CIRJE-F-823

Covariance Tapering for Prediction of Large Spatial Data Sets in Transformed Random Fields

Toshihiro Hirano
Graduate School of Economics, University of Tokyo

Yoshihiro Yajima
University of Tokyo

October 2011; Revised in July and December 2012

Covariance Tapering for Prediction of Large Spatial Data Sets in Transformed Random Fields

Toshihiro Hirano*and Yoshihiro Yajima†

November 29, 2012

Abstract

The best linear unbiased predictor (BLUP) is called a kriging predictor and has been widely used to interpolate a spatially correlated random process in scientific areas such as geostatistics. The BLUP is identical with the conditional expectation if an underlying random field is Gaussian and consequently is the optimal predictor in the mean squared error (MSE) sense. However, if an original data takes a nonnegative value or has a skewed distribution, we frequently apply a nonlinear transformation to it to get a data which is nearer Gaussian. Then the optimality of the BLUP for the original data is unclear because it is not Gaussian. Moreover, in many cases, data sets in spatial problems are often so large that a kriging predictor is impractically time-consuming. To reduce the computational complexity, covariance tapering has been developed by Furrer et al. (2006) for large spatial data sets. In this paper we consider covariance tapering in a class of transformed Gaussian models for random fields and show that the BLUP using covariance tapering, the BLUP and the optimal predictor are asymptotically equivalent in the MSE sense if the underlying Gaussian random field has the Matérn covariance function. This is an extension of Furrer et al. (2006). Monte Carlo simulations support theoretical results.

Key words : Covariance tapering, Hermite polynomials, Kriging, Spatial statistics, Spectral density, Transformed random field

1 Introduction

Interpolation of a spatially correlated random process is widely used in mining, hydrology, forestry and other fields. This method is often called kriging in geostatistical literature. It requires the solution of a linear equation based on the covariance matrix of observations in

*Graduate School of Economics, University of Tokyo, Hongo 7-3-1 Bunkyo-ku, Tokyo 113-0033 Japan. Email: 5820754252@mail.ecc.u-tokyo.ac.jp

†Graduate School of Economics, University of Tokyo, Hongo 7-3-1 Bunkyo-ku, Tokyo 113-0033 Japan. Email: yajima@e.u-tokyo.ac.jp

different spatial points that is the size of the number of observations. The operation count for the direct computation of it is of order n^3 with sample size n. Hence as the sample size is larger, the computation becomes a more formidable one in practice. To deal with this problem Furrer et al. (2006) proposed covariance tapering. The basic idea of covariance tapering is to reduce a spatial covariance function to zero beyond some range by multiplying the true spatial covariance function by a positive definite but compactly supported function. Then the resulting covariance matrix is so sparse that it is much easier and faster to obtain the solution. Furrer et al. (2006) proved the asymptotic efficiency of the BLUP using covariance tapering which we call the tapered BLUP for the original BLUP. Furthermore, Zhu and Wu (2010) investigated properties of covariance tapering for convolution-based nonstationary models and proved that the tapered BLUP is asymptotically efficient in specific assumptions. An alternative approach to reduce the computational time is to calculate a spatial prediction based on a small and manageable number of observations that are given in points close to a prediction point. This approach often shows good performance. However, it is not clear how we may choose samples in a neighborhood of the prediction point and theoretical properties are not derived completely. On the other hand, in covariance tapering it is shown that the MSE ratio of the tapered BLUP and the true BLUP converges to 1 as the sample size goes to infinity regardless of the selection of the taper range (Furrer et al. 2006).

Covariance tapering is also used for the estimation of parameters of a covariance function. The log-likelihood function of Gaussian random fields includes the determinant and the inverse of the covariance matrix between the observations in different spatial points, which is difficult to calculate for large data sets. Kaufman et al. (2008) applied covariance tapering to the log-likelihood function and showed that the estimators maximizing the tapered approximation of the log-likelihood are strongly consistent. Du et al. (2009) proved that this tapered maximum likelihood estimator has the asymptotic normality in one dimensional case. Recently, Wang and Loh (2011) showed the asymptotic normality of the tapered maximum likelihood estimator in multidimensional case by letting the taper range converge to 0 when the sample size goes to infinity.

The BLUP is identical with the conditional expectation if an underlying random field is Gaussian and consequently is the optimal predictor in the MSE sense. However, if an original data takes a nonnegative value or has a skewed distribution, we frequently apply a nonlinear transformation to it to get a data which is nearer Gaussian. Typical ones are a chi-squared process and a lognormal process (Cressie 1993). For example a precipitation data is approximately regarded as a chi-squared process because the standardized square

root values known as anomalies are closer to a Gaussian distribution (Johns et al. 2003; Furrer et al. 2006). On the other hand the variable such as topsoil concentrations of cobalt and copper takes a positive value and has a right skewed sampling distribution. This kind of spatial data is often obtained in large numbers and modeled by the lognormal distribution (Moyeed and Papritz 2002; De Oliveira 2006). However the optimality of the BLUP and the tapered BLUP for an original data is not clear because it is non-Gaussian. In this paper we show that the tapered BLUP and the BLUP in a class of transformed Gaussian models for random fields are asymptotically equivalent to the conditional expectation, which is the optimal predictor in the MSE sense. This is an extension of Furrer et al.(2006).

Granger and Newbold (1976) considered a class of the transformed models in a time series context and calculated the mean, the covariance function and the mean squared error of predictors. Our work can be also regarded as an extension of their results to spatial processes. Moreover since the conditional mean and the BLUP include the inverse of the covariance matrix, covariance tapering is useful to reduce the computational difficulty.

The subsequent sections are organized as follows. We introduce a kriging predictor and asymptotic properties in Section 2. In Section 3, we introduce a class of transformed Gaussian models for random fields and covariance tapering and then prove that the tapered BLUP of transformed random fields has the asymptotic efficiency with respect to the optimal predictor. In Section 4, computer experiments are conducted. A conclusion and future studies are mentioned in Section 5. Supplementary materials and proofs of the lemmas, the corollary and the theorem are given in Appendices A-D.

2 Spatial prediction and its asymptotic properties

Let $\{Z(\boldsymbol{s}), \boldsymbol{s} \in \mathbb{R}^d\}$ be a random field with known constant mean and covariance function. Suppose we have observations from a single realization of the random field $Z(\boldsymbol{s})$, denoted by $\boldsymbol{Z} = (Z(\boldsymbol{s}_1), \ldots, Z(\boldsymbol{s}_n))'$, measured at sampling locations $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n \in \mathbb{R}^d$. The goal is to predict $Z(\boldsymbol{s}_0)$ at an unobserved location $\boldsymbol{s}_0 \in \mathbb{R}^d$ based on $\boldsymbol{Z}$. Then the best linear unbiased predictor (BLUP) and its mean squared error (MSE) are

$$\hat{Z}^{BLUP}(\boldsymbol{s}_0) = \mu_Z + \boldsymbol{c}_Z' \Sigma_Z^{-1} (\boldsymbol{Z} - \mu_Z \cdot \boldsymbol{1}), \tag{1}$$

$$E[\hat{Z}^{BLUP}(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)]^2 = \sigma_Z^2 - \boldsymbol{c}_Z' \Sigma_Z^{-1} \boldsymbol{c}_Z, \tag{2}$$

where $\mu_Z = E[Z(\boldsymbol{s})]$, $\sigma_Z^2 = \mathrm{var}(Z(\boldsymbol{s}_0))$, $(\boldsymbol{c}_Z)_i = \mathrm{cov}(Z(\boldsymbol{s}_0), Z(\boldsymbol{s}_i))$, $(\Sigma_Z)_{ij} = \mathrm{cov}(Z(\boldsymbol{s}_i), Z(\boldsymbol{s}_j))$ $(i, j = 1, \ldots, n)$ and $\boldsymbol{1} = (1, \ldots, 1)'$. $\hat{Z}^{BLUP}(\boldsymbol{s}_0)$ and $E[\hat{Z}^{BLUP}(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)]^2$ are called as the simple kriging predictor and the simple kriging variance of $Z(\boldsymbol{s}_0)$ respectively (see, e.g., Cressie 1993; Stein 1999).

Hereafter we assume that $\{Z(\boldsymbol{s})\}$ is a stationary random field and the covariance function $C(\boldsymbol{h})$ is defined by

$$C(\boldsymbol{h}) \ = \mathrm{cov}(Z(\boldsymbol{s}), Z(\boldsymbol{s}+\boldsymbol{h})), \quad \boldsymbol{h} \in \mathbb{R}^d.$$

In what follows, we write $\boldsymbol{h}_i = \boldsymbol{s}_i - \boldsymbol{s}_0, \boldsymbol{h}_{ij} = \boldsymbol{s}_i - \boldsymbol{s}_j$ $(i, j = 1, \ldots, n)$.

Now we will review some asymptotic properties of the kriging predictor using an incorrect covariance function, which are used in the proof of our main theorem. Hereafter we assume that sampling locations $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n$ are in $D \subset \mathbb{R}^d$ where D is a bounded subset and a prediction point $\boldsymbol{s}_0 \in D$ is an accumulation point of $\{\boldsymbol{s}_i, i = 1, 2, \ldots\}$, that is, $\boldsymbol{s}_0 \in \overline{\{\boldsymbol{s}_i, i = 1, 2, \ldots\}}$ where $\overline{\{\boldsymbol{s}_i, i = 1, 2, \ldots\}}$ is the closure of $\{\boldsymbol{s}_i, i = 1, 2, \ldots\}$. This assumption is called infill asymptotics and often used as an asymptotic framework in spatial statistics (Cressie 1993).

$C_0(\boldsymbol{h})$ and $C_1(\boldsymbol{h})$ denote the covariance functions of the true and the assumed models respectively. Then from (1) the BLUP under $C_i(\boldsymbol{h})$ $(i = 0, 1)$ is $\hat{Z}^{BLUP}(\boldsymbol{s}_0, C_i) = \mu_Z + \boldsymbol{c}_i' \Sigma_i^{-1}(\boldsymbol{Z} - \mu_Z \cdot \boldsymbol{1})$ where $(\boldsymbol{c}_i)_k = C_i(\boldsymbol{h}_k)$ and $(\Sigma_i)_{kl} = C_i(\boldsymbol{h}_{kl})$ $(k, l = 1, \ldots, n)$. Let $Z^*(\boldsymbol{s}_0)$ be a predictor of $Z(\boldsymbol{s}_0)$ and let $E_{C_i}[Z^*(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)]^2$ be the MSE of $Z^*(\boldsymbol{s}_0)$ under C_i $(i = 0, 1)$. This MSE may depend not only on C_i but also on other distributional properties if $Z^*(\boldsymbol{s}_0)$ is a nonlinear predictor. Then $Z^*(\boldsymbol{s}_0)$ is called a consistent predictor if

$$E_{C_0}[Z^*(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)]^2 \to 0 \quad \text{as } n \to \infty.$$

Furthermore, a consistent predictor $Z^*(\boldsymbol{s}_0)$ is called asymptotically efficient with respect to the BLUP $\hat{Z}^{BLUP}(\boldsymbol{s}_0, C_0)$ if

$$\frac{E_{C_0}[Z^*(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)]^2}{E_{C_0}[\hat{Z}^{BLUP}(\boldsymbol{s}_0, C_0) - Z(\boldsymbol{s}_0)]^2} \to 1 \quad \text{as } n \to \infty.$$

Next let $\hat{Z}(\boldsymbol{s}_0)$ be the optimal predictor of $Z(\boldsymbol{s}_0)$ which minimizes the MSE under the true model. Then if

$$\frac{E_{C_0}[Z^*(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)]^2}{E_{C_0}[\hat{Z}(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)]^2} \to 1 \quad \text{as } n \to \infty,$$

$Z^*(\boldsymbol{s}_0)$ is called asymptotically efficient with respect to the optimal predictor $\hat{Z}(\boldsymbol{s}_0)$. If $\{Z(\boldsymbol{s})\}$ is Gaussian, the two definitions of the asymptotic efficiency are equivalent because $\hat{Z}(\boldsymbol{s}_0) = \hat{Z}^{BLUP}(\boldsymbol{s}_0, C_0)$.

Stein (1993) gave the following sufficient conditions for the asymptotic efficiency of the incorrect kriging predictor with respect to the BLUP. Hereafter, $\|\cdot\|$ means the Euclidean norm and a function is called bandlimited if it is the Fourier transform of a function with bounded support.

Theorem 1 (Stein(1993)) Let $f_i(\boldsymbol{\lambda}), \boldsymbol{\lambda} \in \mathbb{R}^d$ be the spectral density function of $C_i(\boldsymbol{h})$ $(i = 0, 1)$. Suppose that

$$0 < \liminf_{\|\boldsymbol{\lambda}\| \to \infty} f_0(\boldsymbol{\lambda})/|\phi(\boldsymbol{\lambda})|^2 \leq \limsup_{\|\boldsymbol{\lambda}\| \to \infty} f_0(\boldsymbol{\lambda})/|\phi(\boldsymbol{\lambda})|^2 < \infty, \tag{3}$$

where $\phi(\boldsymbol{\lambda})$ is bandlimited and $\lim_{\|\boldsymbol{\lambda}\| \to \infty} f_1(\boldsymbol{\lambda})/f_0(\boldsymbol{\lambda}) = c\ (> 0)$. Then as $n \to \infty$

$$\frac{E_{C_0}[\hat{Z}^{BLUP}(\boldsymbol{s}_0, C_1) - Z(\boldsymbol{s}_0)]^2}{E_{C_0}[\hat{Z}^{BLUP}(\boldsymbol{s}_0, C_0) - Z(\boldsymbol{s}_0)]^2} = \frac{C_0(\boldsymbol{0}) - 2\boldsymbol{c}_0{}'{\Sigma_1}^{-1}\boldsymbol{c}_1 + \boldsymbol{c}_1{}'{\Sigma_1}^{-1}\Sigma_0{\Sigma_1}^{-1}\boldsymbol{c}_1}{C_0(\boldsymbol{0}) - \boldsymbol{c}_0{}'\Sigma_0^{-1}\boldsymbol{c}_0} \to 1$$

and

$$\frac{E_{C_1}[\hat{Z}^{BLUP}(\boldsymbol{s}_0, C_1) - Z(\boldsymbol{s}_0)]^2}{E_{C_0}[\hat{Z}^{BLUP}(\boldsymbol{s}_0, C_1) - Z(\boldsymbol{s}_0)]^2} = \frac{C_1(\boldsymbol{0}) - \boldsymbol{c}_1'\Sigma_1^{-1}\boldsymbol{c}_1}{C_0(\boldsymbol{0}) - 2\boldsymbol{c}_0{}'{\Sigma_1}^{-1}\boldsymbol{c}_1 + \boldsymbol{c}_1{}'{\Sigma_1}^{-1}\Sigma_0{\Sigma_1}^{-1}\boldsymbol{c}_1} \to c.$$

It is known that if

$$0 < \liminf_{\|\boldsymbol{\lambda}\| \to \infty} f_0(\boldsymbol{\lambda})\|\boldsymbol{\lambda}\|^r \leq \limsup_{\|\boldsymbol{\lambda}\| \to \infty} f_0(\boldsymbol{\lambda})\|\boldsymbol{\lambda}\|^r < \infty$$

for some $r > d$, then (3) is satisfied (see Stein 1993, p.402). Theorem 1 states that the low frequency behavior of the spectral density function has little impact on the kriging prediction.

3 Covariance tapering in transformed random fields

This section introduces a class of transformed Gaussian models for random fields and covariance tapering and derives the asymptotic optimality of the tapered BLUP.

3.1 Transformed random fields

Assume that $\{Y(\boldsymbol{s}), \boldsymbol{s} \in D \subset \mathbb{R}^d\}$ is an isotropic Gaussian random field with mean $\mu_Y = E[Y(\boldsymbol{s})]$ and Matérn covariance function defined by

$$C_Y(\|\boldsymbol{h}\|) = \frac{\sigma_Y^2}{2^{\nu-1}\Gamma(\nu)}(\alpha\|\boldsymbol{h}\|)^\nu K_\nu(\alpha\|\boldsymbol{h}\|), \quad \alpha > 0,\ \nu > 0,\ \sigma_Y^2 > 0,$$

where $K_\nu(\cdot)$ is the modified Bessel function of the second kind of order ν (see, e.g., Cressie 1993; Stein 1999). The spectral density function of this covariance function is

$$f_Y(\|\boldsymbol{\lambda}\|) = \frac{A\sigma_Y^2}{(\alpha^2 + \|\boldsymbol{\lambda}\|^2)^{\nu+d/2}},$$

where $A = (\Gamma(\nu+d/2)\alpha^{2\nu})/(\pi^{d/2}\Gamma(\nu))$. For example if $\nu = 0.5$, Matérn covariance function is

$$C_Y(\|\boldsymbol{h}\|) = \sigma_Y^2 \exp(-\alpha\|\boldsymbol{h}\|).$$

It is called the exponential covariance function and widely used in many applications. For simplicity, we set $\mu_Y = 0$ and $\sigma_Y^2 = 1$.

Now let $Z(\boldsymbol{s}) = T(Y(\boldsymbol{s}))$ be an observable random field with the underlying field $\{Y(\boldsymbol{s})\}$ where $T(x)$ is an unknown real function, which satisfies $\int_{-\infty}^{\infty} T(x)^2\phi(x)dx < \infty$ and $\phi(x) = e^{-x^2/2}/\sqrt{2\pi}$ is the density function of $N(0,1)$. Then $Z(\boldsymbol{s})$ can be expressed by an infinite sum of Hermite polynomials

$$Z(\boldsymbol{s}) = \sum_{j=0}^{\infty} \alpha_j H_j(Y(\boldsymbol{s})),$$

where $H_j(x)$ $(j = 0, 1, 2, \ldots)$ is the j-th order Hermite polynomial (see Appendix A for properties of Hermite polynomials). Note that α_j's depend on T. By (A.1) and (A.3) of Appendix A,

$$\mu_Z = E[Z(\boldsymbol{s})] = \alpha_0, \tag{4}$$

$$C_Z(\|\boldsymbol{h}\|) = Cov(Z(\boldsymbol{s}), Z(\boldsymbol{s}+\boldsymbol{h})) = \sum_{j\geq 1} \alpha_j^2 j!(C_Y(\|\boldsymbol{h}\|))^j. \tag{5}$$

Next we will derive the optimal predictor in transformed random fields by a similar discussion of Granger and Newbold (1976).

Define

$$W(\boldsymbol{s}_0) = \frac{Y(\boldsymbol{s}_0) - \boldsymbol{c}_Y'\Sigma_Y^{-1}\boldsymbol{Y}}{\sqrt{1 - \boldsymbol{c}_Y'\Sigma_Y^{-1}\boldsymbol{c}_Y}},$$

where $\boldsymbol{Y} = (Y(\boldsymbol{s}_1), \ldots, Y(\boldsymbol{s}_n))'$, $(\boldsymbol{c}_Y)_i = C_Y(\|\boldsymbol{h}_i\|)$ and $(\Sigma_Y)_{ij} = C_Y(\|\boldsymbol{h}_{ij}\|)$ $(i, j = 1, \ldots, n)$. Then the conditional distribution of $W(\boldsymbol{s}_0)$ given $\boldsymbol{Y}$ is $N(0,1)$ and $Y(\boldsymbol{s}_0)$ is expressed by

$$Y(\boldsymbol{s}_0) = \sqrt{1 - \boldsymbol{c}_Y'\Sigma_Y^{-1}\boldsymbol{c}_Y}W(\boldsymbol{s}_0) + \boldsymbol{c}_Y'\Sigma_Y^{-1}\boldsymbol{Y}.$$

Then from (2.7.6) of Brockwell and Davis (1991; page 63), (A.2) and (A.5) of Appendix

A, the optimal predictor $\hat{Z}(\boldsymbol{s}_0)$ of $Z(\boldsymbol{s}_0)$, that is, the conditional mean given $\boldsymbol{Y}$ is

$$\begin{aligned}
\hat{Z}(\boldsymbol{s}_0) &= E[Z(\boldsymbol{s}_0)|\boldsymbol{Y}] \\
&= \sum_{j=0}^{\infty} \alpha_j E\left[H_j\left(\sqrt{1-\boldsymbol{c}_Y'\Sigma_Y^{-1}\boldsymbol{c}_Y}W(\boldsymbol{s}_0) + \sqrt{\boldsymbol{c}_Y'\Sigma_Y^{-1}\boldsymbol{c}_Y}\frac{\boldsymbol{c}_Y'\Sigma_Y^{-1}\boldsymbol{Y}}{\sqrt{\boldsymbol{c}_Y'\Sigma_Y^{-1}\boldsymbol{c}_Y}} \right) \middle| \boldsymbol{Y} \right] \\
&= \sum_{j=0}^{\infty} \alpha_j E\left[\sum_{k=0}^{j} \binom{j}{k} \left(\sqrt{1-\boldsymbol{c}_Y'\Sigma_Y^{-1}\boldsymbol{c}_Y}\right)^k \left(\sqrt{\boldsymbol{c}_Y'\Sigma_Y^{-1}\boldsymbol{c}_Y}\right)^{j-k} H_k(W(\boldsymbol{s}_0)) \right. \\
&\quad \left. \times H_{j-k}\left(\frac{\boldsymbol{c}_Y'\Sigma_Y^{-1}\boldsymbol{Y}}{\sqrt{\boldsymbol{c}_Y'\Sigma_Y^{-1}\boldsymbol{c}_Y}} \right) \middle| \boldsymbol{Y} \right] \\
&= \sum_{j=0}^{\infty} \alpha_j \sum_{k=0}^{j} \binom{j}{k} \left(\sqrt{1-\boldsymbol{c}_Y'\Sigma_Y^{-1}\boldsymbol{c}_Y}\right)^k \left(\sqrt{\boldsymbol{c}_Y'\Sigma_Y^{-1}\boldsymbol{c}_Y}\right)^{j-k} H_{j-k}\left(\frac{\boldsymbol{c}_Y'\Sigma_Y^{-1}\boldsymbol{Y}}{\sqrt{\boldsymbol{c}_Y'\Sigma_Y^{-1}\boldsymbol{c}_Y}} \right) \\
&\quad \times E[H_k(W(\boldsymbol{s}_0))|\boldsymbol{Y}] \\
&= \sum_{j=0}^{\infty} \alpha_j \left(\sqrt{\boldsymbol{c}_Y'\Sigma_Y^{-1}\boldsymbol{c}_Y}\right)^j H_j\left(\frac{\boldsymbol{c}_Y'\Sigma_Y^{-1}\boldsymbol{Y}}{\sqrt{\boldsymbol{c}_Y'\Sigma_Y^{-1}\boldsymbol{c}_Y}} \right), \qquad (6)
\end{aligned}$$

and from (A.1) of Appendix A, its mean squared error is

$$\begin{aligned}
E[\hat{Z}(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)]^2 &= E[Z(\boldsymbol{s}_0)]^2 - E[\hat{Z}(\boldsymbol{s}_0)]^2 \\
&= \sum_{j=1}^{\infty} \alpha_j^2 j! - \sum_{j=1}^{\infty} \alpha_j^2 j! (\boldsymbol{c}_Y'\Sigma_Y^{-1}\boldsymbol{c}_Y)^j. \qquad (7)
\end{aligned}$$

$\hat{Z}(\boldsymbol{s}_0)$ is infeasible because $T(\cdot)$ is unknown. From (1) and (2), the BLUP of $Z(\boldsymbol{s}_0)$ is

$$\hat{Z}^{BLUP}(\boldsymbol{s}_0) = \mu_Z + \boldsymbol{c}_Z'\Sigma_Z{}^{-1}(\boldsymbol{Z} - \mu_Z \cdot \boldsymbol{1}),$$

and its mean squared error is

$$E[\hat{Z}^{BLUP}(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)]^2 = C_Z(0) - \boldsymbol{c}_Z'\Sigma_Z^{-1}\boldsymbol{c}_Z,$$

where $\boldsymbol{Z} = (Z(\boldsymbol{s}_1), \ldots, Z(\boldsymbol{s}_n))' = (T(Y(\boldsymbol{s}_1)), \ldots, T(Y(\boldsymbol{s}_n)))'$, $(\boldsymbol{c}_Z)_i = C_Z(\|\boldsymbol{h}_i\|)$ and $(\Sigma_Z)_{ij} = C_Z(\|\boldsymbol{h}_{ij}\|)$ $(i, j = 1, \ldots, n)$. Hereafter we assume that $\mu_z = \alpha_0$ and $C_Z(\|\boldsymbol{h}\|)$ are known. If they are unknown, the BLUP and the tapered BLUP introduced in the next section are also infeasible as the optimal predictor is. However, the estimators of these predictors are easily constructed by the sample mean and covariance of the observable $\{Z(\boldsymbol{s})\}$. Whereas that of the optimal predictor depends on the unobservable $\{Y(\boldsymbol{s})\}$ and requires a rather difficult procedure. A more rigorous discussion on the BLUP and the tapered BLUP with estimated coefficients is left for a future study. Some comments are given in Section 5.

3.2 Covariance tapering

The computational complexity of $\hat{Z}^{BLUP}(\boldsymbol{s}_0)$ is challenging for large spatial data sets because it includes the inverse of the covariance matrix with the same size as the number of observations. To reduce the computational burden, covariance tapering has been developed by Furrer et al. (2006). They show the asymptotic optimality of covariance tapering in the Gaussian random field with Matérn covariance function, which corresponds to the case of $T(x) = x$ in the framework of the transformed random field.

We will review the result of Furrer et al. (2006). Let $T(x) = x$ and hence $Z(\boldsymbol{s}) = Y(\boldsymbol{s})$ in this subsection. From (1) and (2), the BLUP of $Y(\boldsymbol{s}_0)$ and its MSE are defined by

$$\hat{Y}^{BLUP}(\boldsymbol{s}_0) = \boldsymbol{c}_Y' \Sigma_Y^{-1} \boldsymbol{Y},$$
$$E[\hat{Y}^{BLUP}(\boldsymbol{s}_0) - Y(\boldsymbol{s}_0)]^2 = 1 - \boldsymbol{c}_Y' \Sigma_Y^{-1} \boldsymbol{c}_Y.$$

Let $C_\theta(x)$ be a compactly supported correlation function with $C_\theta(0) = 1$ and $C_\theta(x) = 0$ for $x \geq \theta$. $C_\theta(x)$ is called the taper function with the taper range θ. Then consider the product of the original covariance function and the taper function, that is

$$C_{tap}^Y(\|\boldsymbol{h}\|) = C_Y(\|\boldsymbol{h}\|) C_\theta(\|\boldsymbol{h}\|).$$

Now we replace $C_Y(\|\boldsymbol{h}\|)$ in $\hat{Y}^{BLUP}(\boldsymbol{s}_0)$ with $C_{tap}^Y(\|\boldsymbol{h}\|)$ and obtain the tapered BLUP

$$\hat{Y}^{tapBLUP}(\boldsymbol{s}_0) = \boldsymbol{c}_Y^{tap\prime} \Sigma_Y^{tap^{-1}} \boldsymbol{Y},$$

where $(\boldsymbol{c}_Y^{tap})_i = C_{tap}^Y(\|\boldsymbol{h}_i\|)$ and $(\Sigma_Y^{tap})_{ij} = C_{tap}^Y(\|\boldsymbol{h}_{ij}\|)$ $(i, j = 1, \ldots, n)$. The resulting covariance matrix Σ_Y^{tap} has many zero elements and is called a sparse matrix, so that it is much faster to calculate $\Sigma_Y^{tap^{-1}} \boldsymbol{Y}$ than $\Sigma_Y^{-1} \boldsymbol{Y}$. Next let $f_\theta(\|\boldsymbol{\lambda}\|)$ be the spectral density function of $C_\theta(\|\boldsymbol{h}\|)$. Then Furrer et al. (2006) imposed the following condition.

Taper condition: For some $\epsilon > 0$ and $M_\theta < \infty$

$$0 < f_\theta(\|\boldsymbol{\lambda}\|) \leq \frac{M_\theta}{(1 + \|\boldsymbol{\lambda}\|^2)^{\nu + d/2 + \epsilon}}.$$

Although $\hat{Y}^{tapBLUP}(\boldsymbol{s}_0)$ is not the BLUP under the true covariance function $C_Y(\cdot)$, Furrer et al. (2006) showed that this predictor is asymptotically efficient with respect to the BLUP.

Theorem 2 (Furrer et al. (2006)) If f_θ satisfies the taper condition, Theorem 1 holds with $C_0 = C_Y$, $C_1 = C_Y C_\theta$, that is

$$\frac{E_{C_Y}[\hat{Y}^{tapBLUP}(\boldsymbol{s}_0) - Y(\boldsymbol{s}_0)]^2}{E_{C_Y}[\hat{Y}^{BLUP}(\boldsymbol{s}_0) - Y(\boldsymbol{s}_0)]^2} = \frac{1 - 2\boldsymbol{c}_Y{}' \Sigma_Y^{tap^{-1}} \boldsymbol{c}_Y^{tap} + \boldsymbol{c}_Y^{tap\prime} \Sigma_Y^{tap^{-1}} \Sigma_Y \Sigma_Y^{tap^{-1}} \boldsymbol{c}_Y^{tap}}{1 - \boldsymbol{c}_Y' \Sigma_Y^{-1} \boldsymbol{c}_Y}$$
$$\to 1 \quad \text{as } n \to \infty.$$

Theorem 2 also means the asymptotic efficiency of the tapered BLUP with respect to the optimal predictor because $\hat{Y}^{BLUP}$ is the optimal predictor in the Gaussian random field $\{Y(\boldsymbol{s})\}$.

In this paper, we shall show that for any transformation $T(x)$ the tapered BLUP of $Z(\boldsymbol{s}_0)$ is asymptotically efficient not only with respect to $\hat{Z}^{BLUP}(\boldsymbol{s}_0)$ but also with respect to $\hat{Z}(\boldsymbol{s}_0)$.

3.3 Main result

As in Furrer et al. (2006), the tapered BLUP of $Z(\boldsymbol{s}_0)$ is defined by

$$\hat{Z}^{tapBLUP}(\boldsymbol{s}_0) = \mu_Z + \boldsymbol{c}_Z^{tap\prime}\Sigma_Z^{tap-1}(\boldsymbol{Z} - \mu_Z \cdot \boldsymbol{1}),$$

where $(\boldsymbol{c}_Z^{tap})_i = C_{tap}^Z(\|\boldsymbol{h}_i\|)$, $(\Sigma_Z^{tap})_{ij} = C_{tap}^Z(\|\boldsymbol{h}_{ij}\|)$ and $C_{tap}^Z(\|\boldsymbol{h}\|) = C_Z(\|\boldsymbol{h}\|)C_\theta(\|\boldsymbol{h}\|)$ $(i, j = 1, \ldots, n)$. Hereafter it is assumed that $0 < \sum_{j=1}^{\infty} \alpha_j^2 j! j$ otherwise $\alpha_j = 0$ for any $j \geq 1$ and $Z(\boldsymbol{s}) = T(Y(\boldsymbol{s})) = c$ where c is a constant. $E[\hat{Z}(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)]^2$ of (7) and $E[\hat{Z}^{tapBLUP}(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)]^2$ are denoted by $E_{C_Z}[\hat{Z}(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)]^2$ and $E_{C_Z}[\hat{Z}^{tapBLUP}(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)]^2$ to highlight the expectation under the true model for the observable process $\{Z(\boldsymbol{s})\}$ though $Z(\boldsymbol{s})$ and C_Z are functions of the underlying $Y(\boldsymbol{s})$ and C_Y respectively.

The following theorem is our main result.

Theorem 3 Suppose that $\sum_{j=1}^{\infty} \alpha_j^2 j! j^{2\nu+d+\max\{1,2\epsilon\}} < \infty$ where ϵ is given in the taper condition. If f_θ satisfies the taper condition, the tapered BLUP, the BLUP and the optimal predictor are asymptotically equivalent in the MSE sense, that is

$$\frac{E_{C_Z}[\hat{Z}^{tapBLUP}(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)]^2}{E_{C_Z}[\hat{Z}(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)]^2} = \frac{C_Z(0) - 2\boldsymbol{c}_Z{}'\Sigma_Z^{tap-1}\boldsymbol{c}_Z^{tap} + \boldsymbol{c}_Z^{tap\prime}\Sigma_Z^{tap-1}\Sigma_Z\Sigma_Z^{tap-1}\boldsymbol{c}_Z^{tap}}{\sum_{j=1}^{\infty}\alpha_j^2 j! - \sum_{j=1}^{\infty}\alpha_j^2 j!(\boldsymbol{c}_Y'\Sigma_Y^{-1}\boldsymbol{c}_Y)^j}$$
$$\to 1,$$
$$\frac{E_{C_Z}[\hat{Z}^{tapBLUP}(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)]^2}{E_{C_Z}[\hat{Z}^{BLUP}(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)]^2} \to 1$$

and

$$\frac{E_{C_Z}[\hat{Z}^{BLUP}(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)]^2}{E_{C_Z}[\hat{Z}(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)]^2} \to 1$$

as $n \to \infty$.

Theorem 3 states an extension of Theorem 2 by Furrer et al. (2006) in terms of the asymptotic efficiency of the tapered BLUP not only with respect to the BLUP but also with respect to the optimal nonlinear predictor. Moreover, it is also shown that the BLUP is

asymptotically efficient with respect to the optimal one. One may evaluate the presumed mean squared error $E_{C^Z_{tap}}[\hat{Z}^{tapBLUP}(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)]^2 = C^Z_{tap}(0) - \boldsymbol{c}_Z^{tap\prime}\Sigma_Z^{tap-1}\boldsymbol{c}_Z^{tap}$ to assess a prediction uncertainty. The following corollary shows that in the transformed model it is asymptotically equivalent to the MSE of the optimal predictor.

Corollary 1 Under the conditions of Theorem 3,

$$\frac{E_{C^Z_{tap}}[\hat{Z}^{tapBLUP}(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)]^2}{E_{C_Z}[\hat{Z}(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)]^2} \to 1 \quad \text{as } n \to \infty.$$

The tapered BLUP can be calculated by using the only observable data $\{Z(\boldsymbol{s})\}$ when the transformation $T(\cdot)$ is unknown. However, to justify the practical use of the tapered BLUP, the property of the plug-in predictor must be investigated. It is a future work.

We give two examples, which are often applied to an empirical analysis.

EXAMPLE 1 (The squared transformation). If $Z(\boldsymbol{s}) = (\sigma Y(\boldsymbol{s}) + \mu)^2$, $\sigma > 0$ and $Y(\boldsymbol{s}) \sim N(0,1)$, $Z(\boldsymbol{s})$ has the following expression using Hermite polynomials

$$Z(\boldsymbol{s}) = \sigma^2 H_2(Y(\boldsymbol{s})) + 2\mu\sigma H_1(Y(\boldsymbol{s})) + (\mu^2 + \sigma^2)H_0(Y(\boldsymbol{s})).$$

Since $\alpha_0 = \mu^2 + \sigma^2$, $\alpha_1 = 2\mu\sigma$, $\alpha_2 = \sigma^2$ and $\alpha_j = 0$, $j \geq 3$, clearly the assumption of Theorem 3 is satisfied. Then from (4) - (7),

$$\begin{aligned} \mu_Z &= \mu^2 + \sigma^2, \\ C_Z(\|\boldsymbol{h}\|) &= 4\mu^2\sigma^2 C_Y(\|\boldsymbol{h}\|) + 2\sigma^4 (C_Y(\|\boldsymbol{h}\|))^2, \\ \hat{Z}(\boldsymbol{s}_0) &= (\sigma \boldsymbol{c}_Y{}'\Sigma_Y^{-1}\boldsymbol{Y} + \mu)^2 + \sigma^2(1 - \boldsymbol{c}_Y{}'\Sigma_Y^{-1}\boldsymbol{c}_Y) \end{aligned}$$

and

$$E[\hat{Z}(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)]^2 = 4\mu^2\sigma^2(1 - \boldsymbol{c}_Y{}'\Sigma_Y^{-1}\boldsymbol{c}_Y) + 2\sigma^4\{1 - (\boldsymbol{c}_Y{}'\Sigma_Y^{-1}\boldsymbol{c}_Y)^2\}.$$

In general, for any finite order polynomial $Z(\boldsymbol{s}) = \sum_{j=0}^m a_j(\sigma Y(\boldsymbol{s}) + \mu)^j$ with $a_j \in \mathbb{R}$ and $m < \infty$, the corresponding coefficients α_j of the Hermite polynomials satisfy the assumption of Theorem 3 because $\alpha_j = 0$ $(j > m)$.

EXAMPLE 2 (The exponential transformation). Consider the following transformation

$$Z(\boldsymbol{s}) = \exp(\sigma Y(\boldsymbol{s}) + \mu),$$

where $Y(\boldsymbol{s}) \sim N(0,1)$. This kind of random fields $Z(\boldsymbol{s})$ is called a log-Gaussian random field. Then by (A.4) of Appendix A,

$$Z(\boldsymbol{s}) = \exp\left(\mu + \frac{\sigma^2}{2}\right) \sum_{j=0}^{\infty} \frac{\sigma^j}{j!} H_j(Y(\boldsymbol{s})).$$

Since $\alpha_j = \exp(\mu + \sigma^2/2)\sigma^j/j!$, the assumption of Theorem 3 is satisfied. It follows from (4) - (7) and (A.4) of Appendix A that

$$\begin{aligned}
\mu_Z &= \exp\left(\mu + \frac{\sigma^2}{2}\right), \\
C_Z(\|\boldsymbol{h}\|) &= \mu_Z^2\{\exp(\sigma^2 C_Y(\|\boldsymbol{h}\|)) - 1\}, \\
\hat{Z}(\boldsymbol{s}_0) &= \exp\left(\mu + \frac{\sigma^2}{2}\right) \sum_{j=0}^{\infty} \frac{\sigma^j}{j!} \left(\sqrt{\boldsymbol{c}_Y' \Sigma_Y^{-1} \boldsymbol{c}_Y}\right)^j H_j\left(\frac{\boldsymbol{c}_Y' \Sigma_Y^{-1} \boldsymbol{Y}}{\sqrt{\boldsymbol{c}_Y' \Sigma_Y^{-1} \boldsymbol{c}_Y}}\right) \\
&= \exp\left(\sigma \boldsymbol{c}_Y' \Sigma_Y^{-1} \boldsymbol{Y} + \mu + \frac{\sigma^2 - \sigma^2 \boldsymbol{c}_Y' \Sigma_Y^{-1} \boldsymbol{c}_Y}{2}\right) \qquad (8)
\end{aligned}$$

and

$$E[\hat{Z}(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)]^2 = \mu_Z^2\{\exp(\sigma^2) - \exp(\sigma^2 \boldsymbol{c}_Y' \Sigma_Y^{-1} \boldsymbol{c}_Y)\}.$$

The approach to calculate this nonlinear optimal predictor (8) is known as the lognormal simple kriging (Wackernagel 2003). It is known that the lognormal simple kriging predictor is superior to the BLUP in log-Gaussian random field in the MSE sense (e.g., Cressie 1993; Wackernagel 2003). However, they are asymptotically equivalent under infill asymptotics when the underlying Gaussian random field has the Matérn covariance function.

Remark 1. The assumption $\sum_{j=1}^{\infty} \alpha_j^2 j! j^{2\nu+d+\max\{1,2\epsilon\}} < \infty$ is relaxed for some ν. If $\nu = 0.5$, $f_Z(\|\boldsymbol{\lambda}\|)$ has the analytical form,

$$\begin{aligned}
f_Z(\|\boldsymbol{\lambda}\|) &= \sum_{j \geq 1} \frac{\alpha_j^2 j!}{(2\pi)^d} \int_{\mathbb{R}^d} (C_Y(\|\boldsymbol{h}\|))^j \exp(-i\boldsymbol{\lambda}'\boldsymbol{h}) d\boldsymbol{h} \\
&= \sum_{j=1}^{\infty} \frac{\alpha_j^2 j!}{(2\pi)^d} \int_{\mathbb{R}^d} \exp(-\alpha j \|\boldsymbol{h}\|) \exp(-i\boldsymbol{\lambda}'\boldsymbol{h}) d\boldsymbol{h} \\
&= \frac{\Gamma((d+1)/2)\alpha}{\pi^{(d+1)/2}} \sum_{j=1}^{\infty} \frac{\alpha_j^2 j! j}{((\alpha j)^2 + \|\boldsymbol{\lambda}\|^2)^{(d+1)/2}}.
\end{aligned}$$

Consequently a weaker condition $\sum_{j=1}^{\infty} \alpha_j^2 j! j < \infty$ than $\sum_{j=1}^{\infty} \alpha_j^2 j! j^{2\nu+d+\max\{1,2\epsilon\}} < \infty$ is sufficient for Theorem 3.

Remark 2. The tapered BLUP is based on the observed data $\{Z(\boldsymbol{s})\}$. If $T(\cdot)$ is known, we can apply a tapering technique to the underlying Gaussian random field and have

$$\hat{Z}^{tap}(\boldsymbol{s}_0) = \sum_{j=0}^{\infty} \alpha_j \left(\sqrt{\boldsymbol{c}_Y^{tap\prime} \Sigma_Y^{tap^{-1}} \Sigma_Y \Sigma_Y^{tap^{-1}} \boldsymbol{c}_Y^{tap}} \right)^j H_j \left(\frac{\boldsymbol{c}_Y^{tap\prime} \Sigma_Y^{tap^{-1}} \boldsymbol{Y}}{\sqrt{\boldsymbol{c}_Y^{tap\prime} \Sigma_Y^{tap^{-1}} \Sigma_Y \Sigma_Y^{tap^{-1}} \boldsymbol{c}_Y^{tap}}} \right),$$

where $\{Y(\boldsymbol{s})\}$ is a Gaussian random field with zero-mean and unit variance. If $\hat{Z}^{tap}(\boldsymbol{s}_0)$ converges in mean square, by (A.1) and (A.3) of Appendix A, $\hat{Z}^{tap}(\boldsymbol{s}_0)$ is an unbiased predictor and

$$\begin{aligned} E_{C_Z}[\hat{Z}^{tap}(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)]^2 = & \sum_{j=0}^{\infty} \alpha_j^2 j! - 2 \sum_{j=0}^{\infty} \alpha_j^2 j! (\boldsymbol{c}_Y^{tap\prime} \Sigma_Y^{tap^{-1}} \boldsymbol{c}_Y)^j \\ & + \sum_{j=0}^{\infty} \alpha_j^2 j! (\boldsymbol{c}_Y^{tap\prime} \Sigma_Y^{tap^{-1}} \Sigma_Y \Sigma_Y^{tap^{-1}} \boldsymbol{c}_Y^{tap})^j. \end{aligned}$$

$\hat{Z}^{tap}(\boldsymbol{s}_0)$ is well defined and has the consistency in the previous two examples.

However, the MSE of $\hat{Z}^{tap}(\boldsymbol{s}_0)$ seems to be more erratic than that of $\hat{Z}^{tapBLUP}(\boldsymbol{s}_0)$ in some simulations. Whether the asymptotic efficiency of $\hat{Z}^{tap}(\boldsymbol{s}_0)$ with respect to the optimal predictor holds is a future study.

4 Computational experiments

We conduct Monte Carlo simulation by using MATLAB. Firstly we see the convergence and the finite sample accuracy of Theorem 3 for different sample sizes, taper ranges and smoothness of covariance functions. Since

$$E_{C_Z}[\hat{Z}(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)]^2 \leq E_{C_Z}[\hat{Z}^{BLUP}(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)]^2 \leq E_{C_Z}[\hat{Z}^{tapBLUP}(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)]^2,$$

we see the ratios of the MSE of the tapered BLUP and that of the optimal predictor. Next the time required for one calculation of predictors is measured. Finally, we compare the tapered BLUP with the BLUP obtained by an iterative method, which is more time saving than the direct method. The examples of Section 3 are used for the transformation $T(\cdot)$.

Let $D = [-1, 1]^2$ be the sampling domain. The data locations $\{\boldsymbol{s}_i\}_{i \geq 1}$ are sampled from a uniform distribution over D. The spatial prediction is for the center location $\boldsymbol{s}_0 = (0, 0)$. This sampling scheme satisfies infill asymptotics (see Lemma D. 1 in Appendix D).

For a fixed configuration of sampling locations $\{\boldsymbol{s}_i, i = 1, 2, \ldots, n\}$, each expectation in the ratio of the MSE of the tapered BLUP and that of the optimal predictor is approximated by sample mean of 1000 simulations. We iterate this procedure 100 times and calculate

Table 1: Summary of simulation results in the first experiment for $\theta = 0.8$.

		Squared case			Exponential case		
ν	n	Mean	Stdev	95% interval	Mean	Stdev	95% interval
0.5	100	1.250	0.059	[1.140, 1.349]	1.152	0.031	[1.085, 1.212]
	300	1.057	0.022	[1.015, 1.097]	1.037	0.029	[0.981, 1.093]
	500	1.039	0.013	[1.013, 1.065]	1.024	0.021	[0.981, 1.056]
1.0	100	1.567	0.103	[1.357, 1.767]	1.287	0.060	[1.165, 1.395]
	300	1.098	0.023	[1.057, 1.147]	1.075	0.036	[0.982, 1.113]
	500	1.051	0.015	[1.019, 1.083]	1.045	0.032	[0.971, 1.106]
1.5	100	2.064	0.178	[1.724, 2.421]	1.448	0.099	[1.258, 1.651]
	300	1.168	0.035	[1.095, 1.238]	1.123	0.065	[0.962, 1.236]
	500	1.074	0.019	[1.032, 1.109]	1.070	0.040	[0.996, 1.159]

mean, standard deviation and 95% interval of the approximated ratios to see the variability of the MSE ratio. σ^2 and α are determined so that $C_Z(0) = 1$ and C_Z decreases to 0.05 over the distance 0.8 to examine the influence of different transformations, taper ranges and the smoothness parameter ν of the covariance function. Therefore σ^2 and α depend on each transformation and ν. We use Wendland$_2$ taper function

$$C_\theta(\|\boldsymbol{h}\|) = \left(1 - \frac{\|\boldsymbol{h}\|}{\theta}\right)_+^6 \left(1 + 6\frac{\|\boldsymbol{h}\|}{\theta} + \frac{35\|\boldsymbol{h}\|^2}{3\theta^2}\right),$$

which was introduced by Wendland (1995) and satisfies the taper condition if $d = 2$ and $\nu < 2.5$ (Furrer et al. 2006).

The first experiment examines the convergence and the accuracy of the ratio between the MSE of the tapered BLUP and that of the optimal predictor for different ν. The smoothness parameter is $\nu = 0.5, 1.0$ and 1.5 and the taper range is $\theta = 0.8$. The sample size is 100, 300 and 500. Table 1 shows that as ν is larger, the MSE ratio and fluctuations increase because the discrepancy between the true covariance function and the tapered one is larger in the small distance in this simulation.

The second one examines the convergence and the accuracy of the MSE ratio with different taper ranges. The sample size is 100, 300 and 500 for $\theta = 0.6$ and 0.75, wider ranges and is 300, 1000, 2000 and 3000 for $\theta = 0.15$ and 0.3, narrower ranges respectively. The smoothness parameter is $\nu = 0.5$ and 1.5.

Tables 2 and 3 summarize the results. The number in the parentheses is the average of the number of locations falling within the taper range in 100 sets of the data locations sampled from a uniform distribution over D. As θ decreases, the MSE ratio increases because the discrepancy between the true covariance function and the tapered one is larger and the sample size within the taper range is smaller when θ is smaller. If we choose the

Table 2: Summary of simulation results in the second experiment for $\nu = 0.5$.

		Squared case			Exponential case		
θ	n	Mean	Stdev	95% interval	Mean	Stdev	95% interval
0.6	100	1.411	0.082	[1.243, 1.553]	1.282	0.055	[1.184, 1.403]
	300 (84)	1.075	0.027	[1.023, 1.126]	1.050	0.034	[0.979, 1.125]
	500 (143)	1.046	0.016	[1.013, 1.077]	1.034	0.028	[0.979, 1.089]
0.75	100	1.283	0.064	[1.160, 1.390]	1.175	0.040	[1.102, 1.266]
	300 (132)	1.060	0.023	[1.016, 1.101]	1.039	0.029	[0.990, 1.10]
	500 (222)	1.040	0.014	[1.013, 1.067]	1.026	0.020	[0.988, 1.067]
0.15	300 (5)	1.897	0.153	[1.641, 2.232]	1.736	0.143	[1.475, 2.203]
	1000 (18)	1.313	0.076	[1.175, 1.464]	1.283	0.081	[1.125, 1.422]
	2000 (35)	1.096	0.031	[1.035, 1.149]	1.079	0.047	[0.973, 1.181]
	3000 (52)	1.033	0.018	[0.995, 1.068]	1.031	0.032	[0.971, 1.080]
0.3	300 (21)	1.233	0.066	[1.103, 1.347]	1.191	0.066	[1.032, 1.30]
	1000 (71)	1.069	0.024	[1.027, 1.119]	1.055	0.037	[0.983, 1.118]
	2000 (141)	1.021	0.013	[0.994, 1.043]	1.018	0.022	[0.978, 1.066]
	3000 (213)	1.011	0.008	[0.997, 1.026]	1.006	0.016	[0.962, 1.033]

taper range such that the number of the samples within it is greater than 70 in the case of $\nu = 0.5$, the percentage increase of the MSE of the tapered BLUP over that of the optimal predictor seems to be within about 7%. In the case of $\nu = 1.5$, it seems that the taper range such that the number of the samples within it is greater than 140 is a safe choice. Except the small n the exponential transformation has slightly larger variability than the squared one although the mean of the exponential one is closer to 1 than that of the squared one. This suggests that the convergence rate of Theorem 3 may depend on the transformation. Since the MSE ratio converges to 1 as n is larger regardless of the taper range, Tables 2 and 3 support the result of Theorem 3.

Next we consider the computational times of the true optimal predictor, the BLUP and the tapered BLUP for one fixed realization. The computational time of the optimal predictor is measured as the benchmark, though it is infeasible in practice. We put $\nu = 0.5$ and $\theta = 0.15, 0.3$ and 0.6 for the squared transformation. All computations are carried out by using the MATLAB function `sparse` on Linux powered 3.33GHz Xeon processor with 8 Gbytes RAM. For the calculation of the sparse matrix, there is also the R package `spam` (Furrer and Sain 2010). The sample size n is from 1000 to 9000 with the increment 1000. Figure 1 shows that covariance tapering reduces the computational time substantially for $\theta = 0.15$ and 0.3. However, it does not work well for $\theta = 0.6$. It seems that as θ increases the calculation of some algorithm specific to solve a linear equation of a sparse matrix takes much more time and offsets the computational efficiency. For $n = 6000$, the percentages of

Table 3: Summary of simulation results in the second experiment for $\nu = 1.5$.

		Squared case			Exponential case		
θ	n	Mean	Stdev	95% interval	Mean	Stdev	95% interval
0.6	100	2.986	0.276	[2.503, 3.605]	1.864	0.121	[1.645, 2.107]
	300 (84)	1.201	0.045	[1.112, 1.288]	1.131	0.060	[1.025, 1.244]
	500 (143)	1.091	0.026	[1.034, 1.141]	1.085	0.044	[0.986, 1.169]
0.75	100	2.258	0.205	[1.950, 2.736]	1.535	0.10	[1.321, 1.724]
	300 (132)	1.184	0.038	[1.104, 1.264]	1.121	0.060	[1.032, 1.236]
	500 (222)	1.077	0.019	[1.040, 1.119]	1.079	0.039	[1.012, 1.161]
0.15	300 (5)	28.739	3.123	[23.402, 35.521]	13.566	1.884	[10.228, 18.068]
	1000 (18)	15.488	1.978	[11.871, 19.496]	7.339	1.003	[5.348, 9.221]
	2000 (35)	1.178	0.050	[1.086, 1.275]	1.163	0.066	[1.033, 1.283]
	3000 (52)	1.051	0.025	[1.006, 1.095]	1.046	0.035	[0.972, 1.098]
0.3	300 (21)	4.559	0.467	[3.734, 5.436]	2.746	0.370	[2.116, 3.511]
	1000 (71)	1.152	0.034	[1.078, 1.213]	1.140	0.059	[1.018, 1.261]
	2000 (141)	1.10	0.036	[1.037, 1.176]	1.076	0.048	[0.994, 1.181]
	3000 (213)	1.030	0.016	[1.0, 1.058]	1.021	0.021	[0.979, 1.052]

non-zero entries in the tapered covariance matrix, that is the sparsity of the matrices, are 1.6%, 6.3% and 21.6% for $\theta = 0.15$, 0.3 and 0.6 respectively.

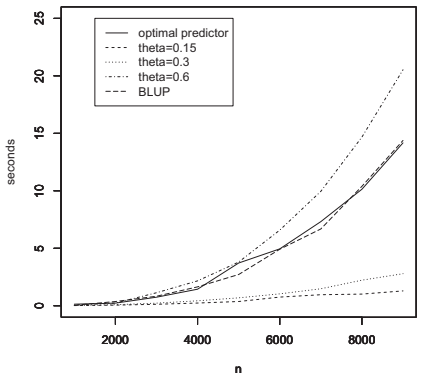

Figure 1: The time required for the calculation of each predictor for various taper ranges.

Finally we consider an iterative method to solve the linear equation of the BLUP. It is known that iterative methods have a lower operation count than direct methods. Here we focus on the conjugate gradient method without preconditioning a linear system (Golub et al. 1996) and apply to $\hat{Z}^{BLUP}$. We put $n = 1000, 3000$, $\nu = 0.5$ and $\theta = 0.15, 0.3$. The count of iterations is 40 and 80. The exponential transformation is used and the number of replications is 1000. From tables 4 and 5 the tapered BLUP by using the direct method is much more time saving than the BLUP with the iterative method. Furthermore, its MSE

Table 4: The time required for the calculation of the BLUP by using the conjugate gradient method.

n	iteration	MSE ratio	Time (sec.)
1000	40	1.045	0.573
	80	1.019	1.139
3000	40	1.044	5.362
	80	1.004	10.827

Table 5: The time required for the calculation of the tapered BLUP by using the direct method.

n	taper range	MSE ratio	Time (sec.)
1000	0.15	1.244	0.015
	0.3	1.072	0.021
3000	0.15	1.036	0.131
	0.3	1.004	0.422

ratio to the optimal predictor is almost equal to that of the BLUP for sufficiently large n. It is a future work to find an efficient preconditioner and compare $\hat{Z}^{BLUP}$ to $\hat{Z}^{tapBLUP}$ by using the conjugate gradient method.

5 Conclusion and future studies

This paper studies covariance tapering for prediction of large spatial data sets in transformed random fields and shows the asymptotic efficiency of the tapered BLUP not only with respect to the BLUP but also with respect to the optimal predictor. Monte Carlo simulations support theoretical results. This result provides a contribution to the analysis of non-Gaussian large spatial data sets and the nonlinear prediction, especially when the transformation is unknown and the optimal predictor is infeasible.

However the case of unknown parameters must be considered in future. Firstly for $\{Z(\boldsymbol{s})\}$, μ_Z and ν are crucial ones by the following reason. Assume that μ_Z and ν are correctly specified and consider the spectral density function

$$f_M(\|\boldsymbol{\lambda}\|) = \frac{\tilde{\sigma}^2}{(\tilde{\alpha}^2 + \|\boldsymbol{\lambda}\|^2)^{\nu+d/2}},$$

for any $\tilde{\alpha} > 0$ and $\tilde{\sigma}^2 > 0$. Then similar to Theorem 3, the tapered BLUP based on μ_Z and the covariance function for $f_M(\|\boldsymbol{\lambda}\|)$ is also asymptotically efficient with respect

to the optimal predictor. The mean μ_Z may be estimated by the sample mean of the observations $\boldsymbol{Z} = (Z(\boldsymbol{s}_1), \ldots, Z(\boldsymbol{s}_n))'$ or a more efficient estimator. Since ν determines the high frequency behavior of $f_M(\|\boldsymbol{\lambda}\|)$, it may be estimated nonparametrically at high frequency bands by using Fourier transforms of $\boldsymbol{Z} = (Z(\boldsymbol{s}_1), \ldots, Z(\boldsymbol{s}_n))'$ (see, e.g., Fuentes and Reich 2010).

Next the estimation of $T(x)$ should be considered. If $T(x)$ is correctly specified, it can be helpful to check the condition in Theorem 3 and determine an appropriate range parameter θ. One candidate for $T^{-1}(x)$ (not $T(x)$) may be the Box-Cox transformation (Box and Cox 1964). Then we have to take account of the instability of its identification which can have a serious effect on the estimation of the parameters (Bickel and Doksum 1981).

Finally how does the tapered BLUP with estimated parameters work well? Putter and Young(2001) considered a related topic. However it seems to need another consideration in our setting as Stein (2010) points out.

Appendix A : Properties of Hermite Polynomials

In this appendix, we state some relevant results on properties of Hermite polynomials (Granger and Newbold 1976; Gradshteyn and Ryzhik 2007; Olver et al. 2010). The system of Hermite polynomials $H_n(x)$ is defined by $H_n(x) = \exp(x^2/2)(-d/dx)^n \exp(-x^2/2)$. For example, $H_0(x) = 1$, $H_1(x) = x$, $H_2(x) = x^2 - 1$, $H_3(x) = x^3 - 3x$, $H_4(x) = x^4 - 6x^2 + 3$, $H_5(x) = x^5 - 10x^3 + 15x$ and so on. The Hermite polynomials are a complete orthogonal system with respect to the standard normal probability density function. Therefore, if $X \sim N(0, 1)$,

$$E\{H_n(X)H_k(X)\} = \begin{cases} 0, & n \neq k, \\ n!, & n = k, \end{cases} \tag{A.1}$$

and since $H_0(x) = 1$,

$$E\{H_n(X)\} = 0, \quad n > 0. \tag{A.2}$$

If X and Y are distributed as bivariate normal random vector with zero means, unit variances and correlation coefficient ρ $(-1 < \rho < 1)$,

$$E\{H_n(X)H_k(Y)\} = \begin{cases} 0, & n \neq k, \\ \rho^n n!, & n = k. \end{cases} \tag{A.3}$$

Finally we have

$$\exp\left(tx - \frac{t^2}{2}\right) = \sum_{n=0}^{\infty} \frac{H_n(x)t^n}{n!} \tag{A.4}$$

and

$$H_n(Ax + By) = \sum_{k=0}^{n} \left(\begin{array}{c} n \\ k \end{array} \right) A^k B^{n-k} H_k(x) H_{n-k}(y) \quad \text{for} \;\; A^2 + B^2 = 1. \tag{A.5}$$

Appendix B : Properties of f_Y and f_Z

To prove the asymptotic optimality of $\hat{Z}^{tapBLUP}(\boldsymbol{s}_0)$, we shall derive some lemmas. Let f_Z denote the spectral density function of C_Z, given by (5).

Lemma B. 1.

$$f_Z(\|\boldsymbol{\lambda}\|) = \sum_{j \geq 1} \alpha_j^2 j! f_Y^{(j-1)}(\|\boldsymbol{\lambda}\|),$$

where

$$f_Y^{(0)}(\|\boldsymbol{\lambda}\|) = f_Y(\|\boldsymbol{\lambda}\|) = \frac{A}{(\alpha^2 + \|\boldsymbol{\lambda}\|^2)^{\nu + d/2}},$$

and $f_Y^{(j-1)}(\|\boldsymbol{\lambda}\|)$ $(j \geq 2)$ *is the* $(j-1)$ *times convolution of* $f_Y(\|\boldsymbol{\lambda}\|)$*, that is*

$$\begin{aligned} f_Y^{(j-1)}(\|\boldsymbol{\lambda}\|) &= A \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} \frac{f_Y(\|\boldsymbol{\omega}_{j-1}\|) \cdots f_Y(\|\boldsymbol{\omega}_1\|)}{(\alpha^2 + \|\boldsymbol{\lambda} - \boldsymbol{\omega}_1 - \cdots - \boldsymbol{\omega}_{j-1}\|^2)^{\nu + d/2}} \prod_{l=1}^{j-1} d\boldsymbol{\omega}_l \\ &= f_Y * \cdots * f_Y(\|\boldsymbol{\lambda}\|) \quad (say). \end{aligned}$$

Proof. Note that $C_Z(\|\boldsymbol{h}\|) = \sum_{j \geq 1} \alpha_j^2 j! (C_Y(\|\boldsymbol{h}\|))^j \leq \sum_{j \geq 1} \alpha_j^2 j! C_Y(\|\boldsymbol{h}\|)$ because $0 \leq C_Y(\|\boldsymbol{h}\|) \leq 1$. Moreover, from 16 of Gradshteyn and Ryzhik (2007; page 676),

$$\begin{aligned} &\int_0^{\infty} C_Y(r) r^{d-1} dr = \frac{\alpha^{\nu}}{2^{\nu-1}\Gamma(\nu)} \int_0^{\infty} r^{\nu+d-1} K_{\nu}(\alpha r) dr \\ &= \frac{\alpha^{\nu}}{2^{\nu-1}\Gamma(\nu)} 2^{\nu+d-2} \alpha^{-\nu-d} \Gamma\left(\frac{d+2\nu}{2}\right) \Gamma\left(\frac{d}{2}\right) < \infty. \end{aligned}$$

Therefore, by using polar coordinates

$$\boldsymbol{h} = \begin{pmatrix} r\cos\theta_1 \\ r\sin\theta_1\cos\theta_2 \\ \vdots \\ r\sin\theta_1 \cdots \sin\theta_{d-2}\cos\theta_{d-1} \\ r\sin\theta_1 \cdots \sin\theta_{d-2}\sin\theta_{d-1} \end{pmatrix} = r\boldsymbol{u} \quad (\|\boldsymbol{u}\| = 1,\; r \geq 0),$$

$$\int_{\mathbb{R}^d} C_Z(\|\boldsymbol{h}\|)d\boldsymbol{h} = \int_{\partial B_d}\int_0^\infty C_Z(r)r^{d-1}drdU(\boldsymbol{u})$$
$$\leq \sum_{j\geq 1}\alpha_j^2 j!\int_{\partial B_d}\int_0^\infty C_Y(r)r^{d-1}drdU(\boldsymbol{u}) < \infty,$$

where ∂B_d is the surface of the unit sphere in $\mathbb{R}^d$ and U is the uniform probability measure on ∂B_d. We have

$$\int_{\mathbb{R}^d}\left|\sum_{j\geq 1}\alpha_j^2 j!(C_Y(\|\boldsymbol{h}\|))^j\exp(-i\boldsymbol{\lambda}'\boldsymbol{h})\right|d\boldsymbol{h} \leq \int_{\mathbb{R}^d}\sum_{j\geq 1}\alpha_j^2 j!(C_Y(\|\boldsymbol{h}\|))^j d\boldsymbol{h}$$
$$= \int_{\mathbb{R}^d} C_Z(\|\boldsymbol{h}\|)d\boldsymbol{h} < \infty,$$

because $0 \leq C_Y(\|\boldsymbol{h}\|)$. Then it follows from Bochner's theorem, the dominated convergence theorem and the inversion formula that

$$\begin{aligned} f_Z(\|\boldsymbol{\lambda}\|) &= \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} C_Z(\|\boldsymbol{h}\|)\exp(-i\boldsymbol{\lambda}'\boldsymbol{h})d\boldsymbol{h} \\ &= \sum_{j\geq 1}\frac{\alpha_j^2 j!}{(2\pi)^d}\int_{\mathbb{R}^d}(C_Y(\|\boldsymbol{h}\|))^j\exp(-i\boldsymbol{\lambda}'\boldsymbol{h})d\boldsymbol{h} \\ &= \sum_{j\geq 1}\frac{\alpha_j^2 j!}{(2\pi)^d}\int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d} f_Y^{(j-1)}(\|\boldsymbol{\omega}\|)\exp(i\boldsymbol{\omega}'\boldsymbol{h})d\boldsymbol{\omega}\right)\exp(-i\boldsymbol{\lambda}'\boldsymbol{h})d\boldsymbol{h} \\ &= \sum_{j\geq 1}\alpha_j^2 j! f_Y^{(j-1)}(\|\boldsymbol{\lambda}\|), \end{aligned}$$

where the third equality is derived by the property that the multiplication of the Fourier transforms of some functions respectively is the Fourier transform of the convolution of these functions. □

Lemma B. 2. *Suppose that* $\sum_{j=1}^\infty \alpha_j^2 j!(j-1)^{2\nu+d+1} < \infty$. *Then*

$$0 < \liminf_{\|\boldsymbol{\lambda}\|\to\infty} f_Z(\|\boldsymbol{\lambda}\|)\|\boldsymbol{\lambda}\|^{d+2\nu} \leq \limsup_{\|\boldsymbol{\lambda}\|\to\infty} f_Z(\|\boldsymbol{\lambda}\|)\|\boldsymbol{\lambda}\|^{d+2\nu} < \infty.$$

Proof. From Lemma B. 1,

$$\begin{aligned} f_Z(\|\boldsymbol{\lambda}\|) =& A\left(\frac{\alpha_1^2}{(\alpha^2+\|\boldsymbol{\lambda}\|^2)^{\nu+d/2}}\right. \\ &\left. + \sum_{j\geq 2}\alpha_j^2 j!\int_{\mathbb{R}^d}\cdots\int_{\mathbb{R}^d}\frac{f_Y(\|\boldsymbol{\omega}_{j-1}\|)\cdots f_Y(\|\boldsymbol{\omega}_1\|)}{(\alpha^2+\|\boldsymbol{\lambda}-\boldsymbol{\omega}_1-\cdots-\boldsymbol{\omega}_{j-1}\|^2)^{\nu+d/2}}\prod_{l=1}^{j-1}d\boldsymbol{\omega}_l\right) \\ =& A(I+II) \quad (say). \end{aligned} \tag{B.1}$$

Then, it suffices to consider the second term II. First consider the lower bound of $f_Z(\|\boldsymbol{\lambda}\|)\|\boldsymbol{\lambda}\|^{d+2\nu}$. We have

$$\|\boldsymbol{\lambda}\|^{d+2\nu} II \geq \sum_{j\geq 2} \alpha_j^2 j! \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} \frac{\|\boldsymbol{\lambda}\|^{d+2\nu} f_Y(\|\boldsymbol{\omega}_{j-1}\|)\cdots f_Y(\|\boldsymbol{\omega}_1\|)}{(\alpha^2 + 2\|\boldsymbol{\lambda}\|^2 + 2\|\boldsymbol{\omega}_1 + \cdots + \boldsymbol{\omega}_{j-1}\|^2)^{\nu+d/2}} \prod_{l=1}^{j-1} d\boldsymbol{\omega}_l.$$

Since

$$\frac{\|\boldsymbol{\lambda}\|^{d+2\nu} f_Y(\|\boldsymbol{\omega}_{j-1}\|)\cdots f_Y(\|\boldsymbol{\omega}_1\|)}{(\alpha^2 + 2\|\boldsymbol{\lambda}\|^2 + 2\|\boldsymbol{\omega}_1 + \cdots + \boldsymbol{\omega}_{j-1}\|^2)^{\nu+d/2}} \leq f_Y(\|\boldsymbol{\omega}_{j-1}\|)\cdots f_Y(\|\boldsymbol{\omega}_1\|)$$

and

$$\sum_{j\geq 2} \alpha_j^2 j! \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} f_Y(\|\boldsymbol{\omega}_{j-1}\|)\cdots f_Y(\|\boldsymbol{\omega}_1\|) \prod_{l=1}^{j-1} d\boldsymbol{\omega}_l < \infty,$$

by the dominated convergence theorem,

$$\begin{aligned}
&\liminf_{\|\boldsymbol{\lambda}\|\to\infty} \|\boldsymbol{\lambda}\|^{d+2\nu} f_Z(\|\boldsymbol{\lambda}\|) \geq \liminf_{\|\boldsymbol{\lambda}\|\to\infty} A\Bigg(\|\boldsymbol{\lambda}\|^{d+2\nu} I \\
&+ \sum_{j\geq 2} \alpha_j^2 j! \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} \frac{\|\boldsymbol{\lambda}\|^{d+2\nu} f_Y(\|\boldsymbol{\omega}_{j-1}\|)\cdots f_Y(\|\boldsymbol{\omega}_1\|)}{(\alpha^2 + 2\|\boldsymbol{\lambda}\|^2 + 2\|\boldsymbol{\omega}_1 + \cdots + \boldsymbol{\omega}_{j-1}\|^2)^{\nu+d/2}} \prod_{l=1}^{j-1} d\boldsymbol{\omega}_l \Bigg) \\
&= A\left(\alpha_1^2 + \frac{1}{2^{\nu+d/2}} \sum_{j\geq 2} \alpha_j^2 j! \right) > 0.
\end{aligned}$$

Next we consider the upper bound of $f_Z(\|\boldsymbol{\lambda}\|)\|\boldsymbol{\lambda}\|^{d+2\nu}$. We set $\boldsymbol{\lambda} = \rho\boldsymbol{v}$, $\boldsymbol{\omega}_1 = r_1\boldsymbol{u}_1$, $\cdots$, $\boldsymbol{\omega}_{j-1} = r_{j-1}\boldsymbol{u}_{j-1}$ with $\|\boldsymbol{v}\| = \|\boldsymbol{u}_1\| = \cdots = \|\boldsymbol{u}_{j-1}\| = 1$. Then for $j \geq 2$, the integral part of (B.1) multiplied by $\|\boldsymbol{\lambda}\|^{d+2\nu}$ reduces to

$$\begin{aligned}
&\int_{\partial B_d} \cdots \int_{\partial B_d} \int_0^\infty \cdots \int_0^\infty \frac{A\rho^{d+2\nu}}{(\alpha^2 + \|\rho\boldsymbol{v} - r_1\boldsymbol{u}_1 - \cdots - r_{j-1}\boldsymbol{u}_{j-1}\|^2)^{\nu+d/2}} \\
&\times \frac{A}{(\alpha^2 + r_1^2)^{\nu+d/2}} \cdots \frac{A}{(\alpha^2 + r_{j-1}^2)^{\nu+d/2}} \prod_{l=1}^{j-1} r_l^{d-1} \prod_{m=1}^{j-1} dr_m \prod_{n=1}^{j-1} dU(\boldsymbol{u}_n) \\
=&\int_{\{r_1,\cdots,r_{j-1},\boldsymbol{u}_1,\cdots,\boldsymbol{u}_{j-1} \mid \|r_1\boldsymbol{u}_1+\cdots+r_{j-1}\boldsymbol{u}_{j-1}\| \leq \rho/2\}} \cdot \prod_{m=1}^{j-1} dr_m \prod_{n=1}^{j-1} dU(\boldsymbol{u}_n) \\
&+ \int_{\{r_1,\cdots,r_{j-1},\boldsymbol{u}_1,\cdots,\boldsymbol{u}_{j-1} \mid \|r_1\boldsymbol{u}_1+\cdots+r_{j-1}\boldsymbol{u}_{j-1}\| > \rho/2\}} \cdot \prod_{m=1}^{j-1} dr_m \prod_{n=1}^{j-1} dU(\boldsymbol{u}_n) \\
=&II_1 + II_2 \quad (say).
\end{aligned}$$

Since in $\{\|r_1\boldsymbol{u}_1+\cdots+r_{j-1}\boldsymbol{u}_{j-1}\|\le\rho/2\}$,

$$\frac{1}{(\alpha^2+\|\rho\boldsymbol{v}-r_1\boldsymbol{u}_1-\cdots-r_{j-1}\boldsymbol{u}_{j-1}\|^2)^{\nu+d/2}}\le\frac{1}{(\alpha^2+\rho^2/4)^{\nu+d/2}},$$

and for $1\le k\le j-1$,

$$\int_{\partial B_d}\int_0^\infty\frac{A}{(\alpha^2+r_k^2)^{\nu+d/2}}r_k^{d-1}dr_k dU(\boldsymbol{u}_k)=1,$$

$$\begin{aligned}II_1\le&\frac{A\rho^{d+2\nu}}{(\alpha^2+\rho^2/4)^{\nu+d/2}}\int_{\partial B_d}\cdots\int_{\partial B_d}\int_0^\infty\cdots\int_0^\infty\frac{A}{(\alpha^2+r_1^2)^{\nu+d/2}}\cdots\frac{A}{(\alpha^2+r_{j-1}^2)^{\nu+d/2}}\\&\times\prod_{l=1}^{j-1}r_l^{d-1}\prod_{m=1}^{j-1}dr_m\prod_{n=1}^{j-1}dU(\boldsymbol{u}_n)\\\le&A2^{2\nu+d}.\end{aligned}$$

On the other hand, $\{\|r_1\boldsymbol{u}_1+\cdots+r_{j-1}\boldsymbol{u}_{j-1}\|>\rho/2\}$ implies that $r_i>\rho/(2(j-1))$ for some i $(1\le i\le j-1)$. Therefore

$$\begin{aligned}II_2\le&\sum_{i=1}^{j-1}\int_{\{r_i>\rho/(2(j-1))\}}\frac{A\rho^{d+2\nu}}{(\alpha^2+\|\rho\boldsymbol{v}-r_1\boldsymbol{u}_1-\cdots-r_{j-1}\boldsymbol{u}_{j-1}\|^2)^{\nu+d/2}}\\&\times\frac{A}{(\alpha^2+r_1^2)^{\nu+d/2}}\cdots\frac{A}{(\alpha^2+r_{j-1}^2)^{\nu+d/2}}\prod_{l=1}^{j-1}r_l^{d-1}\prod_{m=1}^{j-1}dr_m\prod_{n=1}^{j-1}dU(\boldsymbol{u}_n)\\\le&\sum_{i=1}^{j-1}\frac{A\rho^{d+2\nu}}{\left(\alpha^2+\frac{\rho^2}{4(j-1)^2}\right)^{\nu+d/2}}\int_{\partial B_d}\cdots\int_{\partial B_d}\int_0^\infty\cdots\int_0^\infty\\&\times\frac{A}{(\alpha^2+\|\rho\boldsymbol{v}-r_1\boldsymbol{u}_1-\cdots-r_{j-1}\boldsymbol{u}_{j-1}\|^2)^{\nu+d/2}}\frac{A}{(\alpha^2+r_1^2)^{\nu+d/2}}\cdots\frac{A}{(\alpha^2+r_{i-1}^2)^{\nu+d/2}}\\&\times\frac{A}{(\alpha^2+r_{i+1}^2)^{\nu+d/2}}\cdots\frac{A}{(\alpha^2+r_{j-1}^2)^{\nu+d/2}}\prod_{l=1}^{j-1}r_l^{d-1}\prod_{m=1}^{j-1}dr_m\prod_{n=1}^{j-1}dU(\boldsymbol{u}_n)\\=&\sum_{i=1}^{j-1}\frac{A\rho^{d+2\nu}}{\left(\alpha^2+\frac{\rho^2}{4(j-1)^2}\right)^{\nu+d/2}}\int_{\mathbb{R}^d}\cdots\int_{\mathbb{R}^d}f_Y(\|\boldsymbol{\lambda}-\boldsymbol{\omega}_1-\cdots-\boldsymbol{\omega}_{j-1}\|)\\&\times f_Y(\|\boldsymbol{\omega}_1\|)\cdots f_Y(\|\boldsymbol{\omega}_{i-1}\|)f_Y(\|\boldsymbol{\omega}_{i+1}\|)\cdots f_Y(\|\boldsymbol{\omega}_{j-1}\|)d\boldsymbol{\omega}_{j-1}\cdots d\boldsymbol{\omega}_1\\=&\sum_{i=1}^{j-1}\frac{A\rho^{d+2\nu}}{\left(\alpha^2+\frac{\rho^2}{4(j-1)^2}\right)^{\nu+d/2}}\int_{\mathbb{R}^d}f_Y*\cdots*f_Y(\|\boldsymbol{\lambda}-\boldsymbol{\omega}_i\|)d\boldsymbol{\omega}_i\\\le&A2^{2\nu+d}(j-1)^{2\nu+d+1},\end{aligned}$$

because

$$\int_{\mathbb{R}^d} f_Y^{(i)}(\|\boldsymbol{\omega}\|)d\boldsymbol{\omega} = (C_Y(0))^{i+1} = 1. \tag{B.2}$$

Finally we have

$$\begin{aligned}\limsup_{\|\boldsymbol{\lambda}\|\to\infty} \|\boldsymbol{\lambda}\|^{d+2\nu} f_Z(\|\boldsymbol{\lambda}\|) \leq & \limsup_{\|\boldsymbol{\lambda}\|\to\infty} A\left(\|\boldsymbol{\lambda}\|^{d+2\nu} I + \sum_{j\geq 2} \alpha_j^2 j! 2^{2\nu+d}(1+(j-1)^{2\nu+d+1})\right) \\ \leq & A2^{2\nu+d} \sum_{j\geq 1} \alpha_j^2 j!(1+(j-1)^{2\nu+d+1}) < \infty.\end{aligned}$$

□

Lemma B. 3.

$$\lim_{\|\boldsymbol{\lambda}\|\to\infty} \frac{f_Y^{(j)}(\|\boldsymbol{\lambda}\|)}{f_Y(\|\boldsymbol{\lambda}\|)} = j+1 \quad (j \geq 0), \tag{B.3}$$

and if $f_\theta(\|\boldsymbol{\lambda}\|)$ *satisfies the taper condition,*

$$\lim_{\|\boldsymbol{\lambda}\|\to\infty} \frac{f_Y^{(j)} * f_\theta(\|\boldsymbol{\lambda}\|)}{f_Y(\|\boldsymbol{\lambda}\|)} = j+1 \quad (j \geq 0). \tag{B.4}$$

Proof. Firstly consider (B.3). We shall show the assertion by mathematical induction. For $j = 0$, the result holds clearly. Assume that it holds for $j = K$. Consider $j = K+1$. We shall show the assertion in the same way as Proposition 1 of Furrer et al. (2006). Note that $f_Y^{(K+1)} = f_Y * f_Y^{(K)}$. Then

$$\frac{f_Y^{(K+1)}(\|\boldsymbol{\lambda}\|)}{f_Y(\|\boldsymbol{\lambda}\|)} = \frac{\int_{\mathbb{R}^d} f_Y(\|\boldsymbol{x}\|) f_Y^{(K)}(\|\boldsymbol{x}-\boldsymbol{\lambda}\|)d\boldsymbol{x}}{f_Y(\|\boldsymbol{\lambda}\|)}.$$

We divide the numerator into the following five parts

$$\int_{0\leq\|\boldsymbol{x}\|<\Delta'} + \int_{\Delta'\leq\|\boldsymbol{x}\|<\rho/2} + \int_{\rho/2\leq\|\boldsymbol{x}\|<\rho-\Delta} + \int_{\rho-\Delta\leq\|\boldsymbol{x}\|<\rho+\Delta} + \int_{\rho+\Delta\leq\|\boldsymbol{x}\|},$$

where $\boldsymbol{\lambda} = \rho\boldsymbol{v}$, $\|\boldsymbol{v}\| = 1$, $\Delta, \Delta' \to \infty$, $\Delta/\rho \to 0$ and $\Delta'/\rho \to 0$ as $\rho \to \infty$. Define $\Delta = O(\rho^\delta)$ for some $(2\nu+d)/(2\nu+d+2\epsilon) < \delta < 1$ as in Proposition 1 of Furrer et al. (2006).
Case1 : $0 \leq \|\boldsymbol{x}\| < \Delta'$
Note that $\|\boldsymbol{\lambda}-\boldsymbol{x}\| \geq \rho - \Delta'$. Since the result holds for $j = K$,

$$\frac{f_Y^{(K)}(\|\boldsymbol{x}-\boldsymbol{\lambda}\|)}{f_Y(\|\boldsymbol{\lambda}\|)} \to K+1$$

as $\rho \to \infty$. It follows that

$$\lim_{\rho\to\infty} \frac{\int_{0\le\|\boldsymbol{x}\|<\Delta'} f_Y(\|\boldsymbol{x}\|) f_Y^{(K)}(\|\boldsymbol{x}-\boldsymbol{\lambda}\|)d\boldsymbol{x}}{f_Y(\|\boldsymbol{\lambda}\|)} = (K+1)\lim_{\rho\to\infty}\int_{0\le\|\boldsymbol{x}\|<\Delta'} f_Y(\|\boldsymbol{x}\|)d\boldsymbol{x} = K+1.$$

Case2 : $\Delta' \le \|\boldsymbol{x}\| < \rho/2$

$$\begin{aligned}\frac{\int_{\Delta'\le\|\boldsymbol{x}\|<\rho/2} f_Y(\|\boldsymbol{x}\|) f_Y^{(K)}(\|\boldsymbol{x}-\boldsymbol{\lambda}\|)d\boldsymbol{x}}{f_Y(\|\boldsymbol{\lambda}\|)} =& \frac{1}{f_Y(\rho)}\int_{\Delta'\le\|\boldsymbol{x}\|<\rho/2} f_Y(\|\boldsymbol{x}\|) f_Y(\|\boldsymbol{x}-\boldsymbol{\lambda}\|) \\ &\times \frac{f_Y^{(K)}(\|\boldsymbol{x}-\boldsymbol{\lambda}\|)}{f_Y(\|\boldsymbol{x}-\boldsymbol{\lambda}\|)}d\boldsymbol{x} \\ \le& \frac{A' f_Y(\rho/2)}{f_Y(\rho)}\int_{\Delta'\le\|\boldsymbol{x}\|<\rho/2} f_Y(\|\boldsymbol{x}\|)d\boldsymbol{x} \to 0\end{aligned}$$

as $\rho \to \infty$ where A' is a constant being independent of $\boldsymbol{x}$ because $\|\boldsymbol{x}-\boldsymbol{\lambda}\| \ge \rho/2$.
Case3 : $\rho/2 \le \|\boldsymbol{x}\| < \rho - \Delta$

$$\frac{\int_{\rho/2\le\|\boldsymbol{x}\|<\rho-\Delta} f_Y(\|\boldsymbol{x}\|) f_Y^{(K)}(\|\boldsymbol{x}-\boldsymbol{\lambda}\|)d\boldsymbol{x}}{f_Y(\|\boldsymbol{\lambda}\|)} \le \frac{f_Y(\rho/2)}{f_Y(\rho)}\int_{\Delta\le\|\boldsymbol{x}\|} f_Y^{(K)}(\|\boldsymbol{x}\|)d\boldsymbol{x} \to 0$$

as $\rho \to \infty$.
Case4 : $\rho - \Delta \le \|\boldsymbol{x}\| < \rho + \Delta$

$$\frac{\int_{\rho-\Delta\le\|\boldsymbol{x}\|<\rho+\Delta} f_Y(\|\boldsymbol{x}\|) f_Y^{(K)}(\|\boldsymbol{x}-\boldsymbol{\lambda}\|)d\boldsymbol{x}}{f_Y(\|\boldsymbol{\lambda}\|)} \le \frac{f_Y(\rho-\Delta)}{f_Y(\rho)} \to 1$$

as $\rho \to \infty$. Moreover, from (B.2),

$$\frac{\int_{\rho-\Delta\le\|\boldsymbol{x}\|<\rho+\Delta} f_Y(\|\boldsymbol{x}\|) f_Y^{(K)}(\|\boldsymbol{x}-\boldsymbol{\lambda}\|)d\boldsymbol{x}}{f_Y(\|\boldsymbol{\lambda}\|)} \ge \frac{f_Y(\rho+\Delta)}{f_Y(\rho)}\int_{0\le\|\boldsymbol{x}\|<\Delta} f_Y^{(K)}(\|\boldsymbol{x}\|)d\boldsymbol{x} \to 1$$

as $\rho \to \infty$. Thus

$$\frac{\int_{\rho-\Delta\le\|\boldsymbol{x}\|<\rho+\Delta} f_Y(\|\boldsymbol{x}\|) f_Y^{(K)}(\|\boldsymbol{x}-\boldsymbol{\lambda}\|)d\boldsymbol{x}}{f_Y(\|\boldsymbol{\lambda}\|)} \to 1$$

as $\rho \to \infty$.
Case5 : $\rho + \Delta \le \|\boldsymbol{x}\|$

$$\begin{aligned}\int_{\rho+\Delta\le\|\boldsymbol{x}\|} f_Y(\|\boldsymbol{x}\|) f_Y^{(K)}(\|\boldsymbol{x}-\boldsymbol{\lambda}\|)d\boldsymbol{x} =& \int_{\mathbb{R}^d} f_Y(\|\boldsymbol{x}-\boldsymbol{\lambda}\|) f_Y^{(K)}(\|\boldsymbol{x}\|)1\{\rho+\Delta\le\|\boldsymbol{x}-\boldsymbol{\lambda}\|\}d\boldsymbol{x} \\ \le& f_Y(\rho+\Delta)\int_{\partial B_d}\int_0^\infty f_Y^{(K)}(r)1\{\Delta\le r\}r^{d-1}drdU(\boldsymbol{u}).\end{aligned}$$

Hence

$$0 \leq \frac{\int_{\rho+\Delta\leq\|\boldsymbol{x}\|} f_Y(\|\boldsymbol{x}\|) f_Y^{(K)}(\|\boldsymbol{x}-\boldsymbol{\lambda}\|) d\boldsymbol{x}}{f_Y(\|\boldsymbol{\lambda}\|)} \leq \frac{(\alpha^2+\rho^2)^{\nu+d/2}}{(\alpha^2+(\rho+\Delta)^2)^{\nu+d/2}} \int_{\partial B_d} \int_0^\infty f_Y^{(K)}(r) \times 1\{\Delta \leq r\} r^{d-1} dr dU(\boldsymbol{u}).$$

Note that $f_Y^{(K)}(r) 1\{\Delta \leq r\} r^{d-1} \leq f_Y^{(K)}(r) r^{d-1}$ and $\int_{\partial B_d} \int_0^\infty f_Y^{(K)}(r) r^{d-1} dr dU(\boldsymbol{u}) = \int_{\mathbb{R}^d} f_Y^{(K)}(\|\boldsymbol{x}\|) d\boldsymbol{x} = 1$ from (B.2). By the dominated convergence theorem,

$$\int_{\partial B_d} \int_0^\infty f_Y^{(K)}(r) 1\{\Delta \leq r\} r^{d-1} dr dU(\boldsymbol{u}) \to 0$$

as $\rho \to \infty$. Therefore,

$$\frac{\int_{\rho+\Delta\leq\|\boldsymbol{x}\|} f_Y(\|\boldsymbol{x}\|) f_Y^{(K)}(\|\boldsymbol{x}-\boldsymbol{\lambda}\|) d\boldsymbol{x}}{f_Y(\|\boldsymbol{\lambda}\|)} \to 0$$

as $\rho \to \infty$. The result (B.3) holds for $j = K+1$.

Next we consider (B.4). If $j = 0$, the result holds from Proposition 1 of Furrer et al. (2006). Then for any $j > 0$, the assertion is shown in the same way as (B.3). □

Appendix C : Proofs of Theorem 3 and Corollary 1

We first prepare three lemmas.

Lemma C. 1. *Suppose that* $\sum_{j=1}^\infty \alpha_j^2 j! j < \infty$. *Then*

$$\lim_{n\to\infty} \frac{E_{C_Y}[\hat{Y}^{BLUP}(\boldsymbol{s}_0) - Y(\boldsymbol{s}_0)]^2}{E_{C_Z}[\hat{Z}(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)]^2} = \frac{1}{\sum_{j\geq 1} \alpha_j^2 j! j}.$$

Proof. A stationary random field is mean square continuous if and only if the covariance function is continuous at the origin (see Stein 1999; page 20). Hence from Yakowitz and Szidarovszky(1985), $E_{C_Y}[\hat{Y}^{BLUP}(\boldsymbol{s}_0) - Y(\boldsymbol{s}_0)]^2 = 1 - \boldsymbol{c}_Y' \Sigma_Y^{-1} \boldsymbol{c}_Y \geq 0$ and $\boldsymbol{c}_Y' \Sigma_Y^{-1} \boldsymbol{c}_Y \to 1$ as $n \to \infty$. Put $a_n = \boldsymbol{c}_Y' \Sigma_Y^{-1} \boldsymbol{c}_Y$ (≤ 1). From (2) and (7),

$$\frac{E_{C_Y}[\hat{Y}^{BLUP}(\boldsymbol{s}_0) - Y(\boldsymbol{s}_0)]^2}{E_{C_Z}[\hat{Z}(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)]^2} = \frac{1 - a_n}{\sum_{j=1}^\infty \alpha_j^2 j! - \sum_{j=1}^\infty \alpha_j^2 j! (a_n)^j} = \frac{1}{\sum_{j=1}^\infty \alpha_j^2 j! \left(\sum_{i=0}^{j-1} a_n^i\right)}.$$

Since $\alpha_j^2 j! \left(\sum_{i=0}^{j-1} a_n^i\right) \leq \alpha_j^2 j! j$, by the dominated convergence theorem, the assertion is obtained. □

Lemma C. 2. *Suppose that* $\sum_{j=1}^\infty \alpha_j^2 j! j^{2\nu+d+1} < \infty$. *Then*

$$\lim_{n\to\infty} \frac{E_{C_Z}[\hat{Z}^{BLUP}(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)]^2}{E_{C_Y}[\hat{Y}^{BLUP}(\boldsymbol{s}_0) - Y(\boldsymbol{s}_0)]^2} = \sum_{j\geq 1} \alpha_j^2 j! j.$$

Proof. As in the proof of Lemma B. 2 for any $j \geq 1$,

$$\frac{f_Y^{(j-1)}(\|\boldsymbol{\lambda}\|)}{f_Y(\|\boldsymbol{\lambda}\|)} \leq 2^{2\nu+d}(1+(j-1)^{2\nu+d+1}).$$

Then by the dominated convergence theorem, Lemmas B. 1 and B. 3,

$$\lim_{\|\boldsymbol{\lambda}\|\to\infty} \frac{f_Z(\|\boldsymbol{\lambda}\|)}{f_Y(\|\boldsymbol{\lambda}\|)} = \lim_{\|\boldsymbol{\lambda}\|\to\infty} \sum_{j\geq 1} \alpha_j^2 j! \frac{f_Y^{(j-1)}(\|\boldsymbol{\lambda}\|)}{f_Y(\|\boldsymbol{\lambda}\|)} = \sum_{j\geq 1} \alpha_j^2 j! j.$$

By applying Theorem 1 with $C_0 = C_Y$ and $C_1 = C_Z$, we have as $n \to \infty$

$$\frac{1 - 2\boldsymbol{c}_Y{}'\Sigma_Z{}^{-1}\boldsymbol{c}_Z + \boldsymbol{c}_Z{}'\Sigma_Z{}^{-1}\Sigma_Y\Sigma_Z{}^{-1}\boldsymbol{c}_Z}{1 - \boldsymbol{c}_Y'\Sigma_Y^{-1}\boldsymbol{c}_Y} \to 1$$

and

$$\frac{C_Z(0) - \boldsymbol{c}_Z'\Sigma_Z^{-1}\boldsymbol{c}_Z}{1 - 2\boldsymbol{c}_Y{}'\Sigma_Z{}^{-1}\boldsymbol{c}_Z + \boldsymbol{c}_Z{}'\Sigma_Z{}^{-1}\Sigma_Y\Sigma_Z{}^{-1}\boldsymbol{c}_Z} \to \sum_{j\geq 1} \alpha_j^2 j! j.$$

Therefore as $n \to \infty$

$$\frac{E_{C_Z}[\hat{Z}^{BLUP}(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)]^2}{E_{C_Y}[\hat{Y}^{BLUP}(\boldsymbol{s}_0) - Y(\boldsymbol{s}_0)]^2} = \frac{C_Z(0) - \boldsymbol{c}_Z'\Sigma_Z^{-1}\boldsymbol{c}_Z}{1 - \boldsymbol{c}_Y'\Sigma_Y^{-1}\boldsymbol{c}_Y} \to \sum_{j\geq 1} \alpha_j^2 j! j.$$

□

Let f_{tap} denote the spectral density function of the tapered covariance function $C_{tap}^Z(\|\boldsymbol{h}\|) = C_Z(\|\boldsymbol{h}\|)C_\theta(\|\boldsymbol{h}\|)$. Therefore

$$f_{tap}(\|\boldsymbol{\lambda}\|) = f_Z * f_\theta(\|\boldsymbol{\lambda}\|) = \int_{\mathbb{R}^d} f_z(\|\boldsymbol{x}\|) f_\theta(\|\boldsymbol{x} - \boldsymbol{\lambda}\|) d\boldsymbol{x}.$$

Lemma C. 3. *Suppose that* $\sum_{j=1}^\infty \alpha_j^2 j! j^{2\nu+d+\max\{1,2\epsilon\}} < \infty$ *where* ϵ *is given in the taper condition. If* f_θ *satisfies the taper condition,*

$$\lim_{\|\boldsymbol{\lambda}\|\to\infty} \frac{f_{tap}(\|\boldsymbol{\lambda}\|)}{f_Z(\|\boldsymbol{\lambda}\|)} = 1.$$

Proof. From Lemma B. 1, we have

$$\begin{aligned}
\frac{f_{tap}(\|\boldsymbol{\lambda}\|)}{f_Z(\|\boldsymbol{\lambda}\|)} &= \frac{\int_{\mathbb{R}^d} f_z(\|\boldsymbol{x}\|) f_\theta(\|\boldsymbol{x} - \boldsymbol{\lambda}\|) d\boldsymbol{x}}{f_Z(\|\boldsymbol{\lambda}\|)} \\
&= \frac{\sum_{j\geq 1} \alpha_j^2 j! \int_{\mathbb{R}^d} f_Y^{(j-1)}(\|\boldsymbol{x}\|) f_\theta(\|\boldsymbol{x} - \boldsymbol{\lambda}\|) d\boldsymbol{x}}{f_Z(\|\boldsymbol{\lambda}\|)} \\
&= \frac{f_Y(\|\boldsymbol{\lambda}\|)}{f_Z(\|\boldsymbol{\lambda}\|)} \sum_{j\geq 1} \alpha_j^2 j! \frac{f_Y^{(j-1)} * f_\theta(\|\boldsymbol{\lambda}\|)}{f_Y(\|\boldsymbol{\lambda}\|)}.
\end{aligned}$$

We shall evaluate the upper bound of $f_Y^{(j-1)} * f_\theta(\|\boldsymbol{\lambda}\|)/f_Y(\|\boldsymbol{\lambda}\|)$. If $j = 1$, it follows from Proposition 1 of Furrer et al. (2006) that $f_Y * f_\theta(\|\boldsymbol{\lambda}\|)/f_Y(\|\boldsymbol{\lambda}\|) \to 1$ as $\|\boldsymbol{\lambda}\| \to \infty$. Thus $f_Y * f_\theta(\|\boldsymbol{\lambda}\|)/f_Y(\|\boldsymbol{\lambda}\|)$ is bounded. For $j \geq 2$, putting $\boldsymbol{\lambda} = \rho\boldsymbol{v}$, $\boldsymbol{\omega}_1 = r_1\boldsymbol{u}_1$, $\cdots$, $\boldsymbol{\omega}_{j-1} = r_{j-1}\boldsymbol{u}_{j-1}$, $\boldsymbol{\omega}_j = r_j\boldsymbol{u}_j$ with $\|\boldsymbol{v}\| = \|\boldsymbol{u}_1\| = \cdots = \|\boldsymbol{u}_{j-1}\| = \|\boldsymbol{u}_j\| = 1$, we have

$$\begin{aligned}
\frac{f_Y^{(j-1)} * f_\theta(\|\boldsymbol{\lambda}\|)}{f_Y(\|\boldsymbol{\lambda}\|)} =& \int_{\partial B_d} \cdots \int_{\partial B_d} \int_0^\infty \cdots \int_0^\infty \\
& \frac{(\alpha^2+\rho^2)^{\nu+d/2}}{(\alpha^2 + \|\rho\boldsymbol{v} - r_1\boldsymbol{u}_1 - \cdots - r_{j-1}\boldsymbol{u}_{j-1} - r_j\boldsymbol{u}_j\|^2)^{\nu+d/2}} \\
& \times \frac{A}{(\alpha^2+r_1^2)^{\nu+d/2}} \cdots \frac{A}{(\alpha^2+r_{j-1}^2)^{\nu+d/2}} f_\theta(r_j) \prod_{l=1}^{j} r_l^{d-1} \prod_{m=1}^{j} dr_m \prod_{n=1}^{j} dU(\boldsymbol{u}_n) \\
=& \int_{\{r_1,\cdots,r_j,\boldsymbol{u}_1,\cdots,\boldsymbol{u}_j \mid \|r_1\boldsymbol{u}_1+\cdots+r_j\boldsymbol{u}_j\| \leq \rho/2\}} \cdot \prod_{m=1}^{j} dr_m \prod_{n=1}^{j} dU(\boldsymbol{u}_n) \\
& + \int_{\{r_1,\cdots,r_j,\boldsymbol{u}_1,\cdots,\boldsymbol{u}_j \mid \|r_1\boldsymbol{u}_1+\cdots+r_j\boldsymbol{u}_j\| > \rho/2\}} \cdot \prod_{m=1}^{j} dr_m \prod_{n=1}^{j} dU(\boldsymbol{u}_n) \\
=& J_1 + J_2 \quad (say).
\end{aligned}$$

Then similar to I_1 of Lemma B. 2, we have

$$J_1 \leq 2^{2\nu+d}. \tag{C.1}$$

Next we decompose J_2 into the two terms.

$$\begin{aligned}
J_2 \leq& \sum_{i=1}^{j-1} \int_{\{r_i > \rho/(2j)\}} \frac{(\alpha^2+\rho^2)^{\nu+d/2}}{(\alpha^2 + \|\rho\boldsymbol{v} - r_1\boldsymbol{u}_1 - \cdots - r_{j-1}\boldsymbol{u}_{j-1} - r_j\boldsymbol{u}_j\|^2)^{\nu+d/2}} \\
& \times \frac{A}{(\alpha^2+r_1^2)^{\nu+d/2}} \cdots \frac{A}{(\alpha^2+r_{j-1}^2)^{\nu+d/2}} f_\theta(r_j) \prod_{l=1}^{j} r_l^{d-1} \prod_{m=1}^{j} dr_m \prod_{n=1}^{j} dU(\boldsymbol{u}_n) \\
& + \int_{\{r_j > \rho/(2j)\}} \frac{(\alpha^2+\rho^2)^{\nu+d/2}}{(\alpha^2 + \|\rho\boldsymbol{v} - r_1\boldsymbol{u}_1 - \cdots - r_{j-1}\boldsymbol{u}_{j-1} - r_j\boldsymbol{u}_j\|^2)^{\nu+d/2}} \\
& \times \frac{A}{(\alpha^2+r_1^2)^{\nu+d/2}} \cdots \frac{A}{(\alpha^2+r_{j-1}^2)^{\nu+d/2}} f_\theta(r_j) \prod_{l=1}^{j} r_l^{d-1} \prod_{m=1}^{j} dr_m \prod_{n=1}^{j} dU(\boldsymbol{u}_n) \\
=& J_{21} + J_{22} \quad (say).
\end{aligned}$$

As in the proof of Lemma B. 2

$$J_{21} \leq \frac{(\alpha^2+\rho^2)^{\nu+d/2}}{\left(\alpha^2 + \frac{\rho^2}{4j^2}\right)^{\nu+d/2}}(j-1) \leq 2^{2\nu+d} j^{2\nu+d+1} \tag{C.2}$$

and

$$J_{22} \leq \int_{\{r_j > \rho/(2j)\}} \frac{(\alpha^2+\rho^2)^{\nu+d/2}}{(\alpha^2+\|\rho\boldsymbol{v} - r_1\boldsymbol{u}_1 - \cdots - r_{j-1}\boldsymbol{u}_{j-1} - r_j\boldsymbol{u}_j\|^2)^{\nu+d/2}}$$
$$\times \frac{A}{(\alpha^2+r_1^2)^{\nu+d/2}} \cdots \frac{A}{(\alpha^2+r_{j-1}^2)^{\nu+d/2}} \frac{M_\theta}{(1+r_j^2)^{\nu+d/2+\epsilon}} \prod_{l=1}^{j} r_l^{d-1} \prod_{m=1}^{j} dr_m \prod_{n=1}^{j} dU(\boldsymbol{u}_n)$$
$$\leq A_1 2^{2\nu+d+2\epsilon} j^{2\nu+d+2\epsilon}, \tag{C.3}$$

where A_1 is a constant being independent of $\|\boldsymbol{\lambda}\|$. From the case of $j = 1$, (C.1) - (C.3),

$$\frac{f_Y^{(j-1)} * f_\theta(\|\boldsymbol{\lambda}\|)}{f_Y(\|\boldsymbol{\lambda}\|)} \leq A_2 j^{2\nu+d+\max\{1,2\epsilon\}}$$

and

$$\sum_{j\geq 1} \alpha_j^2 j! A_2 j^{2\nu+d+\max\{1,2\epsilon\}} < \infty,$$

where A_2 is a constant being independent of $\|\boldsymbol{\lambda}\|$. Hence by the proof of Lemma C. 2, the dominated convergence theorem and Lemma B. 3,

$$\lim_{\|\boldsymbol{\lambda}\|\to\infty} \frac{f_{tap}(\|\boldsymbol{\lambda}\|)}{f_Z(\|\boldsymbol{\lambda}\|)} = \left(\frac{1}{\sum_{j\geq 1}\alpha_j^2 j! j}\right) \sum_{j\geq 1} \alpha_j^2 j! j = 1.$$

□

Noting that Lemmas B. 2 and C. 3 are the verifications of the conditions of Theorem 1, we have the following proposition.

Proposition C.4. Under the conditions of Theorem 3,

$$\frac{E_{C_Z}[\hat{Z}^{tapBLUP}(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)]^2}{E_{C_Z}[\hat{Z}^{BLUP}(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)]^2} \to 1$$

and

$$\frac{E_{C_{tap}^Z}[\hat{Z}^{tapBLUP}(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)]^2}{E_{C_Z}[\hat{Z}(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)]^2} \to 1$$

as $n \to \infty$.

Proof of Theorem 3

We decompose the ratio of the MSE into the three terms

$$\frac{E_{C_Z}[\hat{Z}^{tapBLUP}(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)]^2}{E_{C_Z}[\hat{Z}(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)]^2} = \frac{E_{C_Y}[\hat{Y}^{BLUP}(\boldsymbol{s}_0) - Y(\boldsymbol{s}_0)]^2}{E_{C_Z}[\hat{Z}(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)]^2} \frac{E_{C_Z}[\hat{Z}^{BLUP}(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)]^2}{E_{C_Y}[\hat{Y}^{BLUP}(\boldsymbol{s}_0) - Y(\boldsymbol{s}_0)]^2} \times \frac{E_{C_Z}[\hat{Z}^{tapBLUP}(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)]^2}{E_{C_Z}[\hat{Z}^{BLUP}(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)]^2}.$$

Then by Lemmas C. 1 and C. 2, the first term and the second term converge to $1/\sum_{j\geq 1}\alpha_j^2 j!j$ and $\sum_{j\geq 1}\alpha_j^2 j!j$ respectively as $n \to \infty$. From Proposition C.4., the third term converges to 1 as $n \to \infty$. The proof is completed.

□

Proof of Corollary 1

We decompose the ratio of the MSE into the two terms

$$\frac{E_{C_{tap}^Z}[\hat{Z}^{tapBLUP}(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)]^2}{E_{C_Z}[\hat{Z}(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)]^2} = \frac{E_{C_Z}[\hat{Z}^{BLUP}(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)]^2}{E_{C_Z}[\hat{Z}(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)]^2} \frac{E_{C_{tap}^Z}[\hat{Z}^{tapBLUP}(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)]^2}{E_{C_Z}[\hat{Z}^{BLUP}(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)]^2}.$$

Then by the proof of Theorem 3, the first term converges to 1 as $n \to \infty$. Finally by Proposition C.4., the second term converges to 1 as $n \to \infty$. The proof is completed.

□

Appendix D : Property of the sampling scheme in Section 4

Lemma D. 1. *Let $\{\boldsymbol{S}_i\}_{i\geq 1}$ be i.i.d. sequence in D. Suppose also that $P(\{\omega | \|\boldsymbol{S}_i - \boldsymbol{s}_0\| \leq \epsilon\}) > 0$ for any $\epsilon > 0$, $P(\boldsymbol{S}_i = \boldsymbol{s}_0) = 0$ and $\boldsymbol{s}_0 \in D$. Then*

$$\boldsymbol{s}_0 \in \overline{\{\boldsymbol{S}_i, i = 1, 2, \ldots\}} \quad \text{a.s.}$$

Proof. Define $A_{ij} = \{\omega | \|\boldsymbol{S}_i - \boldsymbol{s}_0\| \leq 1/j\}$ for $i, j \in \mathbb{N}$. By the assumption, $P(A_{ij}) > 0$. Then

$$P\left(\bigcap_{i=1}^{\infty} A_{ij}^c\right) = \lim_{k\to\infty} (P(A_{ij}^c))^k = 0,$$

because $P(A_{ij}^c) = 1 - P(A_{ij}) < 1$. Thus, $P\left(\bigcup_{i=1}^{\infty} A_{ij}\right) = 1$. Define $B_j = \bigcup_{i=1}^{\infty} A_{ij}$. Then

$$P\left(\bigcap_{j=1}^{\infty} B_j\right) = 1 - P\left(\bigcup_{j=1}^{\infty} B_j^c\right) \geq 1 - \sum_{j=1}^{\infty} P(B_j^c) = 1.$$

Therefore

$$P(\{\omega | \boldsymbol{s}_0 \in \overline{\{\boldsymbol{S}_i, i = 1, 2, \ldots\}}\}) = P\left(\bigcap_{j=1}^{\infty}\bigcup_{i=1}^{\infty} A_{ij}\right) = 1.$$

□

Acknowledgements The authors are grateful to Professor Akimichi Takemura and Professor Mark G. Genton for helpful comments and discussions. We also acknowledge the suggestions from the associate editor and three anonymous referees that refined and improved the manuscript. This work is supported by the Research Fellowship (DC1) from the Japan Society for the Promotion of Science and by the Grants-in-Aid for Scientific Research (A) 23243039 from the Japanese Ministry of Education, Science, Sports, Culture and Technology.

References

[1] Bickel, P. J. and Doksum, K.A. (1981). An analysis of transformations revisited. *Journal of the American Statistical Association*, **76**, 296-311.

[2] Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B*, **26**, 211-252.

[3] Brockwell, P.J. and Davis, R.A. (1991). *Time Series: Theory and Methods*, second ed. Springer, New York.

[4] Cressie, N. (1993). *Statistics for Spatial Data*, revised ed. Wiley, New York.

[5] De Oliveira, V. (2006). On optimal point and block prediction in log-Gaussian random fields. *Scandinavian Journal of Statistics*, **33**, 523-540.

[6] Du, J., Zhang, H. and Mandrekar, V. S. (2009). Fixed-domain asymptotic properties of tapered maximum likelihood estimators. *Annals of Statistics*, **37**, 3330-3361.

[7] Fuentes, M. and Reich, B. (2010). Spectral Domain. In Gelfand, A.E., Diggle, P.J., Fuentes, M. and Guttorp, P. (Eds.), *Handbook of Spatial Statistics*. Chapman and Hall/CRC.

[8] Furrer, R., Genton, M. G. and Nychka, D. (2006). Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics*, **15**, 502-523. *Erratum and Addendum : Journal of Computational and Graphical Statistics*, **21**, 823-824.

[9] Furrer, R. and Sain, S. R. (2010). spam: A sparse matrix R package with emphasis on MCMC methods for Gaussian markov random fields. *Journal of Statistical Software*, **36 (10)**, 1-25.

[10] Golub, G.H. and Van Loan, C.F. (1996). *Matrix Computations*. Johns Hopkins University Press.

[11] Gradshteyn, I.S. and Ryzhik, I.M. (2007). *Table of Integrals, Series, and Products*, seventh ed. Academic Press.

[12] Granger, C.W.J. and Newbold, P. (1976). Forecasting transformed series. *Journal of the Royal Statistical Society: Series B*, **38**, 189-203.

[13] Johns, C., Nychka, D., Kittel, T. and Daly, C. (2003). Infilling sparse records of spatial fields. *Journal of the American Statistical Association*, **98**, 796-806.

[14] Kaufman, C., Schervish, M. and Nychka, D. (2008). Covariance tapering for likelihood-based estimation in large spatial data sets. *Journal of the American Statistical Association*, **103**, 1545-1555.

[15] Moyeed, R. A. and Papritz, A. (2002). An empirical comparison of kriging methods for nonlinear spatial point prediction. *Mathematical Geology*, **34**, 365-386.

[16] Olver, F. W., Daniel, W. L., Boisvert, R. F. and Clark, C. W. (2010). *NIST Handbook of Mathematical Functions*. Cambridge University Press.

[17] Putter, H., and Young, G. A. (2001). On the effect of covariance function estimation on the accuracy of kriging predictors. *Bernoulli*, **7**, 421-438.

[18] Stein, M. L. (1993). A simple condition for asymptotic optimality of linear predictions of random fields. *Statistics & Probability Letters*, **17**, 399-404.

[19] Stein, M. L. (1999). *Interpolation of Spatial Data*. Springer, New York.

[20] Stein, M. L. (2010). Asymptotics for Spatial Processes. In Gelfand, A.E., Diggle, P.J., Fuentes, M. and Guttorp, P. (Eds.), *Handbook of Spatial Statistics*. Chapman and Hall/CRC.

[21] Wackernagel, H. (2003). *Multivariate Geostatistics: An Introduction with Applications*. Springer, Berlin.

[22] Wang, D. and Loh, W.-L. (2011). On fixed-domain asymptotics and covariance tapering in Gaussian random field models. *Electronic Journal of Statistics*, **5**, 238-269.

[23] Wendland, H. (1995). Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. *Advances in Computational Mathematics*, **4**, 389-396.

[24] Yakowitz, S.J. and Szidarovszky, F. (1985). A comparison of kriging with nonparametric regression methods. *Journal of Multivariate Analysis*, **16**, 21-53.

[25] Zhu, Z. and Wu, Y. (2010). Estimation and prediction of a class of convolution-based spatial nonstationary models for large spatial data. *Journal of Computational and Graphical Statistics*, **19**, 74-95.