



Title	ON SOME MATÉRN COVARIANCE FUNCTIONS FOR SPATIO-TEMPORAL RANDOM FIELDS
Author(s)	IP, HL; Li, WK
Citation	Statistica Sinica, 2017, v. 27, p. 805-822
Issued Date	2017
URL	http://hdl.handle.net/10722/240349
Rights	This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

ON SOME MATÉRN COVARIANCE FUNCTIONS FOR SPATIO-TEMPORAL RANDOM FIELDS

Ryan H. L. Ip and W. K. Li

Charles Sturt University and The University of Hong Kong

Abstract: The Matérn class is an important class of covariance functions in spatial statistics. With the recent flourishing trend in modelling spatio-temporal data, in-depth theoretical development of spatio-temporal covariograms is needed. In this paper, theories under the infill asymptotic framework concerning estimation issues of a generally non-separable Matérn class of spatio-temporal covariance function is presented. It is found that not all parameters can be estimated consistently while quantities that can be estimated consistently are found based on equivalence and orthogonality of Gaussian measures. The micro-ergodic parameters are found to be different when the degrees of separability between the space and time components are different. For the computation, an easy-to-implement estimation procedure is given. Simulation studies are conducted to show how well the asymptotic results apply when the sample size is moderate. A set of air pollution data is used to demonstrate the usefulness of the estimation procedure suggested.

Key words and phrases: Gaussian measures, infill asymptotics, micro-ergodic parameters, space-time data.

1. Introduction

Let $X(\mathbf{s})$, $\mathbf{s} \in \mathbb{R}^d$ be a second-order stationary Gaussian process with mean zero and the Matérn class covariogram (Matérn (1986)) such that

$$\text{Cov}(X(\mathbf{s}_1), X(\mathbf{s}_2)) = \sigma^2 M_\nu(h|\alpha) = \sigma^2 \frac{(\alpha h)^\nu}{\Gamma(\nu) 2^{\nu-1}} \mathcal{K}_\nu(\alpha h), \quad (1.1)$$

where \mathcal{K}_ν is the modified Bessel function of the second kind (see Abramowitz and Stegun (1972)), ν is the smoothness parameter, and the scale parameter α governs the rate of decay in correlation of the process. The Matérn class is attractive due to its high flexibility (Stein (1999)), see also Guttorp and Gneiting (2006) for the historical account. Here (1.1) reduces to the exponential covariogram when $\nu = 1/2$. Meanwhile, the process is k times differentiable if $\nu > k$, which is a desirable property in some cases. Given the parameters, the covariogram (1.1) is a function of the Euclidean distance between \mathbf{s}_1 and \mathbf{s}_2 , $h = \|\mathbf{s}_1 - \mathbf{s}_2\|$, only, so is translational and rotational invariant. Estimation of the parameters in (1.1)

have been studied over the past 30 years or so and it has been found that estimators perform very differently under different asymptotic frameworks. In the literature, there are mainly two kinds of asymptotic frameworks: the increasing-domain asymptotics and the infill or fixed-domain asymptotics. For more details on this, see Cressie (1993, Chap. 5) and Stein (1999, Chap. 3). Our results are based on the infill asymptotic framework.

Under some regularity conditions, Mardia and Marshall (1984) showed that the maximum likelihood estimators of model parameters are consistent under the increasing-domain asymptotic framework. However, under the fixed-domain asymptotic framework, Zhang (2004) showed that both σ^2 and α of (1.1) cannot be estimated consistently for $d \leq 3$. Rather, the quantity $\sigma^2 \alpha^{2\nu}$ can be estimated consistently. In addition, if α_1 is a fixed quantity and $\hat{\sigma}^2$ is the maximum likelihood estimator (MLE) estimated through fixing $\alpha = \alpha_1$, then the quantity $\hat{\sigma}^2 \alpha_1^{2\nu}$ converges to the true quantity $\sigma_0^2 \alpha_0^{2\nu}$ almost surely. Similar results were obtained by Ying (1991, 1993) when $\nu = 0.5$ and $d = 1$. Wang and Loh (2011) showed that, for fixed α_1 , $\sqrt{n} (\hat{\sigma}^2 \alpha_1^{2\nu} - \sigma_0^2 \alpha_0^{2\nu}) \rightarrow N(0, 2 (\sigma_0^2 \alpha_0^{2\nu})^2)$ for any $d \leq 3$ as the sample size $n \rightarrow \infty$. As an extension, Kaufman and Shaby (2013) showed that if $\hat{\alpha}$ is a quantity estimated from maximizing the profile likelihood function using the spatial correlation matrix and $\hat{\sigma}^2$ is the MLE estimated through fixing $\alpha = \hat{\alpha}$, then the result $\sqrt{n} (\hat{\sigma}^2 \hat{\alpha}^{2\nu} - \sigma_0^2 \alpha_0^{2\nu}) \rightarrow N(0, 2 (\sigma_0^2 \alpha_0^{2\nu})^2)$ also holds for $d \leq 3$. For higher dimensions, Anderes (2010) showed that both σ^2 and α can be estimated consistently for $d \geq 5$. Recently, Zhang and Cai (2015) provided conditions which lead to equivalence of Gaussian measures for a bivariate Matérn model. Although the estimation issues regarding Gaussian random fields in the purely spatial framework are quite well understood, the estimation issues regarding spatio-temporal random fields are rarely discussed in the literature. We aim to extend the results from the purely spatial framework to the spatio-temporal framework. Specifically, we consider the mean-zero Gaussian process $\{X(\mathbf{s}, t), \mathbf{s} \in \mathbb{R}^d, t \in \mathbb{R}\}$ with the Matérn class covariogram generated by the spectral density of the form

$$f(\omega, \tau) = \gamma (\alpha^2 \beta^2 + \beta^2 \omega^2 + \alpha^2 \tau^2 + \varepsilon^2 \omega^2 \tau^2)^{-\nu}, \quad (1.2)$$

where $\tau \in \mathbb{R}$, $\omega = \|\boldsymbol{\omega}\|$ with $\boldsymbol{\omega} \in \mathbb{R}^d$. This class of covariogram was introduced by Fuentes, Chen and Davis (2008). In (1.2), the scale parameters $\alpha > 0$ and $\beta > 0$ govern the rates of decay in the spatial and temporal correlations, respectively. The greater the values of α and β , the faster the decay in correlations. The parameter ν governs the degree of smoothness of the process while $\varepsilon \in [0, 1]$ measures the degree of separability between the spatial and temporal components. In addition, the parameter $\gamma > 0$ is related to the total variance of the process $X(\mathbf{s}, t)$ through $\sigma^2 = \text{Var}(X(\mathbf{s}, t)) = \gamma \left\{ \int_{\mathbb{R}} \int_{\mathbb{R}^d} [f(\omega, \tau) / \gamma] d\boldsymbol{\omega} d\tau \right\}$.

Let $u = |t_1 - t_2|$ and $h = \|\mathbf{s}_1 - \mathbf{s}_2\|$. The covariogram can be obtained via a Fourier transform,

$$\begin{aligned} \text{Cov}(X(\mathbf{s}_1, t_1), X(\mathbf{s}_2, t_2)) &= \int_{\mathbb{R}} \int_{\mathbb{R}^d} \exp(i\boldsymbol{\omega}^\top \mathbf{h} + i\tau u) f(\boldsymbol{\omega}, \tau) d\boldsymbol{\omega} d\tau \\ &\triangleq \sigma^2 M_\nu^\varepsilon(h, u | \alpha, \beta). \end{aligned} \quad (1.3)$$

For notational convenience, when there is no chance of confusion, we write $\sigma^2 M_\nu^\varepsilon(\alpha, \beta) = \sigma^2 M_\nu^\varepsilon(h, u | \alpha, \beta)$. When $\varepsilon = 1$, by direct integration, the covariance function $\text{Cov}(X(\mathbf{s}_1, t_1), X(\mathbf{s}_2, t_2))$ can be shown to be

$$\begin{aligned} \sigma^2 M_\nu^1(\alpha, \beta) &= \frac{\sigma^2 2^{2-2\nu+\frac{d+1}{2}}}{\Gamma(\nu - \frac{d}{2}) \Gamma(\nu - \frac{1}{2})} (\alpha h)^{\nu - \frac{d}{2}} (\beta u)^{\nu - \frac{1}{2}} \mathcal{K}_{\nu - \frac{d}{2}}(\alpha h) \mathcal{K}_{\nu - \frac{1}{2}}(\beta u) \\ &= \sigma^2 M_{\nu - d/2}(h | \alpha) M_{\nu - 1/2}(u | \beta), \end{aligned} \quad (1.4)$$

which is separable as a product of a spatial and a temporal covariance function where both functions are in the Matérn class. When $\varepsilon = 1$, one requires $\nu > d/2$. When $\varepsilon = 0$, the covariance function $\text{Cov}(X(\mathbf{s}_1, t_1), X(\mathbf{s}_2, t_2))$ is given as

$$\sigma^2 M_\nu^0(\alpha, \beta) = \frac{\sigma^2 2^{1-\nu+\frac{d+1}{2}}}{\Gamma(\nu - \frac{d+1}{2})} \left(\sqrt{\alpha^2 h^2 + \beta^2 u^2}\right)^{\nu - \frac{d+1}{2}} \mathcal{K}_{\nu - \frac{d+1}{2}}\left(\sqrt{\alpha^2 h^2 + \beta^2 u^2}\right), \quad (1.5)$$

a the non-separable model. In this case, it is required that $\nu > (d + 1)/2$. However, the model is separable when $\nu \rightarrow \infty$. The covariogram (1.5) can be regarded as a variant of (1.1) by considering time as an extra axis that allows the rates of decay in correlations to be different in the spatial domain and in the time domain. In general, when $\varepsilon \in (0, 1)$, a closed form of $f(\boldsymbol{\omega}, \tau)$ after Fourier transform does not exist and one must rely on a numerical Fourier transformation. We put our emphasis on the cases $\varepsilon = 0$ and $\varepsilon = 1$, mainly due to the existence of closed form expressions under Fourier transform. We show here that, under certain conditions, two Matérn class spatio-temporal covariograms define an equivalent Gaussian measure. Subsequently, the individual parameters are not consistently estimable. Nevertheless, there exist quantities which are consistent, and we show that, these quantities depend on the degree of space-time separability.

Although we focus on the spatio-temporal case, our results can be analogously applied to the anisotropic spatial case where the rate of decay in correlation in one coordinate is different from that of the remaining d coordinates. Here we assume that there are K (possibly unevenly spaced) locations and T time points. We focus on the estimation issue because any serious attempt at inference and prediction requires proper estimation of the model parameters, especially when the sample size is finite.

The rest of this paper is organized as follows. Section 2 provides the main results and discusses the parameter estimation issues. Simulation results are reported in Section 3. In Section 4, the models discussed in this paper are applied to a set of air pollution data. Conclusions and discussions are in Section 5. Supplementary material is provided online that contains all proofs and detailed simulation results.

2. Main Results

2.1. Preliminary

When ν is fixed and $d \leq 3$, Zhang (2004) showed that the two Gaussian probability measures P_0 and P_1 defined by the corresponding covariogram $\sigma_i^2 M_\nu(h|\alpha_i)$, $i = 0, 1$, are equivalent if and only if $\sigma_0^2 \alpha_0^{2\nu} = \sigma_1^2 \alpha_1^{2\nu}$. Under the equivalence of the Gaussian probability measures, only the quantity $\sigma^2 \alpha^{2\nu}$ is consistently estimable. It is called the micro-ergodic parameter, see Stein (1999, p. 162). The concept of equivalence can be referred to Gikhman and Skorokhod (1974, Chapter VII), for example. Sufficient conditions for equivalence of Gaussian measures were discussed in Skorokhod and Yadrenko (1973), Ibragimov and Rozanov (1978, p. 104–107), Yadrenko (1983, p. 156), Stein (1999, p. 120), Stein (2004, Theorem A.1), and Zhang (2004). For more general discussion of Gaussian measures, see to Bogachev (1998). In particular, consider a spatial process $X(\mathbf{s})$, $\mathbf{s} \in \mathbb{R}^d$ which is stationary Gaussian with mean zero and an isotropic covariogram with a spectral density $f(\omega)$. Denote by P_i , $i = 0, 1$, two probability measures such that under P_i , $X(\mathbf{s})$ has the isotropic spectral density $f_i(\omega)$ (see Zhang (2004, p. 251)). It has been showed that if $f_0(\omega)\omega^\delta$ is bounded away from zero and infinity for some $\delta > 0$ as $\omega \rightarrow \infty$ and, for some finite c , if

$$\int_c^\infty \omega^{d-1} \left\{ \frac{f_1(\omega) - f_0(\omega)}{f_0(\omega)} \right\}^2 d\omega < \infty, \quad (2.1)$$

then $P_0 \equiv P_1$. Now, let $X(\mathbf{s}, t)$ be a Gaussian process with mean zero and a covariogram with a spectral density $f(\omega, \tau)$. If $f_0(\omega, \tau)\|\omega, \tau\|^\delta$ is bounded away from zero and infinity for some $\delta > 0$ as $\|\omega, \tau\| \rightarrow \infty$, where $\|\omega, \tau\|$ is the Euclidean norm defined by $(\omega_1, \dots, \omega_d, \tau)$, and, for some finite c , if

$$\iint_{\|\omega, \tau\| > c} \left\{ \frac{f_1(\omega, \tau) - f_0(\omega, \tau)}{f_0(\omega, \tau)} \right\}^2 d\omega d\tau < \infty \quad (2.2)$$

then $P_0 \equiv P_1$. Hence, for isotropic (in both space and time domain) spectral density f such that $f(\omega, \tau) = f(\omega, \tau)$, Lemma 1 given in the supplementary material shows that (2.2) can be expressed as

$$\iint_{\mathcal{A}} \omega^{d-1} \left\{ \frac{f_1(\omega, \tau) - f_0(\omega, \tau)}{f_0(\omega, \tau)} \right\}^2 d\omega d\tau < \infty. \quad (2.3)$$

2.2 Equivalence of Gaussian probability measures

Based on (2.3), we provide conditions under which two Gaussian probability measures defined by (1.3) are equivalent under different values of ε . It is assumed that $\lim_{\omega, \tau \rightarrow \infty} \omega/\tau = k < \infty$. In addition, we assume the true values of ε and ν are known.

Theorem 1. *Let $P_i, i = 0, 1$, be probability measures so that, under P_i , the process $X(\mathbf{s}, t), \mathbf{s} \in \mathbb{R}^d, t \in \mathbb{R}$, is stationary Gaussian with mean zero and covariance function $\sigma_i^2 M_\nu^{\varepsilon_i}(\alpha_i, \beta_i)$ with ν fixed, $0 \leq \varepsilon_i \leq 1$, and $d = 1, 2$. If $r_i = \beta_i/\alpha_i$, for any bounded infinite set $\mathcal{D}_S \times \mathcal{D}_T \subset \mathbb{R}^d \times \mathbb{R}$, we have*

- (a) *when $\varepsilon_i = 1$, $P_0 \equiv P_1$ if and only if $\sigma_0^2 \alpha_0^{2\nu-d} \beta_0^{2\nu-1} = \sigma_1^2 \alpha_1^{2\nu-d} \beta_1^{2\nu-1}$;*
- (b) *when $0 < \varepsilon_i < 1$, $P_0 \equiv P_1$ if $\gamma_0/\varepsilon_0^{2\nu} = \gamma_1/\varepsilon_1^{2\nu}$;*
- (c) *when $\varepsilon_i = 0$, $P_0 \equiv P_1$ if and only if $\sigma_0^2 \alpha_0^{-d} \beta_0^{2\nu-1} = \sigma_1^2 \alpha_1^{-d} \beta_1^{2\nu-1}$ and $r_0 = r_1$.*

Thus two Matérn covariograms define the same Gaussian measure. This result extends Theorem 2 of Zhang (2004) to the spatio-temporal framework. The condition $0 < d \leq 2$ coincides with the result in Zhang (2004), where the results hold for $d \leq 3$ under the purely spatial framework.

Comparing parts (a) and (c) of Theorem 1, there exists an additional condition $r_0 = r_1$ for the equivalence of P_0 and P_1 to hold when $\varepsilon = 0$. The condition signifies the non-separability of the spatial and temporal components. Meanwhile, if we combine the two conditions $\sigma_0^2 \alpha_0^{-d} \beta_0^{2\nu-1} = \sigma_1^2 \alpha_1^{-d} \beta_1^{2\nu-1}$ and $r_0 = r_1$, we have $\sigma_0^2 r_0^d \beta_0^{2\nu-d-1} = \sigma_1^2 r_1^d \beta_1^{2\nu-d-1}$.

For $\varepsilon \in (0, 1)$, the quantity γ depends heavily on ε, ν and d . There appears no general form. As an example, when $d = 1$ and $\nu = 1.5$, γ takes the form $\sigma^2 \alpha \beta^2 \varepsilon^2 / 4 {}_2F_1(0.5, 1, 1.5, 1 - \varepsilon^{-2})$ where ${}_2F_1$ is the Hypergeometric2F1 function. Then the condition $\gamma_0/\varepsilon_0^{2\nu} = \gamma_1/\varepsilon_1^{2\nu}$ is $\sigma_0^2 \alpha_0 \beta_0^2 / \varepsilon_0^{2\nu-2} {}_2F_1(0.5, 1, 1.5, 1 - \varepsilon_0^{-2}) = \sigma_1^2 \alpha_1 \beta_1^2 / \varepsilon_1^{2\nu-2} {}_2F_1(0.5, 1, 1.5, 1 - \varepsilon_1^{-2})$.

Corollary 1. *Under the conditions of Theorem 1 with a bounded subset of $\mathcal{D}_S \times \mathcal{D}_T \subset \mathbb{R}^d \times \mathbb{R}$ for $d \leq 2$, if the process $X(\mathbf{s}, t), \mathbf{s} \in \mathbb{R}^d, t \in \mathbb{R}$, is stationary Gaussian with mean zero with and covariance function $\sigma^2 M_\nu^\varepsilon(\alpha, \beta), \varepsilon \in [0, 1]$, given observations $X(\mathbf{s}_j, t_j), j = 1, 2, \dots, n$, for $n = KT$, there exist no weakly consistent estimators σ_n^2, α_n , and β_n .*

Under different values of ε , micro-ergodic parameters do exist and it can be shown that these quantities are consistently estimable. Let $\ell_n(\sigma^2, \alpha, \beta; \varepsilon)$ be the log-likelihood function, explicitly

$$\ell_n(\sigma^2, \alpha, \beta; \varepsilon) = -\frac{KT}{2} \log 2\pi - \frac{KT}{2} \log \sigma^2 - \frac{1}{2} \log \det \mathbf{\Gamma}_{\alpha, \beta; \varepsilon} - \frac{1}{2\sigma^2} \mathbf{X}_n^\top \mathbf{\Gamma}_{\alpha, \beta; \varepsilon}^{-1} \mathbf{X}_n, \quad (2.4)$$

where \mathbf{X}_n is the data vector containing all observations $X(\mathbf{s}_j, t_j)$, $j = 1, \dots, n$ and $\mathbf{\Gamma}_{\alpha, \beta; \varepsilon}$ is the correlation matrix independent of σ^2 . With ε fixed, by profiling out σ^2 , the profile log-likelihood function of (α, β) is

$$\tilde{\ell}_n(\alpha, \beta; \varepsilon) = -\frac{KT}{2} \log 2\pi - \frac{KT}{2} \log \left(\frac{1}{KT} \mathbf{X}_n^\top \mathbf{\Gamma}_{\alpha, \beta; \varepsilon}^{-1} \mathbf{X}_n \right) - \frac{1}{2} \log \det \mathbf{\Gamma}_{\alpha, \beta; \varepsilon} - \frac{KT}{2} \quad (2.5)$$

and the MLE of σ^2 is

$$\hat{\sigma}_n^2 = \frac{1}{KT} \mathbf{X}_n^\top \mathbf{\Gamma}_{\alpha, \beta; \varepsilon}^{-1} \mathbf{X}_n. \quad (2.6)$$

2.3. Asymptotic theory under the separable model

We first consider the case of $\varepsilon = 1$.

Theorem 2. *Under the conditions of Theorem 1 with $\varepsilon = 1$, and a bounded region $\mathcal{D}_n \subset \mathcal{D}_S \times \mathcal{D}_T \subset \mathbb{R}^d \times \mathbb{R}$ for $d \leq 2$, for any fixed constants $\alpha_1 > 0$ and $\beta_1 > 0$, if $\hat{\sigma}_n^2$ is obtained from (2.6) with $\mathbf{\Gamma}_{\alpha, \beta}^{-1}$ replaced by $\mathbf{\Gamma}_{\alpha_1, \beta_1}^{-1}$, then $\hat{\sigma}_n^2 \alpha_1^{2\nu-d} \beta_1^{2\nu-1} \rightarrow \sigma_0^2 \alpha_0^{2\nu-d} \beta_0^{2\nu-1}$ almost surely as $K, T \rightarrow \infty$ under P_0 , the Gaussian probability measure defined by the covariogram $\sigma_0^2 M_\nu^1(\alpha_0, \beta_0)$.*

When $\varepsilon = 1$, the covariogram is indeed separable, so Theorem 2 can be considered as an extension of Theorem 3 of Zhang (2004). Although mis-specifying model parameters may have little impact in the asymptotic sense, Kaufman and Shaby (2013) noticed that this idea is not readily transferable to the finite sample case. They also found that although α cannot be estimated consistently, the estimates are usually fairly close to the true value in simulation studies. A similar result can also be obtained under the spatio-temporal setting. Direct maximization of the profile likelihood (2.5) can present a large computational burden owing to the large number of observations and the low efficiency in searching the maximum of a bivariate function. Suggested estimation procedures for different values of ε are given below. They require maximizing functions of a single parameter only, and hence effectively reduce the computational burden.

While estimators of individual parameters are not consistent, the data still contain information about the model parameters. If computational burden is not a concern, one could directly maximize the profile likelihood function (2.5) to obtain $\hat{\alpha}$ and $\hat{\beta}$ jointly. If computational power is limited, one could estimate α and β in the spirit of composite likelihood methods (Varin, Reid and Firth (2011)). Assuming there is no temporal dependence, α can be estimated from

$$\hat{\alpha} = \operatorname{argmax}_{\alpha \in [\alpha_L, \alpha_U]} \sum_{t=1}^T \tilde{\ell}_{n,t}(\alpha), \quad (2.7)$$

where $0 < \alpha_L < \alpha_U < \infty$, and

$$\tilde{\ell}_{n,t}(\alpha) = -\frac{K}{2} \log 2\pi - \frac{K}{2} \log \left(\frac{1}{K} \mathbf{X}_{n,t}^\top \tilde{\Gamma}_\alpha^{-1} \mathbf{X}_{n,t} \right) - \frac{1}{2} \log \det \tilde{\Gamma}_\alpha - \frac{K}{2}. \quad (2.8)$$

In (2.8), $\mathbf{X}_{n,t}$ contains observations of \mathbf{X}_n at time t only and $\tilde{\Gamma}_\alpha$ is the spatial correlation matrix with elements $M_\nu(h|\alpha)$ that do not involve β . Estimation of β can be done similarly. Given no spatial dependence, β can be estimated from

$$\hat{\beta} = \operatorname{argmax}_{\beta \in [\beta_L, \beta_U]} \sum_{k=1}^K \tilde{\ell}_{n,k}(\beta), \quad (2.9)$$

where $0 < \beta_L < \beta_U < \infty$ and

$$\tilde{\ell}_{n,k}(\beta) = -\frac{T}{2} \log 2\pi - \frac{T}{2} \log \left(\frac{1}{T} \mathbf{X}_{n,k}^\top \tilde{\Gamma}_\beta^{-1} \mathbf{X}_{n,k} \right) - \frac{1}{2} \log \det \tilde{\Gamma}_\beta - \frac{T}{2}. \quad (2.10)$$

In (2.10), $\mathbf{X}_{n,k}$ contains those observations of \mathbf{X}_n at location \mathbf{s}_k only, and $\tilde{\Gamma}_\beta$ is the temporal correlation matrix with elements $M_\nu(u|\beta)$ that do not depend on α .

Although the estimates from (2.7) and (2.9) are inconsistent, they are usually fairly close to their true values, as we show in Section 3. With $\hat{\alpha}$ and $\hat{\beta}$, the MLE $\hat{\sigma}_n^2$ can be obtained from (2.6) with $\Gamma_{\alpha,\beta}^{-1}$ replaced by $\Gamma_{\hat{\alpha},\hat{\beta}}^{-1}$. Let

$$\zeta_{\alpha,\beta}^1 = \hat{\sigma}_n^2 \alpha^{2\nu-d} \beta^{2\nu-1} = \frac{\alpha^{2\nu-d} \beta^{2\nu-1}}{KT} \mathbf{X}_n^\top \Gamma_{\alpha,\beta}^{-1} \mathbf{X}_n. \quad (2.11)$$

Lemma 2 given in the supplementary material shows that $\zeta_{\alpha,\beta}^1$ is monotonically decreasing in both α and β and this property leads to the following.

Theorem 3. *Under the conditions in Theorem 1 and $\varepsilon = 1$, if $\hat{\alpha}$ and $\hat{\beta}$ are obtained via (2.7) and (2.9), then $\zeta_{\hat{\alpha},\hat{\beta}}^1 \rightarrow \sigma_0^2 \alpha_0^{2\nu-d} \beta_0^{2\nu-1}$ almost surely as $K, T \rightarrow \infty$.*

Theorem 3 extends Theorem 2 by allowing the model parameters to be estimated rather than fixed. As argued in Kaufman and Shaby (2013), the data contain information about the model parameters although the estimates are inconsistent. Simulation studies in Section 3 reveal that $\hat{\alpha}$ and $\hat{\beta}$ are often close to the true values and their use could prevent arbitrarily fixing these parameters.

2.4. Asymptotic theory under the non-separable model

When $\varepsilon = 0$ an additional condition is needed to establish the consistency of the micro-ergodic parameter.

Theorem 4. *Under the conditions in Theorem 1 and $\varepsilon = 0$, for any fixed constants $\alpha_1 > 0$ and $\beta_1 > 0$, if $\beta_1/\alpha_1 = \beta_0/\alpha_0 = r$ and $\hat{\sigma}_n^2$ is obtained from (2.6) with $\Gamma_{\alpha,\beta}^{-1}$ replaced by $\Gamma_{\alpha_1,\beta_1}^{-1}$, then $\hat{\sigma}_n^2 (\beta_1/\alpha_1)^d \beta_1^{2\nu-d-1} \rightarrow \sigma_0^2 (\beta_0/\alpha_0)^d \beta_0^{2\nu-d-1}$ almost surely as $K, T \rightarrow \infty$ under P_0 , the Gaussian probability measure defined by the covariogram $\sigma_0^2 M_\nu^0(\alpha_0, \beta_0)$.*

We will show that Theorem 4 continues to hold when the model parameters are estimated rather than fixed arbitrarily. Assume α (or, equivalently, β) is fixed as α_2 , then r can be estimated through

$$\begin{aligned} \hat{r}_n &= \operatorname{argmax}_{r \in \mathbb{R}^+} \tilde{\ell}_n(\alpha_2, r\alpha_2; 0) \\ &= \operatorname{argmax}_{r \in \mathbb{R}^+} \left\{ -\frac{KT}{2} \log 2\pi - \frac{KT}{2} \log \left(\frac{\mathbf{X}_n^\top \Gamma_{\alpha_2, r\alpha_2}^{-1} \mathbf{X}_n}{KT} \right) - \frac{1}{2} \log \det \Gamma_{\alpha_2, r\alpha_2} - \frac{KT}{2} \right\}, \end{aligned} \tag{2.12}$$

where $\alpha_2 > 0$ can be any value. In (2.12), \hat{r}_n can be considered as the maximum profile likelihood estimator of r maximized over the positive real line. Hence, the MLE $\hat{\sigma}_n^2$ can be obtained from (2.6) with $\Gamma_{\alpha,\beta}^{-1}$ replaced by $\Gamma_{\alpha_2, \hat{r}_n \alpha_2}^{-1}$. Let

$$\zeta_{\alpha,r}^0 = \hat{\sigma}_n^2 \left(\frac{\beta}{\alpha} \right)^d \beta^{2\nu-d-1} = \frac{r^{2\nu-1} \alpha^{2\nu-d-1}}{KT} \mathbf{X}_n^\top \Gamma_{\alpha,r\alpha}^{-1} \mathbf{X}_n. \tag{2.13}$$

Theorem 5. *Under the conditions in Theorem 1 and $\varepsilon = 0$, if \hat{r}_n is the maximizer of (2.12) for $\alpha_2 > 0$ fixed, and if $\hat{\sigma}_n^2$ is obtained from (2.6), then $\zeta_{\alpha_2, \hat{r}_n}^0 \rightarrow \sigma_0^2 (\beta_0/\alpha_0)^d \beta_0^{2\nu-d-1} = \sigma_0^2 r_0^d \beta_0^{2\nu-d-1}$ almost surely as $K, T \rightarrow \infty$.*

The next result shows that, similar to the separable case, Theorem 5 continues to hold if we estimate α through any method and plug $\hat{\alpha}$ into (2.12) to obtain \hat{r}_n . One way to estimate α is through (2.7), as in the case $\varepsilon = 1$.

Theorem 6. *Under the conditions in Theorem 1 and $\varepsilon = 0$, if $\hat{\alpha}$ is an estimate of α_0 in $[\alpha_L, \alpha_U]$ where $0 < \alpha_L < \alpha_U < \infty$, and \hat{r}_n is the maximizer of (2.12) with $\hat{\alpha}$ plugged in, then if $\hat{\sigma}_n^2$ is obtained from (2.6), $\zeta_{\hat{\alpha}, \hat{r}_n}^0 \rightarrow \sigma_0^2 (\beta_0/\alpha_0)^d \beta_0^{2\nu-d-1} = \sigma_0^2 r_0^d \beta_0^{2\nu-d-1}$ almost surely as $K, T \rightarrow \infty$.*

In the purely spatial setting, Stein (1999), Zhang (2004) and Kaufman and Shaby (2013) showed that interpolations are asymptotically equal when the micro-ergodic parameters are the same, even when the individual parameters are mis-specified. It is conjectured that the interpolation results will also hold under the spatio-temporal framework. We consider it as an important problem and leave it for future study.

3. Simulations

3.1. Simulation set-up

In this simulation study, we set $d = 2$ with the set of true parameters $(\nu_0, \alpha_0, \beta_0, \sigma_0) = (2.0, 0.3, 3.0, 1.0)$. Then the true quantity of $\zeta_{\alpha_0, \beta_0}^1 = \sigma_0^2 \alpha_0^{2\nu-d} \beta_0^{2\nu-1}$ is 2.43, a benchmark for the separable case. Meanwhile, the true ratio r_0 is 10 and $\zeta_{\alpha_0, r_0}^0 = \sigma_0^2 r_0^d \beta_0^{2\nu-d-1}$ is 300, benchmarks for the non-separable case. We fixed the sampling region as $\mathcal{D}_S \times \mathcal{D}_T = [0, 1]^2 \times [0, 1]$.

The spatial coordinates were set as $\{(i/\sqrt{K}, j/\sqrt{K}), i, j = 1, \dots, \sqrt{K}, \sqrt{K} \in \mathbb{N}\}$, and the time coordinates as $\{t/T, t = 1, \dots, T\}$. We took $T = K = \{25, 36, 49\}$, so for each independent replicate there were $n = KT = \{625, 1,296, 2,401\}$ observations. The number of simulation replicates was 500. For each simulation replicate, the observations were generated using the MASS package in R (version 3.2.1) (R Core Team (2015)) developed by Venables and Ripley (2002). The parameters (α, β) were fixed or estimated as given in 11 scenarios.

Scenario 1: $\varepsilon = 1$, fix $(\alpha, \beta) = (0.3, 3.0)$.

Scenario 2: $\varepsilon = 1$, fix $(\alpha, \beta) = (0.4, 2.0)$.

Scenario 3: $\varepsilon = 1$, fix $(\alpha, \beta) = (1.0, 1.0)$.

Scenario 4: $\varepsilon = 1$, estimate (α, β) from (2.7) and (2.9).

Scenario 5: $\varepsilon = 0$, fix $(\alpha, \beta) = (0.3, 3.0)$.

Scenario 6: $\varepsilon = 0$, fix $(\alpha, \beta) = (0.1, 1.0)$ with $\beta/\alpha = r_0$.

Scenario 7: $\varepsilon = 0$, fix $(\alpha, \beta) = (1.0, 1.0)$ with $\beta/\alpha \neq r_0$.

Scenario 8: $\varepsilon = 0$, fix $\alpha = 0.3$ and estimate r from (2.12).

Scenario 9: $\varepsilon = 0$, fix $\alpha = 0.5$ and estimate r from (2.12).

Scenario 10: $\varepsilon = 0$, fix $\alpha = 1.0$ and estimate r from (2.12).

Scenario 11: $\varepsilon = 0$, estimate α and r , respectively, from (2.7) and (2.12).

Scenarios 1 to 4 focus on the separable model while the others focus on the non-separable case.

3.2. Simulation results

Figures 1 and 2 report the distributions of $\hat{\zeta}$ under the different scenarios, via boxplots. The summary statistics of all estimated parameters can be found in Tables 1 to 3 in the supplementary material. Figure 3 is also provided to show the distributions of $\hat{\alpha}$ and \hat{r} under Scenario 11. In general, the simulation results show better performance as sample sizes increase.

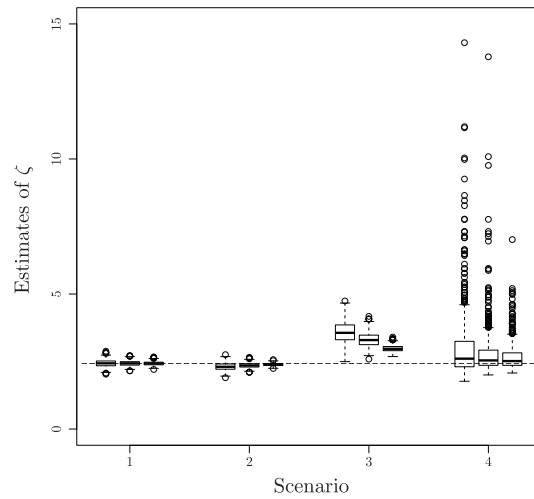


Figure 1. Distributions of $\hat{\zeta}$ under scenarios 1 to 4. For each group of the boxplots, the values of K are 25, 36 and 49 (from left to right). The horizontal dashed line indicates the true value $\sigma_0^2 \alpha_0^{2\nu-d} \beta_0^{2\nu-1} = 2.43$.

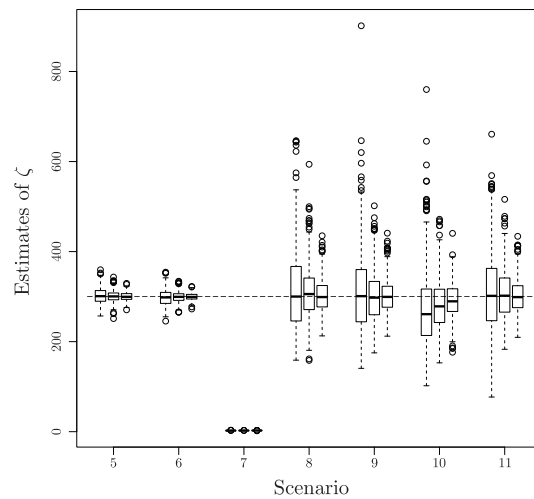


Figure 2. Distributions of $\hat{\zeta}$ under scenarios 5 to 11. For each group of the boxplots, the values of K are 25, 36 and 49 (from left to right). The horizontal dashed line indicates the true value $\sigma_0^2 r_0^d \beta_0^{2\nu-d-1} = 300$.

For Scenarios 1 to 3, it can be seen from Figure 1 that, even when α and β were mis-specified, $\hat{\zeta}$ converges to its true value as sample size increases; the convergence is faster when α and β are closer to the true values. Under Scenario 4, Table 1 in the supplementary material has $\hat{\zeta}$ converging to its true value although $\hat{\alpha}$ and $\hat{\beta}$ are slightly biased and that all the estimates of α , β and σ^2

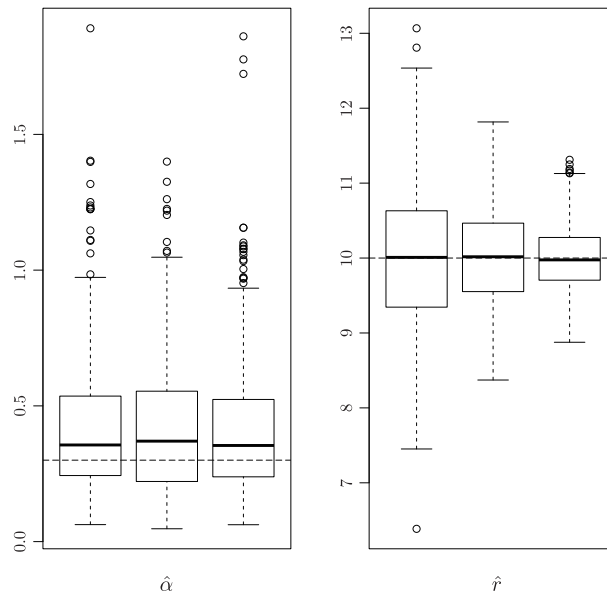


Figure 3. Distributions of $\hat{\alpha}$ (left panel) and \hat{r} (right panel) under scenario 11. For each group of the boxplots, the values of K are 25, 36 and 49 (from left to right). The horizontal dashed lines indicate the true values $\alpha_0 = 0.3$ and $r_0 = 10$.

are inconsistent (the standard deviations of $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\sigma}^2$ show little reductions, if not increased, when K increases from 25 to 49).

Under scenarios 5 to 7, it can be seen that α and β need not be correctly specified, but the ratio β/α needs to be correctly specified (compare Scenarios 6 and 7). Comparing Scenarios 5 and 6, it can be seen that if α and β are closer to the true values, $\hat{\zeta}$ converges more quickly to its true value.

For Scenarios 8 to 10, from Table 3 in the supplementary material, both \hat{r} and $\hat{\zeta}$ appear to be converging to their respective true values as sample size increases. In comparing Scenarios 9 and 10, if α is fixed at a value closer to the true value, both \hat{r} and $\hat{\zeta}$ converge faster.

Under Scenario 11, we allowed α to be estimated from the data. We see from Table 3 in the supplementary material that both \hat{r} and $\hat{\zeta}$ have decreasing standard deviations as sample size increases. Here the standard deviations of $\hat{\alpha}$ remain large as sample size increases. Meanwhile, judging from Figure 3, $\hat{\alpha}$ tends to be over-estimated and the standard deviation seems to be non-decreasing. In practice, unless practitioners have some knowledge regarding the true value of α , it is still reasonable to use $\hat{\alpha}$ to proceed to further estimation. From the right panel of Figure 3, \hat{r} seems to be converging to its true value as sample size increases.

4. Application

In the previous sections, the smoothness parameter ν is known, but knowledge about ν is usually minimal in practice and researchers must estimate it. Not only is the estimation of ν difficult (Bai, Song and Raghunathan (2012, p.820)), the impact on the theoretical results are yet to be investigated if ν was jointly estimated with other parameters. For use of ν as known, we estimate ν based on a grid search, as outlined in Section 4.2. As an illustration of the estimation procedures for the spatio-temporal covariance functions (1.4) and (1.5), we have fitted the models to an air pollution dataset. Section 4.4 studies the effects on estimation when the scale parameters are arbitrarily fixed.

4.1. Data

The dataset was recorded by the California Air Resources Board. It consists of daily averages of nitrogen dioxide (NO₂). The dataset, together with other variables, as recorded on a specific day was studied by Majumdar and Gelfand (2007) and Schmidt and Gelfand (2003) under the purely spatial framework. After removing the stations with missing values over the period from September to October, 2010 (61 days), the final dataset consists of 31 stations (as shown in Figure 4). The maximum distance between stations is approximately 920 km.

The logarithm of the original values were used, in order to achieve approximate normality, as suggested by Schmidt and Gelfand (2003). We also removed the small-scale and monthly effects using a two-way ANOVA model, following Bai, Song and Raghunathan (2012). The mean values of each location over each month were subtracted from the original values to achieve a zero-mean assumption. Moreover, for computational stability, we computed the empirical marginal variance and standardized the variable as suggested by Gneiting, Kleiber and Schlather (2010). Finally, we divided the distance between stations by the maximum distance and the time points by the maximum time point to take the sampling region as $[0, 1]^2 \times [0, 1]$.

4.2. Estimation

We fit both separable and non-separable models to the transformed dataset. For the separable model, (α, β) could be fixed arbitrarily, but we estimated α and β using (2.7) and (2.9) and then estimated σ^2 using (2.6) for fixed values of ν . For the non-separable model, we first estimated α via (2.7), then the ratio r and the variance σ^2 were estimated via (2.12) and (2.6), respectively. We relied on grid search in estimating ν , which turned out to work sufficiently well in our application. We performed grid searches over the interval $(1.0, 2.0]$ for the separable model and over $(1.5, 2.5]$ for the non-separable model, with an

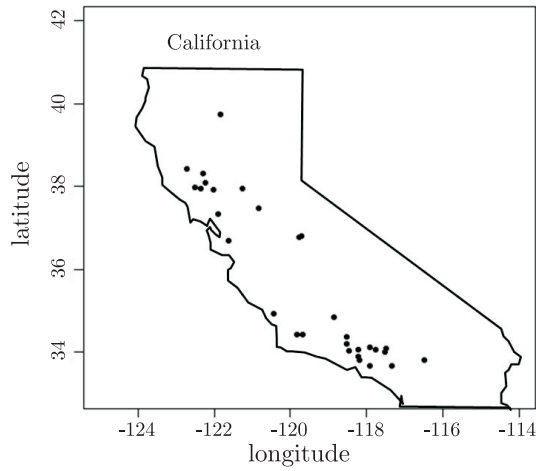


Figure 4. Locations of the sampling stations.

increment of 0.05. Then we performed grid searches with an increment of 0.01 around the values of ν that maximized the log-likelihood values. For each fixed value of ν , α , β , and σ^2 were estimated and the corresponding log-likelihood value was recorded. For each value of $\varepsilon \in \{0, 1\}$, the estimated values of ν were the ones that maximized the log-likelihood values.

4.3. Estimation results

The estimates of the model parameters are in Table 1, while Figure 5 shows the plots of empirical and estimated covariance against distance at different time lags.

From Table 1, the rate of decay in temporal correlation is larger than the rate of decay in spatial correlation, regardless of the degree of separability. Judging from the values of the log-likelihood, the separable model is more favourable. This can also be seen from Figure 5: the separable model (solid line) fits the empirical covariance better than the non-separable model (broken line). The difference between the two models is obvious when time lag is 0 and 1. The difference appeared to be small for higher time lags; we do not show the results for time lags greater than 3 since both empirical and estimated covariances are very close to 0.

4.4. Effect on estimation with arbitrarily fixed parameters

Here we demonstrate the effect on the micro-ergodic parameters when the scale parameters are arbitrarily fixed. The covariance models for NO₂ were refitted. However, instead of estimating α and β , they were fixed arbitrarily.

Table 1. Estimates of parameters for NO₂ under the separable and non-separable models.

ε	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\sigma}^2$	$\hat{\nu}$	$\hat{\zeta}$	log-likelihood
0	1.0420	16.5036	0.8997	1.71	732.640	-1653.06
1	0.3511	32.3782	1.4471	1.15	97.1548	-1549.73

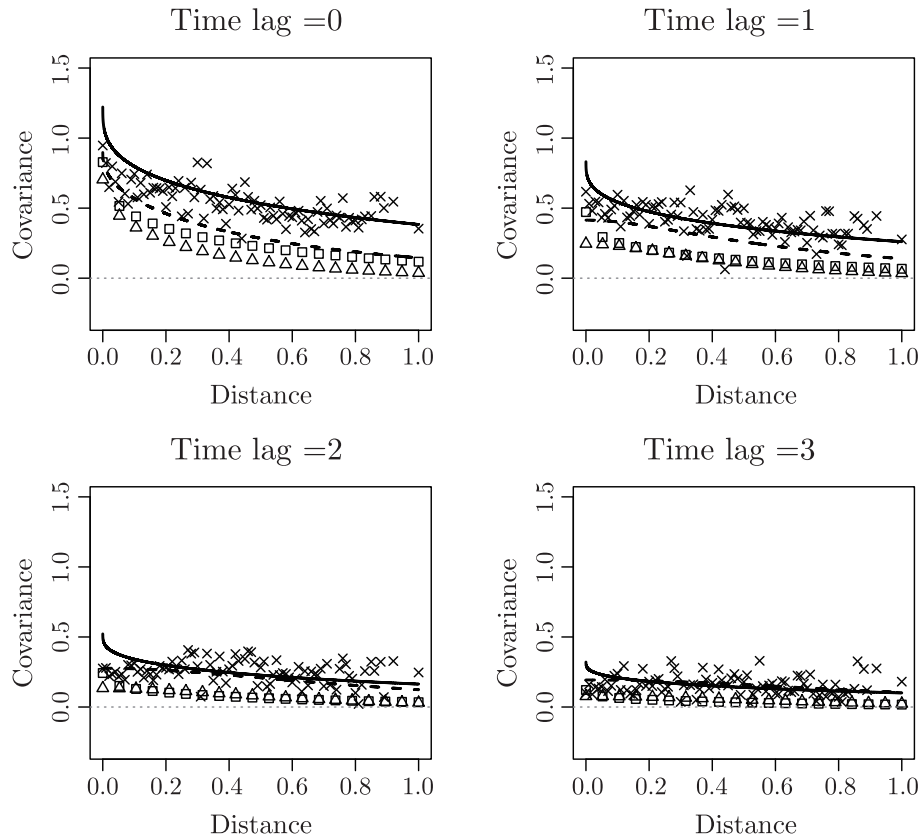


Figure 5. Plots of empirical (cross) and fitted spatial covariance against distance at different time lags (solid line: $\varepsilon = 1, \alpha = \hat{\alpha}, \beta = \hat{\beta}$; broken line: $\varepsilon = 0, \alpha = \hat{\alpha}, \beta = \hat{\beta}$; square: $\varepsilon = 1, (\alpha, \beta)$ fixed as (0.9, 45); triangle: $\varepsilon = 0, \alpha$ fixed as 2.0).

Optimizations were carried out using the estimated ν 's given in Table 1. Under the separable model, we fixed α and β arbitrarily. Under the non-separable model, α was fixed and β was either fixed or estimated via (2.12). Under all cases, σ^2 was estimated via (2.6). The results are summarized in Table 2.

Under the non-separable model, \hat{r} calculated from Table 1 is 15.838 and $\hat{\zeta}$ is 732.64. Table 2 shows that $\hat{\zeta}$ is almost unchanged when (α, β) is fixed

Table 2. Estimates of r , σ^2 and ζ with arbitrarily fixed parameters.

Condition	\hat{r}	$\hat{\sigma}^2$	$\hat{\zeta}$
$\varepsilon = 0$			
(α, β) fixed as (0.5, 7.919)	—	1.2145	726.499
(α, β) fixed as (0.5, 20)	—	1.1797	6642.62
α fixed as 0.5	16.3460	1.2120	782.575
α fixed as 2.0	13.8701	0.7068	548.972
$\varepsilon = 1$			
(α, β) fixed as (0.2, 28)	—	1.9719	92.5749
(α, β) fixed as (0.7, 12)	—	3.7402	84.9873
(α, β) fixed as (0.9, 45)	—	0.8410	114.885

as (0.5, 7.919), where the ratio $r = \beta/\alpha = 15.838$. Yet, when (α, β) is fixed as (0.5, 20), with $r = 40$, $\hat{\zeta}$ was changed by a considerable amount. These observations reflect the conditions required for the convergence of $\hat{\zeta}$ in Theorem 4. Table 2 also shows that when α is fixed arbitrarily, \hat{r} and $\hat{\zeta}$ are fairly close to the corresponding values given in Table 1. The fitted covariance when α is fixed at 2.0 is provided in Figure 5 (triangles). The fitted covariance disagrees with the empirical covariance by a larger extent, compared with the broken line. This demonstrates the disadvantages of fixing the scale parameters arbitrarily rather than estimating it.

Under the separable model, $\hat{\zeta}$ given in Table 1 is 97.1548. From Table 2, it can be seen that, regardless of the arbitrarily fixed (α, β) , $\hat{\zeta}$ remains fairly close to 97.1548. The fitted covariance when (α, β) is fixed at (0.9, 45) is provided in Figure 5 (squares). Similar to the non-separable case, the fitted covariance is poor when compared with the solid line. In practice, it is better to estimate the parameters instead of fixing them, although the estimators are inconsistent.

5. Conclusions and Discussions

In this paper, a Matérn class covariance model in the spatio-temporal framework is considered. Under different degrees of separability, conditions leading to equivalent Gaussian measures are found and practical estimation steps are provided. It is found that not all model parameters can be estimated consistently. Nevertheless, micro-ergodic parameters which are consistently estimable under infill asymptotics are reported. Through simulation, it has been shown that the asymptotic result performs fairly well even when the sample size is moderate. The estimation procedures are found to be useful in modelling a set of data. It is worth emphasizing once again that, although we focus on the spatio-temporal case, the results in this paper can be analogously applied to the anisotropic purely

spatial case where the rate of decay in correlation in one coordinate is different from the remaining d coordinates.

Although the consistency of micro-ergodic parameters has been proven, the asymptotic distributions of the estimators under different values of ε remain open. It is conjectured that the results can be obtained by extending Theorem 3 of Wang and Loh (2011). This will be a potential topic for future research.

Throughout, we have focused on the cases $\varepsilon = 0$ and 1, representing the completely non-separable and separable cases. The case for $\varepsilon \in (0, 1)$ should be practically useful. But, without the closed form expression of the covariance function, implementation of likelihood methods would be difficult. Further studies will be conducted for the case $\varepsilon \in (0, 1)$. We believe that the current results provide insights for study of asymptotics for the spatio-temporal covariance models proposed by Gneiting (2002).

We have estimated the parameters through likelihood approaches with Gaussianity assumed. Still, the method is computationally intensive, especially when K or T is large. The situation worsens when both of them are large. For more efficient computing, one might use the composite or approximate likelihood method as given in Bai, Song and Raghunathan (2012) and Bevilacqua et al. (2012), Curriero and Lele (1999), Varin, Reid and Firth (2011), and Vecchia (1988). Another method introduced recently is the “covariance tapering” method of Kaufman, Schervish and Nychka (2008). Statistical properties of Gaussian random fields using covariance tapering techniques can be found in Du, Zhang and Mandrekar (2009), Shaby and Ruppert (2012), and Wang and Loh (2011). It is expected that similar results can be achieved in the spatio-temporal case. The recent trend of using hierarchical models could also ease the computational burden, but the approach is completely different, see Cressie and Wikle (2011) for more details. When there are strong factors suggesting non-Gaussianity or non-stationarity, one might consult de Luna and Genton (2005), Wikle and Royle (2005) and Fonseca and Stell (2011).

Another issue is the estimation of ν . Here ν is assumed known in the theoretical development and is estimated through a grid search method in applications. As noted by Stein (1999) and Bai, Song and Raghunathan (2012), the estimation of ν requires dense spatial (and probably spatio-temporal) data which are usually unavailable in practice. In addition, Handcock and Wallis (1994) noted that it requires expensive computations if ν is not an integer and hence they resorted to the use of Bayesian methods. Investigation on efficient methods for estimating ν is left for future research.

Supplementary Material

The online supplementary material contains the proofs of the theories and detailed simulation results.

Acknowledgements

The authors are grateful to two anonymous referees, an associate editor and a co-editor for their helpful comments and discussions that have improved this work.

References

- Abramowitz, M. and Stegun, I. A. (1972). Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. National Bureau of Standards Applied Mathematics Series 55, tenth printing. Dover.
- Anderes, E. (2010). On the consistent separation of scale and variance for Gaussian random fields. *Ann. Statist.* **38**, 870-893.
- Bai, Y., Song, P. X.-K. and Raghunathan, T. (2012). Joint composite estimating functions in spatiotemporal models. *J. Roy. Statist. Soc. Ser. B* **74**, 799-824.
- Bevilacqua, M., Gaetan, C., Mateu, J. and Porcu, E. (2012). Estimating space and space-time covariance functions for large data sets: A weighted composite likelihood approach. *J. Amer. Statist. Assoc.* **107**, 268-280.
- Bogachev, V. I. (1998). *Gaussian Measures*. American Mathematical Society.
- Cressie, N. (1993). *Statistics for Spatial Data*. Wiley.
- Cressie, N. and Wikle, C. K. (2011). *Statistics for Spatio-Temporal Data*. Wiley.
- Curriero, F. C. and Lele, S. (1999). A composite likelihood approach to semivariogram estimation. *J. Agric. Biol. Environ. Stat.* **4**, 9-28.
- de Luna, X. and Genton, M. G. (2005). Predictive spatio-temporal models for spatially sparse environmental data. *Statist. Sinica* **15**, 547-568.
- Du, J., Zhang, H. and Mandrekar, V. S. (2009). Fixed-domain asymptotic properties of tapered maximum likelihood estimators. *Ann. Statist.* **37**, 3330-3361.
- Fonseca, T. C. O. and Stell, M. F. J. (2011). Non-Gaussian spatiotemporal modelling through scale mixing. *Biometrika* **98**, 761-774.
- Fuentes, M., Chen, L. and Davis, J. M. (2008). A class of nonseparable and nonstationary spatial temporal covariance functions. *Environmetrics* **19**, 487-507.
- Gikhman, I. I. and Skorokhod, A. V. (1974). *The Theory of Stochastic Processes I*. Springer, Berlin.
- Gneiting, T. (2002). Nonseparable, stationary covariance functions for space-time data. *J. Amer. Statist. Assoc.* **97**, 590-600.
- Gneiting, T., Kleiber, W. and Schlather, M. (2010). Matérn cross-covariance functions for multivariate random fields. *J. Amer. Statist. Assoc.* **105**, 1167-1177.
- Guttorp, P. and Gneiting, T. (2006). Studies in the history of probability and statistics XLIX: On the Matérn correlation family. *Biometrika* **93**, 989-995.
- Handcock, M. S. and Wallis, J. R. (1994). An approach to statistical spatial-temporal modeling of meteorological fields. *J. Amer. Statist. Assoc.* **89**, 368-378.
- Ibragimov, I. A. and Rozanov, Y. A. (1978). *Gaussian Random Processes*. Springer-Verlag, New York.
- Kaufman, C. G., Schervish, M. J. and Nychka, D. W. (2008). Covariance tapering for likelihood-based estimation in large spatial data sets. *J. Amer. Statist. Assoc.* **103**, 1545-1555.

- Kaufman, C. G. and Shaby, B. A. (2013). The role of the range parameter for estimation and prediction in geostatistics. *Biometrika* **100**, 473-484.
- Majumdar, A. and Gelfand, A. E. (2007). Multivariate spatial modeling for geostatistical data using convolved covariance functions. *Math. Geol.* **39**, 225-245.
- Mardia, K. V. and Marshall, R. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika* **71**, 135-146.
- Matérn, B. (1986). *Spatial Variation*. 2nd edition. Springer, New York.
- R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Schmidt, A. M. and Gelfand, A. E. (2003). A Bayesian coregionalization approach for multivariate pollutant data. *J. Geophys. Res.-Atmos.* **108**, STS10-1-8.
- Shaby, B. and Ruppert, D. (2012). Tapered covariance: Bayesian estimation and asymptotics. *J. Comput. Graph. Statist.* **21**, 433-452.
- Skorokhod, A. V. and Yadrenko, M. I. (1973). On absolute continuity of measures corresponding to homogeneous Gaussian fields. *Theor. Probab. Appl.* **18**, 27-40.
- Stein, M. L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. Springer-Verlag, New York.
- Stein, M. L. (2004). Equivalence of Gaussian measures for some nonstationary random fields. *J. Statist. Plann. Inference* **123**, 1-11.
- Varin, C., Reid, N. and Firth, D. (2011). An overview of composite likelihood methods. *Statist. Sinica* **21**, 5-42.
- Vecchia, A. V. (1988). Estimation and model identification for continuous spatial processes. *J. Roy. Statist. Soc. Ser. B* **50**, 297-312.
- Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S*. 4th edition. Springer, New York.
- Wang, D. and Loh, W.-L. (2011). On fixed-domain asymptotics and covariance tapering in Gaussian random field models. *Electron. J. Statist.* **5**, 238-269.
- Wikle, C. K. and Royle, J. A. (2005). Dynamic design of ecological monitoring networks for non-Gaussian spatio-temporal data. *Environmetrics* **16**, 507-522.
- Yadrenko, M. I. (1983). *Spectral Theory of Random Fields*. Optimization Software, New York.
- Ying, Z. (1991). Asymptotic properties of a maximum likelihood estimator with data from a Gaussian process. *J. Multivariate Anal.* **36**, 280-296.
- Ying, Z. (1993). Maximum likelihood estimation of parameters under a spatial sampling scheme. *Ann. Statist.* **21**, 1567-1590.
- Zhang, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *J. Amer. Statist. Assoc.* **99**, 250-261.
- Zhang, H. and Cai, W. (2015). When doesn't cokriging outperform kriging? *Statist. Sci.* **30**, 176-180.

School of Computing and Mathematics, Charles Sturt University, New South Wales, Australia.

E-mail: hoip@csu.edu.au

Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong.

E-mail: hrntlwk@hku.hk

(Received February 2015; accepted April 2016)