Statistica Sinica Preprint No: SS-2020-0092

Title	Mixed Domain Asymptotics for Geostatistical Processes
Manuscript ID	SS-2020-0092
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202020.0092
Complete List of Authors	Tingjin Chu
Corresponding Author	Tingjin Chu
E-mail	tingjin.chu@unimelb.edu.au

Mixed Domain Asymptotics for Geostatistical Processes

Tingjin Chu

School of Mathematics and Statistics, University of Melbourne

Abstract:

Geostatistics is one of the three main branches of spatial statistics, with the maximum likelihood method is widely used for parameter estimation. The asymptotic properties of maximum likelihood estimators are often considered under the increasing domain asymptotic framework or the infill asymptotic framework. A third framework, the mixed domain asymptotic framework, has the advantage of incorporating both local and global properties of the covariance structure. In this study, we establish the asymptotic properties of maximum likelihood estimators under the mixed domain asymptotic framework. In addition to the asymptotic framework, the sampling design and the form of the covariance functions are also important factors for the asymptotic properties of maximum likelihood estimators. Here, general conditions are imposed to ensure the consistency and asymptotic normality of these estimators. The imposed conditions are verified for some commonly used covariance functions. The resulting asymptotics provides novel insights into the convergence rates of parameter estimators under mixed domain asymptotics, as well as some useful guidelines for data analysis in practice. Simulation studies are conducted to examine the finite-sample properties of maximum likelihood estimators, and a yearly precipitation anomaly data set is analyzed for illustration.

Key words and phrases: Asymptotic framework, covariance function, sampling design, spatial dependence parameter, spatial statistics.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0092

1. Introduction

Geostatistics is widely accepted as one of the three main branches of spatial statistics, and various models have been proposed to analyze different types of geostatistical data sets (Cressie, 1993; Schabenberger and Gotway, 2005; Diggle and Ribeiro, 2007). The likelihood-based approach is often used for parameter estimation in geostatistics. The theoretical properties of parameter estimators are typically studied under two asymptotic frameworks, namely, the increasing domain asymptotic framework and the infill asymptotic framework. In the former case, the spatial domain expands, while the density of the sampling locations stays constant. Under increasing domain asymptotics, the consistency and asymptotic normality have been established for both maximum likelihood estimation (MLE) and restricted MLE (Mardia and Marshall, 1984; Cressie and Lahiri, 1993, 1996). The infill asymptotic framework is also known as the fixed domain asymptotic framework, where denser sampling locations are added to a fixed domain and the density of the sampling locations increases. Under infill asymptotics, the estimators of individual parameters can be inconsistent, while those of a combination of parameters are consistent (Ying, 1993; Stein, 1999; Zhang, 2004; Loh, 2005). A detailed comparison between these two frameworks is presented by Zhang and Zimmerman (2005). In addition to these two frameworks, there is a third framework known as the mixed domain framework. Under the mixed domain framework, the sampling domain expands, which enables it to capture the covariance function for locations that are far apart. At the same time, the density of the sampling locations also increases, which enables it to capture the local property of the spatial process (Hall

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0092

and Patil, 1994). The mixed domain framework is widely used for spatial bootstrap and testing in the frequency domain (Lahiri, 2003; Lahiri and Zhu, 2006; Matsuda and Yajima, 2009; Bandyopadhyay and Rao, 2017). However, the asymptotics of the mixed domain framework is underdeveloped for likelihood-based methods. Recently, Chang et al. (2017) established the consistency and derived the limit distribution of maximum likelihood estimators for an Ornstein–Uhlenbeck process under the mixed domain asymptotic framework. In this study, we focus on establishing some general conditions of the mixed domain asymptotics for maximum likelihood estimators.

Another important aspect of spatial asymptotics is the sampling design, which concerns the spatial locations of observations. There are two types of sampling designs: fixed sampling designs and stochastic sampling designs (Lahiri, 2003; Lahiri and Zhu, 2006). Here, we take the fixed sampling design approach, because spatial locations are usually considered to be fixed in geostatistics (Cressie, 1993). In Lahiri (2003), sampling locations lie on a rectangular grid with possibly unequal spacing in different directions, but are not irregularly spaced. We propose a fixed sampling design that includes both a rectangular grid and irregularly spaced locations. For spatial covariance functions, there are many types of spatial covariance functions, including the Matérn class (Matérn, 1960), powered exponential family (Diggle et al., 1998), Cauchy family (Gneiting and Schlather, 2004), and covariance functions with compact support (Gneiting, 2002; Furrer et al., 2006). It is of interest to investigate what types of covariance functions yield sound theoretical properties for maximum likelihood estimators. Because the asymptotics of parameter estimators is related to how fast the

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0092

covariance functions decay with respect to distance, we classify these functions into two categories based on their decay rates. Moreover, we give general conditions for both types of spatial covariance functions, and show that some commonly used covariance functions satisfy these conditions.

We investigate the asymptotic properties of the spatial dependence parameters using three important factors: *the spatial asymptotic framework*, *the sampling design*, and *the form of the covariance function*. For spatial dependence parameters, asymptotic normality is often based on the uniform asymptotic normality of Sweeting (1980), and the result has been applied to different types of spatial processes under increasing domain asymptotics (Mardia and Marshall, 1984; Cressie and Lahiri, 1993, 1996; Xu and Genton, 2017). For mixed domain asymptotics, we take an alternative approach based on the central limit theorem of a generalized quadratic form, which requires weaker assumptions (De Jong, 1987; Shao and Zhang, 2019). Note that although the theoretical results are intended for mixed domain asymptotics, they can also be applied to increasing domain asymptotics. The resulting asymptotic properties show that as the spatial dependence becomes stronger, the convergence rates of the parameter estimators usually become slower. However, this does not necessarily hold for all parameters. One notable exception is the nugget effect. Moreover, in practice, it is not obvious which asymptotic framework should be used for a particular data set. Zhang and Zimmerman (2005) emphasizes that the parametric function of interest needs to be consistently estimable. Our simulation study shows that for mixed domain asymptotics to hold, the parameter estimators also need to be consistently estimable. Moreover, there are

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0092

certain requirements for the sample size and the strength of the spatial dependence.

The remainder of the paper is organized as follows. In Section 2, we introduce the mixed domain asymptotic framework, fixed sampling design, and spatial covariance functions. The asymptotics for the spatial dependence parameter estimators is established in Section 3. A simulation study is conducted in Section 4 and a yearly precipitation anomaly data set is analyzed in Section 5. All technical proofs are given in the Supplementary Material.

2. Frameworks and Spatial Processes

2.1 Spatial Asymptotic Frameworks

To investigate the asymptotics of maximum likelihood estimators, we first review the mixed domain asymptotic framework. Following Lahiri (2003), we denote $\mathcal{U}_0 \subset \mathbb{R}^l$ as an open and connected subset of $(-1/2,\ 1/2]^l$ containing the origin, where $l \in \mathbb{N}$. The prototype of the sampling region $\mathcal{R}_0$ is a Borel set satisfying $\mathcal{U}_0 \subset \mathcal{R}_0 \subset \bar{\mathcal{U}}_0$, where $\bar{\mathcal{U}}_0$ denotes the closure of $\mathcal{U}_0$. Throughout this paper, we refer to n as the stage of asymptotics and N_n as the sample size at the nth stage of asymptotics. Moreover, the sampling region at stage n is

$$\mathcal{R}_n = \lambda_n \mathcal{R}_0,$$

where $\{\lambda_n\}$ is an increasing sequence of positive numbers and $\lambda_n / N_n^{\alpha} \to c$, as $n \to \infty$, for some constant $c > 0$ and $0 \leq \alpha \leq 1/l$. The density at stage n is $N_n \lambda_n^{-l} |\mathcal{R}_0|^{-1}$ and the size of the sampling region $|\mathcal{R}_n|$ is $\lambda_n^l |\mathcal{R}_0|$, which are at the rates of $N_n^{1-\alpha l}$ and $N_n^{\alpha l}$, respectively. Therefore, it corresponds to the increasing domain asymptotic framework

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0092

when $\alpha = 1/l$, and to the infill asymptotic framework when $\alpha = 0$. For $0 < \alpha < 1/l$, both the density and the size of the sampling domain increase, and we refer to it as the α-rate mixed domain asymptotic framework (Lahiri, 2003).

Under the mixed domain asymptotic framework, the sampling domain is allowed to expand, but at a slower rate than that of the increasing domain framework; the density is also allowed to increase, but at a slower rate than that of the infill framework (Hall and Patil, 1994; Lahiri, 2003; Lahiri and Zhu, 2006; Lu and Tjøstheim, 2014). The strength of the mixed domain asymptotic framework is that it obtains the local property, because the density is increasing; at the same time, the covariance function for locations far apart is captured by allowing the spatial domain to expand (Hall and Patil, 1994).

2.2 Fixed Sampling Design

The sampling design also plays an important role in spatial asymptotics. Before formally introducing our fixed sampling design, we first give a brief review of the fixed sampling design (Lahiri, 2003). Let $\Delta = \mathrm{diag}\{\delta_1, \ldots, \delta_l\}$ be an $l \times l$ diagonal matrix, and $\mathcal{Z}^l = \{\Delta \boldsymbol{i} : \boldsymbol{i} \in \mathbb{Z}^l\}$ be a regular grid with an increment δ_j in the jth direction, for $1 \leq j \leq l$. For the mixed domain framework, a geostatistical process is observed at sampling locations

$$\{s_1, \ldots, s_{N_n}\} = \{s \in \eta_n \mathcal{Z}^l : \boldsymbol{s} \in \mathcal{R}_n\}, \tag{2.1}$$

where $\{\eta_n\}$ is a sequence of decreasing positive real numbers and $\eta_n \to 0$ as $n \to \infty$. Consequently, the sampling locations become finer for larger n and fill the sampling

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0092

domain with a larger density. Moreover, the minimum distance between any two adjacent sampling sites is bounded by $\eta_n \min\{\delta_1, \ldots, \delta_l\}$.

Here, we extend the sampling design (2.1) and allow sampling locations observed on $\mathcal{R}_n \cap \mathbb{R}^l$. Thus, we propose the following fixed sampling design:

(S1) There exist constants $d_{min}, d_{max} > 0$, such that for sufficiently large n, any sampling location $\boldsymbol{s}_i \in \mathcal{R}_n \cap \mathbb{R}^l$ satisfies

$$\eta_n d_{min} \le \min_{j:j\neq i} \|\boldsymbol{s}_i - \boldsymbol{s}_j\| \le \eta_n d_{max},$$

where $\eta_n = N_n^{\alpha - 1/l}$.

Here, the sampling locations in (S1) can be on a rectangular grid, similar to (2.1). Moreover, (S1) includes the scenario in which the sampling locations are not on a grid, but on the domain $\mathcal{R}_n \cap \mathbb{R}^l$, provided the condition (S1) is met. Here, we refer to such sampling locations as irregularly spaced locations. Furthermore, the minimum distance requirement is used to avoid having too many sampling locations added to the same region as the sample size N_n increases.

2.3 Covariance Functions

For a Gaussian spatial process $\{y(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^l\}$, let $\gamma(\boldsymbol{s}, \boldsymbol{s}'; \boldsymbol{\theta}) = \text{cov}\{y(\boldsymbol{s}), y(\boldsymbol{s}')\}$ denote the covariance function between $\boldsymbol{s}$ and $\boldsymbol{s}'$, and let $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)^\top$ be a $q \times 1$ vector of spatial dependence parameters. Assuming stationarity and isotropy of the covariance functions, we have $\gamma(\boldsymbol{s}, \boldsymbol{s}'; \boldsymbol{\theta}) = \gamma(\|\boldsymbol{s} - \boldsymbol{s}'\|; \boldsymbol{\theta})$, where $\|\boldsymbol{s} - \boldsymbol{s}'\|$ is a norm defined on a vector space in $\mathbb{R}^q$. For a given distance $d \ge 0$, we further assume

$\gamma(d;\boldsymbol{\theta})$ is twice continuously differentiable with respect to $\boldsymbol{\theta}$, and denote the first-order and second-order partial derivatives of $\gamma(d;\boldsymbol{\theta})$ by $\gamma_k(d;\boldsymbol{\theta}) = \partial\gamma(d;\boldsymbol{\theta})/\partial\theta_k$ and $\gamma_{kk'}(d;\boldsymbol{\theta}) = \partial\gamma(d;\boldsymbol{\theta})/\partial\theta_k\partial\theta_{k'}$, respectively. Moreover, let $\boldsymbol{\theta}_0$ be the true value of the spatial dependence parameter, and let $\mathcal{B}(\boldsymbol{\theta}_0) = \{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|_2 < c\}$ be a neighborhood of $\boldsymbol{\theta}_0$ for some constant $c > 0$, where $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|_2$ is the Euclidean distance between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0$.

In this study, a spatial covariance function $\gamma(d;\boldsymbol{\theta})$ is referred to as a *Type-I covariance function* if it satisfies the following (C1):

(C1) The functions $|\gamma(d;\boldsymbol{\theta})|, |\gamma_k(d;\boldsymbol{\theta})|, |\gamma_{kk'}(d;\boldsymbol{\theta})|$ are bounded and satisfy

$$\max_{\boldsymbol{\theta}\in\mathcal{B}(\boldsymbol{\theta}_0)} \int_0^\infty u^{l-1}|\gamma(u;\boldsymbol{\theta})|du < \infty$$
$$\max_{\boldsymbol{\theta}\in\mathcal{B}(\boldsymbol{\theta}_0)} \int_0^\infty u^{l-1}|\gamma_k(u;\boldsymbol{\theta})|du < \infty$$
$$\max_{\boldsymbol{\theta}\in\mathcal{B}(\boldsymbol{\theta}_0)} \int_0^\infty u^{l-1}|\gamma_{kk'}(u;\boldsymbol{\theta})|du < \infty.$$

In spatial statistics, many covariance functions satisfy (C1), including the following three forms as examples. The first form of covariance functions is the Matérn class,

$$\gamma(d;\boldsymbol{\theta}) = \begin{cases} \theta_3 \frac{2}{\Gamma(\kappa)} \left(\frac{\theta_1 d}{2}\right)^\kappa K_\kappa(\theta_1 d), & \text{if} \quad d > 0, \\ \theta_3 + \theta_2, & \text{if} \quad d = 0, \end{cases} \tag{2.2}$$

where $\theta_1 > 0$ is a scale parameter, $\theta_2 > 0$ is the nugget effect, and $\theta_3 > 0$ is the partial sill parameter representing the variance without the nugget effect. Furthermore, $\kappa > 0$ is a smoothness parameter controlling the smoothness of the Gaussian process, and $K_\kappa(\cdot)$ is a modified Bessel function of the second kind of order κ (Cressie, 1993). In practice, κ is prespecified instead of estimated, because the parameter κ is often poorly

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0092

identified; see Chapter 5.4 of Diggle and Ribeiro (2007) for more details. The second form of covariance functions is the Gaussian covariance function,

$$\gamma(d,\boldsymbol{\theta}) = \begin{cases} \theta_3 \exp\left\{-\left(\frac{d}{\theta_1}\right)^2\right\}, & \text{if} \quad d>0, \\ \theta_3+\theta_2, & \text{if} \quad d=0, \end{cases} \tag{2.3}$$

where $\theta_1 > 0$ is a scale parameter, $\theta_2 > 0$ is the nugget effect, and $\theta_3 > 0$ is the partial sill parameter. The third form of covariance functions is the powered exponential covariance function,

$$\gamma(d,\boldsymbol{\theta}) = \begin{cases} \theta_3 \exp\left\{-\left(\frac{d}{\theta_1}\right)^{\theta_4}\right\}, & \text{if} \quad d>0, \\ \theta_3+\theta_2, & \text{if} \quad d=0, \end{cases} \tag{2.4}$$

where $\theta_1 > 0$ is a scale parameter, $\theta_2 > 0$ is the nugget effect, $\theta_3 > 0$ is the partial sill, and $0 < \theta_4 \leq 2$ is a shape parameter. The covariance functions (2.2)–(2.4) satisfy (C1), as proved in Lemma 1 of the Supplementary Material.

Although many covariance functions satisfy (C1), some do not meet the requirement, such as those with long-range dependence (Beran, 1994; Mikosch and Stărică, 2004). One example of such a covariance function is the Cauchy covariance function,

$$\gamma(d,\boldsymbol{\theta}) = \begin{cases} \theta_3 \left\{1+\left(\frac{d}{\theta_1}\right)^2\right\}^{-\kappa}, & \text{if} \quad d>0, \\ \theta_3+\theta_2, & \text{if} \quad d=0, \end{cases} \tag{2.5}$$

where $\theta_1 > 0$ is a scale parameter, $\theta_2 > 0$ is the nugget effect, and $\theta_3 > 0$ is the partial sill parameter. Moreover, $\kappa > 0$ is the smoothness parameter. If $l - 2\kappa \geq 0$, (C1) is not satisfied, because $\int_0^\infty u^{l-1}|\gamma(u;\boldsymbol{\theta})|du = \infty$. To study this type of covariance function, we propose the following condition:

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0092

(C2) The functions $|\gamma(d;\boldsymbol{\theta})|, |\gamma_k(d;\boldsymbol{\theta})|, |\gamma_{kk'}(d;\boldsymbol{\theta})|$ are bounded. Moreover, there exists some $0 < \zeta \leq l$, such that

$$\max_{\boldsymbol{\theta}\in\mathcal{B}(\boldsymbol{\theta}_0)} \int_0^M u^{l-1}|\gamma(u;\boldsymbol{\theta})|du = \mathcal{O}(M^{l-\zeta}),$$
$$\max_{\boldsymbol{\theta}\in\mathcal{B}(\boldsymbol{\theta}_0)} \int_0^M u^{l-1}|\gamma_k(u;\boldsymbol{\theta})|du = \mathcal{O}(M^{l-\zeta}),$$
$$\max_{\boldsymbol{\theta}\in\mathcal{B}(\boldsymbol{\theta}_0)} \int_0^M u^{l-1}|\gamma_{kk'}(u;\boldsymbol{\theta})|du = \mathcal{O}(M^{l-\zeta}),$$
$$\max_{\boldsymbol{\theta}\in\mathcal{B}(\boldsymbol{\theta}_0)} \int_0^M u^{l-1}\gamma_{kk'}^2(u;\boldsymbol{\theta})du = \mathcal{O}(1),$$

as $M \to \infty$.

A covariance function is referred to as a *Type-II covariance function* if it satisfies (C2). In Lemma 2 of the Supplementary Material, we show that (C2) holds for a Cauchy covariance function with $l/4 \leq \kappa \leq l/2$.

3. Mixed Domain Asymptotics

3.1 Maximum Likelihood Estimation

In this section, we investigate the theoretical properties of maximum likelihood estimators under the mixed domain asymptotic framework. Recall that we denote n as the stage of asymptotics and N_n as the sample size at the nth stage. At the nth stage, let $\boldsymbol{y} = (y(\boldsymbol{s}_1), \ldots, y(\boldsymbol{s}_{N_n}))^T$ denote an $N_n \times 1$ vector of observations and $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = [\gamma(\boldsymbol{s}_i, \boldsymbol{s}_{i'};\boldsymbol{\theta})]_{i,i'=1}^{N_n}$ denote an $N_n \times N_n$ covariance matrix. Here, the spatial process $\{y(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^l\}$ is assumed to be a Gaussian process. Therefore, the negative log-likelihood function of $\boldsymbol{\theta}$ is

$$\ell(\boldsymbol{\theta}) = \frac{N_n}{2}\log(2\pi) + \frac{1}{2}\log|\boldsymbol{\Sigma}(\boldsymbol{\theta})| + \frac{1}{2}\boldsymbol{y}^\top\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{y}. \tag{3.1}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0092

The minimizer of (3.1) is the maximum likelihood estimator of $\boldsymbol{\theta}$ and is denoted as $\hat{\boldsymbol{\theta}}_n$. There is usually no analytical solution for minimizing (3.1), and thus a numerical method is used to find the minimizer. For more details, see Chapter 5 of Diggle and Ribeiro (2007).

In the following sections, we denote $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ and $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$ for ease of presentation. Let $\ell'(\boldsymbol{\theta}) = \partial\ell(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ and $\ell''(\boldsymbol{\theta},\boldsymbol{\theta}) = \partial^2\ell(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$ denote the first-order and second-order derivatives, respectively, of $\ell(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$. The kth element of $\ell'(\boldsymbol{\theta})$ is $\frac{1}{2}\{tr(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_k) + \boldsymbol{y}^\top\boldsymbol{\Sigma}^k\boldsymbol{y}\}$, where $\boldsymbol{\Sigma}_k = \boldsymbol{\Sigma}_k(\boldsymbol{\theta}) = \partial\boldsymbol{\Sigma}/\partial\theta_k$ and $\boldsymbol{\Sigma}^k = \boldsymbol{\Sigma}^k(\boldsymbol{\theta}) = \partial\boldsymbol{\Sigma}^{-1}/\partial\theta_k = -\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_k\boldsymbol{\Sigma}^{-1}$. Moreover, the (k,k')th element of $\ell''(\boldsymbol{\theta},\boldsymbol{\theta})$ is $\frac{1}{2}\{tr(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_{kk'} + \boldsymbol{\Sigma}^k\boldsymbol{\Sigma}_{k'}) + \boldsymbol{y}^\top\boldsymbol{\Sigma}^{kk'}\boldsymbol{y}\}$, where $\boldsymbol{\Sigma}_{kk'} = \boldsymbol{\Sigma}_{kk'}(\boldsymbol{\theta}) = \partial^2\boldsymbol{\Sigma}/\partial\theta_k\partial\theta_{k'}$ and $\boldsymbol{\Sigma}^{kk'} = \boldsymbol{\Sigma}^{kk'}(\boldsymbol{\theta}) = \partial^2\boldsymbol{\Sigma}^{-1}/\partial\theta_k\partial\theta_{k'}$. The information matrix of $\boldsymbol{\theta}$ is

$$\mathcal{J}_{\boldsymbol{\theta}_0} = E\left\{\ell''(\boldsymbol{\theta}_0,\boldsymbol{\theta}_0)\right\} = \left[\frac{t_{kk',n}}{2}\right]_{k,k'=1}^{q,q},$$

where $t_{kk',n} = tr(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_{k0}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_{k'0})$ and $\boldsymbol{\Sigma}_{k0} = \boldsymbol{\Sigma}_k(\boldsymbol{\theta}_0)$.

For any $p \times m$ matrix $\boldsymbol{A}$, let $\|\boldsymbol{A}\|_2 = \sqrt{\lambda_{\max}(\boldsymbol{A}^\top\boldsymbol{A})}$ denote the matrix 2-norm, $\|\boldsymbol{A}\|_\infty = \max_{1\le i\le p}\sum_{j=1}^m |a_{ij}|$ denote the matrix infinity norm, and $\|\boldsymbol{A}\|_F = \sqrt{\mathrm{tr}(\boldsymbol{A}^\top\boldsymbol{A})}$ denote the matrix Frobenius norm, where a_{ij} is the (i,j)th element of $\boldsymbol{A}$ and $\lambda_{\max}(\cdot)$ is the largest eigenvalue. To establish Theorem 1, the following regularity conditions are assumed:

(A1) Given $d \geq 0$, the covariance function $\gamma(d;\boldsymbol{\theta})$ is twice continuously differentiable with respect to $\boldsymbol{\theta}$, for $\boldsymbol{\theta} \in \mathcal{B}(\boldsymbol{\theta}_0)$.

(A2) There exists a constant $c_3 > 0$, such that for any $\boldsymbol{\theta} \in \mathcal{B}(\boldsymbol{\theta}_0)$, we have $\|\boldsymbol{\Sigma}^{-1}\|_2 \leq c_3$

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0092

for sufficiently large N_n.

(A3) The smallest eigenvalue of $\boldsymbol{\Omega}_n = (\omega_{kk',n})_{k,k'=1}^{q}$ is bounded away from zero, where $\omega_{kk',n} = \frac{t_{kk',n}}{(t_{kk,n}t_{k'k',n})^{1/2}}$.

(A4) As $N_n \to \infty$, we have $t_{kk,n} \to \infty$, $\max_{\boldsymbol{\theta}\in\mathcal{B}(\boldsymbol{\theta}_0)} \|\boldsymbol{\Sigma}_k\|_2 = o\left(t_{kk,n}^{1/2}\right)$, and $\max_{\boldsymbol{\theta}\in\mathcal{B}(\boldsymbol{\theta}_0)} \|\boldsymbol{\Sigma}_{kk'}\|_F = o\left(t_{kk,n}^{1/2}t_{k'k',n}^{1/2}\right)$.

Assumption (A1) requires that the covariance functions are twice continuously differentiable with respect to $\boldsymbol{\theta}$, and it is easy to verify that covariance functions (2.2)–(2.5) satisfy (A1). Assumption (A2) requires that the smallest eigenvalue of $\boldsymbol{\Sigma}$ is bounded away from zero and avoids the covariance matrix of $\boldsymbol{y}$ being asymptotically singular as $N_n \to \infty$. If the nugget effect $\theta_2 > 0$, then $\|\boldsymbol{\Sigma}^{-1}\|_2 \leq 1/\theta_2$ and $c_3 = \max_{\boldsymbol{\theta}\in\mathcal{B}(\boldsymbol{\theta}_0)} \{1/\theta_2\}$ for (A2). Assumption (A3) ensures that the information matrix is nonsingular in the limit, and the elements of $\hat{\boldsymbol{\theta}}_n$ are not asymptotically linearly dependent (Mardia and Marshall, 1984). For Assumption (A3), we only require the condition for $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, while in Mardia and Marshall (1984), the condition is imposed for $\boldsymbol{\theta} \in \mathcal{B}(\boldsymbol{\theta}_0)$.

A key assumption for the following Theorem 1 is Assumption (A4), which imposes the bounds for $\boldsymbol{\Sigma}_k$ and $\boldsymbol{\Sigma}_{kk'}$ in terms of $t_{kk,n}$ under mixed domain asymptotics. First, note that under increasing domain asymptotics, the bounds for $\boldsymbol{\Sigma}_k$ and $\boldsymbol{\Sigma}_{kk'}$ are assumed in terms of the sample size N_n. That is, $\max_{\boldsymbol{\theta}\in\mathcal{B}(\boldsymbol{\theta}_0)} \|\boldsymbol{\Sigma}_k\|_2 = \mathcal{O}(1)$, $\max_{\boldsymbol{\theta}\in\mathcal{B}(\boldsymbol{\theta}_0)} \|\boldsymbol{\Sigma}_{kk'}\|_2 = \mathcal{O}(1)$, and $\max_{\boldsymbol{\theta}\in\mathcal{B}(\boldsymbol{\theta}_0)} \|\boldsymbol{\Sigma}_k\|_F^{-2} = \mathcal{O}(N_n^{-1/2-\tau})$, for some $\tau > 0$ (Mardia and Marshall, 1984). However, the above assumptions are too restrictive for mixed domain asymptotics and the different types of covariance functions considered here. Therefore,

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0092

we impose the bounds in terms of $t_{kk,n}$ instead. Second, Assumption (A4) is rather abstract and the sufficient conditions are discussed in Section 3.3. In particular, a sufficient condition for (A4) is given in Theorem 3 for Type-I covariance functions and in Theorem 4 for Type-II covariance functions.

3.2 Asymptotics

Let $\xrightarrow{p}$ and $\xrightarrow{D}$ denote convergence in probability and in distribution, respectively. For $1 \le k \le q$, let $\hat{\theta}_{k,n}$ be the kth element of $\hat{\boldsymbol{\theta}}_n$, and $\theta_{k,0}$ be the kth element of $\boldsymbol{\theta}_0$. First, we establish the consistency of $\hat{\boldsymbol{\theta}}_n$ in Theorem 1:

Theorem 1. *Under Assumptions (A1)–(A4), the estimator $\hat{\boldsymbol{\theta}}_n$ satisfies*

$$\left|\hat{\theta}_{k,n} - \theta_{k,0}\right| = \mathcal{O}_p(t_{kk,n}^{-1/2}), \quad \textit{for } 1 \le k \le q.$$

Theorem 1 shows that the convergence rate of each parameter depends on $t_{kk,n}$. Although the convergence rate of a parameter estimator is usually slower for a smaller α, this does not necessarily hold for all parameters. One notable exception is the nugget effect θ_2. Because $\boldsymbol{\Sigma}_2 = \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_2} = \boldsymbol{I}_{N_n}$,

$$\begin{aligned} t_{22,n} &= \operatorname{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{-1}) = \sum_{i=1}^{N_n} \frac{1}{\lambda_i(\boldsymbol{\Sigma})^2} \ge N_n^{-1} \left\{ \sum_{i=1}^{N_n} \frac{1}{\lambda_i(\boldsymbol{\Sigma})} \right\}^2 \\ &\ge N_n^{-1} \left\{ \frac{N_n^2}{\sum_{i=1}^{N_n} \lambda_i(\boldsymbol{\Sigma})} \right\}^2 = N_n^{-1} \left(\frac{N_n^2}{N_n(\theta_{2,0} + \theta_{3,0})} \right)^2 = \frac{N_n}{(\theta_{2,0} + \theta_{3,0})^2}, \end{aligned}$$

where $\lambda_i(\boldsymbol{\Sigma})$ is the ith largest eigenvalue of $\boldsymbol{\Sigma}$. That is, $t_{22,n}^{-1/2}$ is at the rate of ${N_n}^{-1/2}$ and $\hat{\theta}_2$ has the root-N_n convergence rate under the mixed domain asymptotic framework.

To understand the different convergence rate of $\hat{\theta}_2$ intuitively, we first review the model assumption for an observed spatial process (Cressie, 1993; Schabenberger and

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0092

Gotway, 2005). That is, an observed spatial process $y(\boldsymbol{s})$ is assumed to be the summation of two independent processes, namely, the underlying spatial process of interest $u(\boldsymbol{s})$ and the measurement error $\nu(\boldsymbol{s})$. The measurement errors $\nu(\boldsymbol{s}_1), \ldots, \nu(\boldsymbol{s}_{N_n})$ are modeled by a sequence of independent and identically distributed Gaussian random variables $N(0, \theta_2)$. Because the measurement errors are *independently* realized N_n times, the root-N_n convergence rate is not a surprise. In contrast, the other parameters are for the underlying spatial process $u(\boldsymbol{s})$, and the N_n realizations $u(\boldsymbol{s}_1), \ldots, u(\boldsymbol{s}_{N_n})$ are spatially correlated. As the spatial dependence becomes stronger, N_n realizations provide less information, which leads to a slower convergence rate for these parameter estimators. Note that the root-N_n convergence rate of $\hat{\theta}_2$ can also happen under the infill framework. In particular, Chen et al. (2000) shows that the MLE of the nugget effect has a root-N_n convergence rate for an Ornstein–Uhlenbeck process with a measurement error.

For the mixed domain framework, Theorem 1 does not establish asymptotic normality. To establish the asymptotic normality of $\hat{\boldsymbol{\theta}}_n$, an additional assumption (A5) is needed:

(A5) As $N_n \to \infty$, we have $\max\limits_{\boldsymbol{\theta} \in \mathcal{B}(\boldsymbol{\theta}_0)} \|\boldsymbol{\Sigma}_k\|_2 = o\left(\min\limits_{1 \le k \le q} t_{kk,n}^{1/2}\right)$.

Similarly to (A4), the rate of $\max\limits_{\boldsymbol{\theta} \in \mathcal{B}(\boldsymbol{\theta}_0)} \|\boldsymbol{\Sigma}_k\|_2$ is expressed in terms of $t_{kk,n}$ in (A5), and the sufficient conditions are discussed in Section 3.3.

Theorem 2. *Under Assumptions (A1)–(A5),*

$$\mathcal{J}_{\boldsymbol{\theta}_0}^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \ \xrightarrow{D} \ \mathcal{N}\left(0, I_q\right),$$

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0092

where I_q is an identity matrix of order q. Moreover, let $\mathcal{J}_{\hat{\boldsymbol{\theta}}} = \left[\frac{\hat{t}_{kk',n}}{2}\right]_{k,k'=1}^{q,q}$ be an estimator of $\mathcal{J}_{\boldsymbol{\theta}_0}$. Then, we have

$$\mathcal{J}_{\boldsymbol{\theta}_0}^{-1}\mathcal{J}_{\hat{\boldsymbol{\theta}}} \xrightarrow{p} I_q,$$

where $\hat{t}_{kk',n} = tr(\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\Sigma}}_k\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\Sigma}}_{k'})$, $\hat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_n)$ and $\hat{\boldsymbol{\Sigma}}_k = \boldsymbol{\Sigma}_k(\hat{\boldsymbol{\theta}}_n)$.

Theorem 2 shows that the limiting distribution of $\hat{\boldsymbol{\theta}}_n$ is $\mathcal{N}\left(\boldsymbol{\theta}_0, \mathcal{J}_{\boldsymbol{\theta}_0}^{-1}\right)$. Because $\mathcal{J}_{\boldsymbol{\theta}_0}$ is unknown in practice, $\mathcal{J}_{\hat{\boldsymbol{\theta}}}$ is often used for statistical inference. Theorem 2 also guarantees that the distribution of $\hat{\boldsymbol{\theta}}_n$ is well approximated by $\mathcal{N}\left(\boldsymbol{\theta}_0, \mathcal{J}_{\hat{\boldsymbol{\theta}}}^{-1}\right)$ for sufficiently large N_n. The proofs of Theorems 1 and 2 are provided in Appendix A of the Supplementary Material.

For the increasing domain framework, it is well known that Mardia and Marshall (1984) established the asymptotic normality of $\hat{\boldsymbol{\theta}}_n$. In Appendix C of the Supplementary Material, we show that Theorems 1–2 are compatible with the results under increasing domain asymptotics of Mardia and Marshall (1984).

3.3 Type-I and Type-II Covariance Functions

In this section, Theorems 1–2 are used to investigate the theoretical properties of the Type-I and Type-II covariance functions introduced in Section 2.3. Here, we also require the following condition (C3) for the covariance functions:

(C3) Except for the nugget effect θ_2, the integral $\int_0^\infty u^{l-1}\{\gamma_k(u;\boldsymbol{\theta}_0)\}^2 du > 0$ holds for all $1 \le k \le q$.

Condition (C3) is a mild condition for the first-order derivative $\gamma_k(u;\boldsymbol{\theta}_0)$, and requires that except for the nugget effect θ_2, $\gamma_k(u;\boldsymbol{\theta}_0)$ is not zero in a set with a measure that is greater than zero. It is verified for the covariance functions (2.2)–(2.4) in Lemma 1, and for the covariance function (2.5) in Lemma 2.

First, we investigate the consistency and asymptotic normality of Type-I covariance functions.

Theorem 3. *For Type-I covariance functions and the fixed sampling design (S1), a sufficient condition for (A4) and (A5) is* $\alpha l > 2/3$.

The proof of Theorem 3 can be found in Appendix B of the Supplementary Material. In particular, for the quantities mentioned in Assumptions (A4)–(A5), it is shown that for Type-I covariance functions, $\max_{\boldsymbol{\theta}\in\mathcal{B}(\boldsymbol{\theta}_0)} \|\boldsymbol{\Sigma}_k\|_2 = \mathcal{O}(N_n^{1-\alpha l})$, $\max_{\boldsymbol{\theta}\in\mathcal{B}(\boldsymbol{\theta}_0)} \|\boldsymbol{\Sigma}_{kk'}\|_F = \mathcal{O}(N_n^{1-\alpha l/2})$, and $t_{kk,n}$ is at a rate larger or equal to $N_n^{\alpha l}$. Theorem 3 implies that if (A1)–(A3) hold and $\alpha l > 2/3$, then the consistency and asymptotic normality of the maximum likelihood estimators for Type-I covariance functions are guaranteed under the fixed sampling design (S1). However, $\alpha l > 2/3$ is a lower bound for all Type-I covariance functions and may not be a necessary condition for an individual covariance function.

In the following Corollary 3.1, we establish a theoretical property of the parameter estimator for covariance functions (2.2)–(2.4). Because the covariance functions (2.2)–(2.4) are twice continuously differentiable with respect to $\boldsymbol{\theta}$ in an open set, (A1) is satisfied. Moreover, (A2) holds because the nugget effect $\theta_2 > 0$. Conditions (C1) and (C3) are verified for covariance functions (2.2)–(2.4) in Lemma 1. Thus, we obtain Corollary 3.1, as follows:

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0092

Corollary 3.1. *For covariance functions (2.2)–(2.4), under Assumptions (S1), (A3), and* $\alpha l > 2/3$*, Theorem 1 and Theorem 2 hold.*

Next, we establish the consistency and asymptotic normality of the parameter estimator for Type-II covariance functions.

Theorem 4. *For Type-II covariance functions and the fixed sampling design (S1), a sufficient condition for (A4) and (A5) is* $(4\zeta - l)\alpha > 2$*.*

The proof of Theorem 4 can be found in Appendix B of the Supplementary Material. In particular, for the quantities mentioned in Assumptions (A4)–(A5), it is shown that for Type-II covariance functions, we have $\max_{\boldsymbol{\theta}\in\mathcal{B}(\boldsymbol{\theta}_0)} \|\boldsymbol{\Sigma}_k\|_2 = \mathcal{O}(N_n^{1-\alpha\zeta})$, $\max_{\boldsymbol{\theta}\in\mathcal{B}(\boldsymbol{\theta}_0)} \|\boldsymbol{\Sigma}_{kk'}\|_F = \mathcal{O}(N_n^{1-\alpha l/2})$, and $t_{kk,n}$ is at a rate larger or equal to $N_n^{\alpha(2\zeta-l)}$. Note that the bounds for $\max_{\boldsymbol{\theta}\in\mathcal{B}(\boldsymbol{\theta}_0)} \|\boldsymbol{\Sigma}_k\|_2$, $\max_{\boldsymbol{\theta}\in\mathcal{B}(\boldsymbol{\theta}_0)} \|\boldsymbol{\Sigma}_{kk'}\|_F$, and $t_{kk,n}$ are different for Type-I and Type-II covariance functions, and therefore, different sufficient conditions are needed for Assumption (A4)–(A5). Compared with Type I covariance functions, the sufficient condition for Type-II covariance functions also depends on ζ, which regulates the decay of the covariance functions.

For the Cauchy covariance function (2.5), it is easy to verify that (A1)–(A2) hold. Furthermore, By Lemma 2, conditions (C2) and (C3) hold if $l/4 \le \kappa \le l/2$. Therefore, the theoretical properties of the maximum likelihood estimators of Cauchy covariance functions are established in the following Corollary 4.1:

Corollary 4.1. *For the Cauchy covariance function (2.5) with* $l/4 \le \kappa \le l/2$*, under Assumptions (S1), (A3), and* $(8\kappa - l)\alpha > 2$*, Theorem 1 and Theorem 2 hold.*

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0092

4. Simulation Studies

4.1 Type-I Covariance Function and Effect of Frameworks

Here, the finite-sample properties of maximum likelihood estimators are investigated for Type-I covariance functions. For the mixed domain asymptotic framework, we set $\alpha = 0.4$ and 0.3. For each choice of α, the sample sizes are set to $N_n = 200$, 400, 800, and 1600. The spatial domain of interest is $\mathcal{R}_n = [-\iota_n/2, \iota_n/2]^2$, with the side length $\iota_n = 10(N_n/100)^{\alpha}$. The minimum distance between any two sampling locations is $0.2(N_n/100)^{\alpha-1/2}$. For spatial covariance functions, we consider the Gaussian covariance function and the exponential covariance function. For the Gaussian covariance function (2.3), we set the range parameter $\theta_{1,0} = 5$, the nugget effect $\theta_{2,0} = 0.5$, and the partial sill $\theta_{3,0} = 2$. For the exponential covariance function

$$\gamma(d;\boldsymbol{\theta}) = \begin{cases} \theta_3 \exp(-d/\theta_1), & \text{if} \quad d > 0, \\ \theta_3 + \theta_2, & \text{if} \quad d = 0, \end{cases} \tag{4.1}$$

we set the range parameter $\theta_{1,0} = 3$, the nugget effect $\theta_{2,0} = 0.5$, and the partial sill $\theta_{3,0} = 2$.

For each choice of the sample size N_n and α, a total of 500 data sets are simulated. The parameter estimation is carried out by minimizing (3.1) for each simulated data set. The standard deviation of $\hat{\boldsymbol{\theta}}$ is obtained as $sd(\hat{\boldsymbol{\theta}}) = \mathrm{diag}(\mathcal{J}_{\boldsymbol{\theta}_0}^{-1})^{1/2}$, and the standard error of $\hat{\boldsymbol{\theta}}$ is obtained as $se(\hat{\boldsymbol{\theta}}) = \mathrm{diag}\{\mathcal{J}_{\hat{\boldsymbol{\theta}}}^{-1}\}^{1/2}$. Moreover, the $100(1-\alpha)\%$ confidence interval of θ_i is

$$\left(\hat{\theta}_i - z_{1-\alpha/2} se(\hat{\theta}_i), \hat{\theta}_i + z_{1-\alpha/2} se(\hat{\theta}_i)\right),$$

where $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of a standard normal distribution and $se(\hat{\theta}_i)$ is

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0092

Table 1: The mean, root mean squared error (RMSE), standard deviation (SD), and coverage probability of $\hat{\boldsymbol{\theta}}$ and the mean of the standard errors (SEm) of $se(\hat{\boldsymbol{\theta}})$ for the Gaussian covariance function.

N_n		α=0.4			α=0.3		
		θ_1	θ_2	θ_3	θ_1	θ_2	θ_3
	True value	5.00	0.50	2.00	5.00	0.50	2.00
200	Mean	5.03	0.50	2.07	5.10	0.50	2.11
	RMSE	0.84	0.06	0.98	1.54	0.05	1.07
	SD	0.70	0.05	0.94	0.73	0.05	0.99
	SEm	0.73	0.05	0.99	0.81	0.05	1.07
	CP90	0.87	0.87	0.83	0.88	0.90	0.83
400	Mean	5.01	0.50	2.05	5.06	0.50	2.10
	RMSE	0.52	0.04	0.77	0.62	0.04	0.89
	SD	0.49	0.04	0.74	0.53	0.04	0.82
	SEm	0.50	0.04	0.77	0.55	0.04	0.88
	CP90	0.90	0.88	0.87	0.88	0.91	0.86
800	Mean	5.03	0.50	2.06	5.05	0.50	2.11
	RMSE	0.37	0.03	0.61	0.43	0.03	0.73
	SD	0.35	0.03	0.58	0.40	0.03	0.68
	SEm	0.36	0.03	0.61	0.40	0.03	0.73
	CP90	0.89	0.90	0.89	0.91	0.90	0.89
1600	Mean	5.02	0.50	2.03	5.01	0.50	2.03
	RMSE	0.26	0.02	0.44	0.31	0.02	0.58
	SD	0.25	0.02	0.45	0.30	0.02	0.56
	SEm	0.26	0.02	0.46	0.30	0.02	0.57
	CP90	0.90	0.89	0.91	0.89	0.90	0.88

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0092

Table 2: The mean, root mean squared error (RMSE), standard deviation (SD), and coverage probability of $\hat{\boldsymbol{\theta}}$ and the mean of the standard errors (SEm) of $se(\hat{\boldsymbol{\theta}})$ for the exponential covariance function.

N_n		α=0.4			α=0.3		
		θ_1	θ_2	θ_3	θ_1	θ_2	θ_3
	True value	3.00	0.50	2.00	3.00	0.50	2.00
200	Mean	3.18	0.47	2.09	3.14	0.49	2.03
	RMSE	1.45	0.14	0.65	1.61	0.13	0.77
	SD	1.36	0.14	0.67	1.39	0.13	0.71
	SEm	1.57	0.14	0.76	1.62	0.14	0.78
	CP90	0.86	0.91	0.89	0.83	0.91	0.84
400	Mean	3.09	0.49	2.02	3.01	0.49	1.98
	RMSE	1.22	0.10	0.56	1.22	0.09	0.61
	SD	1.03	0.10	0.53	1.11	0.09	0.59
	SEm	1.14	0.10	0.56	1.19	0.09	0.61
	CP90	0.86	0.88	0.87	0.81	0.90	0.83
800	Mean	3.07	0.50	2.00	2.99	0.49	2.01
	RMSE	0.81	0.06	0.41	0.86	0.06	0.49
	SD	0.78	0.06	0.41	0.89	0.06	0.48
	SEm	0.84	0.06	0.43	0.92	0.06	0.50
	CP90	0.89	0.90	0.88	0.87	0.88	0.87
1600	Mean	3.02	0.50	2.01	3.01	0.49	2.02
	RMSE	0.57	0.04	0.31	0.75	0.04	0.40
	SD	0.59	0.04	0.32	0.71	0.04	0.40
	SEm	0.61	0.04	0.32	0.73	0.04	0.41
	CP90	0.89	0.91	0.90	0.86	0.89	0.89

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0092

the ith element of $se(\hat{\boldsymbol{\theta}})$. The simulation results for the Gaussian covariance function are reported in Table 1, and the results for the exponential covariance function are reported in Table 2. Specifically, "Mean" and "RMSE" represent the mean and root mean squared error, respectively, of the parameter estimates from the 500 simulated data sets, "SD" represents the standard deviation of $\hat{\boldsymbol{\theta}}$, "SEm" represents the mean of the standard errors from the 500 simulated data sets, and "CP90" represents the coverage probability of the nominal 90% confidence intervals.

For both covariance functions, as the sample size increases, the mean of the parameter estimates becomes closer to the true value, and the root mean squared error becomes smaller. For the convergence rate, $\hat{\theta}_1$ and $\hat{\theta}_3$ have smaller root mean squared errors for $\alpha = 0.4$ than those of $\alpha = 0.3$, indicating a faster convergence of $\hat{\theta}_1$ and $\hat{\theta}_3$ for a bigger α. On the other hand, the standard deviations of the nugget effect $\hat{\theta}_2$ are similar for both $\alpha = 0.4$ and $\alpha = 0.3$, because the convergence rate of $\hat{\theta}_2$ is root-N_n, as shown in Section 3.2. Moreover, as the sample size increases, the root mean squared error of $\hat{\boldsymbol{\theta}}$ becomes closer to the standard deviation, and the mean of the standard error of $se(\hat{\boldsymbol{\theta}})$ becomes closer to standard deviation.

For statistical inference with $\alpha = 0.4$, the confidence interval coverage probabilities become closer to the nominal ones as the sample size increases. Similarly to Zhang and Zimmerman (2005), we plot the $0.05 + 0.1(i-1)$ quantiles for each parameter estimate, where the horizontal axis represents the empirical quantiles and the vertical axis represents the theoretical quantiles derived from Theorem 2. These plots can be found in Figures A–D of the Supplementary Material. Overall, the normal approximations

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0092

provided by Theorem 2 and Corollary 3.1 appear to be appropriate when $\alpha = 0.4$. Note that $\alpha = 0.4$ satisfies the sufficient condition in Corollary 3.1, while $\alpha = 0.3$ does not. For $\alpha = 0.3$, the simulation results show that the normal approximations appear to be appropriate for the Gaussian covariance function, but less than satisfactory for the exponential covariance function.

4.2 Type-II Covariance Function

Here, we investigate the finite-sample properties for Type-II covariance functions. A simulation study is conducted for $\alpha = 0.4$. The sampling design is the same as that in Section 4.1. For the covariance function, we use the Cauchy covariance function with $\kappa = 0.8$ and set the range parameter $\theta_{1,0} = 4$, the nugget effect $\theta_{2,0} = 0.5$, and the partial sill $\theta_{3,0} = 2$.

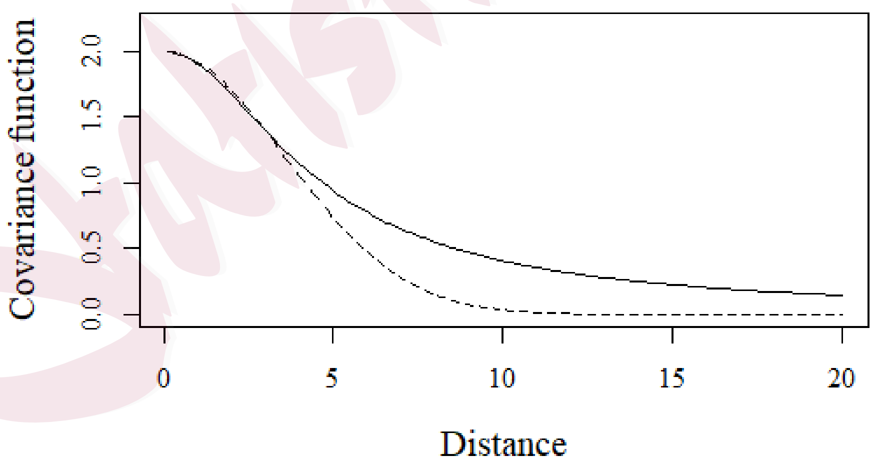

Figure 1: Cauchy (solid line) and Gaussian (dashed line) Covariance Functions

In Figure 1, the Cauchy covariance function is plotted with the Gaussian covari-

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0092

Table 3: The mean, root mean squared error (RMSE), standard deviation (SD), and coverage probability of $\hat{\boldsymbol{\theta}}$ and the mean of the standard errors (SEm) of $se(\hat{\boldsymbol{\theta}})$ for the Cauchy covariance function.

	N_n=200			N_n=400			N_n=800			N_n=1600		
	θ_1	θ_2	θ_3	θ_1	θ_2	θ_3	θ_1	θ_2	θ_3	θ_1	θ_2	θ_3
True value	4.00	0.50	2.00	4.00	0.50	2.00	4.00	0.50	2.00	4.00	0.50	2.00
Mean	4.07	0.50	2.17	4.05	0.50	2.10	4.05	0.50	2.07	4.00	0.50	2.03
RMSE	1.01	0.06	1.05	0.66	0.04	0.73	0.49	0.03	0.53	0.36	0.02	0.43
SD	0.91	0.06	0.92	0.65	0.04	0.69	0.46	0.03	0.50	0.36	0.02	0.44
SEm	0.94	0.06	1.03	0.67	0.04	0.74	0.47	0.03	0.53	0.36	0.02	0.45
CP90	0.88	0.89	0.84	0.89	0.88	0.88	0.89	0.89	0.88	0.90	0.91	0.90

ance function considered in Section 4.1. For smaller distances, the Cauchy covariance function has a similar spatial dependence to that of the Gaussian covariance function. As the distance increases, the spatial dependence of the Cauchy covariance function decays more slowly. The results of the simulation study are reported in Table 3. As the sample size increases, the mean of the parameter estimates becomes closer to the true value and the root mean squared error becomes smaller. Moreover, the mean of the standard errors of $se(\hat{\boldsymbol{\theta}})$ is closer to the standard deviation for larger sample sizes. Furthermore, as the sample size increases, the confidence interval coverage probabilities become closer to the nominal ones. Therefore, despite the stronger spatial dependence of Type-II covariance functions, the consistency and asymptotic normality of the parameter estimates still hold.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0092

4.3 Practical Consideration for Mixed Domain Asymptotics

In practice, there is usually only one data set and it is not obvious which asymptotic framework should be used. Zhang and Zimmerman (2005) studied the choice of asymptotic frameworks both theoretically and empirically, pointing out the importance of microergodicity. That is, the parametric function of interest needs to be consistently estimable. Moreover, they conducted simulation studies with sample sizes between 40 and 160 for the exponential covariance function. In Section 4.1, our simulation studies are conducted for sample sizes between 200 and 1600. The QQ plots of each individual parameter and $\phi = \theta_3/\theta_1$ are presented in Figures C, D, and F of the Supplementary Material for $\alpha = 0.4$, 0.3, and 0. The theoretical quantiles of $\hat{\phi}$ are derived from $\hat{\phi} \sim \mathcal{N}(\phi_0, sd(\hat{\phi}))$. To the best of our knowledge, the asymptotic standard deviation of $\hat{\phi}$ is not available for irregularly spaced locations and the exponential covariance function with the nugget effect. Therefore, $sd(\hat{\phi})$ is replaced by the sample standard deviation. Under the infill asymptotics, Figure F shows that $\hat{\phi}$ appears to be normally distributed, while $\hat{\boldsymbol{\theta}}$ is not. Moreover, $\hat{\phi}$ appears to be normally distributed under the mixed domain asymptotic framework. Therefore, the infill asymptotics is a better choice if ϕ is of interest or the goal is spatial interpolation (Stein, 1999).

If the individual parameter is of interest, mixed domain asymptotics or increasing domain asymptotics is used. In the following, we attempt to provide some guidelines for mixed domain asymptotics. First, it is important that the individual parameter is consistently estimable, which we have established in Section 3. Second, any asymptotic result requires certain sample sizes. However, a larger sample alone cannot guarantee

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0092

Table 4: The mean, root mean squared error (RMSE), standard deviation (SD), and coverage probability of $\hat{\boldsymbol{\theta}}$ and the mean of the standard errors (SEm) of $se(\hat{\boldsymbol{\theta}})$ for Gaussian, exponential, and Cauchy covariance functions with sample size $N_n = 800$ and various values of ι/d_p.

ι/d_q		Gaussian			Exponential			Cauchy		
		θ_1	θ_2	θ_3	θ_1	θ_2	θ_3	θ_1	θ_2	θ_3
	True value	5.00	0.50	2.00	3.00	0.50	2.00	4.00	0.50	2.00
1	Mean	5.11	0.50	2.24	3.29	0.50	2.15	4.03	0.50	2.11
	RMSE	1.27	0.02	1.63	2.48	0.04	1.46	0.71	0.03	1.03
	SD	0.85	0.03	1.33	1.95	0.04	1.20	0.66	0.03	0.97
	SEm	0.93	0.03	1.54	2.46	0.04	1.48	0.68	0.03	1.05
	CP90	0.87	0.91	0.80	0.76	0.90	0.76	0.89	0.92	0.85
2	Mean	5.10	0.50	2.13	3.05	0.49	2.04	4.04	0.50	2.09
	RMSE	0.60	0.03	0.97	1.37	0.05	0.78	0.54	0.03	0.67
	SD	0.50	0.03	0.88	1.32	0.05	0.77	0.51	0.03	0.62
	SEm	0.53	0.03	0.96	1.46	0.05	0.85	0.52	0.03	0.66
	CP90	0.90	0.90	0.88	0.81	0.92	0.82	0.90	0.89	0.89
3	Mean	5.01	0.50	2.04	3.08	0.49	2.03	3.98	0.50	2.00
	RMSE	0.40	0.03	0.68	1.18	0.06	0.63	0.47	0.03	0.47
	SD	0.39	0.03	0.66	1.04	0.06	0.58	0.45	0.03	0.47
	SEm	0.39	0.03	0.68	1.16	0.06	0.63	0.45	0.03	0.47
	CP90	0.88	0.89	0.88	0.83	0.91	0.85	0.90	0.89	0.88
4	Mean	5.00	0.50	2.00	3.04	0.49	2.00	3.98	0.50	1.99
	RMSE	0.35	0.03	0.53	0.97	0.07	0.46	0.44	0.03	0.39
	SD	0.33	0.03	0.54	0.88	0.07	0.46	0.43	0.03	0.38
	SEm	0.33	0.03	0.54	0.94	0.07	0.49	0.43	0.03	0.38
	CP90	0.89	0.91	0.89	0.86	0.88	0.87	0.89	0.90	0.88
5	Mean	5.02	0.50	2.03	3.05	0.49	2.02	4.02	0.50	2.02
	RMSE	0.31	0.03	0.48	0.80	0.07	0.41	0.44	0.04	0.36
	SD	0.30	0.03	0.45	0.77	0.07	0.39	0.42	0.04	0.33
	SEm	0.30	0.03	0.46	0.81	0.07	0.41	0.42	0.04	0.33
	CP90	0.89	0.90	0.87	0.89	0.92	0.91	0.89	0.88	0.87

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0092

good normal approximations for parameter estimators. In Section 5 of Zhang and Zimmerman (2005), they cautioned "certain covariogram parameters may be hard to estimate and that the estimates may be badly nonnormal even with a large sample size." In Figure F, the distributions of $\hat{\theta}_1$ and $\hat{\theta}_3$ are hardly normal distributed, even when the sample size $n = 1600$. This observation prompts us to consider the strength of spatial dependence in mixed domain asymptotics. Here, we suggest using the ratio ι/d_q, with $\gamma(d_q; \boldsymbol{\theta}_0) = 0.25(\theta_{2,0} + \theta_{3,0})$, where a larger ratio means weaker spatial dependence.

Simulation studies are conducted with different values of ι/d_q for the covariance functions in Section 4.1. The sample size is fixed at $N = 800$ and the results are given in Table 4. For $\iota/d_q = 1$, the coverage probabilities for all three covariance functions are well below the nominal 90% confidence intervals, especially for $\hat{\theta}_3$. As ι/d_q increases, the bias of the parameter estimates becomes smaller and the coverage probabilities improve. For the Gaussian and Cauchy covariance functions, $\iota/d_q \geq 3$ provides a good normal approximation, whereas $\iota/d_q \geq 4$ is required for the exponential covariance function. Therefore, for statistical inference of the individual parameters, a relatively large sample is needed and the spatial dependence cannot be too strong.

5. Data Example

Here, yearly precipitation anomalies in 1962 at a set of US weather stations are analyzed for illustration. The yearly precipitation anomaly is the departure of the precipitation from its long-period precipitation average value, standardized with respect to the long-period mean and standard deviation of each station (Johns et al., 2003; Kaufman

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0092

et al., 2008). There are 7352 observations across the United States, and we consider the observations in the Great Plains that are north of 40 degrees latitude. The area measures around 1000 km from north to south, and the distance from east to west varies from 1000 to 1300 km. There are 926 weather stations in this area; their locations are shown in Figure 2.

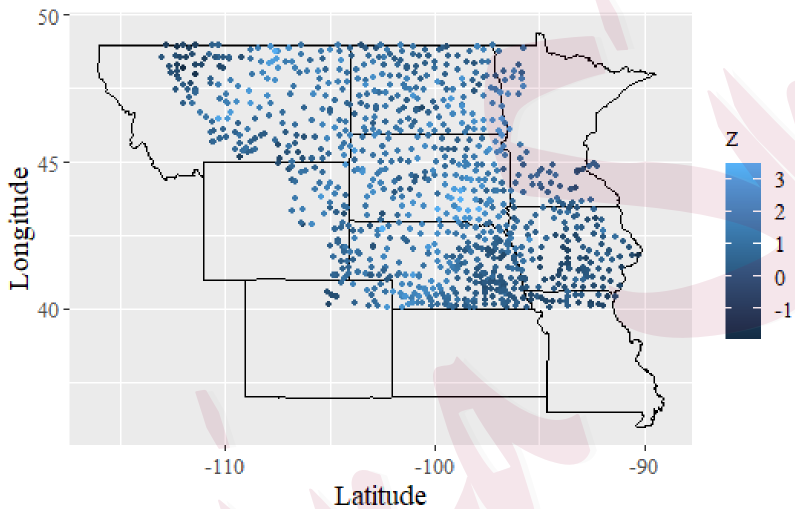

Figure 2: Locations of Weather Stations

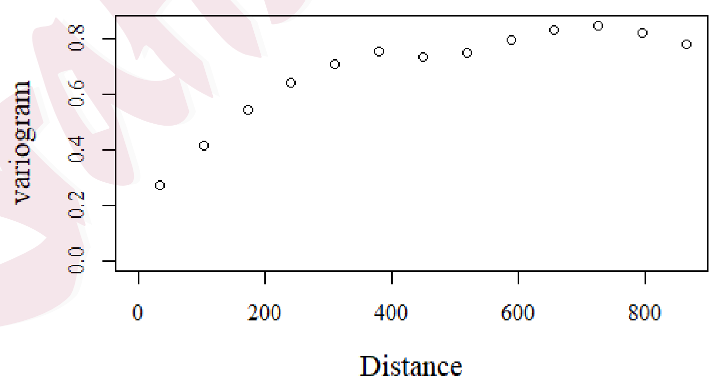

Figure 3: Variogram of Yearly Precipitation Anomaly Data

The variogram is presented in Figure 3. The estimated $\hat{d}_q$ is around 250 km and

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0092

the ratio $\iota/\hat{d}_q \geq 5$. The maximized log-likelihood function is 895 for the Gaussian covariance function, 922 for the exponential covariance function, and 911 for the Cauchy covariance function. Therefore, among the three covariance functions, the exponential covariance function performs the best. From the exponential covariance function, we obtain $\hat{\boldsymbol{\theta}} = (130, 0.09, 0.72)$ and the standard error $se(\hat{\boldsymbol{\theta}}) = (29, 0.016, 0.13)$. The estimated range parameter $\hat{\theta}_1 = 130$ km and the estimated practical range is 374 km. Although the nugget effect $\hat{\theta}_2 = 0.09$ is much smaller than the partial sill $\hat{\theta}_3 = 0.72$, the measurement error is not negligible, because the standard error of $\hat{\theta}_2$ is 0.016.

Supplementary Material

The online Supplementary Material contains all proofs, as well as additional simulation studies.

Acknowledgments

The author thanks the editor, associate editor and referees for their helpful comments.

References

Bandyopadhyay, S. and S. S. Rao (2017). A test for stationarity for irregularly spaced spatial data. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 79, 95–123.

Beran, J. (1994). Statistics for Long-Memory Processes, Volume 61. CRC Press.

Chang, C.-H., H.-C. Huang, C.-K. Ing, et al. (2017). Mixed domain asymptotics for a stochastic process model with time trend and measurement error. Bernoulli 23, 159–190.

Chen, H.-S., D. G. Simpson, and Z. Ying (2000). Infill asymptotics for a stochastic process model with measurement error. Statistica Sinica 10, 141–156.

Cressie, N. (1993). Statistics for Spatial Data (Revised ed.). Wiley, New York.

Cressie, N. and S. Lahiri (1993). The asymptotic distribution of REML estimators. Journal of Multivariate

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0092

Analysis 45, 217–233.

Cressie, N. and S. N. Lahiri (1996). Asymptotics for REML estimation of spatial covariance parameters. Journal of Statistical Planning and Inference 50, 327–341.

De Jong, P. (1987). A central limit theorem for generalized quadratic forms. Probability Theory and Related Fields 75, 261–277.

Diggle, P. and P. Ribeiro (2007). Model-based Geostatistics. Springer, New York.

Diggle, P. J., J. Tawn, and R. Moyeed (1998). Model-based geostatistics. Journal of the Royal Statistical Society: Series C (Applied Statistics) 47, 299–350.

Furrer, R., M. G. Genton, and D. Nychka (2006). Covariance tapering for interpolation of large spatial datasets. Journal of Computational and Graphical Statistics 15, 502–523.

Gneiting, T. (2002). Compactly supported correlation functions. Journal of Multivariate Analysis 83, 493–508.

Gneiting, T. and M. Schlather (2004). Stochastic models that separate fractal dimension and the Hurst effect. SIAM review 46, 269–282.

Hall, P. and P. Patil (1994). Properties of nonparametric estimators of autocovariance for stationary random fields. Probability Theory and Related Fields 99, 399–424.

Johns, C. J., D. Nychka, T. G. F. Kittel, and C. Daly (2003). Infilling sparse records of spatial fields. Journal of the American Statistical Association 98, 796–806.

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

Lahiri, S. (2003). Central limit theorems for weighted sums of a spatial process under a class of stochastic and fixed designs. Sankhyā: The Indian Journal of Statistics 65, 356–388.

Lahiri, S. N. and J. Zhu (2006). Resampling methods for spatial regression models under a class of stochastic designs. The Annals of Statistics 34, 1774–1813.

Loh, W.-L. (2005). Fixed-domain asymptotics for a subclass of Matérn-type Gaussian random fields. The Annals of Statistics 33, 2344–2394.

Lu, Z. and D. Tjøstheim (2014). Nonparametric estimation of probability density functions for irregularly observed spatial data. Journal of the American Statistical Association 109, 1546–1564.

Mardia, K. V. and R. J. Marshall (1984). Maximum likelihood estimation of models for residual covariance

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0092

in spatial regression. Biometrika 71, 135–146.

Matérn, B. (1960). Spatial Variation: Stachastic Models and Their Application to Some Problems in Forest Surveys and Other Sampling Investigations. Statens skogsforskningsinstitut.

Matsuda, Y. and Y. Yajima (2009). Fourier analysis of irregularly spaced data on rd. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 71, 191–217.

Mikosch, T. and C. Stărică (2004). Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects. Review of Economics and Statistics 86, 378–390.

Schabenberger, O. and C. A. Gotway (2005). Statistical Methods for Spatial Data Analysis. Chapman & Hill/CRC, Boca Raton.

Shao, Q.-M. and Z.-S. Zhang (2019). Berry–Esseen bounds of normal and nonnormal approximation for unbounded exchangeable pairs. The Annals of Probability 47, 61–108.

Stein, M. L. (1999). Interpolation of Spatial Data: Some Theory for Kriging. New York: Springer.

Sweeting, T. (1980). Uniform asymptotic normality of the maximum likelihood estimator. The Annals of Statistics 8, 1375–1381.

Xu, G. and M. G. Genton (2017). Tukey g-and-h random fields. Journal of the American Statistical Association 112, 1236–1249.

Ying, Z. (1993). Maximum likelihood estimation of parameters under a spatial sampling scheme. The Annals of Statistics 21, 1567–1590.

Zhang, H. (2004). Inconsistent estimation and asymptotically equivalent interpolations in model-based geostatistics. Journal of the American Statistical Association 99, 250–261.

Zhang, H. and D. L. Zimmerman (2005). Towards reconciling two asymptotic frameworks in spatial statistics. Biometrika 92, 921–936.

Tingjin Chu

School of Mathematics and Statistics

University of Melbourne

E-mail: tingjin.chu@unimelb.edu.au