Statistica Sinica **25** (2015), 151-171
doi:http://dx.doi.org/10.5705/ss.2013.211w

NON-STATIONARY MULTIVARIATE SPATIAL COVARIANCE ESTIMATION VIA LOW-RANK REGULARIZATION

ShengLi Tzeng and Hsin-Cheng Huang

China Medical University and Academia Sinica

Abstract: We introduce a regularization approach to multivariate spatial covariance estimation based on a spatial random effect model. The proposed method is flexible to incorporate not only spatial non-stationarity but also asymmetry in spatial cross-covariances. By introducing a regularization term in the objective function, our method automatically produces a low-rank covariance estimate that effectively controls estimation variability even when the number of parameters is large. In addition, we offer a computationally efficient method for solving the regularization problem and obtaining the optimal spatial predictions that require no high-dimensional matrix inversion. Some numerical examples are provided to demonstrate the effectiveness of the proposed method.

Key words and phrases: Fixed rank kriging, large data, spatial prediction.

1. Introduction

Multivariate spatial data are more and more common. Multivariate geostatistical models are increasingly required in many environmental, atmospheric, and geophysical sciences. In many situations, data are sampled at several locations and, at each location, multiple variables are observed. Not only does each variable exhibit some spatial dependence, but spatial interactions among different variables are often seen. Clearly, it is of interest to develop a flexible multivariate geostatistical model to account for spatial covariances among observations of individual variables, and spatial cross-covariances among observations of different variables. Consider a p-variate spatial process $\{\mathbf{y}(\mathbf{s}) : \mathbf{s} \in D\}$ defined on a spatial domain $D \subset \mathbb{R}^d$, where $\mathbf{y}(\mathbf{s}) = (y_1(\mathbf{s}), \ldots, y_p(\mathbf{s}))'$ and d is a positive integer. The central issue of multivariate geostatistical models lies in modeling the spatial cross-covariance function:

$$C_{ij}(\mathbf{s}, \mathbf{s}^*) \equiv \mathrm{cov}(y_i(\mathbf{s}), y_j(\mathbf{s}^*)); \quad i, j = 1, \ldots, p, \, \mathbf{s}, \mathbf{s}^* \in D.$$

Many stationary multivariate spatial cross-covariance estimation methods have been developed. A popular construction strategy is the linear model of coregionalization (Goulard and Voltz (1992); Schmidt and Gelfand (2003);

Marchant and Lark (2007); Zhang (2007)), which forms a stationary multivariate process as a sum of several basic univariate processes. Apanasovich and Genton (2010) developed a latent-dimension approach. Recently, a parametric multivariate Matérn covariance model (Gneiting, Kleiber and Schlather (2010); Apanasovich, Genton and Sun (2012)) was developed with its cross-covariances given by

$$C_{ij}^{M}(\mathbf{s}-\mathbf{s}^{*}) \equiv \operatorname{cov}(y_i(\mathbf{s}), y_j(\mathbf{s}^{*})) = \sigma_{ij} M(\mathbf{s}-\mathbf{s}^{*}; \nu_{ij}, \alpha_{ij}), \tag{1.1}$$

where

$$M(\mathbf{h}; \nu, \alpha) = \frac{1}{2^{\nu-1}\Gamma(\nu)} (\alpha\|\mathbf{h}\|)^{\nu} K_{\nu}(\alpha\|\mathbf{h}\|); \quad \mathbf{h} \in \mathbb{R}^{d},$$

and K_ν is a modified Bessel function of the second kind with order ν. The σ_{ij}'s, ν_{ij}'s, and α_{ij}'s have to satisfy some conditions to ensure a valid cross-covariance model. An approach to develop an asymmetric multivariate spatial model from a symmetric multivariate spatial model was provided by Li and Zhang (2011).

On the other hand, many non-stationary multivariate spatial models have also been developed. For example, Gelfand et al. (2004) incorporated non-stationarity via spatial varying coregionalizations. Kleiber and Nychka (2012) allowed the parameters of the multivariate Matérn covariance class to vary with locations. Approaches based on factor analysis can be found in Wang and Wall (2003); Christensen and Amemiya (2003); Krzanowski and Bailey (2007); Furrer and Genton (2011). Recently, a multivariate version of predictive processes (Banerjee et al. (2008)) was carried out by Ren and Banerjee (2013). However, neither of these methods is computationally very efficient. Kernel convolution is another widely used approach (Ver Hoef and Barry (1998); Ver Hoef, Cressie and Barry (2004); Majumdar and Gelfand (2007)). Although these methods require no matrix inversion, their performance depends on locations of knots and the form of the kernel, and choices of them are not clear (see e.g., Fanshawe and Diggle (2012)).

Multivariate geostatistical methods that are both flexible and computationally fast are still in need of improvement. In this article, we develop a non-stationary multivariate spatial covariance estimation method in a spatial random effect model framework (Banerjee, Carlin and Gelfand (2004)). We propose to estimate the model parameters via a regularization approach when a sample covariance matrix corresponding to some spatial locations is available. Our method in the univariate case can be regarded as a regularized version of fixed rank kriging (Cressie and Johannesson (2008)). The regularization helps to reduce estimation variability effectively, which in turn allows for a more flexible model with a larger number of parameters. The proposed method is able to well approximate many non-separable, asymmetric, or non-stationary multivariate spatial covariance functions. It is computationally fast, avoiding the direct inversion of a high-dimensional matrix.

The rest of the paper is organized as follows. Section 2 introduces our model and the proposed regularization method for univariate processes, which is then extended to multivariate processes in Section 3. Some simulation examples are described in Section 4.

2. Univariate Models

Consider a sequence of independent spatial processes, $\{y(\mathbf{s},t) : \mathbf{s} \in D\}$, $t = 1,\ldots,T$, defined on a d-dimensional spatial domain $D \subset \mathbb{R}^d$, $T > 1$. The processes are assumed to have a common spatial covariance function, $C(\mathbf{s},\mathbf{s}^*) = \mathrm{cov}(y(\mathbf{s},t), y(\mathbf{s}^*,t))$. Suppose that we observe data $\mathbf{z}_t \equiv (z(\mathbf{s}_1,t),\ldots,z(\mathbf{s}_n,t))'$, $t = 1,\ldots,T$, at $\mathbf{s}_1,\ldots,\mathbf{s}_n \in D$ with additive noise $\boldsymbol{\varepsilon}_t$ according to

$$\mathbf{z}_t = \mathbf{y}_t + \boldsymbol{\varepsilon}_t; \quad t = 1,\ldots,T, \tag{2.1}$$

where $\mathbf{y}_t \equiv (y(\mathbf{s}_1,t),\ldots,y(\mathbf{s}_n,t))'$, $\boldsymbol{\varepsilon}_t \sim \left(\mathbf{0},\sigma^2\mathbf{I}_n\right)$ is uncorrelated with $\mathbf{y}_t$, and the $\boldsymbol{\varepsilon}_t$'s are mutually uncorrelated. We assume that σ^2 is known. The goal is to estimate $C(\cdot,\cdot)$ based on the data $\mathbf{z}_1,\ldots,\mathbf{z}_T$, from which the best linear unbiased prediction of $\{y(\mathbf{s},t) : \mathbf{s} \in D\}$ for $t = 1,\ldots,T$, can be obtained. Here a stationary assumption is not made nor is a parametric structure is assumed for $C(\cdot,\cdot)$. Throughout the paper, the mean of $y(\cdot,\cdot)$ is assumed known and, without loss of generality, to be 0.

2.1. The proposed model

We consider a spatial random effect model for the latent process $y(\cdot,\cdot)$:

$$y(\mathbf{s},t) = \mathbf{w}_t'\mathbf{f}(\mathbf{s}) + \xi(\mathbf{s},t) = \sum_{k=1}^{K} w_k(t) f_k(\mathbf{s}) + \xi(\mathbf{s},t), \tag{2.2}$$

where $\mathbf{f}(\mathbf{s}) = (f_1(\mathbf{s}),\ldots,f_K(\mathbf{s}))'$ consists of $K \le n$ known basis functions, $f_1(\cdot),\ldots,f_K(\cdot)$ such that $\mathbf{F} \equiv (\mathbf{f}(\mathbf{s}_1),\ldots,\mathbf{f}(\mathbf{s}_n))'$ is an $n \times K$ matrix of rank K, $\mathbf{w}_t = (w_1(t),\ldots,w_K(t))'$, $t = 1,\ldots,T$, are uncorrelated random effects with $\mathrm{E}(\mathbf{w}_t) = \mathbf{0}$ and $\mathrm{var}(\mathbf{w}_t) = \mathbf{M}$, and $\xi(\cdot,\cdot)$ is a white-noise process with variance v^2. The $\xi(\cdot,\cdot)$ process captures the variation not explained by $\mathbf{w}_t'\mathbf{f}(\mathbf{s})$ when K is small. It follows that $\mathrm{var}(\mathbf{y}_t) = \mathbf{FMF}' + v^2\mathbf{I}_n$ and $\mathbf{V} \equiv \mathrm{var}(\mathbf{z}_t) = \mathbf{FMF}' + v^2\mathbf{I}_n + \sigma^2\mathbf{I}_n$, for $t = 1,\ldots,T$. The parameters that need to be estimated are $\mathbf{M}$ and v^2, where $\mathbf{M}$ is required to be non-negative definite, denoted by $\mathbf{M} \succeq \mathbf{0}$.

2.2. Parameter estimation

With $\mathbf{Z} = (\mathbf{z}_1,\ldots,\mathbf{z}_T)$, the sample covariance matrix $\mathbf{S} = \mathbf{ZZ}'/T$ is an unbiased estimate of $\mathbf{V}$, but a poor estimate of $\mathbf{V}$ when n is large or T is small. Although the proposed model helps to reduce estimation variability, the unknown

parameters in $\mathbf{M}$ still have high complexity unless K is very small. To reduce the estimation variability further while suitably controlling the bias, we consider a regularization approach that has been proven effective in a wide range of statistical problems; see Bickel and Li (2006) and the references therein. We propose to estimate $\mathbf{M}$ and v^2 by regularizing the eigenvalues of $\mathbf{FMF}'$ in terms of the objective function,

$$\phi(\mathbf{M}, v^2) = \frac{1}{2}\left\|\mathbf{FMF}' + v^2\mathbf{I}_n + \sigma^2\mathbf{I}_n - \mathbf{S}\right\|_F^2 + \tau\left\|\mathbf{FMF}'\right\|_*, \tag{2.3}$$

over all $\mathbf{M} \succeq \mathbf{0}$ and $v^2 \geq 0$, where $\|\mathbf{X}\|_F^2 = \operatorname{trace}(\mathbf{X}'\mathbf{X})$ is the squared Frobenius norm of $\mathbf{X}$, and $\|\mathbf{X}\|_* = \operatorname{trace}\big((\mathbf{X}'\mathbf{X})^{1/2}\big)$ is the nuclear norm of $\mathbf{X}$. As the penalty $\|\mathbf{FMF}'\|_*$ is the sum of the eigenvalues of $\mathbf{FMF}'$, it shrinks eigenvalues and forces some small eigenvalues to zero. The minimizers of (2.3) are denoted as $\big(\hat{\mathbf{M}}_\tau, \hat{v}_\tau^2\big)$.

Let $\mathbf{H}_F = \mathbf{F}(\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'$, and let $\mathbf{QDQ}'$ be the eigen-decomposition of $\mathbf{H}_F(\mathbf{S} - \sigma^2\mathbf{I}_n)\mathbf{H}_F$, where $\mathbf{D} = \operatorname{diag}(d_1, \ldots, d_n)$ is the diagonal matrix with diagonal elements $d_1, \ldots, d_n$, and $|d_1| \geq \cdots \geq |d_{K^*}| > 0 = |d_{K^*+1}| = \cdots = |d_n|$. Also, let $\mathbf{Q}_{K^*}$ be the submatrix of $\mathbf{Q}$ consisting of its first K^* columns. Then $\mathbf{QDQ}' = \mathbf{Q}_{K^*}\mathbf{D}_{K^*}\mathbf{Q}'_{K^*}$, where $\mathbf{D}_{K^*} = \operatorname{diag}(d_1, \ldots, d_{K^*})$. Hence $\mathbf{Q}_{K^*} = \mathbf{H}_F(\mathbf{S} - \sigma^2\mathbf{I}_n)\mathbf{H}_F\mathbf{Q}_{K^*}\mathbf{D}_{K^*}^{-1}$, and

$$\mathbf{H}_F\mathbf{Q}_{K^*} = \mathbf{Q}_{K^*}\,. \tag{2.4}$$

As shown below with its proof given in the Appendix, $\hat{\mathbf{M}}_\tau$ and $\hat{v}_\tau^2$ can be expressed in closed form.

Proposition 1. *Let $\mathbf{H}_F = \mathbf{F}(\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'$ and $\mathbf{QDQ}'$ be the eigen-decomposition of $\mathbf{H}_F(\mathbf{S} - \sigma^2\mathbf{I}_n)\mathbf{H}_F$, where $\mathbf{D} = \operatorname{diag}(d_1, \ldots, d_n)$ and $|d_1| \geq \cdots \geq |d_{K^*}| > 0 = |d_{K^*+1}| \cdots = |d_n|$. Then for any $\tau \geq 0$, the minimizers of (2.3) are*

$$\hat{v}_\tau^2 = \operatorname*{arg\,min}_{v^2 \geq 0}\Big\{v^2(nv^2 - 2\operatorname{trace}(\mathbf{S} - \sigma^2\mathbf{I}_n)) - \sum_{k=1}^{K^*}(d_k - \tau - v^2)_+^2\Big\}, \tag{2.5}$$

$$\hat{\mathbf{M}}_\tau = (\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'\mathbf{Q}_{K^*}\operatorname{diag}\left((d_1 - \tau - \hat{v}_\tau^2)_+, \ldots, (d_{K^*} - \tau - \hat{v}_\tau^2)_+\right)\mathbf{Q}'_{K^*}\mathbf{F}(\mathbf{F}'\mathbf{F})^{-1}, \tag{2.6}$$

where $x_+ \equiv \max(x, 0)$.

The $\hat{v}_\tau^2$ in (2.5) can be efficiently computed by a golden section search in $[0,\ d_1]$ when $d_1, \ldots, d_{K^*}$ are available. As shown below, there is a more efficient way to compute $d_1, \ldots, d_{K^*}$ and $\hat{\mathbf{M}}_\tau$ in (2.6) without the need to go through the eigen-decomposition of $\mathbf{H}_F(\mathbf{S} - \sigma^2\mathbf{I}_n)\mathbf{H}_F$. Its proof is in the Appendix.

Corollary 1. *Let* $\mathbf{L} = \mathbf{F}(\mathbf{F}'\mathbf{F})^{-1/2}$. *Under the conditions of Proposition* 1, *the eigen-decomposition of* $\mathbf{L}'(\mathbf{S} - \sigma^2\mathbf{I}_n)\mathbf{L}$ *can be written as* $\mathbf{P}\mathbf{D}_K\mathbf{P}'$ *for some orthogonal matrix* $\mathbf{P}$, *where* $\mathbf{D}_K = \mathrm{diag}(d_1, \ldots, d_K)$. *In addition,* $\mathbf{Q}_{K^*} = \mathbf{L}\mathbf{P}_{K^*}$ *and*

$$\hat{\mathbf{M}}_\tau = (\mathbf{F}'\mathbf{F})^{-1/2}\mathbf{P}_{K^*}\,\mathrm{diag}\left((d_1 - \tau - \hat{v}_\tau^2)_+, \ldots, (d_{K^*} - \tau - \hat{v}_\tau^2)_+\right)\mathbf{P}'_{K^*}(\mathbf{F}'\mathbf{F})^{-1/2}, \tag{2.7}$$

where $\mathbf{P}_{K^*}$ *is the submatrix of* $\mathbf{P}$ *consisting of its first* K^* *columns.*

Here $\mathbf{L}'(\mathbf{S} - \sigma^2\mathbf{I}_n)\mathbf{L} = (\mathbf{L}'\mathbf{Z})(\mathbf{Z}'\mathbf{L})/T - \sigma^2\mathbf{L}'\mathbf{L}$ can be obtained in $O(KnT) + O(K^2n)$ computations, from which we can compute $\mathbf{P}$, $\mathbf{Q}_{K^*}$ and $\hat{\mathbf{M}}_\tau$ in $O(K^2n)$ operations. The overall computational complexity for parameter estimation is $O(KnT) + O(K^2n)$.

2.3. Spatial prediction

Suppose $\mathbf{M}$ is estimated by $\hat{\mathbf{M}}_\tau$ and v^2 is estimated by $\hat{v}_\tau^2$ for some $\tau \geq 0$. Then for $\mathbf{s} \in D$ and $t = 1, \ldots, T$, the estimated best linear unbiased predictor (EBLUP) of $y(\mathbf{s}, t)$ and the corresponding estimated mean squared prediction error (MSPE) are given by

$$\hat{y}(\mathbf{s}, t) = \left(\mathbf{f}(\mathbf{s})'\hat{\mathbf{M}}_\tau\mathbf{F}' + \hat{v}_\tau^2\boldsymbol{\delta}'\right)\hat{\mathbf{V}}_\tau^-\mathbf{z}_t, \tag{2.8}$$

$$\begin{aligned}\widehat{\mathrm{var}}\left(\hat{y}(\mathbf{s}, t) - y(\mathbf{s}, t)\right) &= \mathbf{f}(\mathbf{s})'\hat{\mathbf{M}}_\tau\mathbf{f}(\mathbf{s}) + \hat{v}_\tau^2 \\ &\quad - \left(\mathbf{f}(\mathbf{s})'\hat{\mathbf{M}}_\tau\mathbf{F}' + \hat{v}_\tau^2\boldsymbol{\delta}'\right)\hat{\mathbf{V}}_\tau^-\left(\mathbf{F}\hat{\mathbf{M}}_\tau\mathbf{f}(\mathbf{s}) + \hat{v}_\tau^2\boldsymbol{\delta}\right),\end{aligned} \tag{2.9}$$

where $\boldsymbol{\delta} \equiv (I(\mathbf{s} = \mathbf{s}_1), \ldots, I(\mathbf{s} = \mathbf{s}_n))'$ and $\hat{\mathbf{V}}_\tau = \mathbf{F}\hat{\mathbf{M}}_\tau\mathbf{F}' + \hat{v}_\tau^2\mathbf{I}_n + \sigma^2\mathbf{I}_n$.

As shown below, direct computation of $\mathbf{V}_\tau^-$ can be avoided. From (2.4) and (2.6), we have

$$\mathbf{F}\hat{\mathbf{M}}_\tau\mathbf{F}' = \mathbf{Q}_{K^*}\,\mathrm{diag}\left((d_1 - \tau - \hat{v}_\tau^2)_+, \ldots, (d_{K^*} - \tau - \hat{v}_\tau^2)_+\right)\mathbf{Q}'_{K^*}, \tag{2.10}$$

which implies

$$\hat{\mathbf{V}}_\tau = \mathbf{Q}_{K^*}\,\mathrm{diag}\big((d_1 - \tau - \hat{v}_\tau^2)_+, \ldots, (d_{K^*} - \tau - \hat{v}_\tau^2)_+\big)\mathbf{Q}'_{K^*} + \hat{v}_\tau^2\mathbf{I}_n + \sigma^2\mathbf{I}_n.$$

Applying the Sherman-Morrison-Woodbury formula (Woodbury (1950)), the Moore-Penrose inverse of $\hat{\mathbf{V}}_\tau$ is

$$\begin{cases} \dfrac{1}{\hat{v}_\tau^2 + \sigma^2}\mathbf{I}_n - \dfrac{1}{\hat{v}_\tau^2 + \sigma^2}\mathbf{Q}_{K^*}\mathrm{diag}\left(\dfrac{\hat{d}_1}{\hat{d}_1 + \hat{v}_\tau^2 + \sigma^2}, \ldots, \dfrac{\hat{d}_{K^*}}{\hat{d}_{K^*} + \hat{v}_\tau^2 + \sigma^2}\right)\mathbf{Q}'_{K^*}; \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if } \hat{v}_\tau^2 + \sigma^2 > 0, \\ \mathbf{Q}_{K^*}\left(\mathrm{diag}\big(\hat{d}_1, \ldots, \hat{d}_{K^*}\big)\right)^-\mathbf{Q}'_{K^*}; \qquad\qquad\qquad\qquad\qquad\qquad \text{if } \hat{v}_\tau^2 + \sigma^2 = 0, \end{cases}$$

where $\hat{d}_k \equiv (d_k - \tau - \hat{v}_\tau^2)_+$; $k = 1, \ldots K^*$. Using this form for $\hat{\mathbf{V}}_\tau^-$, we can compute $\hat{y}(\mathbf{s}, t)$ and $\widehat{\text{var}}\,(\hat{y}(\mathbf{s}, t) - y(\mathbf{s}, t))$ in only $O(K^2 n)$ computations after $\mathbf{Q}_{K^*}$ is obtained. Therefore, $\hat{y}(\mathbf{s}, t)$ and $\widehat{\text{var}}\,(\hat{y}(\mathbf{s}, t) - y(\mathbf{s}, t))$, including parameter estimation, can be computed in $O(KnT) + O(K^2 n)$ operations. The computational complexity is largely reduced, particularly when $K << n$.

When σ^2 is unknown, we propose to estimate it by minimizing $\|\mathbf{FMF}' + \sigma^2 \mathbf{I}_n - \mathbf{S}\|_F^2$ over all $\mathbf{M} \succeq \mathbf{0}$. Applying arguments similar to those in proofs of Proposition 1 and Corollary 1, we obtain

$$\hat{\sigma}^2 = \arg\min_{\sigma^2 \geq 0} \frac{1}{2} \Big\{ \sigma^2 (n\sigma^2 - 2\,\text{trace}(\mathbf{S})) - \sum_{k=1}^{K} (\gamma_k - \sigma^2)_+^2 \Big\}, \tag{2.11}$$

where γ_k's are the eigenvalues of $(\mathbf{F}'\mathbf{F})^{-1/2}\mathbf{F}'\mathbf{S}\mathbf{F}(\mathbf{F}'\mathbf{F})^{-1/2}$ with $\gamma_1 \geq \cdots \geq \gamma_K$.

2.4. Choice of the regularization parameter

An L-fold cross-validation is applied to choose the regularization parameter τ. We first randomly decompose $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$ into L disjoint subsets $\mathcal{D}_1, \ldots, \mathcal{D}_L$ with $|\mathcal{D}_1|, \ldots, |\mathcal{D}_L|$ being as close to equal as possible. Let $\hat{y}^{(\ell)}(\mathbf{s}, t; \tau)$ be the EBLUP of $y(\mathbf{s}, t)$ based on only the data observed at $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\} \setminus \mathcal{D}_\ell$ for $\mathbf{s} \in D$, $t = 1, \ldots, T$ and $\ell = 1, \ldots, L$. The proposed L-fold CV criterion is

$$\text{CV}(\tau) = \sum_{\ell=1}^{L} \sum_{t=1}^{T} \sum_{\mathbf{s} \in \mathcal{D}_\ell} \Big\{ z(\mathbf{s}, t) - \hat{y}^{(\ell)}(\mathbf{s}, t; \tau) \Big\}^2 .$$

The final regularization parameter is $\hat{\tau}$, which minimizes $\text{CV}(\tau)$ over $\tau \geq 0$.

3. Multivariate Models

Consider a sequence of p-variate spatial processes, $\{\mathbf{y}(\mathbf{s}, t);\ \mathbf{s} \in D\}$ that are independent for $t = 1, \ldots, T$, and defined on a d-dimensional spatial domain $D \subset \mathbb{R}^d$, with $\mathbf{y}(\mathbf{s}, t) \equiv (y_1(\mathbf{s}, t), \ldots, y_p(\mathbf{s}, t))'$. The processes are assumed to have mean zero and a common spatial cross-covariance function $C_{ij}(\mathbf{s}, \mathbf{s}^*) = \text{cov}\,(y_i(\mathbf{s}, t),\ y_j(\mathbf{s}^*, t))$; $i, j = 1, \ldots, p$, $\mathbf{s}, \mathbf{s}^* \in D$. Suppose we observe data $\{\mathbf{z}(\mathbf{s}_i, t) : i = 1, \ldots, n,\ t = 1, \ldots, T\}$ at n locations $\mathbf{s}_1, \ldots, \mathbf{s}_n \in D$ with additive noise $\{\boldsymbol{\varepsilon}(\mathbf{s}_i, t)\}$ according to

$$\mathbf{z}(\mathbf{s}_i, t) = \mathbf{y}(\mathbf{s}_i, t) + \boldsymbol{\varepsilon}(\mathbf{s}_i, t); \quad i = 1, \ldots, n,\ t = 1, \ldots, T, \tag{3.1}$$

where $\boldsymbol{\varepsilon}(\mathbf{s}_i, t) \sim (\mathbf{0}, \boldsymbol{\Sigma})$ is uncorrelated with $\mathbf{y}(\mathbf{s}_i, t)$, the $\boldsymbol{\varepsilon}(\mathbf{s}_i, t)$'s are mutually uncorrelated, $\boldsymbol{\Sigma} = \text{diag}(\sigma_1^2, \ldots, \sigma_p^2)$, and $\sigma_j^2 \geq 0$; $j = 1, \ldots, p$. As in the univariate model, $\boldsymbol{\Sigma}$ is assumed known. The goal is to estimate $C_{ij}(\cdot, \cdot)$'s based on the data $\mathbf{Z}_t \equiv (\mathbf{z}(\mathbf{s}_1, t), \ldots, \mathbf{z}(\mathbf{s}_n, t))'$; $t = 1, \ldots, T$.

3.1. The proposed model

As to (2.2), we propose a spatial random effect model for the latent process $\mathbf{y}(\cdot,\cdot)$ in terms of $K \le n$ known basis functions, $f_1(\cdot),\ldots,f_K(\cdot)$:

$$\mathbf{y}(\mathbf{s},t) = \mathbf{W}_t'\mathbf{f}(\mathbf{s}) + \xi(\mathbf{s},t) = \sum_{k=1}^{K} f_k(\mathbf{s})\mathbf{w}_k(t) + \xi(\mathbf{s},t), \tag{3.2}$$

where $\mathbf{w}_k(t) = (w_{k1}(t),\ldots,w_{kp}(t))'$, $\mathbf{W}_t \equiv (\mathbf{w}_1(t),\ldots,\mathbf{w}_K(t))'$, $t = 1,\ldots,T$, are $K \times p$ uncorrelated random matrices with $\mathrm{E}(\mathbf{W}_t) = \mathbf{0}$ and $\mathrm{var}(\mathrm{vec}(\mathbf{W}_t)) = \mathbf{M}$, and $\xi(\mathbf{s},t) \sim (\mathbf{0}, v^2\mathbf{I}_p)$; $\mathbf{s} \in D$, $t = 1,\ldots,T$.

For $t = 1,\ldots,T$, let $\mathbf{Y}_t \equiv (\mathbf{y}(\mathbf{s}_1,t),\ldots,\mathbf{y}(\mathbf{s}_n,t))'$ and $\xi_t \equiv (\xi(\mathbf{s}_1,t),\ldots,\xi(\mathbf{s}_n,t))'$. Then $\mathbf{Y}_t = \mathbf{F}\mathbf{W}_t + \xi_t$, $\mathrm{var}(\mathrm{vec}(\mathbf{Y}_t)) = (\mathbf{I}_p \otimes \mathbf{F})\mathbf{M}(\mathbf{I}_p \otimes \mathbf{F}') + v^2\mathbf{I}_{np}$, and

$$\mathbf{V} \equiv \ \mathrm{var}(\mathrm{vec}(\mathbf{Z}_t)) = (\mathbf{I}_p \otimes \mathbf{F})\mathbf{M}(\mathbf{I}_p \otimes \mathbf{F}') + v^2\mathbf{I}_{np} + \mathbf{\Sigma} \otimes \mathbf{I}_n.$$

The parameters that need to be estimated are $\mathbf{M}$ and v^2, where $\mathbf{M} \succeq \mathbf{0}$. Here $C_{ij}(\mathbf{s},\mathbf{s}^*) = \mathbf{f}(\mathbf{s})'\mathbf{M}_{ij}\mathbf{f}(\mathbf{s}^*) + v^2 I(\mathbf{s} = \mathbf{s}^*)$, and

$$\mathbf{M}_{ij} = \mathrm{cov}\big((w_{1i}(t),\ldots,w_{Ki}(t))',\ (w_{1j}(t),\ldots,\ w_{Kj}(t))'\big), \tag{3.3}$$

is the (i,j)th sub-block of $\mathbf{M}$ of dimension $K \times K$.

The proposed model is flexible enough to approximate many non-stationary spatial covariance functions well, and it allows asymmetric cross-covariance function, $C_{ij}(\mathbf{s},\mathbf{s}^*) \neq C_{ij}(\mathbf{s}^*,\mathbf{s})$ when $\mathbf{M}_{ij} \neq \mathbf{M}_{ji}$ for some $i \neq j$. For example, as shown in Figure 1, it well approximates a multivariate stationary Matérn covariance function with $\begin{pmatrix}\sigma_{11}\ \sigma_{12}\\ \sigma_{21}\ \sigma_{22}\end{pmatrix} = \begin{pmatrix}1 & 0.4\\ 0.4 & 1\end{pmatrix}$, $\begin{pmatrix}\alpha_{11}\ \alpha_{12}\\ \alpha_{21}\ \alpha_{22}\end{pmatrix} = \begin{pmatrix}0.5 & 1.6\\ 1.6 & 2\end{pmatrix}$ and $\nu_{11} = \nu_{22} = \nu_{12} = 0.5$, when K is only moderately large, where the $f_k(\cdot)$'s are cubic B-splines with equally spaced knots at $\big(0,0,0,0,1/(K-3),\ldots,(K-4)/(K-3),1,1,1,1\big)$, and the $(\mathbf{M},v^2)$ of the approximated covariance function are obtained by minimizing

$$\sum_{1\le i,j\le 2}\int_0^1\int_0^1 \big|C_{ij}(s,s^*;\mathbf{M},v^2) - C_{ij}^M(s-s^*)\big|^2 ds ds^*. \tag{3.4}$$

Another bivariate example of $C_{ij}(\mathbf{s},\mathbf{s}^*) = \mathbf{f}(\mathbf{s})'\mathbf{M}_{ij}\mathbf{f}(\mathbf{s}^*)$ is based on the same cubic B-splines with $K = 20$, where the (k,k^*)th elements of $\mathbf{M}_{12}$, $\mathbf{M}_{21}$ and $\mathbf{M}_{11} = \mathbf{M}_{22}$ are $\exp(-0.1|k-k^*-\rho|)$, $\exp(-0.1|k-k^*+\rho|)$ and $\exp(-0.1|k-k^*|)$, respectively. Here $\rho \in \mathbb{R}$ is a tuning parameter such that a large value of $|\rho|$ corresponding to a higher degree of asymmetry. The cross-covariance functions based on different ρ values are shown in Figure 2, where the asymmetric features are apparent.

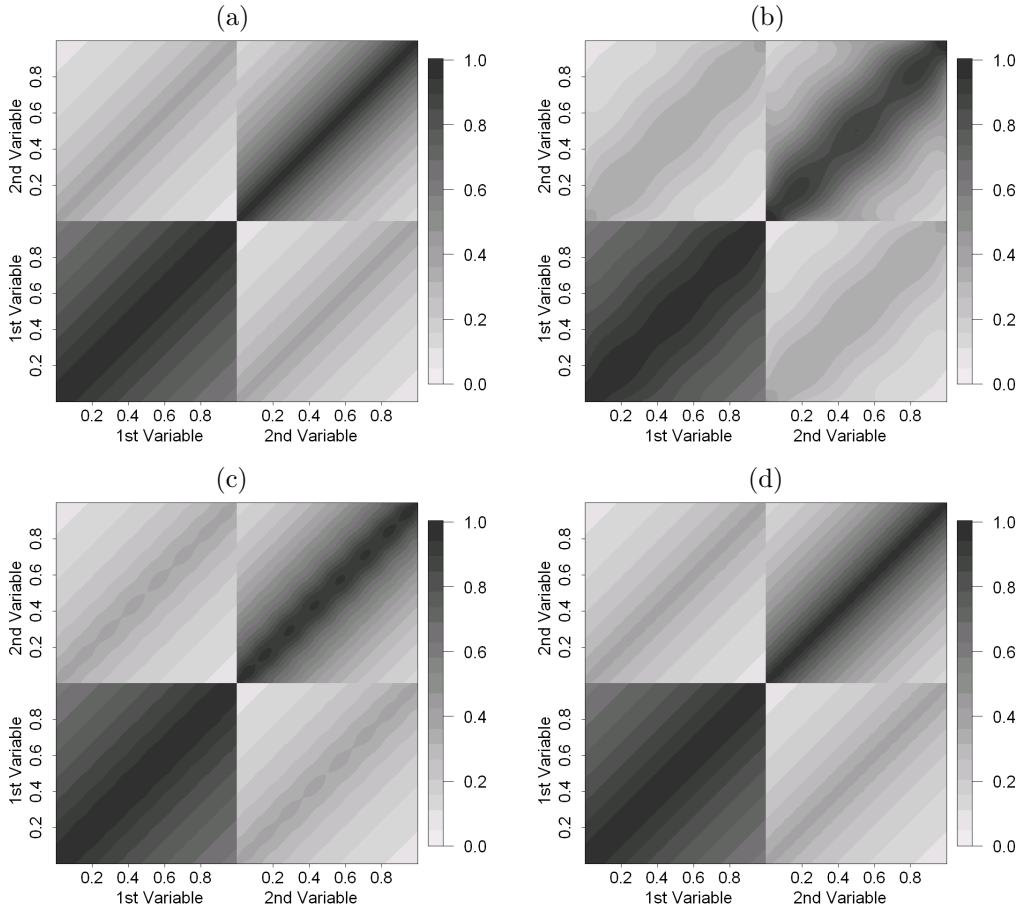

Figure 1. (a) A multivariate Matérn covariance function; (b) An approximated covariance function obtained by minimizing (3.4) with $K = 5$; (c) An approximated covariance function obtained by minimizing (3.4) with $K = 10$; (d) An approximated covariance function obtained by minimizing (3.4) with $K = 20$.

3.2. Parameter estimation

Take $T > 1$ and let $\mathbf{S} = \sum_{t=1}^{T} \mathrm{vec}(\mathbf{Z}_t)\mathrm{vec}(\mathbf{Z}_t)'/T$. We propose to estimate $\mathbf{M}$ and v^2 by minimizing the objective function,

$$\begin{aligned}\phi(\mathbf{M}, v^2) = \frac{1}{2} &\left\|(\mathbf{I}_p \otimes \mathbf{F})\mathbf{M}(\mathbf{I}_p \otimes \mathbf{F}') + v^2\mathbf{I}_{np} + \mathbf{\Sigma} \otimes \mathbf{I}_n - \mathbf{S}\right\|_F^2 \\ &+\tau \left\|(\mathbf{I}_p \otimes \mathbf{F})\mathbf{M}(\mathbf{I}_p \otimes \mathbf{F}')\right\|_*, \end{aligned} \tag{3.5}$$

over all $\mathbf{M} \succeq \mathbf{0}$ and $v^2 \geq 0$. Let $\mathbf{H}_F = \mathbf{F}(\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'$, and let $\mathbf{QDQ}$ be the eigen-decomposition of $(\mathbf{I}_p \otimes \mathbf{H}_F)(\mathbf{S} - \mathbf{\Sigma} \otimes \mathbf{I}_n)(\mathbf{I}_p \otimes \mathbf{H}_F)$, where $\mathbf{D} = \mathrm{diag}(d_1, \ldots, d_{np})$

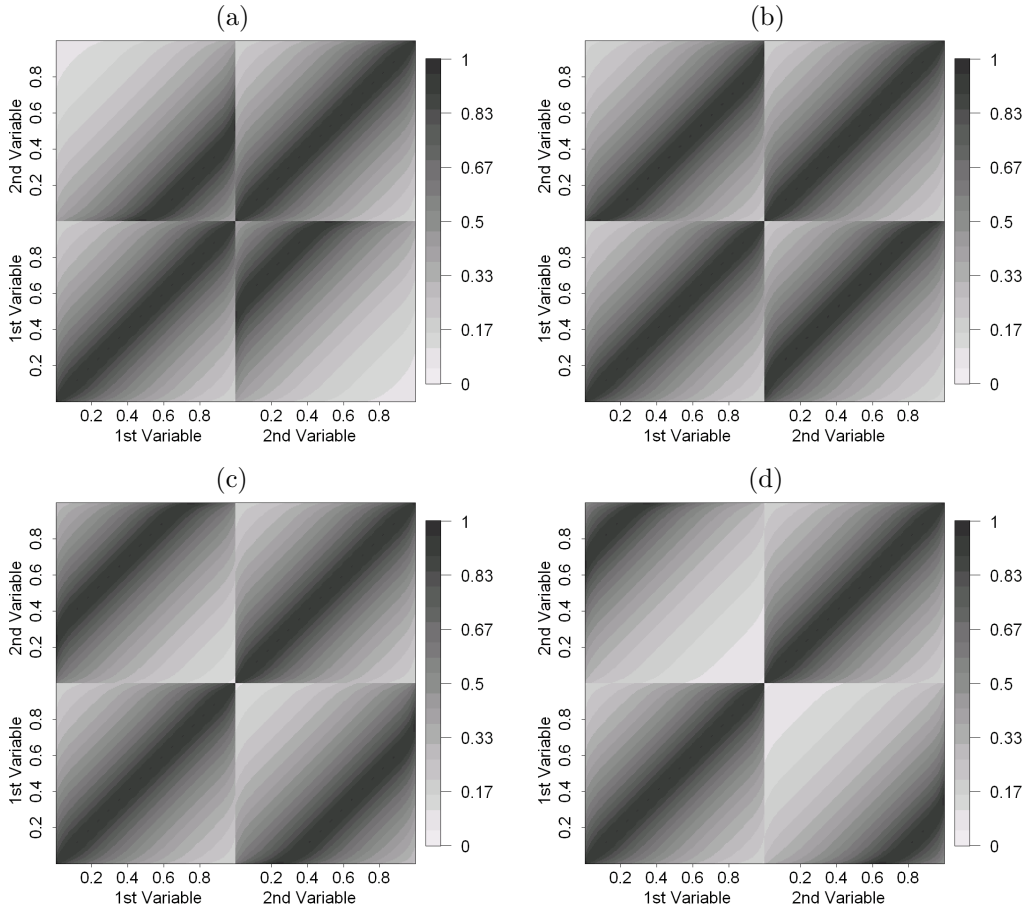

Figure 2. An asymmetric covariance function with different degrees of asymmetry. (a) $\rho = -9$; (b) $\rho = -2$; (c) $\rho = 5$; (d) $\rho = 12$.

with $|d_1| \geq \cdots \geq |d_{K^*}| > 0 = |d_{K^*+1}| = \cdots = |d_{np}|$, and $K^* \leq pK$. According to Proposition 1, the minimizers of $(\mathbf{M}, v^2)$ in (3.5) for a given τ are

$$\begin{aligned}
\hat{v}_\tau^2 = & \underset{v^2 \geq 0}{\arg\min} \frac{1}{2} \left\{ v^2(nv^2 - 2\,\mathrm{trace}(\mathbf{S} - \boldsymbol{\Sigma} \otimes \mathbf{I}_n)) - \sum_{k=1}^{K^*} (d_k - \tau - v^2)_+^2 \right\}, \\
\hat{\mathbf{M}}_\tau = & \left(\mathbf{I}_p \otimes (\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'\right) \mathbf{Q}_{K^*} \,\mathrm{diag}\left((d_1 - \tau - \hat{v}_\tau^2)_+, \ldots, (d_{K^*} - \tau - \hat{v}_\tau^2)_+\right) \mathbf{Q}_{K^*}' \\
& \times \left(\mathbf{I}_p \otimes \mathbf{F}(\mathbf{F}'\mathbf{F})^{-1}\right), \qquad (3.6)
\end{aligned}$$

where $\mathbf{Q}_{K^*}$ is the submatrix of $\mathbf{Q}$ consisting of its first K^* columns. Applying a similar trick as in the proof of Corollary 1, the eigen-decomposition of $(\mathbf{I}_p \otimes \mathbf{L}')(\mathbf{S} - \boldsymbol{\Sigma} \otimes \mathbf{I}_n)(\mathbf{I}_p \otimes \mathbf{L})$ can be written as $\mathbf{P}\mathbf{D}_{pK}\mathbf{P}'$, where $\mathbf{L} = \mathbf{F}(\mathbf{F}'\mathbf{F})^{-1/2}$ and $\mathbf{D}_{pK} = \mathrm{diag}(d_1, \ldots, d_{pK})$. Then $d_1, \ldots, d_{K^*}$ and $\mathbf{Q}^*$ can be efficiently computed

by $(\mathbf{I}_p \otimes \mathbf{L})\,\mathbf{P}_{K^*}$, where $\mathbf{P}_{K^*}$ is the submatrix of $\mathbf{P}$ consisting of its first K^* columns.

When $\boldsymbol{\Sigma}$ is unknown, we propose to estimate $\boldsymbol{\Sigma}$ by minimizing

$$\frac{1}{2}\left\|(\mathbf{I}_p \otimes \mathbf{F})\mathbf{M}(\mathbf{I}_p \otimes \mathbf{F}') + \boldsymbol{\Sigma} \otimes \mathbf{I}_n - \mathbf{S}\right\|_F^2, \tag{3.7}$$

obtained from (3.5) with $v^2 = 0$ and $\tau = 0$ over all $\mathbf{M} \succeq \mathbf{0}$ and $\sigma_j^2 \geq 0$; $j = 1, \ldots, p$. The resulting estimate is denoted by $\hat{\boldsymbol{\Sigma}}$, and can be solved iteratively. Given some initial estimate $\hat{\boldsymbol{\Sigma}}^{(0)}$, we successively compute the following for $i = 0, 1, \ldots$, until convergence:

$$\begin{aligned}\hat{\mathbf{M}}^{(i+1)} &= \left(\mathbf{I}_p \otimes (\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'\right)\mathbf{Q}^{(i)}_{K^*(i)}\,\mathrm{diag}\left(\left(d_1^{(i)}\right)_+, \ldots, \left(d^{(i)}_{K^*(i)}\right)_+\right)\left(\mathbf{Q}^{(i)}_{K^*(i)}\right)' \\ &\quad \times \left(\mathbf{I}_p \otimes \mathbf{F}(\mathbf{F}'\mathbf{F})^{-1}\right), \qquad (3.8)\\ \hat{\boldsymbol{\Sigma}}^{(i+1)} &= \frac{1}{n}\mathrm{diag}\Big(\left(\mathrm{trace}\left(\boldsymbol{\Delta}_{11}^{(i+1)}\right)\right)_+, \ldots, \left(\mathrm{trace}\left(\boldsymbol{\Delta}_{pp}^{(i+1)}\right)\right)_+\Big), \qquad (3.9)\end{aligned}$$

where

$$\begin{aligned}\boldsymbol{\Delta}^{(i+1)} &= \mathbf{S} - (\mathbf{I}_p \otimes \mathbf{F})\hat{\mathbf{M}}^{(i+1)}(\mathbf{I}_p \otimes \mathbf{F}') \\ &= \mathbf{S} - \mathbf{Q}^{(i)}_{K^*(i)}\,\mathrm{diag}\left(\left(d_1^{(i)}\right)_+, \ldots, \left(d^{(i)}_{K^*(i)}\right)_+\right)\left(\mathbf{Q}^{(i)}_{K^*(i)}\right)',\end{aligned}$$

$\mathbf{Q}^{(i)}\mathrm{diag}\left(d_1^{(i)}, \ldots, d_{np}^{(i)}\right)\mathbf{Q}^{(i)}$ is the eigen-decomposition of $(\mathbf{I}_p \otimes \mathbf{H}_F)(\mathbf{S} - \hat{\boldsymbol{\Sigma}}^{(i)} \otimes \mathbf{I}_n)(\mathbf{I}_p \otimes \mathbf{H}_F)$ with $\left|d_1^{(i)}\right| \geq \cdots \geq |d^{(i)}_{K^*(i)}| > 0 = |d^{(i)}_{K^*(i)+1}| = \cdots = \left|d_{np}^{(i)}\right|$, $\mathbf{Q}^{(i)}_{K^*(i)}$ is the submatrix of $\mathbf{Q}^{(i)}$ consisting of its first $K^*(i)$ columns, and $\boldsymbol{\Delta}^{(i)}_{jj}$ is the jth diagonal block of $\boldsymbol{\Delta}^{(i)}$ with dimension n.

For $T = 1$, we write $z_j(\mathbf{s}_i, t)$ as $z_j(\mathbf{s}_i)$. We adopt the data binning approach of Cressie and Johannesson (2008) and first partition the index set $\{1, \ldots, n\}$ into n^* subsets, $\{\mathcal{I}_1, \ldots, \mathcal{I}_{n^*}\}$ such that the sampling locations corresponding to each subset are close in space. Let $\mathbf{s}_k^* = |\mathcal{I}_k|^{-1}\sum_{i \in \mathcal{I}_k}\mathbf{s}_i$ be the representative location corresponding to the kth subset, for $k = 1, \ldots, n^*$, and let $\bar{\mathbf{Z}}$ be an $n^* \times p$ matrix with the (k, j)th element $\bar{z}_{kj} = |\mathcal{I}_k|^{-1}\sum_{i \in \mathcal{I}_k} z_j(\mathbf{s}_i)$. Following Cressie and Johannesson (2008), $\mathbf{S}^* = \mathrm{vec}(\bar{\mathbf{Z}})\mathrm{vec}(\bar{\mathbf{Z}})' + \mathrm{diag}(\mathrm{vec}(\mathbf{B}))$ is used as an initial estimate of $\mathrm{var}\left(\mathrm{vec}(\mathbf{Z}^*)\right)$, where $\mathbf{Z}^* = \left(z_j(\mathbf{s}_k^*)\right)_{n^* \times p}$, and $\mathbf{B}$ is an $n^* \times p$ matrix with its (k, j)th element $b_{kj} = |\mathcal{I}_k|^{-1}\sum_{i \in \mathcal{I}_k}(z_j(\mathbf{s}_i) - \bar{z}_{kj})^2$. Then $\mathbf{M}$, v^2, and $\boldsymbol{\Sigma}$ can be estimated in the same way as in (3.5)−(3.9), except that $\mathbf{S}$ is replaced by $\mathbf{S}^*$, n is replaced by n^*, and $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$ are replaced by $\{\mathbf{s}_1^*, \ldots, \mathbf{s}_{n^*}^*\}$.

3.3. Spatial prediction and cross validation

Suppose that $\mathbf{M}$ is estimated by $\hat{\mathbf{M}}_\tau$ and v^2 is estimated by $\hat{v}_\tau^2$ for some $\tau \geq 0$. Then, for $\mathbf{s} \in D$, the EBLUP of $\mathbf{y}(\mathbf{s}, t)$ and the corresponding estimated MSPE are

$$\begin{aligned}\hat{\mathbf{y}}(\mathbf{s},t) &= \big\{(\mathbf{I}_p \otimes \mathbf{f}(\mathbf{s})')\hat{\mathbf{M}}_\tau(\mathbf{I}_p \otimes \mathbf{F}') + \hat{v}_\tau^2\mathbf{I}_p \otimes \boldsymbol{\delta}'\big\}\hat{\mathbf{V}}_\tau^- \mathrm{vec}(\mathbf{Z}_t),\\ \widehat{\mathrm{var}}\,(\hat{\mathbf{y}}(\mathbf{s},t) - \mathbf{y}(\mathbf{s},t)) &= (\mathbf{I}_p \otimes \mathbf{f}(\mathbf{s})')\hat{\mathbf{M}}_\tau(\mathbf{I}_p \otimes \mathbf{f}(\mathbf{s})) + \hat{v}_\tau^2\mathbf{I}_{np}\\ &\quad - \big\{(\mathbf{I}_p \otimes \mathbf{f}(\mathbf{s})')\hat{\mathbf{M}}_\tau(\mathbf{I}_p \otimes \mathbf{F}') + \hat{v}_\tau^2\mathbf{I}_p \otimes \boldsymbol{\delta}'\big\}\hat{\mathbf{V}}_\tau^-\\ &\quad \times \big\{(\mathbf{I}_p \otimes \mathbf{F})\hat{\mathbf{M}}_\tau(\mathbf{I}_p \otimes \mathbf{f}(\mathbf{s})) + \hat{v}_\tau^2\mathbf{I}_p \otimes \boldsymbol{\delta}\big\},\end{aligned}$$

where $\boldsymbol{\delta} \equiv (I(\mathbf{s} = \mathbf{s}_1), \ldots, I(\mathbf{s} = \mathbf{s}_n))'$ and

$$\hat{\mathbf{V}}_\tau = (\mathbf{I}_p \otimes \mathbf{F})\hat{\mathbf{M}}_\tau(\mathbf{I}_p \otimes \mathbf{F}') + (\hat{v}_\tau^2\mathbf{I}_p + \boldsymbol{\Sigma}) \otimes \mathbf{I}_n.$$

Applying the same argument as that for (2.10), we obtain

$$\hat{\mathbf{V}}_\tau = \mathbf{Q}_{K^*}\mathrm{diag}\left((d_1 - \tau - \hat{v}_\tau^2)_+, \ldots, (d_{K^*} - \tau - \hat{v}_\tau^2)_+\right)\mathbf{Q}'_{K^*} + (\hat{v}_\tau^2\mathbf{I}_p + \boldsymbol{\Sigma}) \otimes \mathbf{I}_n.$$

Consequently, if $\hat{v}_\tau^2\mathbf{I}_p + \boldsymbol{\Sigma}$ is non-singular, applying the Sherman-Morrison-Woodbury formula, we have

$$\hat{\mathbf{V}}_\tau^{-1} = \boldsymbol{\Lambda} - \boldsymbol{\Lambda}\mathbf{U}(\mathbf{I}_{K^*} + \mathbf{U}'\boldsymbol{\Lambda}\mathbf{U})^{-1}\mathbf{U}'\boldsymbol{\Lambda},$$

where $\boldsymbol{\Lambda} \equiv \left(\hat{v}_\tau^2\mathbf{I}_p + \boldsymbol{\Sigma}\right)^{-1} \otimes \mathbf{I}_n$ and $\mathbf{U} = \mathbf{Q}_{K^*}\,\mathrm{diag}\big((d_1 - \tau - \hat{v}_\tau^2)_+^{1/2}, \ldots, (d_{K^*} - \tau - \hat{v}_\tau^2)_+^{1/2}\big)$. Hence $\hat{\mathbf{y}}(\mathbf{s},t)$ and $\widehat{\mathrm{var}}\,(\hat{\mathbf{y}}(\mathbf{s},t) - \mathbf{y}(\mathbf{s},t))$, including parameter estimation, can be computed in $O(K^2np^3) + O(KnTp^2)$ operations.

An L-fold CV is applied to choose the regularization parameter τ, with

$$\mathrm{CV}(\tau) = \sum_{\ell=1}^{L}\sum_{t=1}^{T}\sum_{\mathbf{s}\in\mathcal{D}_\ell}\left\|\mathbf{z}(\mathbf{s},t) - \hat{\mathbf{y}}^{(\ell)}(\mathbf{s},t;\tau)\right\|^2, \tag{3.10}$$

where $\{\mathcal{D}_1, \ldots, \mathcal{D}_L\}$ is a partition of $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$ with roughly the same size, and $\hat{\mathbf{y}}^{(\ell)}(\mathbf{s},t;\tau)$ is the EBLUP of $\mathbf{y}(\mathbf{s},t)$ based on only the data observed at $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}\backslash\mathcal{D}_\ell$, for $\mathbf{s} \in D$, $t = 1, \ldots, T$ and $\ell = 1, \ldots, L$. The final regularization parameter is $\hat{\tau}$, which minimizes $\mathrm{CV}(\tau)$ over $\tau \geq 0$.

4. Simulation

4.1. A univariate example

We generated the process $y(\mathbf{s},t)$ on $\mathbf{s} \in D = [0,1]^2$ for $t = 1, \ldots, T$, according to (2.2) with $K = 2$ and $v^2 = 0$:

$$y(\mathbf{s},t) = \cos(\pi\|\mathbf{s} - (0,1)\|)w_1(t) + \cos\left(2\pi\left\|\mathbf{s} - \left(\frac{3}{4}, \frac{1}{4}\right)\right\|\right)w_2(t), \tag{4.1}$$

where $w_1(t) \sim N(0, 25)$ and $w_2(t) \sim N(0, 9)$. The cosine functions in (4.1) are shown in Figure 3. The data $\{\mathbf{z}_t : t = 1, \ldots, T\}$ were simulated according to (3.2) with $n = 50$ and $\sigma^2 = 3$, where $\mathbf{s}_1, \ldots, \mathbf{s}_{50}$ were sampled from D using simple random sampling.

Table 1. Medians of $(d_k - \tau - \hat{v}_\tau^2)_+$ for the proposed methods with $\tau = 0$ and $\hat{\tau}$ selected by CV.

T	Methods	k									
		1	2	3	4	5	6	7	8	9	10
20	$\tau = 0$	708.9	118.0	7.1	5.2	4.1	2.9	2.0	1.3	0.8	0.1
	CV	599.5	63.3	0	0	0	0	0	0	0	0
50	$\tau = 0$	772.2	137.1	4.3	3.4	2.6	2.1	1.6	1.2	0.8	0.5
	CV	711.7	80.8	0	0	0	0	0	0	0	0

Similar to Cressie and Johannesson (2008), we used $K = 21$ local bisquare functions for estimating $C(\cdot, \cdot)$:

$$f_k(\mathbf{s}) = \left\{1 - \frac{\|\mathbf{s} - \mathbf{c}_k\|^2}{r_k^2}\right\}^2 I(\|\mathbf{s} - \mathbf{c}_k\| < r_k); \quad k = 1, \ldots, 21, \tag{4.2}$$

at two spatial resolutions as our basis functions, where $\{\mathbf{c}_1, \ldots, \mathbf{c}_{16}\} = \{0, 1/3, 2/3, 1\}^2$ are regular grid points in D at the finer resolution with a common radius $r_1 = \cdots = r_{16} = 1/2$, and $\mathbf{c}_{17} = (1/6, 1/6)$, $\mathbf{c}_{18} = (1/6, 5/6)$, $\mathbf{c}_{19} = (1/2, 1/2)$, $\mathbf{c}_{20} = (5/6, 1/6)$ and $\mathbf{c}_{21} = (5/6, 5/6)$ are regular grid points at the coarser resolution with a common radius $r_{17} = \cdots = r_{21} = 1/2^{1/2}$. We assumed σ^2 unknown and estimated by $\hat{\sigma}^2$ in (2.11).

We considered $T = 20, 50$. For each T, we applied the proposed method with the regularization parameter $\hat{\tau}$ selected by 4-fold CV. The proposed method was compared with a simple kriging method that estimates $C(\cdot, \cdot)$ and σ^2 by maximum likelihood (ML) based on the stationary exponential covariance model. The results are summarized in Figure 4 in terms of

$$\text{MSPE} = \frac{1}{T|D|} \sum_{t=1}^{T} \int_{s \in D} |\hat{y}(s, t) - y(s, t)|^2, \tag{4.3}$$

based on 50 simulation replications. The MSPEs based on the proposed method with $\tau = 0$ and those based on the true covariance function are also shown, for comparison.

It is a bit surprising that the proposed method with no regularization is only comparable with the exponential model when $T = 20$. Nevertheless, the proposed method with $\hat{\tau}$ selected by CV outperforms the other methods by a large margin. Table 1 shows the medians of $(d_k - \tau - \hat{v}_\tau^2)_+$ among the 50 simulations with $\tau = 0$ and $\hat{\tau}$ selected by CV. Clearly, regularization helps reduce variance and recover the true rank of the covariance matrix.

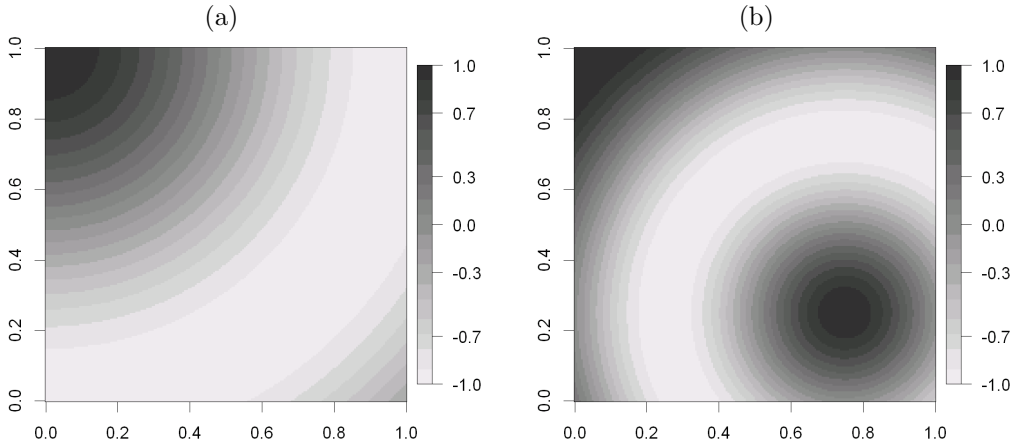

Figure 3. The two cosine functions of (4.1).

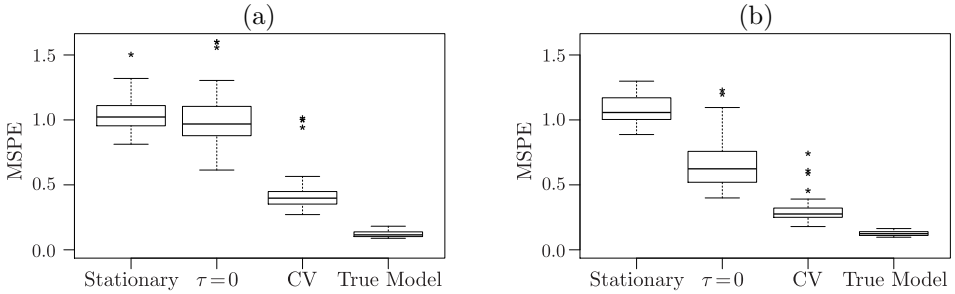

Figure 4. Boxplots of MSPEs for various methods based on (a) $T = 20$; (b) $T = 50$.

4.2. A bivariate one-dimensional example

Consider the bivariate process $\mathbf{y}(s,t)$ defined on $s \in D = [0,1]$, with $t = 1,\ldots,T$,

$$\mathbf{y}(s,t) = \sum_{k=1}^{5} \left(1 - 4\,\|s - c_k\|^2\right)^2 I\left(\|s - c_k\| < \frac{1}{2}\right) \mathbf{w}_k(t), \tag{4.4}$$

where $(c_1,\ldots,c_5) = (0, 1/4, 1/2, 3/4, 1)$, and

$$\mathbf{w}_k(t) \sim \mathrm{N}\left(\mathbf{0}, \begin{pmatrix} \eta_{k1}^2 & 0.5\eta_{k1}\eta_{k2} \\ 0.5\eta_{k1}\eta_{k2} & \eta_{k2}^2 \end{pmatrix}\right); \quad k = 1,\ldots,5,\ t = 1,\ldots,T,$$

are uncorrelated with $(\eta_{11}^2,\ldots,\eta_{51}^2) = (8, 10, 12, 14, 16)$, and $(\eta_{12}^2,\ldots,\eta_{52}^2) = (17, 14, 11, 8, 5)$. The data were generated at $\mathbf{s}_i = (i-1)/9$, $i = 1,\ldots,10$, according to (3.1) with $(\sigma_1^2, \sigma_2^2) = (2.5, 3.5)$. The spatial cross-covariance function of $\mathbf{y}(\cdot, t)$ is shown in Figure 5(a). We assumed $\mathbf{\Sigma} = \mathrm{diag}(\sigma_1^2, \sigma_2^2)$ unknown and estimated by minimizing (3.7) through (3.8) and (3.9).

We applied the proposed method with $K \in \{4, 7\}$, where for $K = 4$, $f_k(\cdot)$'s are cubic B-splines with knots $(0, 0, 0, 0, 1, 1, 1, 1)$, and for $K = 7$, $f_k(\cdot)$'s are cubic B-splines with knots $(0, 0, 0, 0, 1/4, 1/2, 3/4, 1, 1, 1, 1)$. The regularization parameter τ was selected using 4-fold CV. To avoid the boundary effect caused by a small number of data points, we always included $s_1 = 0$ and $s_{10} = 1$ in the training sets for CV. Figure (5)(b) shows the approximated covariance function obtained by minimizing

$$\sum_{1\le i,j\le 2} \int_0^1 \int_0^1 \left|\mathbf{f}(s)'\mathbf{M}_{ij}\mathbf{f}(s^*) + v^2 I(s = s^*) - C_{ij}(s, s^*)\right|^2 ds ds^*, \tag{4.5}$$

over $\{\mathbf{M}_{ij}\}$ and v^2, where $f_k(\cdot)$'s are the cubic B-splines with $K = 7$, $\{\mathbf{M}_{ij} : i, j = 1, 2\}$ are $K \times K$ matrices in (3.3), and $C_{ij}(s, s^*) = \mathrm{cov}\left(y_i(\mathbf{s}, t),\ y_j(\mathbf{s}^*, t)\right)$, $i, j = 1, 2$.

Our method was compared with a simple kriging method that estimates $C_{ij}(\cdot, \cdot)$ and σ^2 using ML based on the bivariate Matérn model of (1.1) with all the smoothness parameters fixed at 0.5. The results for various methods are summarized in Figure 6 in terms of

$$\mathrm{MSPE} = \frac{1}{T|D|} \sum_{t=1}^{T} \int_{s\in D} \left\|\hat{\mathbf{y}}(s, t) - \mathbf{y}(s, t)\right\|^2, \tag{4.6}$$

based on 50 simulation replications, where $\hat{\mathbf{y}}(s, t)$ is a generic estimate of $\mathbf{y}(s, t)$. The MSPEs based on our method with $\tau = 0$ and those based on the true covariance function $C_{ij}(s, s^*)$ are also shown for comparison.

From Figure 6, the proposed method with either $\tau = 0$ or $\hat{\tau}$ selected by CV outperforms the bivariate Matérn model in terms of MSPE. Overall, the proposed method with $\hat{\tau}$ performs slightly better than that with $\tau = 0$. As expected, the proposed method with $K = 4$ tends to perform better than that with $K = 7$ when T is small, and vice versa. For $K = 7$ and $T = 20$, the means of estimated $C_{ij}(s, s^*)$'s based on $\tau = 0$ and $\hat{\tau}$ are shown in Figures 5(c) and (d), respectively. Some slight shrinkage can be seen for the estimated cross-covariance functions when $\hat{\tau}$ is used, compared to $\tau = 0$.

4.3. A bivariate two-dimensional example with $T = 1$

We took $T = 1$, and generated the process, $\mathbf{y}(\mathbf{s}) = \mathbf{y}(\mathbf{s}, 1)$, on $\mathbf{s} \in D = [0, 1]^2$ according to (3.2) and (3.3) with $K = 3$ and $v^2 = 0$. That is,

$$\mathbf{y}(\mathbf{s}) = \cos(\pi\|\mathbf{s} - (0, 1)\|)\mathbf{w}_1 + \cos(2\pi\|\mathbf{s} - (\frac{3}{4}, \frac{1}{4})\|)\mathbf{w}_2 + \|\mathbf{s}\|\|(1 - \mathbf{s})\|\mathbf{w}_3$$

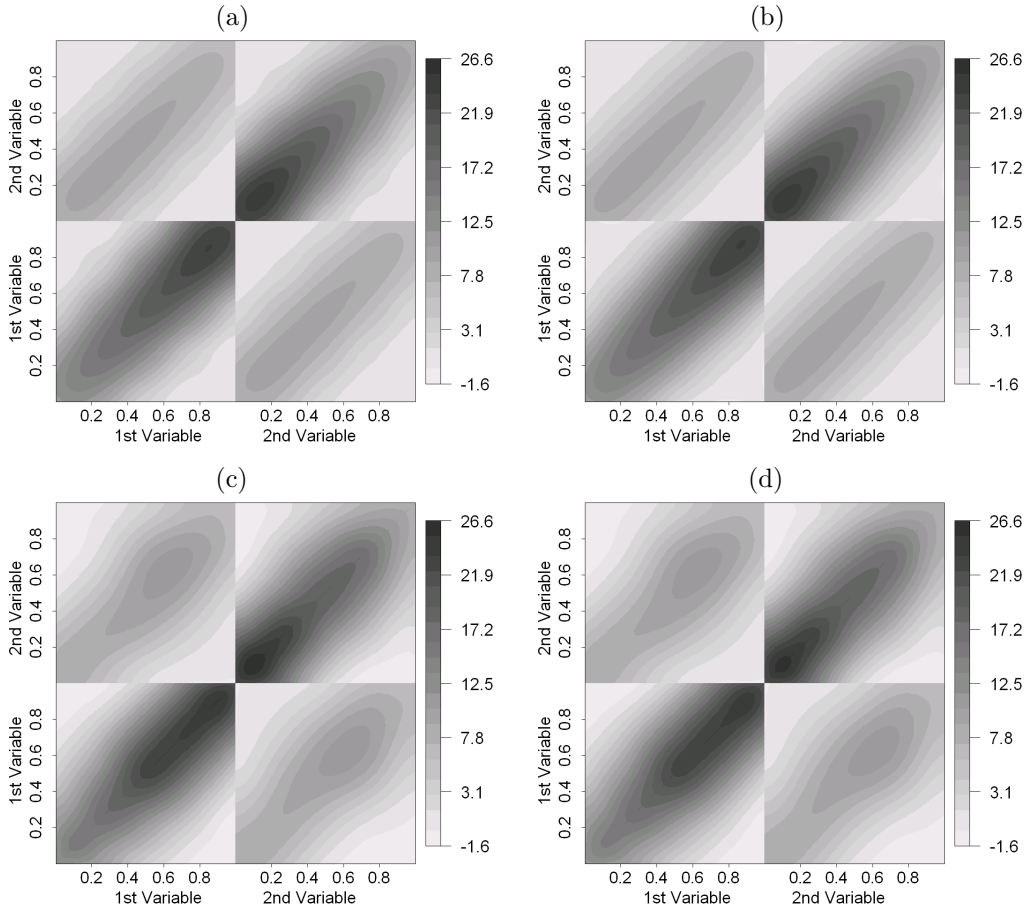

Figure 5. (a) Spatial cross-covariance functions of (4.4); (b) Approximated cross-covariance functions based on (4.5) with $K = 7$ cubic B-splines; (c) Means of estimated cross-covariance functions with $\tau = 0$ based on 50 simulation replicates; (d) Means of estimated cross-covariance with $\hat{\tau}$ selected by 4-fold CV based on 50 simulation replicates.

where, for simplicity, $\mathbf{w}_k(1)$ is written as $\mathbf{w}_k$, and

$$\mathbf{M}_{11} = \begin{pmatrix} 15 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{M}_{12} = \mathbf{M}_{21}' = \begin{pmatrix} 0 & 0 & 10 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{M}_{22} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 5 & 6 \\ 0 & 6 & 18 \end{pmatrix}.$$

Since $\mathbf{M}_{12} \neq \mathbf{M}_{21}$, the cross-covariance function is asymmetric. The data $\mathbf{Z}_1$ were generated at $n = 48^2$ grid points, $\{0, 1/47, 2/47, \ldots, 1\}^2$, according to (3.1) with $(\sigma_1^2, \sigma_2^2) = (2, 3)$. We assumed $\mathbf{\Sigma} = (\sigma_1^2, \sigma_2^2)$ known, and considered four classes of basis functions. The first class consisted of bisquare functions of (4.2) used in Section 4.1, with $K = 16$ and $\{\mathbf{c}_1, \ldots, \mathbf{c}_{16}\} = \{0, 1/3, 2/3, 1\}^2$. The second class

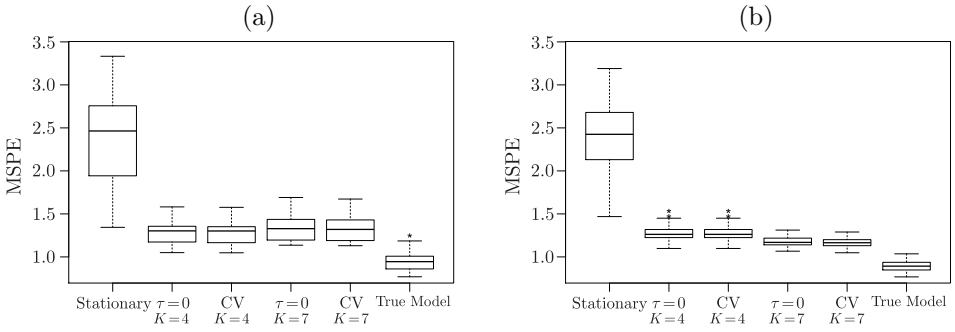

Figure 6. MSPE performance of various methods for the bivariate one-dimensional example: (a) $T = 20$; (b) $T = 50$.

consisted of the same $K = 21$ bisquare functions considered in Section 4.1 at two spatial resolutions. The other two classes were composed of thin-plate spline (TPS) functions with $f_1(\mathbf{s}) = 1$, $f_2(\mathbf{s}) = x_1$, $f_3(\mathbf{s}) = x_2$, and

$$f_k(\mathbf{s}) = \frac{1}{16\pi}\|\mathbf{s} - \mathbf{c}_k\|^2 \log(\|\mathbf{s} - \mathbf{c}_k\|^2); \quad k = 4, \ldots, K,$$

where $\mathbf{s} = (x_1, x_2)$. We considered $K = 19$ with $\{\mathbf{c}_4, \ldots, \mathbf{c}_{19}\} = \{0, 1/3, 2/3, 1\}^2$ and $K = 28$ with $\{\mathbf{c}_4, \ldots, \mathbf{c}_{28}\} = \{0, 1/4, 1/2, 3/4, 1\}^2$ as our third and fourth classes of basis functions.

For each of the basis-function classes, we merged the 48×48 spatial locations into 24×24 and 16×16 subsets having 2×2 and 3×3 locations, respectively, in each subset. The regularization parameters were selected by 4-fold CV. We further applied CV to select among the eight combinations as our final method.

The proposed method was compared with a simple kriging method that estimates $C_{ij}(\cdot, \cdot)$ and σ^2 based on the bivariate Matérn model of (1.1) with all the smoothness parameters fixed at 0.5. Instead of estimating the covariance parameter vector $\boldsymbol{\theta}$ using ML, computationally too intensive, we chose $\boldsymbol{\theta}$ to minimize the Kullback-Leibler loss between the distribution of the true variance covariance matrix $\mathbf{V}_0$ and that of the Matérn model $\mathbf{V}(\boldsymbol{\theta})$:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \operatorname{tr}(\mathbf{V}^{-1}(\boldsymbol{\theta})\mathbf{V}_0) - \log(|\mathbf{V}_0|) + \log(|\mathbf{V}(\boldsymbol{\theta})|).$$

This gives some advantage to the bivariate Matérn method, because the resulting estimate can be regarded as the ML estimate based on an infinite sample size; see Chapter 12 of Cover and Thomas (1991). Figures 7 and 8 show the predicted surfaces based on a randomly selected data set using our method (using TPS with $K = 19$ and bin size 2×2) and the Matérn model for the first and second processes, respectively.

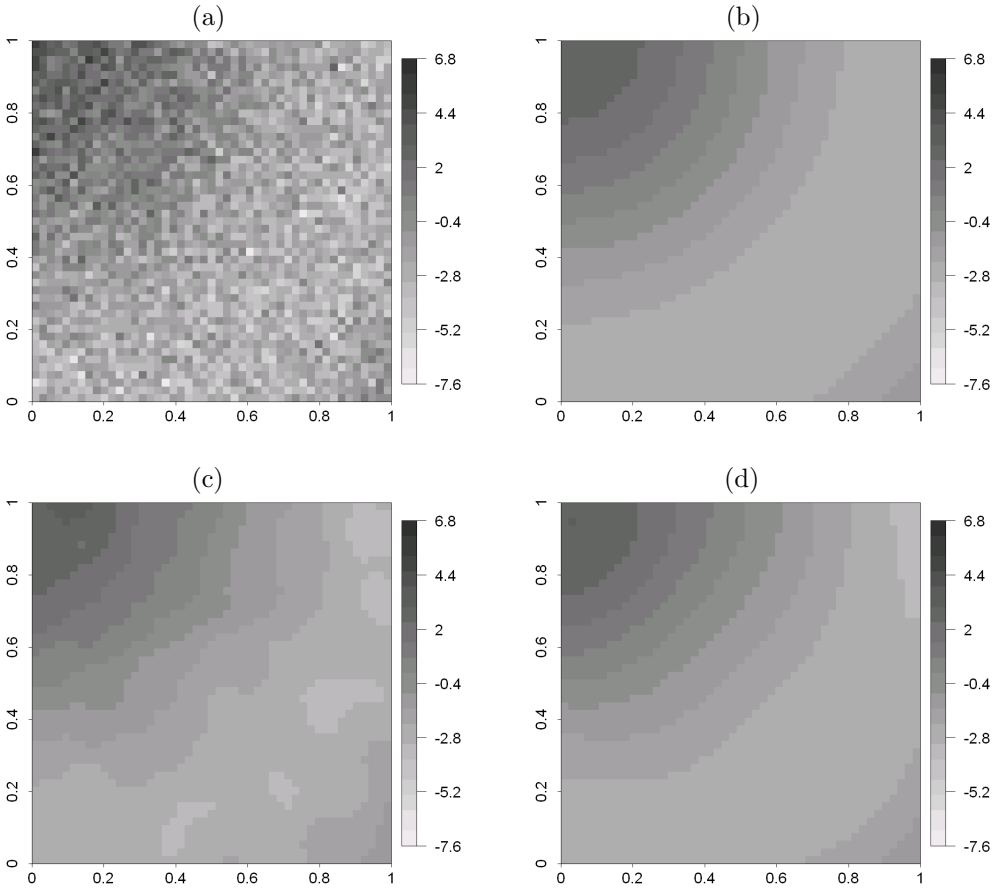

Figure 7. (a) Noisy data observed at 48×48 locations for the first variable; (b) the underlying spatial process $y_1(\cdot)$; (c) Predicted surface based on a bivariate Matérn model; (d) Predicted surface based on the proposed method using TPS with 19 basis functions and bin size 2×2.

The results for various methods are summarized in Table 2 in terms of MSPE of (4.6) with $T = 1$ based on 100 simulation replications. The MSPEs based on the true covariance function $C_{ij}(s, s^*)$ are shown for comparison. In general, our method based on TPS basis functions outperformed that based on bisquare basis functions, and TPS with $K = 19$ performed better than that with $K = 28$. Nevertheless, the bin sizes do not appear to play a major role in terms of MSPE. The final method, selected among eight methods by CV, performed better than the individual methods, indicating that CV is sensible. The running time averaged over 100 simulation replicates is also reported for each method based on R code implemented on a Windows PC. As expected, our method has clear advantage over the stationary Matérn model.

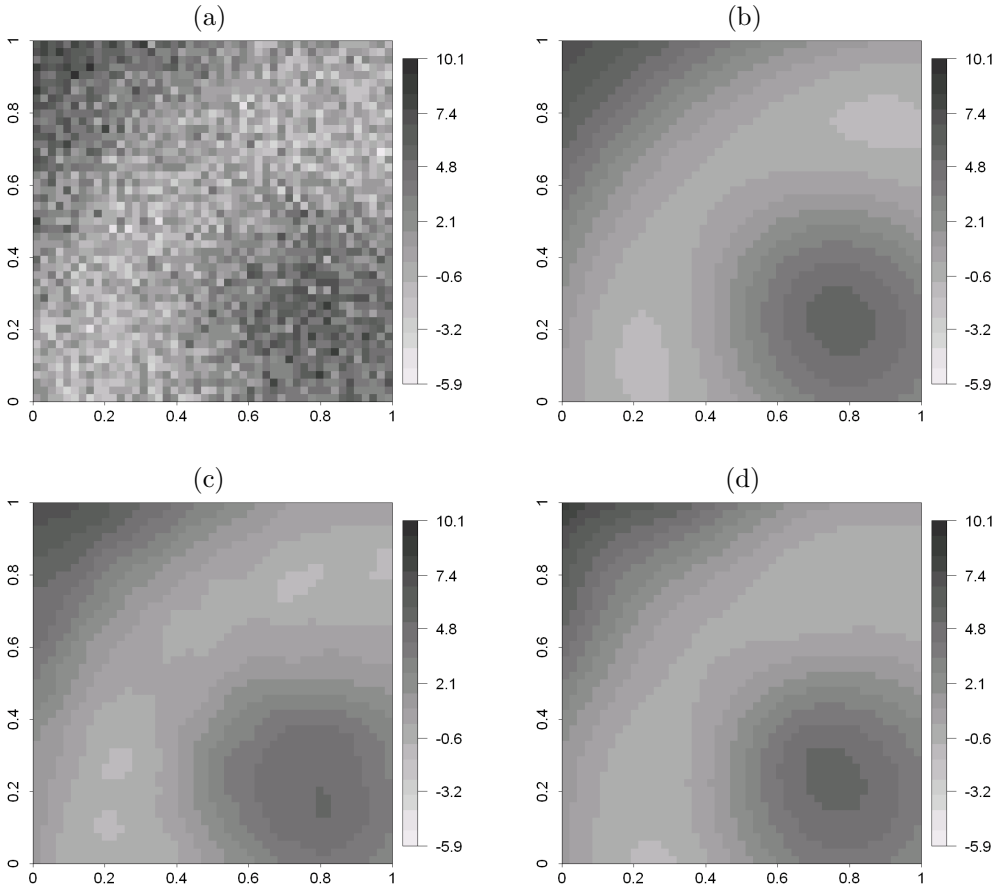

Figure 8. (a) Noisy data observed at 48×48 locations for the second variable; (b) the underlying spatial process $y_2(\cdot)$; (c) Predicted surface based on a bivariate Matérn model; (d) Predicted surface based on the proposed method using TPS with 19 basis functions and bin size 2×2.

4.4. Sensitivity analysis for the number of folds in CV

Throughout the paper, 4-fold CV was applied. It is of interest to see how the MSPE is affected by the number of folds. Table 3 shows the MSPE performance for different numbers of folds in combination with seven T values using the proposed method with $K = 19$ TPS basis functions given in the previous subsection. The results show that our method is relatively insensitive to the number of folds in CV. Comparing the results based on CV with those based on $\tau = 0$ indicates the effectiveness of using regularization, particularly for moderate T values where variances can be considerably reduced without introducing much bias.

Table 2. MSPE and average running time performance of various methods based on 100 simulation replicates, where the values in parentheses are the corresponding standard errors.

Methods		MSPE		Time
True		0.0020	(0.0003)	0.9 sec
Stationary		0.0479	(0.0013)	14.2 hr
Our		0.0250	(0.0008)	9.0 min
Bisquare	$K = 16$, bin size=2^2	0.0470	(0.0054)	56.2 sec
Bisquare	$K = 16$, bin size=3^2	0.0471	(0.0054)	59.7 sec
Bisquare	$K = 21$, bin size=2^2	0.0327	(0.0021)	64.9 sec
Bisquare	$K = 21$, bin size=3^2	0.0329	(0.0022)	70.2 sec
TPS	$K = 19$, bin size=2^2	0.0256	(0.0017)	60.3 sec
TPS	$K = 19$, bin size=3^2	0.0258	(0.0017)	63.1 sec
TPS	$K = 28$, bin size=2^2	0.0295	(0.0011)	88.2 sec
TPS	$K = 28$, bin size=3^2	0.0299	(0.0011)	89.9 sec

Table 3. MSPE performance of the various methods for different T values, where the values in parentheses are the corresponding standard errors.

Methods	$T = 2$	$T = 3$	$T = 6$	$T = 9$	$T = 27$	$T = 81$	$T = 243$
True	0.0019	0.0019	0.0018	0.0017	0.0017	0.0017	0.0016
	(0.0003)	(0.0003)	(0.0002)	(0.0001)	(0.0001)	(0.00005)	(0.00002)
Stationary	0.0486	0.0484	0.0484	0.0482	0.0480	0.0480	0.0478
	(0.0009)	(0.0006)	(0.0005)	(0.0005)	(0.0003)	(0.0001)	(0.0001)
$\tau = 0$	0.0256	0.0240	0.0239	0.0206	0.0159	0.0124	0.0107
	(0.0012)	(0.0009)	(0.0009)	(0.0007)	(0.0004)	(0.0002)	(0.0001)
2-fold CV	0.0244	0.0232	0.0207	0.0151	0.0117	0.0102	0.0097
	(0.0010)	(0.0009)	(0.0008)	(0.0006)	(0.0004)	(0.0002)	(0.0001)
4-fold CV	0.0255	0.0227	0.0196	0.0147	0.0117	0.0103	0.0098
	(0.0012)	(0.0009)	(0.0008)	(0.0007)	(0.0004)	(0.0002)	(0.0001)
8-fold CV	0.0248	0.0231	0.0191	0.0148	0.0120	0.0106	0.0099
	(0.0011)	(0.0010)	(0.0009)	(0.0007)	(0.0004)	(0.0002)	(0.0001)
12-fold CV	0.0251	0.0225	0.0194	0.0145	0.0116	0.0105	0.0098
	(0.0012)	(0.008)	(0.0008)	(0.0006)	(0.0005)	(0.0002)	(0.0001)

Acknowledgement

This research was supported in part by Taiwan National Science Council grant NSC 97-2118-M-001-001-MY3. The authors thank the editors and the reviewers for their helpful comments and suggestions.

Supplementary Material

Supplementary material available at Statistica Sinica online contains the proofs of Proposition 1 and Corollary 1.

References

Apanasovich, T. V. and Genton, M. G. (2010). Cross-covariance functions for multivariate random fields based on latent dimensions. *Biometrika* **97**, 15-30.

Apanasovich, T. V., Genton, M. G. and Sun, Y. (2012). A valid matérn class of cross-covariance functions for multivariate random fields with any number of components. *J. Amer. Statist. Assoc.* **107**, 180-193.

Banerjee, S., Carlin, B. P. and Gelfand, A. E. (2004). *Hierarchical Modeling and Analysis for Spatial Data*. Chapman & Hall/CRC, Boca Raton, FL.

Banerjee, S., Gelfand, A. E., Finley, A. O. and Sang, H. (2008). Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society B* **70**, 825-848.

Bickel, P. J. and Li, B. (2006). Regularization in statistics. *Test* **15**, 271-344.

Christensen, W. F. and Amemiya, Y. (2003). Modeling and prediction for multivariate spatial factor analysis. *J. Statist. Plann. Inference* **115**, 543-564.

Cover, T. M. and Thomas, J. A. (1991). *Elements of Information Theory*. Wiley, New York.

Cressie, N. and Johannesson, G. (2008). Fixed rank kriging for very large spatial data sets. *J. Roy. Statist. Soc. Ser. B* **70**, 209-226.

Fanshawe, T. R. and Diggle, P. J. (2012). Bivariate geostatistical modelling: a review and an application to spatial variation in radon concentrations. *Environmental and Ecolog. Statist.* **19**, 139-160.

Furrer, R. and Genton, M. G. (2011). Aggregation-cokriging for highly multivariate spatial data. *Biometrika* **98**, 615-631.

Gelfand, A. E., Schmidt, A. M., Banerjee, S. and Sirmans, C. (2004). Nonstationary multivariate process modeling through spatially varying coregionalization. *Test* **13**, 263-312.

Gneiting, T., Kleiber, W. and Schlather, M. (2010). Matérn cross-covariance functions for multivariate random fields. *J. Amer. Statist. Assoc.* **105**, 1167-1177.

Goulard, M. and Voltz, M. (1992). Linear coregionalization model: tools for estimation and choice of cross-variogram matrix. *Math. Geology* **24**, 269-286.

Kleiber, W. and Nychka, D. (2012). Nonstationary modeling for multivariate spatial processes. *J. Multivariate Anal.* **112**, 76-91.

Krzanowski, W. and Bailey, T. (2007). Extraction of spatial features using factor methods, illustrated on stream sediment data. *Math. Geology* **39**, 69-85.

Li, B. and Zhang, H. (2011). An approach to modeling asymmetric multivariate spatial covariance structures. *J. Multivariate Anal.* **102**, 1445-1453.

Majumdar, A. and Gelfand, A. E. (2007). Multivariate spatial modeling for geostatistical data using convolved covariance functions. *Math. Geology* **39**, 225-245.

Marchant, B. and Lark, R. (2007). Estimation of linear models of coregionalization by residual maximum likelihood. *European J. Soil Sci.* **58**, 1506-1513.

Ren, Q. and Banerjee, S. (2013). Hierarchical factor models for large spatially misaligned data: A low-rank predictive process approach. *Biometrics* **69**, 19-30.

Schmidt, A. M. and Gelfand, A. E. (2003). A bayesian coregionalization approach for multivariate pollutant data. *J. Geophys. Res.: Atmospheres* **108**.

Ver Hoef, J. M. and Barry, R. P. (1998). Constructing and fitting models for cokriging and multivariable spatial prediction. *J. Statist. Plann. Inference* **69**, 275-294.

Ver Hoef, J. M., Cressie, N. and Barry, R. P. (2004). Flexible spatial models for kriging and cokriging using moving averages and the fast fourier transform (fft). *J. Comput. Graph. Statist.* **13**, 265-282.

Wang, F. and Wall, M. M. (2003). Generalized common spatial factor model. *Biostatistics* **4**, 569-582.

Woodbury, M. A. (1950). Inverting modified matrices. Technical report, Memorandum report 42, Statistical Research Group, Princeton University, Princeton.

Zhang, H. (2007). Maximum-likelihood estimation for multivariate spatial linear coregionalization models. *Environmetrics* **18**, 125-139.

Department of Public Health, China Medical University, Taichung 40402, Taiwan.

E-mail: slt.cmu@gmail.com

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

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

(Received July 2013; accepted February 2014)