta}, \boldsymbol{\theta}) \equiv T \log |\boldsymbol{\Sigma}_Z(\boldsymbol{\theta})| + \sum_{t=1}^T (\mathbf{Z}_t - \mathbf{X}_t \boldsymbol{\beta})' \boldsymbol{\Sigma}_Z^{-1}(\boldsymbol{\theta}) (\mathbf{Z}_t - \mathbf{X}_t \boldsymbol{\beta}) + \alpha \text{tr}(\boldsymbol{\Phi}'_K \boldsymbol{\Omega} \boldsymbol{\Phi}_K), \quad (2.5)$$

subject to  $\phi_k' (\mathbf{V}_\xi(\boldsymbol{\theta}_\xi) + \mathbf{I}_n)^{-1} \phi_k = 1$ ,  $k = 1, \dots, K$ . We estimate both  $\boldsymbol{\beta}$  and  $\boldsymbol{\theta}$  simultaneously by minimizing (2.5) using an ECM algorithm (Meng and Rubin (1993)) with details given in Section 3. The smoothness parameter  $\alpha$  can be selected using cross-validation (CV, see examples in Sections 4 and 5).

## 2.2. Spatial prediction

When both  $\boldsymbol{\beta}$  and  $\boldsymbol{\theta}$  are known, the minimum mean squared error predictor of  $Y_t(\mathbf{s})$  is

$$\widehat{Y}_t(\mathbf{s}) \equiv E(Y_t(\mathbf{s}) | \mathbf{Z}_t) = \mathbf{x}_t(\mathbf{s})' \boldsymbol{\beta} + \boldsymbol{\sigma}_Y(\mathbf{s})' \boldsymbol{\Sigma}_Z^{-1}(\boldsymbol{\theta}) (\mathbf{Z}_t - \mathbf{X}_t \boldsymbol{\beta}), \quad \mathbf{s} \in D, \quad (2.6)$$

where

$$\boldsymbol{\sigma}_Y(\mathbf{s}) \equiv \text{cov}(\mathbf{Y}_t, Y_t(\mathbf{s})) = \sum_{k=1}^K \lambda_k \phi_k(\mathbf{s}) \phi_k + \sigma_\varepsilon^2 \mathbf{v}_\xi(\mathbf{s})$$

and

$$\mathbf{v}_\xi(\mathbf{s}) \equiv \frac{1}{\sigma_\varepsilon^2} \text{cov}(\boldsymbol{\xi}_t, \xi_t(\mathbf{s})).$$

The corresponding mean squared prediction error (MSPE) of  $\widehat{Y}_t(\mathbf{s})$  is

$$\text{MSPE}(\mathbf{s}) \equiv E(\widehat{Y}_t(\mathbf{s}) - Y_t(\mathbf{s}))^2 = C_Y(\mathbf{s}, \mathbf{s}) - \boldsymbol{\sigma}_Y(\mathbf{s})' \boldsymbol{\Sigma}_Z^{-1}(\boldsymbol{\theta}) \boldsymbol{\sigma}_Y(\mathbf{s}), \quad (2.7)$$

where

$$C_Y(\mathbf{s}, \mathbf{s}') \equiv \sum_{k=1}^K \lambda_k \phi_k(\mathbf{s}) \phi_k(\mathbf{s}') + C_\xi(\mathbf{s}, \mathbf{s}')$$

and

$$C_\xi(\mathbf{s}, \mathbf{s}') \equiv \text{cov}(\xi_t(\mathbf{s}), \xi_t(\mathbf{s}')).$$

Alternatively, for  $\mathbf{s} \in D$ ,  $\widehat{Y}_t(\mathbf{s})$  and  $E(\widehat{Y}_t(\mathbf{s}) - Y_t(\mathbf{s}))^2$  can be written as:

$$\begin{aligned} \widehat{Y}_t(\mathbf{s}) &= E(E(Y_t(\mathbf{s})|\mathbf{Z}_t, \mathbf{w}_t)|\mathbf{Z}_t) \\ &= \mathbf{x}_t(\mathbf{s})'\boldsymbol{\beta} + \boldsymbol{\phi}(\mathbf{s})'E(\mathbf{w}_t|\mathbf{Z}_t) \\ &\quad + \mathbf{v}_\xi(\mathbf{s})'(\mathbf{V}_\xi + \mathbf{I}_n)^{-1}(\mathbf{Z}_t - \mathbf{X}_t\boldsymbol{\beta} - \boldsymbol{\Phi}_K E(\mathbf{w}_t|\mathbf{Z}_t)), \end{aligned} \tag{2.8}$$

$$\begin{aligned} \text{MSPE}(\mathbf{s}) &= \text{var}(Y_t(\mathbf{s})|\mathbf{Z}_t) \\ &= E(\text{var}(Y_t(\mathbf{s})|\mathbf{Z}_t, \mathbf{w}_t)|\mathbf{Z}_t) + \text{var}(E(Y_t(\mathbf{s})|\mathbf{Z}_t, \mathbf{w}_t)|\mathbf{Z}_t) \\ &= E(\text{var}(\xi_t(\mathbf{s})|\mathbf{Z}_t, \mathbf{w}_t)|\mathbf{Z}_t) \\ &\quad + \text{var}(\boldsymbol{\phi}(\mathbf{s})'\mathbf{w}_t + \mathbf{v}_\xi(\mathbf{s})'(\mathbf{V}_\xi + \mathbf{I}_n)^{-1}(\mathbf{Z}_t - \mathbf{X}_t\boldsymbol{\beta} - \boldsymbol{\Phi}_K\mathbf{w}_t)|\mathbf{Z}_t) \\ &= E(\text{var}(\xi_t(\mathbf{s})|\mathbf{Z}_t, \mathbf{w}_t)|\mathbf{Z}_t) \\ &\quad + \text{var}((\boldsymbol{\phi}(\mathbf{s})' - \mathbf{v}_\xi(\mathbf{s})'(\mathbf{V}_\xi + \mathbf{I}_n)^{-1}\boldsymbol{\Phi}_K)\mathbf{w}_t|\mathbf{Z}_t) \\ &= \text{var}(\xi_t(\mathbf{s})) - \sigma_\varepsilon^2\mathbf{v}_\xi(\mathbf{s})'(\mathbf{V}_\xi + \mathbf{I}_n)^{-1}\mathbf{v}_\xi(\mathbf{s}) \\ &\quad + (\boldsymbol{\phi}(\mathbf{s}) - \boldsymbol{\Phi}'_K(\mathbf{V}_\xi + \mathbf{I}_n)^{-1}\mathbf{v}_\xi(\mathbf{s}))' \\ &\quad \times \text{var}(\mathbf{w}_t|\mathbf{Z}_t)(\boldsymbol{\phi}(\mathbf{s}) - \boldsymbol{\Phi}'_K(\mathbf{V}_\xi + \mathbf{I}_n)^{-1}\mathbf{v}_\xi(\mathbf{s})), \end{aligned} \tag{2.9}$$

where  $\boldsymbol{\phi}(\mathbf{s}) \equiv (\phi_1(\mathbf{s}), \dots, \phi_K(\mathbf{s}))'$  and, for simplicity,  $\mathbf{V}_\xi(\boldsymbol{\theta}_\xi)$  is written as  $\mathbf{V}_\xi$ . Equations (2.8) and (2.9) allow us to compute  $\widehat{Y}_t(\mathbf{s})$  more efficiently than those given by (2.6) and (2.7), because  $E(\mathbf{w}_t|\mathbf{Z}_t)$ ,  $\text{var}(\mathbf{w}_t|\mathbf{Z}_t)$  and  $(\mathbf{V}_\xi + \mathbf{I}_n)^{-1}$  are byproducts of the proposed ECM algorithm (see Section 3).

When  $\boldsymbol{\beta}$  is unknown, we can plug-in the generalized least squares estimate

$$\widehat{\boldsymbol{\beta}}_{gls} = \left( \sum_{t=1}^T \mathbf{X}'_t \boldsymbol{\Sigma}_Z^{-1}(\boldsymbol{\theta}) \mathbf{X}_t \right)^{-1} \left( \sum_{t=1}^T \mathbf{X}'_t \boldsymbol{\Sigma}_Z^{-1}(\boldsymbol{\theta}) \mathbf{Z}_t \right),$$

for  $\boldsymbol{\beta}$  in (2.8), resulting in the universal kriging predictor (see Cressie (1993)) of  $Y_t(\mathbf{s})$  given by

$$\begin{aligned} \widehat{Y}_t(\mathbf{s}) &= \mathbf{x}_t(\mathbf{s})'\widehat{\boldsymbol{\beta}}_{gls} + \boldsymbol{\phi}(\mathbf{s})'E(\mathbf{w}_t|\mathbf{Z}_t) \\ &\quad + \mathbf{v}_\xi(\mathbf{s})'(\mathbf{V}_\xi + \mathbf{I}_n)^{-1}(\mathbf{Z}_t - \mathbf{X}_t\widehat{\boldsymbol{\beta}}_{gls} - \boldsymbol{\Phi}_K E(\mathbf{w}_t|\mathbf{Z}_t)). \end{aligned} \tag{2.10}$$

The corresponding MSPE can be written as

$$\begin{aligned} \text{MSPE}_{uk}(\mathbf{s}) &= \text{MSPE}(\mathbf{s}) + (\mathbf{x}_t(\mathbf{s})' - \boldsymbol{\sigma}_Y(\mathbf{s})'\boldsymbol{\Sigma}_Z^{-1}(\boldsymbol{\theta})\mathbf{X}_t) \left( \sum_{t=1}^T \mathbf{X}'_t \boldsymbol{\Sigma}_Z^{-1}(\boldsymbol{\theta}) \mathbf{X}_t \right)^{-1} \\ &\quad \times (\mathbf{x}_t(\mathbf{s})' - \boldsymbol{\sigma}_Y(\mathbf{s})'\boldsymbol{\Sigma}_Z^{-1}(\boldsymbol{\theta})\mathbf{X}_t)'. \end{aligned} \tag{2.11}$$

### 3. Multicycle ECM Algorithm

The penalized log-likelihood function of (2.5) is difficult to minimize directly. We instead apply the ECM algorithm (Meng and Rubin (1993)) by treating  $\mathbf{w}_t$ 's as missing data. The full negative penalized log-likelihood of  $\mathbf{Z}_1, \dots, \mathbf{Z}_T$  and  $\mathbf{w}_1, \dots, \mathbf{w}_T$ , up to an additive constant, can be written as

$$\begin{aligned} \ell_{Z,w}(\boldsymbol{\beta}, \boldsymbol{\theta}) &= Tn \log \sigma_\varepsilon^2 + T \left( \log |\mathbf{V}_\xi(\boldsymbol{\theta}_\xi) + \mathbf{I}_n| + \log |\boldsymbol{\Lambda}_K| \right) \\ &\quad + \frac{1}{\sigma_\varepsilon^2} \sum_{t=1}^T (\mathbf{Z}_t - \mathbf{X}_t \boldsymbol{\beta} - \boldsymbol{\Phi}_K \mathbf{w}_t)' (\mathbf{V}_\xi(\boldsymbol{\theta}_\xi) + \mathbf{I}_n)^{-1} (\mathbf{Z}_t - \mathbf{X}_t \boldsymbol{\beta} - \boldsymbol{\Phi}_K \mathbf{w}_t) \\ &\quad + \sum_{t=1}^T \mathbf{w}_t' \boldsymbol{\Lambda}_K^{-1} \mathbf{w}_t + \alpha \operatorname{tr}(\boldsymbol{\Phi}'_K \boldsymbol{\Omega} \boldsymbol{\Phi}_K), \end{aligned} \quad (3.1)$$

where  $\boldsymbol{\Lambda}_K$  is a diagonal matrix with diagonal elements  $\lambda_1, \dots, \lambda_K$ . Basically, the ECM algorithm approaches the problem of minimizing (2.5) by iteratively performing an expectation (E) step that computes the expectation of  $\ell_{Z,w}(\boldsymbol{\beta}, \boldsymbol{\theta})$  conditioned on  $\mathbf{Z}_1, \dots, \mathbf{Z}_T$  based on the current fits of  $\boldsymbol{\beta}$  and  $\boldsymbol{\theta}$ , followed by a sequence of conditional maximization (CM) steps, which successively maximize the conditional penalized log-likelihood obtained in the E-step with respect to each component of  $\boldsymbol{\beta}$  and  $\boldsymbol{\theta}$ .

In the E-step, we calculate  $\hat{\mathbf{w}}_t \equiv \mathbb{E}(\mathbf{w}_t | \mathbf{Z}_t)$ ,  $t = 1, \dots, T$ , and  $\boldsymbol{\Sigma}_{w|Z} \equiv \operatorname{var}(\mathbf{w}_t | \mathbf{Z}_t)$  based on the current fits,  $\tilde{\boldsymbol{\beta}}$  and  $\tilde{\boldsymbol{\theta}}$  of  $\boldsymbol{\beta}$  and  $\boldsymbol{\theta}$  as

$$\hat{\mathbf{w}}_t = (\tilde{\sigma}_\varepsilon^2 \tilde{\boldsymbol{\Lambda}}_K^{-1} + \tilde{\boldsymbol{\Phi}}'_K (\mathbf{V}_\xi(\tilde{\boldsymbol{\theta}}_\xi) + \mathbf{I}_n)^{-1} \tilde{\boldsymbol{\Phi}}_K)^{-1} \tilde{\boldsymbol{\Phi}}'_K (\mathbf{V}_\xi(\tilde{\boldsymbol{\theta}}_\xi) + \mathbf{I}_n)^{-1} (\mathbf{Z}_t - \mathbf{X}_t \tilde{\boldsymbol{\beta}}), \quad (3.2)$$

$$\boldsymbol{\Sigma}_{w|Z} = \left( \tilde{\boldsymbol{\Lambda}}_K^{-1} + \frac{1}{\tilde{\sigma}_\varepsilon^2} \tilde{\boldsymbol{\Phi}}'_K (\mathbf{V}_\xi(\tilde{\boldsymbol{\theta}}_\xi) + \mathbf{I}_n)^{-1} \tilde{\boldsymbol{\Phi}}_K \right)^{-1}, \quad (3.3)$$

from which we obtain

$$\begin{aligned} \mathbb{E}(\ell_{Z,w}(\boldsymbol{\beta}, \boldsymbol{\theta}) | \mathbf{Z}_1, \dots, \mathbf{Z}_T) &= Tn \log \sigma_\varepsilon^2 + T \left( \log |\mathbf{V}_\xi(\boldsymbol{\theta}_\xi) + \mathbf{I}_n| + \log |\boldsymbol{\Lambda}_K| \right) \\ &\quad + \frac{1}{\sigma_\varepsilon^2} \sum_{t=1}^T (\mathbf{Z}_t - \mathbf{X}_t \boldsymbol{\beta} - \boldsymbol{\Phi}_K \hat{\mathbf{w}}_t)' (\mathbf{V}_\xi(\boldsymbol{\theta}_\xi) + \mathbf{I}_n)^{-1} (\mathbf{Z}_t - \mathbf{X}_t \boldsymbol{\beta} - \boldsymbol{\Phi}_K \hat{\mathbf{w}}_t) \\ &\quad + \frac{T}{\sigma_\varepsilon^2} \operatorname{tr}((\mathbf{V}_\xi(\boldsymbol{\theta}_\xi) + \mathbf{I}_n)^{-1/2} \boldsymbol{\Phi}_K \boldsymbol{\Sigma}_{w|Z} \boldsymbol{\Phi}'_K (\mathbf{V}_\xi(\boldsymbol{\theta}_\xi) + \mathbf{I}_n)^{-1/2}) \\ &\quad + \sum_{t=1}^T \hat{\mathbf{w}}_t' \boldsymbol{\Lambda}_K^{-1} \hat{\mathbf{w}}_t + T \operatorname{tr}(\boldsymbol{\Lambda}_K^{-1} \boldsymbol{\Sigma}_{w|Z}) + \alpha \operatorname{tr}(\boldsymbol{\Phi}'_K \boldsymbol{\Omega} \boldsymbol{\Phi}_K). \end{aligned} \quad (3.4)$$

We then successively update each component of  $\boldsymbol{\beta}$  and  $\boldsymbol{\theta}$  by minimizing (3.4) with the other components held fixed in a sequence of CM steps.

Let  $\mathbf{R} \equiv ((\mathbf{Z}_1 - \mathbf{X}_1\tilde{\boldsymbol{\beta}}), \dots, (\mathbf{Z}_T - \mathbf{X}_T\tilde{\boldsymbol{\beta}}))$  and  $\widehat{\mathbf{W}} \equiv (\widehat{\mathbf{w}}_1, \dots, \widehat{\mathbf{w}}_T)'$ . First,  $\Phi_K$  is updated by minimizing

$$\sum_{t=1}^T \|(\mathbf{V}_\xi(\tilde{\boldsymbol{\theta}}_\xi) + \mathbf{I}_n)^{-1/2}(\mathbf{Z}_t - \mathbf{X}_t\tilde{\boldsymbol{\beta}} - \Phi_K\widehat{\mathbf{w}}_t)\|^2 + T \operatorname{tr}((\mathbf{V}_\xi(\tilde{\boldsymbol{\theta}}_\xi) + \mathbf{I}_n)^{-1/2}\Phi_K\Sigma_{w|Z}\Phi_K'(\mathbf{V}_\xi(\tilde{\boldsymbol{\theta}}_\xi) + \mathbf{I}_n)^{-1/2}) + \alpha\tilde{\sigma}_\varepsilon^2 \operatorname{tr}(\Phi_K'\Omega\Phi_K),$$

which is equivalent to minimizing

$$-2\operatorname{vec}(\mathbf{R})'(\widehat{\mathbf{W}} \otimes (\mathbf{V}_\xi(\tilde{\boldsymbol{\theta}}_\xi) + \mathbf{I}_n)^{-1})\operatorname{vec}(\Phi_K) + \operatorname{vec}(\Phi_K)'((\widehat{\mathbf{W}}'\widehat{\mathbf{W}} + T\Sigma_{w|Z}) \otimes (\mathbf{V}_\xi(\tilde{\boldsymbol{\theta}}_\xi) + \mathbf{I}_n)^{-1} + \alpha\tilde{\sigma}_\varepsilon^2\mathbf{I}_K \otimes \Omega)\operatorname{vec}(\Phi_K),$$

resulting in

$$\operatorname{vec}(\tilde{\Phi}_K) = \left( (\widehat{\mathbf{W}}'\widehat{\mathbf{W}} + T\Sigma_{w|Z}) \otimes (\mathbf{V}_\xi(\tilde{\boldsymbol{\theta}}_\xi) + \mathbf{I}_n)^{-1} + \alpha\tilde{\sigma}_\varepsilon^2\mathbf{I}_K \otimes \Omega \right)^{-1} \times \left( \widehat{\mathbf{W}} \otimes (\mathbf{V}_\xi(\tilde{\boldsymbol{\theta}}_\xi) + \mathbf{I}_n)^{-1} \right)' \operatorname{vec}(\mathbf{R}), \tag{3.5}$$

where  $\tilde{\Phi}_K = (\tilde{\phi}_1, \dots, \tilde{\phi}_K)$ . We then rescale  $\tilde{\phi}_k$  for  $k = 1, \dots, K$ , by dividing it by  $(\tilde{\phi}_k'(\mathbf{V}_\xi(\tilde{\boldsymbol{\theta}}_\xi) + \mathbf{I}_n)^{-1}\tilde{\phi}_k)^{1/2}$  so that  $\tilde{\phi}_k'(\mathbf{V}_\xi(\tilde{\boldsymbol{\theta}}_\xi) + \mathbf{I}_n)^{-1}\tilde{\phi}_k = 1$ . Notice that we update  $\tilde{\Phi}_K$  in this CM step only if the value of (3.4) decreases after rescaling. The CM updating formulae for  $\boldsymbol{\beta}$ ,  $\sigma_\varepsilon^2$ , and the  $\lambda_k$ 's have simple closed forms given by

$$\tilde{\boldsymbol{\beta}} = \left( \sum_{t=1}^T \mathbf{X}_t'(\mathbf{V}_\xi(\tilde{\boldsymbol{\theta}}_\xi) + \mathbf{I}_n)^{-1}\mathbf{X}_t \right)^{-1} \left( \sum_{t=1}^T \mathbf{X}_t'(\mathbf{V}_\xi(\tilde{\boldsymbol{\theta}}_\xi) + \mathbf{I}_n)^{-1}(\mathbf{Z}_t - \tilde{\Phi}_K\widehat{\mathbf{w}}_t) \right), \tag{3.6}$$

$$\tilde{\sigma}_\varepsilon^2 = \frac{1}{Tn} \sum_{t=1}^T (\mathbf{Z}_t - \mathbf{X}_t\tilde{\boldsymbol{\beta}} - \tilde{\Phi}_K\widehat{\mathbf{w}}_t)'(\mathbf{V}_\xi(\tilde{\boldsymbol{\theta}}_\xi) + \mathbf{I}_n)^{-1}(\mathbf{Z}_t - \mathbf{X}_t\tilde{\boldsymbol{\beta}} - \tilde{\Phi}_K\widehat{\mathbf{w}}_t) + \frac{1}{n} \operatorname{tr}((\mathbf{V}_\xi(\tilde{\boldsymbol{\theta}}_\xi) + \mathbf{I}_n)^{-1/2}\tilde{\Phi}_K\Sigma_{w|Z}\tilde{\Phi}_K'(\mathbf{V}_\xi(\tilde{\boldsymbol{\theta}}_\xi) + \mathbf{I}_n)^{-1/2}), \tag{3.7}$$

$$\tilde{\lambda}_k = \frac{1}{T} \sum_{t=1}^T \widehat{w}_{t,k}^2 + \sigma_{w|Z,k}^2, \quad k = 1, \dots, K, \tag{3.8}$$

where  $\widehat{w}_{t,k}$  is the  $k$ th entry of  $\widehat{\mathbf{w}}_t$ , and  $\sigma_{w|Z,k}^2$  is the  $k$ th diagonal element of  $\Sigma_{w|Z}$ . Finally, the updating formula for  $\boldsymbol{\theta}_\xi$  in the CM step is

$$\tilde{\boldsymbol{\theta}}_\xi = \arg \min_{\boldsymbol{\theta}_\xi} \left\{ T \log |\mathbf{V}_\xi(\boldsymbol{\theta}_\xi) + \mathbf{I}_n| + \frac{1}{\tilde{\sigma}_\varepsilon^2} \sum_{t=1}^T (\mathbf{Z}_t - \mathbf{X}_t\tilde{\boldsymbol{\beta}} - \tilde{\Phi}_K\widehat{\mathbf{w}}_t)'(\mathbf{V}_\xi(\boldsymbol{\theta}_\xi) + \mathbf{I}_n)^{-1}(\mathbf{Z}_t - \mathbf{X}_t\tilde{\boldsymbol{\beta}} - \tilde{\Phi}_K\widehat{\mathbf{w}}_t) \right\}$$



$$+ \frac{T}{\tilde{\sigma}_\varepsilon^2} \text{tr} \left\{ \left( \mathbf{V}_\xi(\boldsymbol{\theta}_\xi) + \mathbf{I}_n \right)^{-1/2} \tilde{\boldsymbol{\Phi}}_K \boldsymbol{\Sigma}_{w|Z} \tilde{\boldsymbol{\Phi}}_K' \left( \mathbf{V}_\xi(\boldsymbol{\theta}_\xi) + \mathbf{I}_n \right)^{-1/2} \right\}. \quad (3.9)$$

Unfortunately,  $\tilde{\boldsymbol{\theta}}_\xi$  in (3.9) does not have a closed-form expression, and hence has to be computed using some optimization routine. Therefore, it appears computationally more efficient to apply a multicycle ECM algorithm by updating  $\boldsymbol{\theta}_\xi$  only after all the other parameters converge. Detailed steps of our multicycle ECM algorithm are given below.

1. Start with some initial estimates of  $\boldsymbol{\beta}$  and  $\boldsymbol{\theta}$ .
2. E-step: Compute  $\hat{\boldsymbol{w}}_t$ ,  $t = 1, \dots, T$ , and  $\boldsymbol{\Sigma}_{w|Z}$  based on the current values of  $\boldsymbol{\beta}$  and  $\boldsymbol{\theta}$ .
3. Inner CM-step: Update  $\boldsymbol{\Phi}_K$ ,  $\boldsymbol{\beta}$ ,  $\sigma_\varepsilon^2$ , and  $\lambda_k$ ,  $k = 1, \dots, K$  according to (3.5)-(3.8).
4. Repeat Steps 2 and 3 until convergence.
5. Outer CM-step: Update  $\boldsymbol{\theta}_\xi$  according to (3.9).
6. Repeat Steps 2-5 until convergence.

Throughout the paper, the initial estimates of  $\boldsymbol{\beta}$ ,  $\boldsymbol{\theta}_\xi$ ,  $\lambda_1, \dots, \lambda_{K'-1}$ , and  $\sigma_\varepsilon^2$  for  $K = K' \geq 1$  are given by the corresponding PML estimates for  $K = K' - 1$ . The initial estimate of  $\lambda_{K'}$  is set to be  $\lambda_{K'}/2$ , and the initial estimate of  $\boldsymbol{\Phi}_K$  is given by  $(\mathbf{V}_\xi(\tilde{\boldsymbol{\theta}}_\xi) + \mathbf{I}_n)^{1/2} \mathbf{M}$ , where  $\mathbf{M}$  is an  $n \times K$  matrix formed by the first  $K$  left singular vectors of  $(\mathbf{V}_\xi(\tilde{\boldsymbol{\theta}}_\xi) + \mathbf{I}_n)^{-1/2} \mathbf{R}$  and  $\tilde{\boldsymbol{\theta}}_\xi$  denotes the initial estimate of  $\boldsymbol{\theta}_\xi$ . Note that the PML estimates of  $\boldsymbol{\beta}$  and  $\boldsymbol{\theta}$  for  $K = 0$  are nothing but the maximum likelihood (ML) estimates  $\hat{\boldsymbol{\beta}}_0$  and  $\hat{\boldsymbol{\theta}}_0$  for the stationary model given by (2.3) without the  $\boldsymbol{\Phi}_K \boldsymbol{w}_t$  term, where

$$\hat{\boldsymbol{\beta}}_0 = \left( \sum_{t=1}^T \mathbf{X}_t' (\mathbf{V}_\xi(\hat{\boldsymbol{\theta}}_{\xi,0}) + \mathbf{I}_n)^{-1} \mathbf{X}_t \right)^{-1} \left( \sum_{t=1}^T \mathbf{X}_t' (\mathbf{V}_\xi(\hat{\boldsymbol{\theta}}_{\xi,0}) + \mathbf{I}_n)^{-1} \mathbf{Z}_t \right),$$

$$\hat{\sigma}_{\varepsilon,0}^2 = \frac{1}{Tn} \sum_{t=1}^T (\mathbf{Z}_t - \mathbf{X}_t \hat{\boldsymbol{\beta}}_0)' (\mathbf{V}_\xi(\hat{\boldsymbol{\theta}}_{\xi,0}) + \mathbf{I}_n)^{-1} (\mathbf{Z}_t - \mathbf{X}_t \hat{\boldsymbol{\beta}}_0),$$

and  $\hat{\boldsymbol{\theta}}_{\xi,0}$  can be obtained by maximizing the profile likelihood of  $\boldsymbol{\theta}_\xi$  using some optimization algorithm. According to Meng and Rubin (1993), the cost function value of our multicycle ECM algorithm is guaranteed to decrease at each iteration. In addition, they also show that a multicycle ECM algorithm shares the same theoretical properties as the corresponding ECM algorithm. Some examples regarding the rates of convergence of EM, ECM, and multicycle ECM can be found in Meng (1994).

With  $\widehat{\sigma}_\varepsilon^2$ ,  $\widehat{\boldsymbol{\theta}}_\xi$ ,  $\widehat{\boldsymbol{\Lambda}}_K$ , and  $\widehat{\boldsymbol{\Phi}}_K$  estimated by the proposed PML method, the estimated spatial covariance function of  $C_Y(\cdot, \cdot)$  can be written as

$$\widehat{C}_Y(\mathbf{s}, \mathbf{s}') \equiv \sum_{k=1}^K \widehat{\lambda}_k \widehat{\phi}_k(\mathbf{s}) \widehat{\phi}_k(\mathbf{s}') + \widehat{C}_\xi(\mathbf{s}, \mathbf{s}'), \quad \mathbf{s}, \mathbf{s}' \in D,$$

where  $\widehat{C}_\xi(\mathbf{s}, \mathbf{s}')$  is the estimate of  $C_\xi(\mathbf{s}, \mathbf{s}')$  by plugging  $\widehat{\boldsymbol{\theta}}_\xi$  in for  $\boldsymbol{\theta}_\xi$  in  $C_\xi(\mathbf{s}, \mathbf{s}')$ . When  $\boldsymbol{\beta}$  is either known or zero, an estimate of  $E(\widehat{Y}_t(\mathbf{s}) - Y_t(\mathbf{s}))^2$  can be obtained by plugging  $\widehat{\boldsymbol{\theta}}$  in for  $\boldsymbol{\theta}$  in (2.9)

$$\begin{aligned} \widehat{\text{MSPE}}(\mathbf{s}) &= \widehat{C}_\xi(\mathbf{s}, \mathbf{s}) - \widehat{\sigma}_\varepsilon^2 \widehat{\mathbf{v}}_\xi(\mathbf{s})' (\widehat{\mathbf{V}}_\xi + \mathbf{I}_n)^{-1} \widehat{\mathbf{v}}_\xi(\mathbf{s}) \\ &\quad + (\widehat{\boldsymbol{\phi}}(\mathbf{s}) - \widehat{\boldsymbol{\Phi}}_K' (\widehat{\mathbf{V}}_\xi + \mathbf{I}_n)^{-1} \widehat{\mathbf{v}}_\xi(\mathbf{s}))' \\ &\quad \times \widehat{\boldsymbol{\Sigma}}_{w|Z} (\widehat{\boldsymbol{\phi}}(\mathbf{s}) - \widehat{\boldsymbol{\Phi}}_K' (\widehat{\mathbf{V}}_\xi + \mathbf{I}_n)^{-1} \widehat{\mathbf{v}}_\xi(\mathbf{s})), \end{aligned} \tag{3.10}$$

where

$$\begin{aligned} \widehat{\mathbf{v}}_\xi(\mathbf{s}) &= \frac{1}{\widehat{\sigma}_\varepsilon^2} (\widehat{C}_\xi(\mathbf{s}_1, \mathbf{s}), \dots, \widehat{C}_\xi(\mathbf{s}_n, \mathbf{s}))', \\ \widehat{\mathbf{V}}_\xi &= (\widehat{\mathbf{v}}_\xi(\mathbf{s}_1), \dots, \widehat{\mathbf{v}}_\xi(\mathbf{s}_n)), \\ \widehat{\boldsymbol{\Sigma}}_{w|Z} &= \left( \widehat{\boldsymbol{\Lambda}}_K^{-1} + \frac{1}{\widehat{\sigma}_\varepsilon^2} \widehat{\boldsymbol{\Phi}}_K' (\widehat{\mathbf{V}}_\xi + \mathbf{I}_n)^{-1} \widehat{\boldsymbol{\Phi}}_K \right)^{-1}. \end{aligned}$$

Similarly from (2.11), an estimate of  $E(\widehat{Y}_t(\mathbf{s}) - Y_t(\mathbf{s}))^2$  for the universal kriging predictor is

$$\begin{aligned} \widehat{\text{MSPE}}_{uk}(\mathbf{s}) &= \widehat{\text{MSPE}}(\mathbf{s}) + (\mathbf{x}_t(\mathbf{s})' - \widehat{\boldsymbol{\sigma}}_Y(\mathbf{s})' \boldsymbol{\Sigma}_Z^{-1}(\widehat{\boldsymbol{\theta}}) \mathbf{X}_t) \left( \sum_{t=1}^T \mathbf{X}_t' \boldsymbol{\Sigma}_Z^{-1}(\widehat{\boldsymbol{\theta}}) \mathbf{X}_t \right)^{-1} \\ &\quad \times (\mathbf{x}_t(\mathbf{s})' - \widehat{\boldsymbol{\sigma}}_Y(\mathbf{s})' \boldsymbol{\Sigma}_Z^{-1}(\widehat{\boldsymbol{\theta}}) \mathbf{X}_t)'. \end{aligned}$$

#### 4. Simulation Studies

To assess the performance of the proposed model, we conducted four simulation experiments in one- and two-dimensional spatial domains. Here we focus on covariance function estimation, and hence we consider no regressor with  $\mathbf{X}_t(\mathbf{s}) = \mathbf{0}$ ,  $t = 1, \dots, T$ ,  $\mathbf{s} \in D$  for the first three experiments.

The prediction performance of various methods is compared using the averaged MSPE criterion

$$\text{AMSPE} \equiv \frac{1}{T|D|} \sum_{t=1}^T \int_{\mathbf{s} \in D} E(\widehat{Y}_t^*(\mathbf{s}) - Y_t(\mathbf{s}))^2 d\mathbf{s},$$

where  $|D| = \int_{\mathbf{s} \in D} d\mathbf{s}$ , and  $\widehat{Y}_t^*(\mathbf{s})$  denotes a generic predictor of  $Y_t(\mathbf{s})$ , for  $t = 1, \dots, T$ ,  $\mathbf{s} \in D$ . Similar to the plug-in estimate  $\widehat{\text{MSPE}}(\mathbf{s})$  of  $E(\widehat{Y}_t(\mathbf{s}) - Y_t(\mathbf{s}))^2$

in (3.10), the plug-in estimate of the averaged MSPE is

$$\widehat{\text{AMSPE}} \equiv \frac{1}{T|D|} \sum_{t=1}^T \int_{\mathbf{s} \in D} \widehat{\text{MSPE}}^*(\mathbf{s}) d\mathbf{s},$$

where  $\widehat{\text{MSPE}}^*(\mathbf{s})$  denotes a generic plug-in estimate of  $E(\widehat{Y}_t^*(\mathbf{s}) - Y_t(\mathbf{s}))^2$ . In addition, the performance of various methods in covariance function estimation can be compared using the averaged mean squared error criterion

$$\text{AMSE} \equiv \frac{1}{|D|^2} \int \int_{\mathbf{s}, \mathbf{s}' \in D} E(\widehat{C}_Y^*(\mathbf{s}, \mathbf{s}') - C_Y(\mathbf{s}, \mathbf{s}'))^2 d\mathbf{s} d\mathbf{s}',$$

where  $\widehat{C}_Y^*(\mathbf{s}, \mathbf{s}')$  denotes a generic estimate of  $C_Y(\mathbf{s}, \mathbf{s}')$  for  $\mathbf{s}, \mathbf{s}' \in D$ .

### 4.1. Experiment I

In the first experiment, we consider an example brought up by Nott and Dunsmuir (2002) in a one-dimensional spatial domain with  $D = [-4, 4]$ . The spatial process  $Y_t(\cdot)$  was generated according to  $Y_t(\mathbf{s}) = \sqrt{\Psi(\mathbf{s})}Y_{1,t}(\mathbf{s}) + \sqrt{1 - \Psi(\mathbf{s})}Y_{2,t}(\mathbf{s})$ , where  $\Psi(\cdot)$  is the cumulative distribution function of the standard normal distribution, and  $Y_{j,t}(\cdot)$  is a zero-mean Gaussian process with the exponential covariance function

$$C_{Y_j}(\mathbf{s}, \mathbf{s}') = \exp(-\rho_j \|\mathbf{s} - \mathbf{s}'\|), \quad j = 1, 2. \tag{4.1}$$

The data were generated on  $D = [-4, 4]$  according to (2.1) with  $n = 20$  and  $\sigma_\varepsilon^2 = 0.04$ , where  $\mathbf{s}_1 = -4$ ,  $\mathbf{s}_{20} = 4$ , and  $\mathbf{s}_2, \dots, \mathbf{s}_{n-1}$  were sampled on  $D$  using simple random sampling, and then rearranged so that  $\mathbf{s}_1 < \dots < \mathbf{s}_n$ . The resulting spatial covariance function of  $Y_t(\cdot)$  is given by  $C_Y(\mathbf{s}, \mathbf{s}') = \sqrt{\Psi(\mathbf{s})}\sqrt{\Psi(\mathbf{s}')}C_{Y_1}(\mathbf{s}, \mathbf{s}') + \sqrt{1 - \Psi(\mathbf{s})}\sqrt{1 - \Psi(\mathbf{s}')}C_{Y_2}(\mathbf{s}, \mathbf{s}')$ .

We considered time points  $T = 50, 100$  in combination with  $(\rho_1, \rho_2) = (2, 1/20), (2, 1/2), (2, 2)$ , corresponding to high nonstationarity, mild nonstationarity, and near stationarity, respectively, for  $Y_{j,t}(\cdot)$ 's in (4.1). The corresponding spatial covariance functions are shown in Figure 1.

We applied the proposed PML method and considered the exponential covariance family for  $\xi_t(\cdot)$  in (2.2) parametrized by  $\boldsymbol{\theta}_\xi = (\tau_\xi, \rho_\xi)'$ :

$$\frac{1}{\sigma_\varepsilon^2} \text{cov}(\xi_t(\mathbf{s}), \xi_t(\mathbf{s}')) = \tau_\xi \exp(-\rho_\xi \|\mathbf{s} - \mathbf{s}'\|). \tag{4.2}$$

We chose  $K \in \{0, 1, \dots, 3\}$  and  $\alpha \in \mathcal{A}$  in (2.5) via 6-fold CV, where  $\mathcal{A} = \{2^{-8}, 2^{-7}, \dots, 2^8\}$ . Specifically, we randomly partitioned  $\mathbf{s}_2, \dots, \mathbf{s}_{19}$  into six disjoint subsets,  $\mathcal{D}_1, \dots, \mathcal{D}_6$ , each with size 3. For each  $\mathcal{D}_j$ , we estimated the parameters using the proposed PML based on data  $\{Z_t(\mathbf{s}) : \mathbf{s} \in \{\mathbf{s}_1, \dots, \mathbf{s}_{20}\} \setminus \mathcal{D}_j, t =$

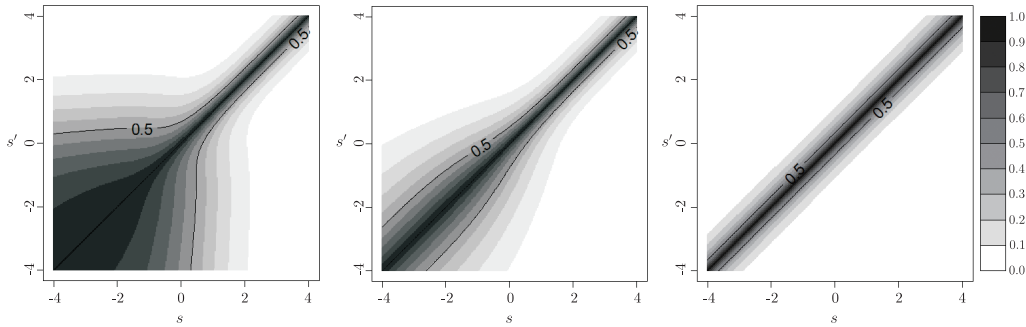


Figure 1. True spatial covariance functions of  $C_Y(\cdot, \cdot)$  in Experiment I with three different pairs of parameters: (a)  $(\rho_1, \rho_2) = (2, 1/20)$ ; (b)  $(\rho_1, \rho_2) = (2, 1/2)$ ; (c)  $(\rho_1, \rho_2) = (2, 2)$ .

$1, \dots, T$ ] at the remaining locations, and predicted  $Y_t(\mathbf{s})$  by  $\widehat{Y}_t^{(K, \alpha)}(\mathbf{s})$ , for  $\mathbf{s} \in \mathcal{D}_j$ ,  $K \in \{0, 1, \dots, 3\}$ , and  $\alpha \in \mathcal{A}$ . Notice that  $\mathbf{s}_1$  and  $\mathbf{s}_n$  were always included as data locations in each fitting sample of CV to avoid the boundary effect. We then chose  $K$  and  $\alpha$  in (2.5) simultaneously such that

$$\text{CV}(K, \alpha) = \sum_{t=1}^T \sum_{j=1}^{\nu} \sum_{\mathbf{s} \in \mathcal{D}_j} \|\widehat{Y}_t^{(K, \alpha)}(\mathbf{s}) - Z_t(\mathbf{s})\|^2, \quad (4.3)$$

was minimized among  $K \in \{0, 1, 2, 3\}$  and  $\alpha \in \mathcal{A}$ , where  $\nu = 6$ . Note that for  $K = 0$ , the model reduces to a stationary model independent of  $\alpha$ , where the parameters are estimated by ML and  $Y_t(\mathbf{s})$  is predicted using simple kriging.

The proposed method is compared with the following methods.

1. **Method I:** A reduced model similar to (2.1) and (2.2) was considered except that there was no stationary process term  $\xi_t(\cdot)$  in (2.2). The proposed PML method was used for parameter estimation, and the parameters were computed using the ECM algorithm introduced in Section 3 but with no  $\boldsymbol{\theta}_\xi$  term. The tuning parameters  $K$  and  $\alpha$  were selected by CV, (4.3), but with  $\widehat{Y}_t^{(K, \alpha)}(\mathbf{s})$  obtained from the reduced model. This method, avoiding taking the inverse of  $\mathbf{V}_\xi(\boldsymbol{\theta}_\xi) + \mathbf{I}_n$  in the proposed multicycle ECM algorithm, has the advantage of being computationally more efficient.
2. **Method II:** This method applied a pre-smoothing (or pre-gridding) step commonly used in EOF analysis (e.g., Karl, Koscielny, and Diaz (1982); Wikle and Cressie (1999)) when data are sampled at irregularly spaced locations. First,  $\{Y_t(\mathbf{s}) : \mathbf{s} \in D\}$  were imputed based on data  $\mathbf{Z}_t$  using natural cubic splines with the smoothing parameter chosen by generalized cross-validation

(GCV) for each  $t = 1, \dots, T$ . Then we obtained the sample covariance function  $\left\{ \sum_{t=1}^T Y_t(\mathbf{s})Y_t(\mathbf{s}')/T : \mathbf{s}, \mathbf{s}' \in D \right\}$ , whose first  $K$  eigenfunctions were used as estimates of  $\phi_1(\cdot), \dots, \phi_K(\cdot)$  and the corresponding eigenvalues were used as estimates of  $\lambda_1, \dots, \lambda_K$ . With the parameters given by these estimates being assumed known, the remaining parameters,  $\sigma_\varepsilon^2$  and  $\boldsymbol{\theta}_\xi$ , were estimated by ML.

3. **Method III:** This method used the estimates of  $\phi_1(\cdot), \dots, \phi_K(\cdot)$  from Method II, but  $\lambda_1, \dots, \lambda_K$  were further re-estimated along with  $\sigma_\varepsilon^2$  and  $\boldsymbol{\theta}_\xi$  using ML by applying a multicycle ECM algorithm similar to the one described in Section 3.

The AMSPE performance of the four methods is shown in Table 1, where “Ideal” refers to ideal selection for the number of basis functions, which corresponds to the smallest AMSPE among those based on  $K \in \{0, 1, 2, 3\}$  for each replicate. The performance of the four methods for covariance function estimation in terms of AMSE is shown in Table 2. In general, the proposed PML method performed significantly better than the other three methods for almost all cases in terms of covariance function estimation. Regarding spatial prediction, although the proposed method also outperformed Method III in all cases, the differences were only significant for some highly nonstationary cases. For highly and mildly nonstationary cases with  $\rho_1 \neq \rho_2$ , the proposed PML method achieved the smallest AMSPE at  $K = 1$  for all cases, whereas the other three methods all achieved the smallest AMSPE at  $K = 0$  in all cases, indicating that the nonstationary feature is not well captured. We found that the surfaces imputed from the pre-smoothing step in Methods II and III tend to be under-smoothed, causing instability in covariance function estimation. In particular, Method III performed miserably due to poor estimation of  $\phi_k(\cdot)$ 's, which in turn caused high variation in CV selection. In addition, the proposed PML method tended to have a smaller bias in estimating AMSPE than the other three methods.

## 4.2. Experiment II

In this experiment, we consider a two-dimensional example of Chang, Hsu, and Huang (2010):

$$Y_t(\mathbf{s}) = \sqrt{x}Y_{1,t}(\mathbf{s}) + \sqrt{1-x}Y_{2,t}(\mathbf{s}),$$

where  $(x, y) = \mathbf{s} \in D = [0, 1] \times [0, 1]$ , and  $Y_{1,t}(\cdot)$  and  $Y_{2,t}(\cdot)$  are independent Gaussian processes with zero means and covariance functions given by (4.1). Then the covariance function of  $Y_t$  depending on  $\rho_1$  and  $\rho_2$  in (4.1) is

Table 1. AMSPEs of various methods for spatial prediction in Experiment I based on 500 simulation replicates, where a boldface value indicates that the corresponding method performs significantly better than the other three methods, and the values given in parentheses are the corresponding standard errors.

$(\rho_1, \rho_2)$	$T$	Method	Ideal	CV	K=0	K=1	K=2	K=3	
(2, 1/20)	50	PML	<b>0.2292</b> (0.0015)	<b>0.2302</b> (0.0015)	0.2369 (0.0017)	<b>0.2299</b> (0.0015)	<b>0.2364</b> (0.0016)	<b>0.2376</b> (0.0018)	
		Method I	.2369 (0.0017)	0.2372 (0.0017)	0.2369 (0.0017)	0.5299 (0.0017)	0.4417 (0.0014)	0.3834 (0.0014)	
		Method II	.2348 (0.0016)	0.2370 (0.0017)	0.2369 (0.0017)	0.2456 (0.0017)	0.2484 (0.0018)	0.2525 (0.0023)	
		Method III	.2335 (0.0016)	0.2694 (0.0054)	0.2369 (0.0017)	0.2648 (0.0055)	0.3051 (0.0061)	0.3532 (0.0072)	
	100	PML	<b>.2290</b> (0.0014)	0.2295 (0.0015)	<b>0.2366</b> (0.0016)	<b>0.2292</b> (0.0014)	<b>0.2359</b> (0.0015)	<b>0.2387</b> (0.0017)	
		Method I	.2366 (0.0016)	0.2369 (0.0016)	0.2366 (0.0016)	0.5306 (0.0012)	0.4463 (0.0011)	0.3875 (0.0010)	
		Method II	.2343 (0.0016)	0.2363 (0.0016)	0.2366 (0.0016)	0.2453 (0.0016)	0.2454 (0.0017)	0.2472 (0.0020)	
		Method III	.2332 (0.0016)	0.2550 (0.0040)	0.2366 (0.0016)	0.2603 (0.0050)	0.2865 (0.0047)	0.3274 (0.0054)	
	(2, 1/2)	50	PML	.2867 (0.0016)	0.2885 (0.0016)	0.2894 (0.0017)	<b>0.2882</b> (0.0016)	<b>0.2886</b> (0.0016)	<b>0.2886</b> (0.0016)
			Method I	.2894 (0.0017)	0.2894 (0.0017)	0.2894 (0.0017)	0.7447 (0.0024)	0.6247 (0.0020)	0.5393 (0.0017)
			Method II	.2875 (0.0016)	0.2896 (0.0017)	0.2894 (0.0017)	0.2921 (0.0017)	0.2946 (0.0017)	0.2955 (0.0018)
			Method III	.2888 (0.0017)	0.3444 (0.0056)	0.2894 (0.0017)	0.3296 (0.0052)	0.3473 (0.0053)	0.3711 (0.0055)
100		PML	.2864 (0.0015)	0.2874 (0.0015)	0.2889 (0.0016)	<b>0.2872</b> (0.0015)	<b>0.2881</b> (0.0015)	<b>0.2881</b> (0.0015)	
		Method I	.2889 (0.0016)	0.2889 (0.0016)	0.2889 (0.0016)	0.7445 (0.0019)	0.6310 (0.0015)	0.5480 (0.0013)	
		Method II	.2873 (0.0015)	0.2888 (0.0016)	0.2889 (0.0016)	0.2919 (0.0016)	0.2940 (0.0016)	0.2955 (0.0017)	
		Method III	.2881 (0.0016)	0.3167 (0.0036)	0.2889 (0.0016)	0.3132 (0.0036)	0.3239 (0.0037)	0.3389 (0.0036)	
(2, 2)		50	PML	.4162 (0.0017)	0.4177 (0.0018)	0.4167 (0.0017)	0.4186 (0.0018)	0.4185 (0.0018)	<b>0.4189</b> (0.0018)
			Method I	.4167 (0.0017)	0.4167 (0.0017)	0.4167 (0.0017)	0.8898 (0.0020)	0.8041 (0.0020)	0.7314 (0.0019)
			Method II	.4161 (0.0017)	0.4169 (0.0017)	0.4167 (0.0017)	0.4181 (0.0017)	0.4207 (0.0017)	0.4225 (0.0018)
			Method III	.4166 (0.0017)	0.5088 (0.0064)	0.4167 (0.0017)	0.4797 (0.0054)	0.5021 (0.0062)	0.5204 (0.0063)
	100	PML	.4153 (0.0016)	0.4163 (0.0016)	0.4155 (0.0016)	0.4168 (0.0016)	0.4169 (0.0017)	<b>0.4170</b> (0.0017)	
		Method I	.4155 (0.0016)	0.4155 (0.0016)	0.4155 (0.0016)	0.8974 (0.0014)	0.8141 (0.0013)	0.7446 (0.0013)	
		Method II	.4152 (0.0016)	0.4157 (0.0016)	0.4155 (0.0016)	0.4169 (0.0016)	0.4189 (0.0016)	0.4208 (0.0016)	
		Method III	.4155 (0.0016)	0.4585 (0.0039)	0.4155 (0.0016)	0.4470 (0.0038)	0.4555 (0.0039)	0.4625 (0.0038)	

Table 2. AMSEs of various methods for covariance function estimation in Experiment I based on 500 simulation replicates, where a boldface value indicates that the corresponding method performs significantly better than the other three methods, and the values given in parentheses are the corresponding standard errors.

$(\rho_1, \rho_2)$	Method	$T = 50$				$T = 100$			
		$K = 0$	$K = 1$	$K = 2$	$K = 3$	$K = 0$	$K = 1$	$K = 2$	$K = 3$
$(2, \frac{1}{20})$	PML	0.0817 (0.0006)	<b>0.0245</b> (0.0008)	<b>0.0278</b> (0.0008)	<b>0.0281</b> (0.0010)	0.0789 (0.0005)	<b>0.0154</b> (0.0004)	<b>0.0191</b> (0.0004)	<b>0.0200</b> (0.0005)
	Method I	0.0817 (0.0006)	0.0482 (0.0009)	0.0408 (0.0012)	0.0593 (0.0034)	0.0789 (0.0005)	0.0373 (0.0004)	0.0281 (0.0004)	0.0325 (0.0012)
	Method II	0.0817 (0.0006)	1.1627 (0.0599)	1.3613 (0.0656)	1.4558 (0.0690)	0.0789 (0.0005)	1.0551 (0.0548)	1.2948 (0.0658)	1.3925 (0.0691)
	Method III	0.0817 (0.0006)	0.1097 (0.0189)	0.1084 (0.0142)	0.1257 (0.0175)	0.0789 (0.0005)	0.1118 (0.0368)	0.0845 (0.0151)	0.0995 (0.0195)
$(2, \frac{1}{2})$	PML	0.0168 (0.0002)	<b>0.0116</b> (0.0004)	<b>0.0111</b> (0.0003)	<b>0.0104</b> (0.0003)	0.0159 (0.0001)	<b>0.0078</b> (0.0002)	<b>0.0091</b> (0.0006)	<b>0.0089</b> (0.0006)
	Method I	0.0168 (0.0002)	0.0637 (0.0007)	0.0467 (0.0005)	0.0389 (0.0007)	0.0159 (0.0001)	0.0570 (0.0006)	0.0403 (0.0003)	0.0303 (0.0003)
	Method II	0.0168 (0.0002)	0.3478 (0.0203)	0.5179 (0.0258)	0.6255 (0.0292)	0.0159 (0.0001)	0.3712 (0.0397)	0.5467 (0.0418)	0.6667 (0.0434)
	Method III	0.0168 (0.0002)	0.0884 (0.0119)	0.0928 (0.0112)	0.1015 (0.0114)	0.0159 (0.0001)	0.0547 (0.0065)	0.0571 (0.0067)	0.0613 (0.0071)
$(2, 2)$	PML	0.0006 (0.0000)	<b>0.0048</b> (0.0002)	<b>0.0060</b> (0.0006)	<b>0.0073</b> (0.0008)	.0003 (0.0000)	<b>0.0031</b> (0.0001)	<b>0.0031</b> (0.0002)	<b>0.0034</b> (0.0002)
	Method I	0.0006 (0.0000)	0.0511 (0.0002)	0.0432 (0.0002)	0.0379 (0.0003)	0.0003 (0.0000)	0.0498 (0.0001)	0.0407 (0.0001)	0.0337 (0.0002)
	Method II	0.0006 (0.0000)	0.2603 (0.0256)	0.3565 (0.0268)	0.4356 (0.0285)	0.0003 (0.0000)	0.3042 (0.0969)	0.3844 (0.0956)	0.4539 (0.0951)
	Method III	0.0006 (0.0000)	0.0990 (0.0141)	0.1187 (0.0157)	0.1269 (0.0148)	0.0003 (0.0000)	0.0608 (0.0177)	0.0658 (0.0159)	0.0672 (0.0158)

$C_Y(\mathbf{s}, \mathbf{s}') = \sqrt{xx'}C_{Y_1}(\mathbf{s}, \mathbf{s}') + \sqrt{(1-x)(1-x')}C_{Y_2}(\mathbf{s}, \mathbf{s}')$ . The data were generated on  $D$  according to (2.1) with  $n = 25$  and  $\sigma_\varepsilon^2 = 0.04$ , where  $\mathbf{s}_1 = (0, 0)$ ,  $\mathbf{s}_2 = (0, 1)$ ,  $\mathbf{s}_3 = (1, 0)$ ,  $\mathbf{s}_4 = (1, 1)$ , and the remaining 21 points,  $\mathbf{s}_5, \dots, \mathbf{s}_{25}$ , were sampled on  $D$  using simple random sampling.

Similar to Experiment I, we considered time points  $T = 50, 100$  in combination with  $(\rho_1, \rho_2) = (10, 1), (4, 1), (1, 1)$ , corresponding to high nonstationarity, mild nonstationarity, and near stationarity, respectively, for  $Y_{j,t}(\cdot)$ 's in (4.1). The contour plots of  $C_Y(\mathbf{s}_0, \mathbf{s})$  at  $\mathbf{s}_0 = (1/2, 1/2)$  and  $(1/4, 1/4)$  for the three models are provided in Figure 2.

We applied the proposed PML method and considered the exponential covariance family of (4.2) for  $\xi_t(\cdot)$  in (2.2). Similar to Experiment I, we chose  $K \in \{0, 1, 2, 3\}$  and  $\alpha \in \mathcal{A}$  simultaneously via 7-fold CV, where  $\{\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, \mathbf{s}_4\}$

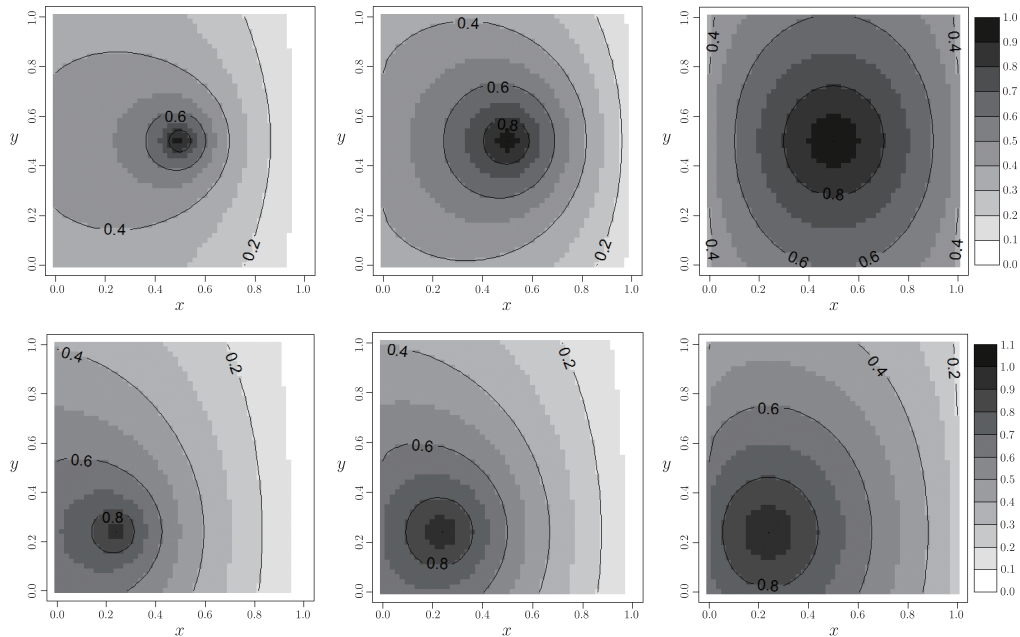


Figure 2. True spatial covariance functions of  $\{C_Y(\mathbf{s}_0, \mathbf{s}) : \mathbf{s} \in D\}$  in Experiment II with three different pairs of parameters,  $(\rho_1, \rho_2) = (10, 1), (4, 1), (1, 1)$  (corresponding to the three columns), and two locations,  $\mathbf{s}_0 = (1/2, 1/2), (1/4, 1/4)$  (corresponding to the two rows).

were always included as data locations in each fitting sample of CV to avoid the boundary effect, and the remaining 21 locations were randomly partitioned into seven disjoint subsets for computing CV as described in Experiment I.

The proposed method was compared with the same methods as in Experiment I. For Methods II and III,  $\{Y_t(\mathbf{s}) : \mathbf{s} \in D\}$  were imputed based on data  $\mathbf{Z}_t$  using thin-plate splines (Duchon (1977); Wahba (1990)) with the smoothing parameter chosen by GCV, for  $t = 1, \dots, T$ . The AMSPE performance of the four methods for spatial prediction is summarized in Table 3, and the AMSE performance for covariance function estimation is shown in Table 4. Overall, the results are similar to those in Experiment I except that Method III outperformed Methods I and II in covariance function estimation. In fact, it even performed better than the proposed PML method in some nearly stationary cases because, unlike Experiment I, the surfaces were well imputed by the pre-smoothing step for Methods II and III. A randomly selected example of the estimated spatial covariance function,  $\widehat{C}_Y(\mathbf{s}_0, \mathbf{s})$ , at  $\mathbf{s}_0 = (1/2, 1/2), (1/4, 1/4)$  based on the proposed PML method for  $T = 50$  is shown in Figure 3. Our method appears to catch nonstationary features quite well.



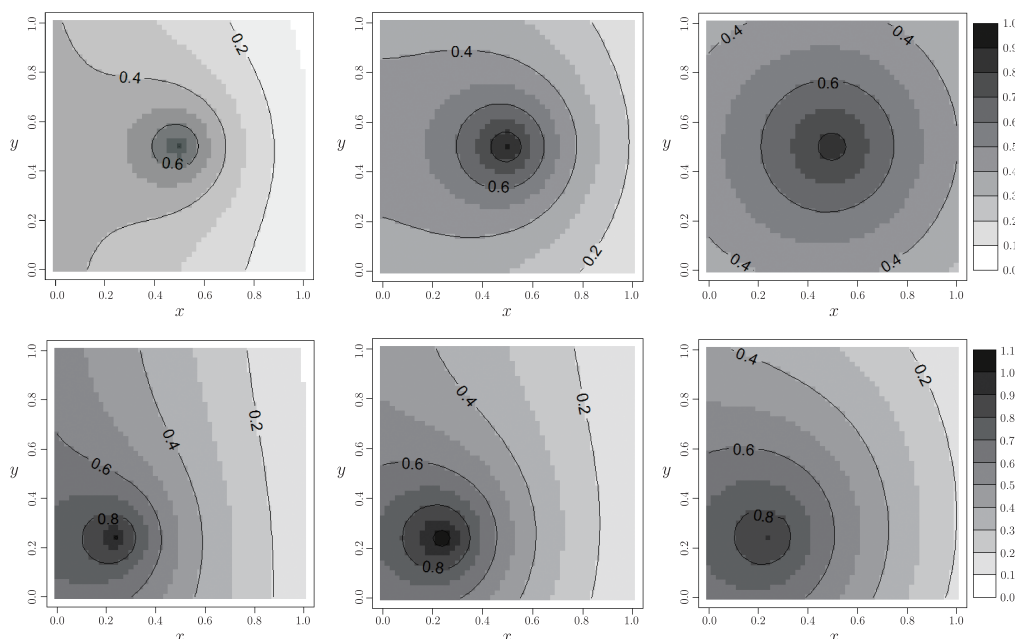


Figure 3. Estimated spatial covariance function  $\{\widehat{C}_Y(\mathbf{s}_0, \mathbf{s}) : \mathbf{s} \in D\}$  using the proposed PML method with  $K = 1$  in Experiment II with three different pairs of parameters,  $(\rho_1, \rho_2) = (10, 1), (4, 1), (1, 1)$  (corresponding to the three columns), and two locations,  $\mathbf{s}_0 = (1/2, 1/2), (1/4, 1/4)$  (corresponding to the two rows).

### 4.3. Experiment III

For the third simulation experiment, we generated data  $\{\mathbf{Z}_1, \dots, \mathbf{Z}_T\}$  directly from (2.3) with  $\phi_1(\mathbf{s}) = \sqrt{3}x$ ,  $\phi_2(\mathbf{s}) = \sqrt{3}(2y - 1)$ , where  $\mathbf{s} = (x, y) \in D = [0, 1] \times [0, 1]$ ,  $\mathbf{w}_t \sim N(\mathbf{0}, \text{diag}(0.25, 0.16))$ ,  $\xi_t(\cdot)$  is a Gaussian process with zero mean and covariance function,  $\text{cov}(\xi_t(\mathbf{s}), \xi_t(\mathbf{s}')) = 0.16 \exp\{-\|\mathbf{s} - \mathbf{s}'\|\}$ , and  $\sigma_\varepsilon^2 = 0.04$ . We took  $n = 25$  and selected the data locations as in Experiment II. We considered time points  $T = 50, 100$ . The contour plots of  $\{C_Y(\mathbf{s}_0, \mathbf{s}) : \mathbf{s} \in D\}$  at  $\mathbf{s}_0 = (1/2, 1/2)$  and  $(1/4, 1/4)$  are shown in Figure 4.

We applied the proposed PML method as in Experiment II. Here we compared our method only with Method III, because Method I contains no stationary process term,  $\xi_t(\cdot)$ , found to be important in the previous two experiments, and Method II had performed worse than Method III in Experiment II. The AMSPE performance of the two methods for spatial prediction is summarized in Table 5, and the AMSE performance of the methods for covariance function estimation is displayed in Table 6. Clearly, our method outperformed Method III in most situations. Notice that both methods achieved the smallest AMSPE value at  $K = 2$ , which agrees with the true number of components. A randomly selected

Table 3. AMSPEs of various methods for spatial prediction in Experiment II based on 500 simulation replicates, where a boldface value indicates that the corresponding method performs significantly better than the other three methods, and the values given in parentheses are the corresponding standard errors.

$(\rho_1, \rho_2)$	$T$	Method	Ideal	CV	$K = 0$	$K = 1$	$K = 2$	$K = 3$	
(10, 1)	50	PML	0.4975 (0.0008)	<b>0.4996</b> (0.0008)	0.5040 (0.0009)	<b>0.4991</b> (0.0008)	<b>0.5055</b> (0.0009)	<b>0.5132</b> (0.0013)	
		Method I	0.5040 (0.0009)	0.5040 (0.0009)	0.5040 (0.0009)	0.6855 (0.0011)	0.6409 (0.0010)	0.6113 (0.0009)	
		Method II	0.5026 (0.0008)	0.5044 (0.0009)	0.5040 (0.0009)	0.5359 (0.0039)	0.5546 (0.0056)	0.5798 (0.0077)	
		Method III	0.4990 (0.0008)	0.5058 (0.0010)	0.5040 (0.0009)	0.5010 (0.0008)	0.5226 (0.0011)	0.5607 (0.0015)	
	100	PML	<b>0.4981</b> (0.0006)	<b>0.4989</b> (0.0006)	0.5047 (0.0007)	0.4988 (0.0006)	<b>0.5034</b> (0.0007)	<b>0.5136</b> (0.0011)	
		Method I	0.5047 (0.0007)	0.5049 (0.0008)	0.5047 (0.0007)	0.6882 (0.0008)	0.6455 (0.0008)	0.6154 (0.0008)	
		Method II	0.5034 (0.0007)	0.5049 (0.0007)	0.5047 (0.0007)	0.5332 (0.0036)	0.5424 (0.0048)	0.5632 (0.0068)	
		Method III	0.4994 (0.0007)	0.5010 (0.0007)	0.5047 (0.0007)	0.5006 (0.0007)	0.5153 (0.0007)	0.5545 (0.0014)	
	(4, 1)	50	PML	0.3195 (0.0009)	0.3203 (0.0009)	0.3205 (0.0009)	0.3202 (0.0009)	<b>0.3225</b> (0.0009)	<b>0.3256</b> (0.0010)
			Method I	0.3205 (0.0009)	0.3205 (0.0009)	0.3205 (0.0009)	0.6331 (0.0015)	0.5448 (0.0012)	0.4826 (0.0011)
			Method II	0.3202 (0.0009)	0.3208 (0.0009)	0.3205 (0.0009)	0.3254 (0.0009)	0.3316 (0.0018)	0.3587 (0.0062)
			Method III	0.3198 (0.0009)	0.3229 (0.0009)	0.3205 (0.0009)	0.3213 (0.0009)	0.3310 (0.0009)	0.3686 (0.0016)
100		PML	0.3200 (0.0008)	0.3204 (0.0008)	0.3210 (0.0008)	0.3203 (0.0008)	<b>0.3216</b> (0.0008)	<b>0.3248</b> (0.0009)	
		Method I	0.3210 (0.0008)	0.3210 (0.0008)	0.3210 (0.0008)	0.6380 (0.0012)	0.5534 (0.0010)	0.4902 (0.0010)	
		Method II	0.3208 (0.0008)	0.3212 (0.0008)	0.3210 (0.0008)	0.3255 (0.0008)	0.3292 (0.0011)	0.3474 (0.0048)	
		Method III	0.3203 (0.0008)	0.3209 (0.0008)	0.3210 (0.0008)	0.3211 (0.0008)	0.3278 (0.0008)	0.3637 (0.0014)	
(1, 1)		50	PML	0.1534 (0.0005)	0.1538 (0.0005)	0.1537 (0.0005)	0.1537 (0.0005)	<b>0.1550</b> (0.0005)	<b>0.1613</b> (0.0008)
			Method I	0.1537 (0.0005)	0.1537 (0.0005)	0.1537 (0.0005)	0.4614 (0.0016)	0.3336 (0.0010)	0.2621 (0.0007)
			Method II	0.1536 (0.0005)	0.1538 (0.0005)	0.1537 (0.0005)	0.1548 (0.0005)	0.1590 (0.0007)	0.1653 (0.0014)
			Method III	0.1536 (0.0005)	0.1546 (0.0005)	0.1537 (0.0005)	0.1539 (0.0005)	0.1582 (0.0005)	0.1845 (0.0008)
	100	PML	0.1542 (0.0004)	0.1544 (0.0004)	0.1544 (0.0004)	0.1543 (0.0004)	<b>0.1553</b> (0.0004)	<b>0.1616</b> (0.0006)	
		Method I	0.1544 (0.0004)	0.1544 (0.0004)	0.1544 (0.0004)	0.4666 (0.0012)	0.3391 (0.0008)	0.2666 (0.0006)	
		Method II	0.1543 (0.0004)	0.1544 (0.0004)	0.1544 (0.0004)	0.1551 (0.0004)	0.1598 (0.0007)	0.1710 (0.0029)	
		Method III	0.1543 (0.0004)	0.1546 (0.0004)	0.1544 (0.0004)	0.1544 (0.0004)	0.1584 (0.0004)	0.1849 (0.0007)	

Table 4. AMSEs of various methods for covariance function estimation in Experiment II based on 500 simulation replicates, where a boldface value indicates that the corresponding method performs significantly better than the other three methods, and the values given in parentheses are the corresponding standard errors.

$(\rho_1, \rho_2)$	Method	$T = 50$				$T = 100$			
		$K = 0$	$K = 1$	$K = 2$	$K = 3$	$K = 0$	$K = 1$	$K = 2$	$K = 3$
(10, 1)	PML	0.0266 (0.0003)	<b>0.0104</b> (0.0003)	<b>0.0112</b> (0.0003)	<b>0.0123</b> (0.0005)	0.0249 (0.0002)	<b>0.0069</b> (0.0002)	<b>0.0076</b> (0.0002)	<b>0.0083</b> (0.0002)
	Method I	0.0266 (0.0003)	0.0225 (0.0004)	0.0205 (0.0004)	0.0257 (0.0010)	0.0249 (0.0002)	0.0174 (0.0002)	0.0146 (0.0002)	0.0140 (0.0003)
	Method II	0.0266 (0.0003)	0.5094 (0.0270)	0.5891 (0.0302)	0.6427 (0.0325)	0.0249 (0.0002)	0.5276 (0.0246)	0.6163 (0.0287)	0.6980 (0.0320)
	Method III	0.0266 (0.0003)	0.0125 (0.0004)	0.0153 (0.0004)	0.0165 (0.0004)	0.0249 (0.0002)	0.0074 (0.0002)	0.0088 (0.0002)	0.0097 (0.0002)
(4, 1)	PML	0.0144 (0.0002)	<b>0.0115</b> (0.0004)	<b>0.0143</b> (0.0006)	<b>0.0166</b> (0.0012)	0.0129 (0.0001)	<b>0.0070</b> (0.0002)	0.0113 (0.0019)	0.0144 (0.0033)
	Method I	0.0144 (0.0002)	0.0405 (0.0005)	0.0317 (0.0005)	0.0297 (0.0010)	0.0129 (0.0001)	0.0331 (0.0003)	0.0240 (0.0003)	0.0178 (0.0003)
	Method II	0.0144 (0.0002)	0.5813 (0.0292)	0.7469 (0.0367)	0.8893 (0.0435)	0.0129 (0.0001)	0.5902 (0.0264)	0.7666 (0.0350)	0.8998 (0.0411)
	Method III	0.0144 (0.0002)	0.0196 (0.0008)	0.0193 (0.0006)	0.0191 (0.0005)	0.0129 (0.0001)	0.0119 (0.0004)	0.0109 (0.0003)	<b>0.0106</b> (0.0002)
(1, 1)	PML	0.0107 (0.0005)	<b>0.0248</b> (0.0009)	0.0270 (0.0008)	0.0288 (0.0008)	0.0083 (0.0003)	<b>0.0172</b> (0.0005)	0.0207 (0.0012)	0.0250 (0.0029)
	Method I	0.0107 (0.0005)	0.0550 (0.0009)	0.0362 (0.0008)	0.0305 (0.0010)	0.0083 (0.0003)	0.0422 (0.0005)	0.0236 (0.0004)	0.0162 (0.0004)
	Method II	0.0107 (0.0005)	1.0131 (0.0513)	1.7709 (0.0904)	2.0795 (0.1020)	0.0083 (0.0003)	0.9867 (0.0465)	1.7504 (0.0870)	2.0841 (0.1019)
	Method III	0.0107 (0.0005)	0.0513 (0.0023)	<b>0.0245</b> (0.0009)	<b>0.0249</b> (0.0008)	0.0083 (0.0003)	0.0322 (0.0013)	<b>0.0124</b> (0.0004)	<b>0.0124</b> (0.0004)

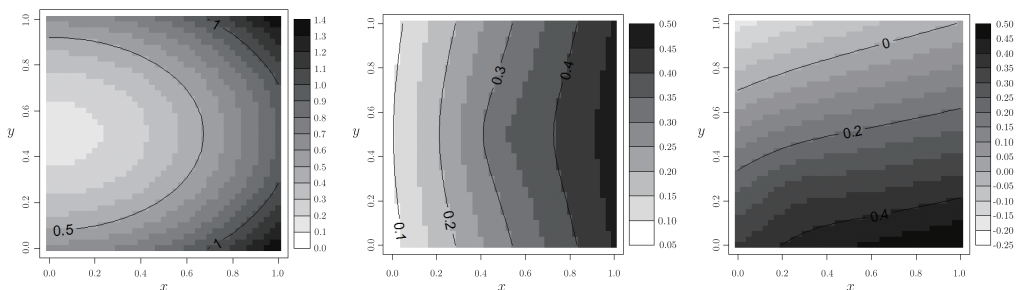


Figure 4. True spatial covariance functions in Experiment III: (a)  $\{C_Y(\mathbf{s}, \mathbf{s}) : \mathbf{s} \in D\}$ ; (b)  $\{C_Y(\mathbf{s}_0, \mathbf{s}) : \mathbf{s} \in D\}$  for  $\mathbf{s}_0 = (1/2, 1/2)$ ; (c)  $\{C_Y(\mathbf{s}_0, \mathbf{s}) : \mathbf{s} \in D\}$  for  $\mathbf{s}_0 = (1/4, 1/4)$ .

Table 5. AMSPEs of various methods for spatial prediction in Experiment III based on 500 simulation replicates, where a boldface value indicates that the corresponding method performs significantly better than the other method, and the values given in parentheses are the corresponding standard errors.

$T$	Method	Ideal	CV	$K = 0$	$K = 1$	$K = 2$	$K = 3$
50	PML	<b>0.0327</b> (0.0001)	<b>0.0330</b> (0.0001)	0.0351 (0.0001)	<b>0.0336</b> (0.0001)	<b>0.0330</b> (0.0001)	0.0389 (0.0001)
	Method III	0.0332 (0.0001)	0.0335 (0.0001)	0.0351 (0.0001)	0.0345 (0.0001)	0.0334 (0.0001)	<b>0.0385</b> (0.0001)
100	PML	<b>0.0327</b> (0.0001)	<b>0.0328</b> (0.0001)	0.0352 (0.0001)	<b>0.0336</b> (0.0001)	<b>0.0328</b> (0.0001)	0.0389 (0.0001)
	Method III	0.0331 (0.0001)	0.0332 (0.0001)	0.0352 (0.0001)	0.0347 (0.0001)	0.0332 (0.0001)	<b>0.0380</b> (0.0001)

Table 6. AMSEs of various methods for covariance function estimation in Experiment III based on 500 simulation replicates, where a boldface value indicates that the corresponding method performs significantly better than the other method, and the values given in parentheses are the corresponding standard errors.

Method	$T = 50$				$T = 100$			
	$K = 0$	$K = 1$	$K = 2$	$K = 3$	$K = 0$	$K = 1$	$K = 2$	$K = 3$
PML	0.6705 (0.0132)	<b>0.0613</b> (0.0016)	<b>0.0096</b> (0.0003)	0.0092 (0.0004)	0.6533 (0.0093)	<b>0.0582</b> (0.0011)	<b>0.0053</b> (0.0002)	0.0048 (0.0002)
Method III	0.6705 (0.0132)	0.3407 (0.0084)	0.0104 (0.0004)	0.0094 (0.0004)	0.6533 (0.0093)	0.3609 (0.0061)	0.0057 (0.0002)	0.0049 (0.0002)

example of the estimated spatial covariance function,  $\{\widehat{C}_Y(\mathbf{s}_0, \mathbf{s}) : \mathbf{s} \in D\}$ , at  $\mathbf{s}_0 = (1/2, 1/2), (1/4, 1/4)$ , based on the proposed PML method for  $T = 50$ , is shown in Figure 5. Again, our method can be seen to catch nonstationary features quite well.

#### 4.4. Experiment IV

Here we generated data using the model fitted by our PML method to the monthly precipitation dataset in Section 5, in order to more closely reflect a real-world situation. Specifically, we considered the model of (2.1) and (2.2) with mean function given by (5.1), and with the parameters of  $\boldsymbol{\beta}$  and  $\boldsymbol{\theta}$  estimated by our PML method (see Table 10) based on the data observed at 79 stations with complete records. We further investigated how the proposed method is affected by the temporal correlation of  $\{\mathbf{w}_t : t \in \mathbb{N}\}$  by assuming a multivariate stationary autoregressive time series of order 1:

$$\mathbf{w}_t = a\mathbf{w}_{t-1} + \mathbf{v}_t, \quad (4.4)$$

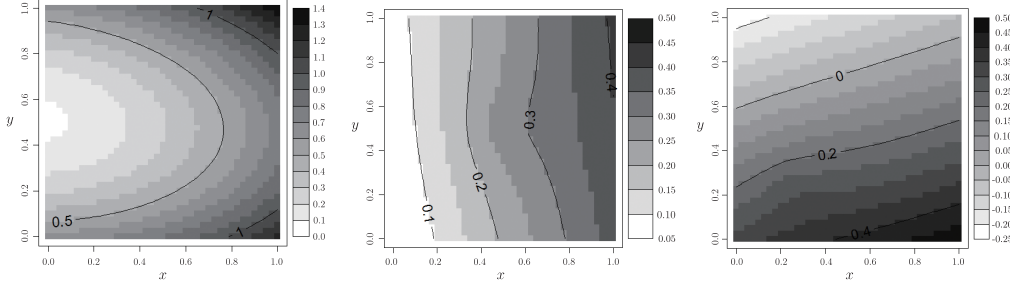


Figure 5. Estimated spatial covariance functions using the proposed PML method with  $K = 2$  in Experiment III: (a)  $\{\hat{C}_Y(\mathbf{s}, \mathbf{s}) : \mathbf{s} \in D\}$ ; (b)  $\{\hat{C}_Y(\mathbf{s}_0, \mathbf{s}) : \mathbf{s} \in D\}$  for  $\mathbf{s}_0 = (1/2, 1/2)$ ; (c)  $\{\hat{C}_Y(\mathbf{s}_0, \mathbf{s}) : \mathbf{s} \in D\}$  for  $\mathbf{s}_0 = (1/4, 1/4)$ .

where  $\mathbf{v}_t \sim N(\mathbf{0}, (1 - a^2)\mathbf{\Lambda}_K)$ . We considered  $a \in \{0, 0.1, 0.5, 0.9\}$  corresponding to different strengths of temporal dependence, and generated data according to (2.3) and (4.4) at the same 79 locations and the same number of time points. For locations at which  $x_{t,13}(\mathbf{s}), \dots, x_{t,15}(\mathbf{s})$  were not observed, we estimated the surface  $x_{t,j}(\cdot)$  for  $j = 13, 14, 15$ , at each time  $t$  separately, using thin plate splines with the tuning parameter chosen by GCV.

We applied the proposed PML method in the same way as in Experiments II and III, and chose  $K \in \{0, 1, 2, 3\}$  and  $\alpha \in \{2^{-10}, 2^{-9}, \dots, 2^{10}\}$  using 5-fold CV by randomly partitioning the sampling locations into five disjoint subsets of sizes 16, 16, 16 and 15. The AMSPE performance of our method for spatial prediction and the AMSE performance of our method for covariance function estimation are displayed in Tables 7 and 8, respectively. We note that our covariance function estimate deteriorated only slightly when the temporal dependence parameter  $a$  increased from 0 to 0.5. However, it deteriorated more severely when  $a = 0.9$  with a significantly higher variance for  $\hat{\lambda}_1$ . This is not much of a surprise since the effective sample size (i.e., the equivalent number of independent repeated measurements) is smaller when  $a$  is larger. Similar behavior can also be seen in spatial prediction but with less significant differences. As expected, the model based on  $K = 1$  performed the best in both spatial prediction and covariance function estimation. On the other hand, it is interesting to see that the model based on either  $K = 2$  or  $K = 3$  performed almost as well, because  $\hat{\lambda}_2$  and  $\hat{\lambda}_3$  were nearly negligible. Additionally, Table 9 reports the means and the standard deviations for the estimated parameters based on the model selected by CV. All the estimates can be seen to perform reasonably well with relatively small biases regardless of values of  $a$ . Note that except for  $\hat{\lambda}_1$ , which is more directly influenced by  $a$ , the standard errors of the other parameter estimates increase only slightly with  $a$ . This is not much a surprise, because the variance contributed

Table 7. AMSPEs ( $\times 100$ ) of the proposed method for spatial prediction in Experiment IV for data with different strengths of temporal dependence based on 500 simulation replicates, where the values given in parentheses are the corresponding standard errors.

$a$	Ideal	CV	$K = 0$	$K = 1$	$K = 2$	$K = 3$
0.0	6.239	6.243	6.299	6.241	6.243	6.243
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
0.1	6.239	6.243	6.299	6.241	6.243	6.243
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
0.5	6.239	6.244	6.302	6.241	6.244	6.244
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
0.9	6.241	6.245	6.310	6.243	6.245	6.245
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)

Table 8. AMSEs ( $\times 100$ ) of the proposed method for covariance function estimation in Experiment IV for data with different strengths of temporal dependence based on 500 simulation replicates, where the values given in parentheses are the corresponding standard errors.

$a$	CV	$K = 0$	$K = 1$	$K = 2$	$K = 3$
0.0	0.049	0.379	0.043	0.045	0.045
	(0.002)	(0.005)	(0.002)	(0.002)	(0.002)
0.1	0.050	0.379	0.043	0.045	0.045
	(0.002)	(0.005)	(0.002)	(0.002)	(0.002)
0.5	0.055	0.384	0.048	0.050	0.051
	(0.003)	(0.006)	(0.002)	(0.002)	(0.002)
0.9	0.146	0.496	0.136	0.139	0.141
	(0.014)	(0.032)	(0.014)	(0.014)	(0.014)

by the (temporally independent) stationary component ( $3.137 \times 0.056$ ) is about three times larger than that contributed by the (autoregressive) nonstationary component (0.057), and hence the effective sample size is not too much smaller than  $T$  even when  $a = 0.9$ .

Although our method ignores temporal dependence, the performance is not much affected as long as that is not strong. This is consistent with the theoretical result of Hörmann and Kokoszka (2010) for weakly dependent functional data. In addition, CV appears to be effective in selecting  $K$  for spatial prediction and covariance function estimation.

## 5. Application

We applied the proposed method to a precipitation dataset obtained from the National Climatic Data Center (available at <http://www.cgd.ucar.edu/stats/Data/US.monthly.met/>). The data consist of monthly total precipitation at 376 stations, mostly located in Colorado, from November 1993 to December 1997

Table 9. Means and the standard deviations (in parentheses) of the estimated parameters for the PML method in Experiment IV based on 500 simulation replicates.

	True	$a = 0.0$	$a = 0.1$	$a = 0.5$	$a = 0.9$
$\tau_\xi$	3.137	3.006(0.220)	3.009(0.228)	3.008(0.237)	3.014(0.228)
$\rho_\xi$	39.050	43.428(4.219)	43.397(4.329)	43.436(4.397)	43.284(4.431)
$\sigma_\varepsilon^2$	0.056	0.055(0.002)	0.055(0.002)	0.055(0.002)	0.055(0.002)
$\lambda_1 \int_D \phi_1^2(\mathbf{s}) d\mathbf{s} /  D $	0.057	0.054(0.018)	0.054(0.018)	0.054(0.019)	0.053(0.035)
$\lambda_2 \int_D \phi_2^2(\mathbf{s}) d\mathbf{s} /  D $	0.000	0.004(0.007)	0.004(0.008)	0.004(0.008)	0.004(0.008)
$\lambda_3 \int_D \phi_3^2(\mathbf{s}) d\mathbf{s} /  D $	0.000	0.001(0.003)	0.001(0.003)	0.001(0.003)	0.001(0.003)
$\beta_1$	1.485	1.478(0.117)	1.477(0.117)	1.476(0.117)	1.473(0.126)
$\beta_2$	0.139	0.144(0.129)	0.144(0.128)	0.143(0.128)	0.143(0.129)
$\beta_3$	0.458	0.461(0.133)	0.461(0.133)	0.460(0.133)	0.458(0.134)
$\beta_4$	1.271	1.272(0.144)	1.272(0.144)	1.271(0.143)	1.269(0.145)
$\beta_5$	1.863	1.873(0.153)	1.873(0.153)	1.872(0.154)	1.868(0.161)
$\beta_6$	2.119	2.115(0.172)	2.115(0.172)	2.114(0.173)	2.110(0.183)
$\beta_7$	2.230	2.224(0.178)	2.224(0.178)	2.222(0.180)	2.217(0.190)
$\beta_8$	2.444	2.448(0.174)	2.448(0.174)	2.446(0.176)	2.441(0.186)
$\beta_9$	1.980	1.982(0.160)	1.982(0.160)	1.981(0.161)	1.978(0.168)
$\beta_{10}$	1.257	1.258(0.145)	1.258(0.145)	1.258(0.145)	1.255(0.151)
$\beta_{11}$	0.404	0.405(0.130)	0.405(0.129)	0.405(0.129)	0.403(0.128)
$\beta_{12}$	-0.151	-0.153(0.125)	-0.153(0.125)	-0.152(0.124)	-0.153(0.123)
$\beta_{13}$	-0.139	-0.136(0.038)	-0.136(0.038)	-0.136(0.040)	-0.133(0.048)
$\beta_{14}$	-0.065	-0.065(0.005)	-0.065(0.005)	-0.065(0.005)	-0.065(0.005)
$\beta_{15}$	-0.003	-0.003(0.003)	-0.003(0.003)	-0.003(0.003)	-0.003(0.003)

between latitudes 101 degrees and 109.5 degrees west and between longitudes 36.5 degrees and 41.5 degrees north. Among 376 stations, 79 of them have complete data, which were used as training data. The remaining data at 297 stations were used for validation purpose. Figure 6(a) displays the locations of all 376 stations and the elevations for the 79 sampling (training) locations on a rectangular region  $D$  based on the polyconic projection. Because monthly total rainfalls take only positive values and tend to have a distribution that is skewed to the right, they were first transformed using the log transformation

$$Y_t(\mathbf{s}) = \log(y_t(\mathbf{s}) + 1), \quad \mathbf{s} \in D, t = 1, \dots, 50,$$

where  $y_t(\mathbf{s})$  denotes the monthly total precipitation (in millimeters) at location  $\mathbf{s}$  and time  $t$ . We applied the propose PML method with  $\xi_t(\cdot)$  in (2.2) modeled by the exponential covariance family of (4.2). In addition, the mean function in (2.2) was modeled as

$$\mathbf{x}_t(\mathbf{s})' \boldsymbol{\beta} = \beta_1 + \sum_{j=2}^{12} \beta_j x_{t,j}(\mathbf{s}) + \beta_{13} x_{t,13}(\mathbf{s}) + \beta_{14} x_{t,14}(\mathbf{s}) + \beta_{15} x_{t,15}(\mathbf{s}),$$

$$\mathbf{s} \in D, t = 1, \dots, 50, \quad (5.1)$$

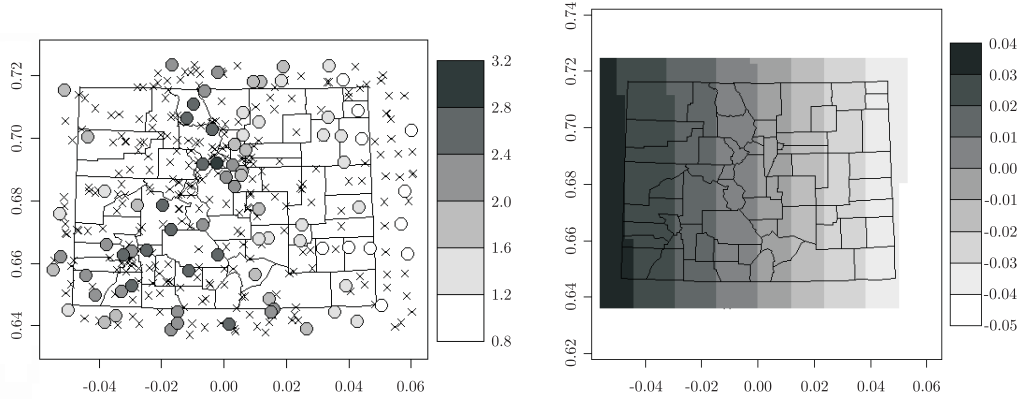


Figure 6. (a) Elevations (in kilometers) of 79 sample stations (denoted as “o”) and the locations of the 297 validation stations (denoted as “x”) for the precipitation dataset in Colorado. (b) The first eigenfunction  $\phi_1(\cdot)$  estimated by PML based on  $K = 1$ .

where  $x_{t,2}(\mathbf{s}), \dots, x_{t,12}(\mathbf{s})$  are dummy variables for February,  $\dots$ , December, independent of  $\mathbf{s}$ , such that  $x_{t,2}(\mathbf{s}) = 1$  if  $t$  falls in February and 0 otherwise, and so on,  $x_{t,13}(\mathbf{s})$  denotes the elevation (in kilometers) at location  $\mathbf{s}$  independent of  $t$ ,  $x_{t,14}(\mathbf{s})$  denotes the average daily maximum temperature (in degrees Celsius) at location  $\mathbf{s}$  and time  $t$ , and  $x_{t,15}(\mathbf{s})$  denotes the extent of the average daily temperatures (in degrees Celsius) at location  $\mathbf{s}$  and time  $t$ . We chose  $K \in \{0, 1, 2, 3\}$  and  $\alpha \in [2^{-10}, 2^5]$  using 5-fold CV by randomly partitioning the sampling locations into five disjoint subsets of sizes 16, 16, 16, 16 and 15.

For known  $\boldsymbol{\theta}$ , an unbiased predictor of  $y_t(\mathbf{s})$  obtained from the universal kriging predictor  $\hat{Y}_t(\mathbf{s})$  of (2.10) is (Journal (1980))

$$\hat{y}_t(\mathbf{s}) \equiv \exp \left\{ \hat{Y}_t(\mathbf{s}) + \text{MSPE}_{uk}(\mathbf{s})/2 + \mathbf{x}(\mathbf{s})' \mathbf{m}(\mathbf{s}) \right\} - 1, \quad (5.2)$$

where  $\mathbf{m}(\mathbf{s}) \equiv (\mathbf{X}'\boldsymbol{\Sigma}_Z^{-1}\mathbf{X})^{-1}(\mathbf{X}'\boldsymbol{\Sigma}_Z^{-1}\boldsymbol{\sigma}_Y(\mathbf{s}) - \mathbf{x}(\mathbf{s}))$ . The corresponding conditional MSPE is (Journal (1980); Chilès and Delfiner (1999, p.191))

$$\begin{aligned} E(\hat{y}_t(\mathbf{s}) - y_t(\mathbf{s}))^2 &= \exp \left\{ 2\mathbf{x}(\mathbf{s})' \boldsymbol{\beta} + 2C_Y(\mathbf{s}, \mathbf{s}) \right\} \\ &\quad \times \left[ 1 + \exp \left\{ -\text{MSPE}_{uk}(\mathbf{s}) - 2\mathbf{x}(\mathbf{s})' \mathbf{m}(\mathbf{s}) \right\} \right. \\ &\quad \left. \times \left\{ \exp \left( -2\mathbf{x}(\mathbf{s})' \mathbf{m}(\mathbf{s}) \right) - 2 \right\} \right]. \end{aligned} \quad (5.3)$$

Our final predictor of  $\hat{y}_t(\mathbf{s})$  and the corresponding conditional MSPE were obtained by plugging the PML estimates of  $\boldsymbol{\beta}$  and  $\boldsymbol{\theta}$  into (5.2) and (5.3).

For the proposed PML method, CV selected  $K = 1$  with the estimated eigenfunction shown in Figure 6(b). Evidently, a nearly linear pattern from west



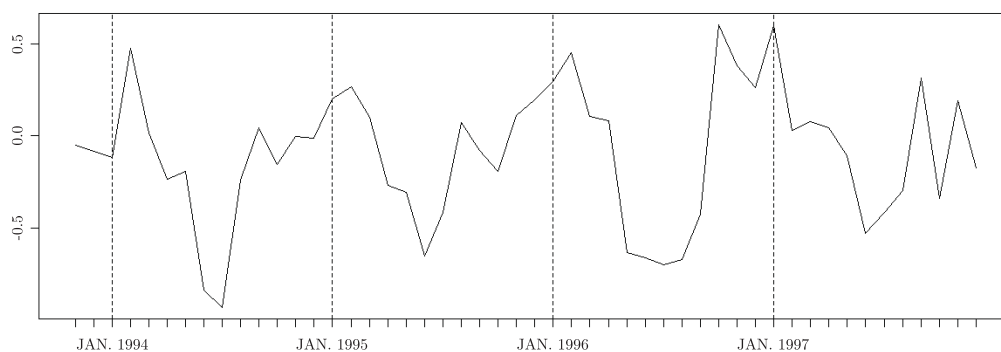


Figure 7. The predicted scores  $\{\hat{w}_{t,1} : t = 1, \dots, 50\}$  for  $\phi_1(\cdot)$  obtained from PML with  $K = 1$  for the precipitation dataset in Colorado.

to east can be seen. As shown by the predicted scores,  $\{\hat{w}_{t,1} : t = 1, \dots, 50\}$ , in Figure 7, the precipitation tends to be higher in the west (east) during winter (summer). Note that in Colorado the mountains and ranges extend north and south approximately through the middle of the state and, hence, stop much of moisture from the Pacific Ocean and the Gulf of Mexico, particularly in winter when the prevailing winds come from the west with much of rain or snow falling on the mountaintops and westward-facing slopes, causing lighter precipitation in the eastern plain. On the other hand, the eastern plain tends to have more showers and thunderstorms in spring and summer due to warm and moist air from the south. In addition to the seasonal pattern, the  $\hat{w}_{t,1}$ 's also showed some temporal dependence structure with the lag 1 sample autocorrelation 0.529, which could be modeled further but that is beyond the scope of this paper.

The data were also analyzed by recursively applying Method III to the residuals obtained from generalized least squares fit of  $\beta$  for estimating  $\theta$  until convergence, where the initial estimate of  $\beta$  was obtained using ordinary least squares. Different from the PML method, Method III chooses  $K = 2$  eigenfunctions from CV. Table 10 displays the estimated parameters for both methods, where the estimated standard deviations for the elements of  $\hat{\beta}_{gls}$  shown in parentheses were obtained from the square roots of the diagonal of  $(\sum_{t=1}^T \mathbf{X}_t' \hat{\Sigma}_Z^{-1} \mathbf{X}_t)^{-1}$  and  $\hat{\Sigma}_Z \equiv \hat{\Phi}_K \hat{\Lambda}_K \hat{\Phi}_K' + \hat{\sigma}_\varepsilon^2 (\hat{\mathbf{V}}_\xi(\hat{\theta}_\xi) + \mathbf{I}_n)$  since  $\text{var}(\hat{\beta}_{gls}) \approx (\sum_{t=1}^T \mathbf{X}_t' \hat{\Sigma}_Z^{-1} \mathbf{X}_t)^{-1}$ . The results show that the amounts of rainfall are affected by elevation, average maximum daily temperature, and some monthly effects. As we expect, higher temperatures tend to produce larger amounts of rainfall. Although the estimated coefficient for elevation is negative, higher elevations tend to have lower temperatures and hence smaller amounts of rainfall. Consequently, the overall effect of elevation on rainfall is not clear. Comparing these two methods, we notice that

Table 10. Estimated parameters based on PML and Method III for the Colorado precipitation data, where values in parentheses are the corresponding estimated standard deviations.

	PML		Method III	
$\tau_\xi$	3.137		1.349	
$\rho_\xi$	39.050		47.614	
$\sigma_\varepsilon^2$	0.056		0.059	
$\lambda_1 \int_D \phi_1^2(\mathbf{s}) d\mathbf{s} /  D $	0.057		4.575	
$\lambda_2 \int_D \phi_2^2(\mathbf{s}) d\mathbf{s} /  D $	-		0.080	
$\lambda_3 \int_D \phi_3^2(\mathbf{s}) d\mathbf{s} /  D $	-		-	
$\beta_1$	1.485	(0.117)	-0.458	(0.211)
$\beta_2$	0.139	(0.126)	0.770	(0.295)
$\beta_3$	0.458	(0.130)	1.524	(0.299)
$\beta_4$	1.271	(0.134)	3.074	(0.303)
$\beta_5$	1.863	(0.147)	4.522	(0.310)
$\beta_6$	2.119	(0.164)	5.123	(0.322)
$\beta_7$	2.230	(0.172)	4.578	(0.328)
$\beta_8$	2.444	(0.171)	5.050	(0.327)
$\beta_9$	1.980	(0.155)	4.357	(0.315)
$\beta_{10}$	1.257	(0.139)	3.556	(0.303)
$\beta_{11}$	0.404	(0.121)	0.737	(0.281)
$\beta_{12}$	-0.151	(0.119)	0.225	(0.280)
$\beta_{13}$	-0.139	(0.037)	-0.643	(0.035)
$\beta_{14}$	-0.065	(0.005)	-0.111	(0.004)
$\beta_{15}$	-0.003	(0.003)	-0.000	(0.003)

the variation of our method was contributed mainly by the stationary component, whereas that of Method III is contributed mainly by nonstationary components.

We used the validation data to compare the performance between the two methods based on the averaged squared prediction errors

$$\text{ASPE} \equiv \frac{1}{\sum_{t=1}^{50} n_t} \sum_{t=1}^{50} \sum_{j=1}^{n_t} (z_t(\mathbf{s}_{t,j}) - \hat{y}_t(\mathbf{s}_{t,j}))^2,$$

where  $z_t(\mathbf{s}_{t,j})$  denotes the observed precipitation at time  $t$  and site  $\mathbf{s}_{t,j}$ , and  $n_t$  is the total number of observed stations for the validation data at time  $t$ . The proposed method (based on  $K = 1$ ) achieved an ASPE value of 9.815, smaller than the 11.751 obtained from Method III (based on  $K = 2$ ).

## 6. Discussion

In this paper, we consider nonstationary spatial modeling and develop a new EOF-type estimation method in the penalized likelihood framework. The

proposed method is applicable to sparse data and is shown to perform well in spatial prediction and covariance function estimation. Although the methodology is developed for complete data, the same PML formulation can be extended to allow for different missing observations at different time points, leading to more complicated formulae by having both missing observations and latent variables as missing data in the ECM framework.

Specifically, and without loss of generality, we can assume that for each  $t = 1, \dots, T$ , we observe data at a subset of  $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$  with size  $n_t \leq n$ . In other words, we can decompose  $\mathbf{Z}_t$  into the observed data  $\mathbf{Z}_t^{(obs)}$  and the missing data  $\mathbf{Z}_t^{(mis)}$ ,  $t = 1, \dots, T$ . We also consider the same full negative penalized log-likelihood of (3.1), but treat both  $\mathbf{w}_t$ 's and  $\mathbf{Z}_t^{(mis)}$ 's as missing data. Then the PML estimates for  $\boldsymbol{\beta}$  and  $\boldsymbol{\theta}$  can be obtained by applying a similar ECM algorithm while conditioning only on  $\mathbf{Z}_1^{(obs)}, \dots, \mathbf{Z}_T^{(obs)}$  in the E-step.

Note that the proposed method is designed for small to moderately large  $n$  (number of data locations) since it requires taking the inverse of  $\mathbf{V}_\xi(\boldsymbol{\theta}_\xi) + \mathbf{I}_n$  for various values of  $\boldsymbol{\theta}_\xi$ ; that is difficult to handle if  $n$  is larger than a few thousands. Further research is needed to accommodate very large  $n$  using, for example, the tapering technique (Furrer, Genton, and Nychka (2006)), and to incorporate temporal dependence structure, say in terms of (4.4). Although CV appears to work well in selecting the number of eigenfunctions, we might further adopt some regularization method such as Lasso (Tibshirani (1996)) by adding a specific penalty to the proposed PML; this would allow simultaneous selection of  $K$  and parameter estimation.

## References

- Alishouse, J. C., Crone, L. J., Fleming, H. E., Van Cleef, F. L. and Wark, D. Q. (1967). A discussion of empirical orthogonal functions and their application to vertical temperature profiles. *Tellus* **19**, 477-482.
- Braud, I., Obled, Ch. and Phamdinhtuan, A. (1993). Empirical orthogonal function (EOF) analysis of spatial random fields: Theory, accuracy of the numerical approximations and sampling effects. *Stochastic Hydrology and Hydraulics* **7**, 146-160.
- Chilès, J. P. and Delfiner, P. (1999). *Geostatistics: Modeling Spatial Uncertainty*. Wiley, New York.
- Cohen, A. and Jones, R. H. (1969). Regression on a random field. *J. Amer. Statist. Assoc.* **64**, 1172-1182.
- Cressie, N. (1993). *Statistics for Spatial Data*. Revised edition. Wiley, New York.
- Cressie, N. and Wikle, C. K. (2011). *Statistics for Spatio-Temporal Data*. Wiley, New York.
- Creutin, J. D. and Obled, C. (1982). Objective analysis and mapping techniques for rainfall fields: an objective comparison. *Water Resources Research* **18**, 413-431.
- Deville, J. C. (1974). Méthodes statistiques et numériques de l'analyse harmonique. *Annales de l'INSEE* **15**, 1-101.

- Duchon, J. (1977). Splines minimizing rotation-invariant semi-norms in Sobolev spaces. *Constructive Theory of Functions of Several Variables* (Edited by W. Schempp and Z. Keller), 85-100. Springer-Verlag, Berlin-Heidelberg.
- Furrer, R., Genton, M. G. and Nychka, D. (2006). Covariance tapering for interpolation of large spatial datasets. *J. Comput. Graph. Statist.* **15**, 502-523.
- Green, P. J. and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*. Chapman & Hall.
- Hannachi, A., Jolliffe, I. T. and Stephenson, D. B. (2007). Empirical orthogonal functions and related techniques in atmospheric science: a review. *Internat. J. Climatology* **27**, 1119-1152.
- Holmström, I. (1963). On a method for parametric representation of the state of the atmosphere. *Tellus* **15**, 127-149.
- Hörmann, S. and Kokoszka, P. (2010). Weakly dependent functional data. *Ann. Statist.* **38**, 1845-1884.
- Chang, Y.-M., Hsu, N.-J. and Huang, H.-C. (2010). Semiparametric estimation and selection for nonstationary spatial covariance functions. *J. Comput. Graph. Statist.* **19**, 117-139.
- Huang, J., Shen, H. and Buja, A. (2008). Functional principal components analysis via penalized rank one approximation. *Electronic J. Statist.* **2**, 678-695.
- Jolliffe, I. T. (2002). *Principal Component Analysis*. 2nd edition. Springer, New York.
- Journel, A. G. (1980). The lognormal approach to predicting local distribution of selective mining unit grades. *Math. Geology* **12**, 285-303.
- Karhunen, K. (1947). Über lineare Methoden in der Wahrscheinlichkeitsrechnung. *Annales Academiae Scientiarum Fennicae* **37**, 1-79.
- Karl, T. R., Koscielny, A. J. and Diaz, H. F. (1982). Potential errors in the application of principal component (eigenvector) analysis to geophysical data. *J. Appl. Meteorology* **21**, 1183-1186.
- Loève, M. (1978). *Probability Theory*. Springer-Verlag, New York.
- Lorenz, E. N. (1956). Empirical orthogonal functions and statistical weather prediction. Scientific Report No. 1, Statistical Forecasting Project, Department of Meteorology, Massachusetts Institute of Technology.
- Meng, X.-L. (1994). On the Rate of Convergence of the ECM Algorithm. *Ann. Statist.* **22**, 326-339.
- Meng, X.-L. and Rubin, D. B. (1993). Maximum likelihood estimation via the ECM algorithm: a general framework. *Biometrika* **80**, 267-278.
- Nott, D. J. and Dunsmuir, T. M. (2002). Estimation of nonstationary spatial covariance structure. *Biometrika* **89**, 819-829.
- Obled, C. and Creutin, J. D. (1986). Some developments in the use of empirical orthogonal functions for mapping meteorological fields. *J. Climate and Applied Meteorology* **25**, 1189-1204.
- Obukhov, A. M. (1960). The statistically orthogonal expansion of empirical functions. *Bulletin of the Academy of Sciences of the U.S.S.R.: Geophysics Series* **1**, 288-291.
- Preisendorfer, R. W. (1988). *Principal Component Analysis in Meteorology and Oceanography*. Elsevier, Amsterdam.
- Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *J. Roy. Statist. Soc. Ser. B* **58**, 267-288.

- Wahba, G. (1990). *Spline Models for Observational Data*. Society for Industrial and Applied Mathematics, Philadelphia.
- Wikle, C. K. and Cressie, N. (1999). A dimension-reduced approach to space-time Kalman filtering. *Biometrika* **86**, 815-829.
- Wilks, D. S. (1995). *Statistical Methods in the Atmospheric Sciences*. Academic Press, San Diego.
- Yao, F., Müller, H. G. and Wang, J. L. (2005). Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc.* **100**, 577-590.

Institute of Statistics, National Central University, Taoyuan 32049, Taiwan.

E-mail: ypingchn@gmail.com

Institute of Statistical Science, Academia Sinica, Taipei 11529, Taiwan.

E-mail: hchuang@stat.sinica.edu.tw

(Received October 2011; accepted September 2012)