Statistica Sinica **34** (2024), 1863-1881
doi:https://doi.org/10.5705/ss.202021.0428

FIXED-DOMAIN ASYMPTOTICS UNDER VECCHIA'S APPROXIMATION OF SPATIAL PROCESS LIKELIHOODS

Lu Zhang, Wenpin Tang and Sudipto Banerjee*

University of Southern California, Columbia University and University of California, Los Angeles

Abstract: Statistical modeling for massive spatial data sets has generated a substantial body of literature on scalable spatial processes based on Vecchia's approximation. Vecchia's approximation for Gaussian process models enables fast evaluation of the likelihood by restricting dependencies at a location to its neighbors. We establish inferential properties of microergodic spatial covariance parameters within the paradigm of fixed-domain asymptotics when the parameters are estimated using Vecchia's approximation. We explore the conditions required to formally establish these properties, theoretically and empirically. Our results further corroborate the effectiveness of Vecchia's approximation from the standpoint of fixed-domain asymptotics.

Key words and phrases: Fixed-domain asymptotics, Gaussian processes, Matérn covariance function, microergodic parameters, spatial statistics.

1. Introduction

Geostatististical data are often modeled by treating observations as partial realizations of a spatial random field. We customarily model the random field $\{Y(s) : s \in \mathcal{D}\}$ over a bounded region $\mathcal{D} \in \mathbb{R}^d$ as a Gaussian process (GP), denoted as $Y(s) \sim GP(\mu_\beta(s), K_\theta(\cdot,\cdot))$, with mean $\mu_\beta(s)$ and covariance function $K_\theta(s, s') = \mathrm{cov}(y(s_i), y(s_j))$. The probability law for a finite set $\chi = \{s_1, s_2, \ldots, s_n\}$ is given by $y \sim N(\mu_\beta, K_\theta)$, where $y = (y(s_i))$ and $\mu_\beta = (\mu_\beta(s_i))$ are $n \times 1$ vectors with elements $y(s_i)$ and $\mu_\beta(s_i)$, respectively, and $K_\theta = (K_\theta(s_i, s_j))$ is an $n \times n$ spatial covariance matrix in which the (i, j)th element is the value of the covariance function $K_\theta(s_i, s_j)$. We consider the widely employed stationary Matérn covariance function (Matérn (1986); Stein (1999b)) given by

$$K_\theta(s, s') := \frac{\sigma^2(\phi\|h\|)^\nu}{\Gamma(\nu)2^{\nu-1}}\mathcal{K}_\nu(\phi\|h\|), \quad \|h\| \geq 0 \,, \tag{1.1}$$

where $h = s - s'$, $\sigma^2 > 0$ is called the *partial sill* or spatial variance, $\phi > 0$ is the scale or decay parameter, $\nu > 0$ is a smoothness parameter, $\Gamma(\cdot)$ is the gamma function, $\mathcal{K}_\nu(\cdot)$ is the modified Bessel function of order ν Abramowitz and Stegun (1965, Sec. 10), and $\theta = \{\sigma^2, \phi, \nu\}$. The spectral density corresponding to (1.1),

*Corresponding author.

which we use later, is

$$f(u) = C \frac{\sigma^2 \phi^{2\nu}}{(\phi^2 + u^2)^{\nu + d/2}} , \quad \text{for some } C > 0. \tag{1.2}$$

Likelihood-based inference for θ requires matrix computations in the order of $\sim n^3$ floating point operations (flops), which can become impractical when the number of spatial locations, n, is very large. Writing $Y_n = (y_1, y_2, \dots, y_n)^\top$, where $y_i := y(s_i)$, for $i = 1, 2, \dots, n$, are the n sampled measurements, we write $p(Y_n \,|\, \theta) := N(Y_n;\ \mu_\beta, K_\theta)$ as

$$p(Y_n \,|\, \theta) = p(y_1; \theta) \prod_{i=2}^{n} p(y_i \,|\, y_{(i-1)}; \theta) , \tag{1.3}$$

where $y_{(i)} = (y_1, \dots y_i)$. Vecchia (1988) suggested a simple approximation to (1.3) based on the notion that it may not be critical to use all components of $y_{(i-1)}$ in $p(y_i \,|\, y_{(i-1)}; \theta)$. Instead, the joint density $p(Y_n \,|\, \theta)$ in (1.3) is approximated by

$$\tilde{p}(Y_n \,|\, \theta) = p(y_1 \,|\, \theta) \prod_{i=2}^{n} p(y_i \,|\, S_{(i-1)}; \theta) , \tag{1.4}$$

where $S_{(i)}$ is a subvector of $y_{(i)}$, for $i = 1, \dots, n$. The density $\tilde{p}(Y_n \,|\, \theta)$ in (1.4) is called Vecchia's approximation and can be regarded as a quasi-likelihood or composite likelihood (Zhang (2012); Eidsvik et al. (2014); Bachoc and Lagnoux (2020)). Vecchia's approximation has now appeared in a large body of literature (see, e.g., Stein, Chi and Welty (2004); Datta et al. (2016a,b); Guinness (2018); Katzfuss et al. (2020); Katzfuss and Guinness (2021); Peruzzi, Banerjee and Finley (2022)). Algorithmic developments in Bayesian and frequentist settings (Finley et al. (2019); Zhang, Datta and Banerjee (2019); Katzfuss et al. (2020)) have enabled scalability to massive data sets (with $n \sim 10^7$ locations), and (1.4) lies at the core of several methods that tackle such "big data" problems in geospatial analysis (Sun, Li and Genton (2012); Banerjee (2017); Heaton et al. (2019)).

Vecchia's approximation has recently garnered substantial attention in the spatial statistics literature for building massively scalable GP models. However, although substantial methodological innovations have been generated using this approach, theoretical understanding of inference and identifiability of the spatial process parameters remains largely unaddressed. This is because Vecchia's approximation distorts the stationarity of the parent process and, hence, loses the theoretical tractability of spatial processes. Our current approach is a first attempt based upon Zhang (2012) to formally introduce methods that can be used to study the asymptotic properties of inference from Vecchia's approximation. Although we provide a rigorous development only in the one-dimensional setting, our proposed approach is novel, and should generate subsequent theoretical research in two dimensions. Therefore, we limit the formal theory to one dimension, but present some numerical experiments in two dimensions to show that the inferential behavior secured over the real line carries over to spatial domains.

Following the fixed-domain (infill) asymptotic paradigm for spatial inference (Stein (1999a); Zhang and Zimmerman (2005)), we discuss inferential properties for the parameters in (1.1). In this setting, Zhang (2004) shows that not all parameters in θ admit consistent maximum likelihood estimators from the full Gaussian likelihood in (1.3) constructed using a stationary Matérn covariance function, but that certain *microergodic* parameters are estimated consistently. Du, Zhang and Mandrekar (2009, Thm. 5) formally establish the asymptotic distributions of these microergodic parameters. Kaufman and Shaby (2013) jointly estimate the decay and the variance parameters in the Matérn family, and the effect of a prefixed decay on inference based on relatively small sample sizes. The aforementioned works all use (1.3). Here, we formally establish inferential properties for the estimates of microergodic parameters obtained from Vecchia's approximate likelihood in (1.4). Our work is motivated by a discussion in Section 10.5.3 of Zhang (2012) regarding the inferential behavior arising from (1.4). To the best of our knowledge, this is the first work to formally examine this topic. Following the aforementioned works in spatial asymptotics, we restrict our attention to the infill or fixed-domain setting, and focus on the inferential properties of the microergodic parameters for any given value of the smoothness parameter. Specifically, we examine the criteria for the asymptotic normality of the maximum likelihood estimates of the microergodic parameters obtained using Vecchia's approximation. In this regard, our work follows the paradigm laid out in Zhang (2012) in that we do not assume that the conditioning set is bounded. We provide conditions under which inferences under Vecchia's approximation of the Matérn process are asymptotically equivalent to the full model. This distinguishes our contribution from that of Bachoc and Lagnoux (2020), where bounded conditional sets are exploited to establish consistency results for some selected values of the smoothness parameter. In contrast, we show that a different set of conditions can yield a closed-form asymptotic distribution for any given value of the smoothness parameter. For the subsequent development, it suffices to assume that $\mu_\beta(s) = 0$, that is, the data have been detrended. Hence, we work with a zero-centered stationary GP with the Matérn covariance function in (1.1), a fixed smoothness parameter ν, and with the sampling locations χ_n restricted to a bounded region.

The remainder of this paper proceeds as follows. One of our key results, Theorem 1, is presented in Section 2, providing general criteria for the asymptotic normality of maximum likelihood estimates of microergodic parameters obtained using Vecchia's approximation. In Section 3, we demonstrate that these general criteria are implied by a condition on the conditioning size, which grows much slower than the sample size. We numerically check the conclusions for one-dimensional cases, and extend the discussion to two-dimensional cases in Section 4.

2. Infill Asymptotics for Vecchia's Approximation

2.1. Microergodic parameters

Identifiability and consistent estimation of θ in (1.1) rely on the equivalence and orthogonality of Gaussian measures. Two probability measures P_1 and P_2 on a measurable space $(\Omega, \mathcal{F})$ are said to be *equivalent*, denoted as $P_1 \equiv P_2$, if they are absolutely continuous with respect to each other. Thus, $P_1 \equiv P_2$ implies that for all $A \in \mathcal{F}$, $P_1(A) = 0$ if and only if $P_2(A) = 0$. On the other hand, P_1 and P_2 are orthogonal, denoted as $P_1 \perp P_2$, if there exists $A \in \mathcal{F}$ for which $P_1(A) = 1$ and $P_2(A) = 0$. Although measures might not be equivalent or orthogonal, Gaussian measures are, in general, one or the other. For a Gaussian probability measure P_θ indexed by a set of parameters θ and κ, a function of θ, we say that $\kappa(\theta)$ is *microergodic* if $\kappa(\theta_1) \neq \kappa(\theta_2)$ implies $P_{\theta_1} \perp P_{\theta_2}$ (see, e.g., Stein (1999b); Zhang (2012)). Two Gaussian probability measures defined by Matérn covariance functions $K_{\theta_1}(h)$ and $K_{\theta_2}(h)$, respectively, where $\theta_1 = \{\sigma_1^2, \phi_1, \nu\}$ and $\theta_2 = \{\sigma_2^2, \phi_2, \nu\}$ are equivalent if and only if $\sigma_1^2\phi_1^{2\nu} = \sigma_2^2\phi_2^{2\nu}$ (Theorem 2 in Zhang (2004)). Consequently, although one cannot consistently estimate σ^2 or ϕ (Corollary 1 in Zhang (2004)) from full GP likelihood functions, $\sigma^2\phi^{2\nu}$ is a microergodic parameter that can be estimated consistently.

If the oracle (data-generating) values of ϕ and σ^2 are ϕ_0 and σ_0^2, respectively, then for any fixed value of the decay $\phi = \phi_1$, we know from Du, Zhang and Mandrekar (2009, Thm. 5) that

$$\sqrt{n}(\hat{\sigma}_n^2\phi_1^{2\nu} - \sigma_0^2\phi_0^{2\nu}) \xrightarrow{\mathcal{L}} N(0, 2(\sigma_0^2\phi_0^{2\nu})^2)\ , \tag{2.1}$$

where $\hat{\sigma}_n^2$ is the maximum likelihood estimator from the full likelihood (1.3).

2.2. Parameter estimation

Let $\hat{\sigma}_{n,vecch}^2$ be the maximum likelihood estimate of the variance σ^2,

$$\hat{\sigma}_{n,vecch}^2 = \underset{\sigma^2}{\operatorname{argmax}}\{\tilde{p}(Y_n \mid \phi_1, \sigma^2),\ \sigma^2 \in \mathbb{R}^+\}\ , \tag{2.2}$$

where $\tilde{p}(\cdot)$ is the density (1.4). We develop the asymptotic equivalence of $\hat{\sigma}_{n,vecch}^2$ using $\hat{\sigma}_n^2$. Before proceeding further, we introduce some notation. Assume that the target process $y(s) \sim GP(0, K_\theta(\cdot))$, where $K_\theta(h)$ is defined in (1.1), has a fixed ν. Let P_j, for $j = 0, 1$, denote probability measures for $y(s) \sim GP(0, K_{\theta_i})$ with $\theta_j = \{\sigma_j^2, \phi_j, \nu\}$. Assume that $\sigma_1^2 = \sigma_0^2\phi_0^{2\nu}/\phi_1^{2\nu}$ and let $E_j(\cdot)$ denote the expectation with respect to probability measure P_j, for $j = 0, 1$. We define

$$\begin{aligned} &e_{0,j} := y_1,\ \mu_{i,j} := E_j(y_i \,|\, y_{(i-1)}),\ e_{i,j} := y_i - \mu_{i,j},\ i = 2, \ldots, n \\ &\tilde{e}_{0,j} := y_1,\ \tilde{\mu}_{i,j} := E_j(y_i \,|\, S_{(i-1)}),\ \tilde{e}_{i,j} := y_i - \tilde{\mu}_{i,j},\ i = 2, \ldots, n\ . \end{aligned} \tag{2.3}$$

In Lemma 1, we derive a useful expression for $\hat{\sigma}^2_{n,vecch}$ using the quantities in (2.3).

Lemma 1. *The estimate of σ^2 from Vecchia's likelihood approximation with fixed ν and $\phi = \phi_1$ can be expressed as*

$$\hat{\sigma}^2_{n,vecch} = \frac{\sigma_1^2}{n} \sum_{i=1}^{n} \frac{\tilde{e}^2_{i,1}}{E_1 \tilde{e}^2_{i,1}}\ . \tag{2.4}$$

Proof. In Vecchia's approximation (1.4) with fixed ν, $\phi = \phi_1$, and unknown σ^2 in $K_\theta(\cdot)$, $p(y_i \,|\, S_{(i-1)})$ is Gaussian with mean $\widetilde{\mu}_{i,1} = \widetilde{\Sigma}^{12}_{i,1}(\widetilde{\Sigma}^{22}_{i,1})^{-1} S_{(i-1)}$ and variance $\widetilde{\Sigma}_{i,1} := \widetilde{\Sigma}^{11}_{i,1} - \widetilde{\Sigma}^{12}_{i,1}(\widetilde{\Sigma}^{22}_{i,1})^{-1}\widetilde{\Sigma}^{21}_{i,1}$, where $\begin{pmatrix} \widetilde{\Sigma}^{11}_{i,1} & \widetilde{\Sigma}^{12}_{i,1} \\ \widetilde{\Sigma}^{21}_{i,1} & \widetilde{\Sigma}^{22}_{i,1} \end{pmatrix}$ is the covariance matrix of $\begin{pmatrix} y_i \\ S_{(i-1)} \end{pmatrix}$ under $\tilde{p}(\cdot;\phi_1,\sigma^2)$. Because $\widetilde{\mu}_{i,1}$ does not depend on σ^2, and $\widetilde{\Sigma}_{i,1}$ can be expressed as $\sigma^2 \widetilde{\Sigma}^\dagger_{i,1}$, where $\widetilde{\Sigma}^\dagger_{i,1}$ does not depend on σ^2, the conditional distribution $p(y_i \,|\, S_{(i-1)})$ under Vecchia's approximation is

$$p(y_i \,|\, S_{(i-1)}) = \frac{1}{\sqrt{2\pi\sigma^2 \widetilde{\Sigma}^\dagger_{i,1}}} \exp\left(-\frac{\tilde{e}^2_{i,1}}{2\sigma^2 \widetilde{\Sigma}^\dagger_{i,1}} \right).$$

A direct computation of (2.2) with any fixed ϕ_1 yields (2.4), where we use the fact $E_1 \tilde{e}^2_{i,1} = \sigma_1^2 \widetilde{\Sigma}^\dagger_{i,1}$ on the right-hand side of (2.4).

Our main result builds on the discussion in Section 10.5.3 of Zhang (2012), yielding the following theorem that explores the asymptotic distribution of $\hat{\sigma}^2_{n,vecch}$.

Theorem 1. *Assume that one of the following conditions holds:*

$$\sum_{i=1}^{n} \frac{E_0(\tilde{e}_{i,1} - e_{i,0})^2}{E_1 \tilde{e}^2_{i,1}} = \mathcal{O}(1) \quad \textit{and} \quad \sum_{i=1}^{n} \left(\frac{E_0 e^2_{i,0}}{E_1 \tilde{e}^2_{i,1}} - 1 \right)^2 = \mathcal{O}(1) \tag{2.5}$$

or

$$\sum_{i=1}^{n} \frac{E_1(\tilde{e}_{i,1} - e_{i,0})^2}{E_0 e^2_{i,0}} = \mathcal{O}(1) \quad \textit{and} \quad \sum_{i=1}^{n} \left(\frac{E_1 \tilde{e}^2_{i,1}}{E_0 e^2_{i,0}} - 1 \right)^2 = \mathcal{O}(1). \tag{2.6}$$

Then,

$$\sqrt{n}(\hat{\sigma}^2_{n,vecch}\phi_1^{2\nu} - \sigma_0^2\phi_0^{2\nu}) \xrightarrow{\mathcal{L}} N(0, 2(\sigma_0^2\phi_0^{2\nu})^2)\ . \tag{2.7}$$

Before presenting the proof of Theorem 1, we state and prove the following lemma.

Lemma 2. *The assumptions in* (2.5) *imply that*

$$E_0\left[\sum_{i=1}^{n}\frac{\tilde{e}_{i,1}^2}{E_1\tilde{e}_{i,1}^2}-\sum_{i=1}^{n}\frac{e_{i,0}^2}{E_0e_{i,0}^2}\right]=o(\sqrt{n}). \tag{2.8}$$

Proof. We first prove that (2.5) implies (2.8). Note that

$$\begin{aligned}\left|E_0\left[\sum_{i=1}^{n}\frac{\tilde{e}_{i,1}^2}{E_1\tilde{e}_{i,1}^2}-\sum_{i=1}^{n}\frac{e_{i,0}^2}{E_0e_{i,0}^2}\right]\right| &= \left|\sum_{i=1}^{n}\left(\frac{E_0(\tilde{e}_{i,1}-e_{i,0}+e_{i,0})^2}{E_1\tilde{e}_{i,1}^2}-1\right)\right| \\ &= \left|\sum_{i=1}^{n}\frac{E_0(\tilde{e}_{i,1}-e_{i,0})^2}{E_1\tilde{e}_{i,1}^2}+\sum_{i=1}^{n}\left(\frac{E_0e_{i,0}^2}{E_1\tilde{e}_{i,1}^2}-1\right)\right| \\ &\le \sum_{i=1}^{n}\frac{E_0(\tilde{e}_{i,1}-e_{i,0})^2}{E_1\tilde{e}_{i,1}^2}+\sum_{i=1}^{n}\left|\frac{E_0e_{i,0}^2}{E_1\tilde{e}_{i,1}^2}-1\right|,\end{aligned}$$

where the second equality follows because $\tilde{e}_{i,1}-e_{i,0}$ and $e_{i,0}$ are independent under P_0. By the first condition in (2.5), we get $\sum_{i=1}^{n}E_0(\tilde{e}_{i,1}-e_{i,0})^2/E_1\tilde{e}_{i,1}^2=\mathcal{O}(1)=o(\sqrt{n})$. Fix $\varepsilon>0$. By the second condition in (2.5), there is $M>0$ such that $\sum_{i>M}(E_0e_{i,0}^2/E_1\tilde{e}_{i,1}^2-1)^2<\varepsilon$. Thus, for $n>M$, we can use the Cauchy–Schwarz inequality to obtain

$$\begin{aligned}\sum_{i=1}^{n}\left|\frac{E_0e_{i,0}^2}{E_1\tilde{e}_{i,1}^2}-1\right| &= \sum_{i=1}^{M}\left|\frac{E_0e_{i,0}^2}{E_1\tilde{e}_{i,1}^2}-1\right|+\sum_{i=M+1}^{n}\left|\frac{E_0e_{i,0}^2}{E_1\tilde{e}_{i,1}^2}-1\right| \\ &\le \sum_{i=1}^{M}\left|\frac{E_0e_{i,0}^2}{E_1\tilde{e}_{i,1}^2}-1\right|+\sqrt{(n-M)\varepsilon}\ .\end{aligned}$$

Dividing both sides by $\sqrt{n}$ reveals that $\limsup_{n\to\infty}(1/\sqrt{n})\sum_{i=1}^{n}|E_0e_{i,0}^2/E_1\tilde{e}_{i,1}^2-1|\le\sqrt{\varepsilon}$. Since $\varepsilon>0$ is arbitrary, it follows that $\sum_{i=1}^{n}|E_0e_{i,0}^2/E_1\tilde{e}_{i,1}^2-1|=o(\sqrt{n})$, and we obtain (2.8).

We now present a proof of Theorem 1.

Proof of Theorem 1. Recall that $\sum_{i=1}^{n}e_{i,0}^2/(E_0e_{i,0}^2)=Y_n^\top V_{n,0}^{-1}Y_n$, where $V_{n,0}$ is the covariance matrix of Y_n under P_0. Using (2.4) derived in Lemma 1, we obtain

$$\sqrt{n}\left(\frac{\hat{\sigma}_{n,vecch}^2}{\sigma_1^2}-1\right)=\frac{1}{\sqrt{n}}\left[\sum_{i=1}^{n}\frac{\tilde{e}_{i,1}^2}{E_1\tilde{e}_{i,1}^2}-\sum_{i=1}^{n}\frac{e_{i,0}^2}{E_0e_{i,0}^2}\right]+\sqrt{n}\left(\frac{1}{n}Y_n^\top V_{n,0}^{-1}Y_n-1\right). \tag{2.9}$$

By the central limit theorem, we have $\sqrt{n}((1/n)Y_n^\top V_{n,0}^{-1}Y_n-1)\xrightarrow{\mathcal{L}}N(0,2)$.

We next show that condition (2.5) implies that

$$\sum_{i=1}^{n} \frac{\tilde{e}_{i,1}^2}{E_1 \tilde{e}_{i,1}^2} - \sum_{i=1}^{n} \frac{e_{i,0}^2}{E_0 e_{i,0}^2} = o(\sqrt{n}). \tag{2.10}$$

To prove this, it is sufficient to show that

$$\sum_{i=1}^{n} \frac{\tilde{e}_{i,1}^2}{E_1 \tilde{e}_{i,1}^2} - \sum_{i=1}^{n} \frac{e_{i,0}^2}{E_0 e_{i,0}^2} - E_0 \left[\sum_{i=1}^{n} \frac{\tilde{e}_{i,1}^2}{E_1 \tilde{e}_{i,1}^2} - \sum_{i=1}^{n} \frac{e_{i,0}^2}{E_0 e_{i,0}^2} \right] = O(1). \tag{2.11}$$

The result in (2.10) then follows from Lemma 2. Next we prove (2.11). Our argument relies on the equivalence of Gaussian sequences. Let $\widetilde{P}_{1,n}$ be the probability distribution corresponding to $\tilde{p}(Y_n; \phi_1, \sigma_1^2)$, and let $\rho_n := \tilde{p}(Y_n; \phi_1, \sigma_1^2)/p(Y_n; \phi_0, \sigma_0^2)$ be the Radon–Nikodym derivative of $\widetilde{P}_{1,n}$ with respect to P_0 on the realization Y_n for a given n. Write $\widetilde{P}_{1,\infty}$ (respectively, $P_{0,\infty}$) for the probability distribution corresponding to $\tilde{p}(\cdot; \phi_1, \sigma_1^2)$ (respectively, $p(\cdot; \phi_0, \sigma_0^2)$) on the infinite sequence $(Y_1, Y_2, \ldots)$. By Kakutani's dichotomy, $\widetilde{P}_{1,\infty}$ and $P_{0,\infty}$ are either equivalent or mutually singular to each other. If $\widetilde{P}_{1,\infty}$ is equivalent to $P_{0,\infty}$, then $\lim_{n\to\infty} \rho_n = \rho_\infty =: d\widetilde{P}_{1,\infty}/dP_{0,\infty}$ with P_0-probability one (see, e.g., Ibragimov and Rozanov (1978, Sec. III.2.1)). In addition, $\mathbb{P}_0(0 < \rho_\infty < \infty) = 1$ and $-\infty < E_0(\log \rho_\infty) < \infty$. As a result,

$$\log \rho_n = -\frac{1}{2} \log \frac{\det \widetilde{V}_{n,1}}{\det V_{n,0}} - \frac{1}{2} \left(\sum_{i=1}^{n} \frac{\tilde{e}_{i,1}^2}{E_1 \tilde{e}_{i,1}^2} - \sum_{i=1}^{n} \frac{e_{i,0}^2}{E_0 e_{i,0}^2} \right) = \mathcal{O}(1),$$
$$E_0(\log \rho_n) = -\frac{1}{2} \log \frac{\det \widetilde{V}_{n,1}}{\det V_{n,0}} - \frac{1}{2} E_0 \left[\sum_{i=1}^{n} \frac{\tilde{e}_{i,1}^2}{E_1 \tilde{e}_{i,1}^2} - \sum_{i=1}^{n} \frac{e_{i,0}^2}{E_0 e_{i,0}^2} \right] = \mathcal{O}(1),$$

where $\widetilde{V}_{n,1}$ (respectively, $V_{n,0}$) is the covariance matrix of Y_n under $\widetilde{P}_{1,\infty}$ (respectively, P_0). By taking the difference of the above two equations, we get (2.11) under the condition that $\widetilde{P}_{1,\infty}$ is equivalent to $P_{0,\infty}$. Using Theorem 5, Section VII.6 of Shiryaev (1996), we conclude that

$$\begin{aligned} \widetilde{P}_{1,\infty} \text{ is equivalent to } P_{0,\infty} &\Longleftrightarrow \sum_{i=1}^{\infty} \left[\frac{E_0(\tilde{e}_{i,1} - e_{i,0})^2}{E_1 \tilde{e}_{i,1}^2} + \left(\frac{E_0 e_{i,0}^2}{E_1 \tilde{e}_{i,1}^2} - 1 \right)^2 \right] < \infty \\ &\Longleftrightarrow \sum_{i=1}^{\infty} \left[\frac{E_1(\tilde{e}_{i,1} - e_{i,0})^2}{E_0 e_{i,0}^2} + \left(\frac{E_1 \tilde{e}_{i,1}^2}{E_0 e_{i,0}^2} - 1 \right)^2 \right] < \infty. \end{aligned} \tag{2.12}$$

Because the first equivalence in (2.12) is simply a reformulation of (2.5), we have established that (2.5) implies (2.11), and hence the result in (2.10).

The proof from condition (2.6) to (2.10) is established following the proof from condition (2.5) to (2.10). We now break the quantity $\sum_{i=1}^{n} \tilde{e}_{i,1}^2/E_1 \tilde{e}_{i,1}^2 -$

$\sum_{i=1}^n e_{i,0}^2/E_0 e_{i,0}^2$ in (2.10) into

$$\sum_{i=1}^n \frac{\tilde{e}_{i,1}^2}{E_1\tilde{e}_{i,1}^2} - \sum_{i=1}^n \frac{e_{i,0}^2}{E_0 e_{i,0}^2} - E_1\left[\sum_{i=1}^n \frac{\tilde{e}_{i,1}^2}{E_1\tilde{e}_{i,1}^2} - \sum_{i=1}^n \frac{e_{i,0}^2}{E_0 e_{i,0}^2}\right]$$
$$+E_1\left[\sum_{i=1}^n \frac{\tilde{e}_{i,1}^2}{E_1\tilde{e}_{i,1}^2} - \sum_{i=1}^n \frac{e_{i,0}^2}{E_0 e_{i,0}^2}\right],$$

and replace the left-hand sides of (2.11) and (2.8) with the two quantities, respectively. Then, the equivalence of P_1 and P_0, along with Lemma 2, shows that the replaced (2.8) holds. The proof that (2.6) implies the replaced (2.11) remains the same, except that we now use the second equivalence in (2.12). Thus, we complete the proof of Theorem 1.

Turning to the connection between Theorem 1 and the predictive consistency of Vecchia's approximation in the sense of Kaufman and Shaby (2013, p.478), note that $e_{i,0}$ and $\tilde{e}_{i,1}$ are the predictive errors for y_i under the full model with correct parameters and under (1.4) with possibly incorrectly specified parameters $(\phi, \sigma) = (\phi_1, \sigma_1)$, respectively. A consequence of (2.5) or (2.6) is that $E_1\tilde{e}_{i,1}^2/E_0 e_{i,0}^2 \to 1$ as i (and hence n) increases. Hence, (2.5) or (2.6) implies the asymptotic normality of the estimates and the predictive consistency.

3. Infill Asymptotics for Vecchia's Approximation on the Line

We can obtain further insights into Theorem 1 by considering the asymptotic normality of $\hat{\sigma}_{n,vecch}$ for Matérn models with observations on the real line. Although the conditions (2.5) and (2.6) are, in general, analytically intractable, owing to the presence of $E_0\tilde{e}_{i,1}^2$, we will show that (2.6) holds for Matérn models on $\mathbb{R}$.

To simplify the presentation, we consider the fixed domain $\mathcal{D} = [0, 1]$, and the sampled locations $\chi = \{i/n : 0 \le i \le n\}$. Denote $\delta = 1/n$ for the spacing of χ, and $y_i = y(i\delta)$, for $0 \le i \le n$, for the observations. We define $S_{(i)} = S_{(i)}[k] := (y_i, y_{i-1}, \ldots, y_{i-k+1})$ for a positive integer k, where $S_{(i)}[k]$ is a vector of k consecutive observations backward from y_i. The integer k is capped by i because $S_{(i)}[k]$ is a subvector of $y_{(i)}$.

Assumption 1. *Let $\mathcal{D} = [0, 1]$, and $\chi = \{i\delta : 0 \le i \le n\}$ with $\delta = 1/n$. Then,*

$$\sum_{i=1}^n \frac{E_1(e_{i,1} - e_{i,0})^2}{E_1 e_{i,1}^2} = \mathcal{O}(1). \tag{3.1}$$

Before stating the main result on $\mathbb{R}$, we demonstrate why this assumption is reasonable. We empirically investigate $\sum_{i=1}^n E_1(e_{i,1} - e_{i,0})^2/E_1 e_{i,1}^2$ for increasing values of n. Figure 1 plots the values of $\sum_{i=1}^n E_1(e_{i,1} - e_{i,0})^2/E_1 e_{i,1}^2$, with $\chi = \{i\delta : 0 \le i \le n\}$, for $\nu = 0.25, 0.5, 1.0, 1.5, 2.0$, and n ranging from 100 to 1,200.

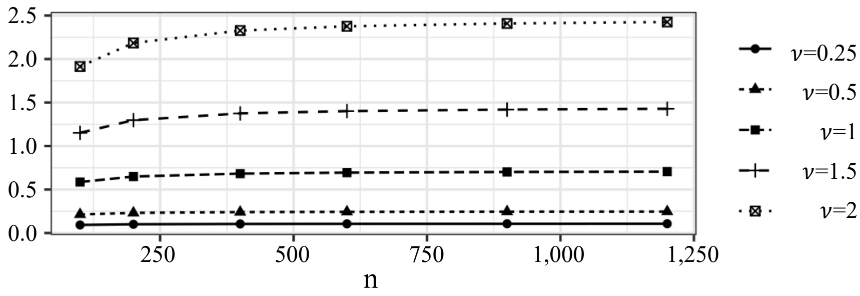


Figure 1. Trend of $\sum_{i=1}^{n} E_1(e_{i,1} - e_{i,0})^2/E_1 e_{i,1}^2$ for the Matérn model when χ is a regular grid $\chi = \{i\delta : 0 \le i \le n\}$. The parameter σ^2 in the Matérn covariogram is equal to 1.0 and we set the decay ϕ for different ν to make the correlation of two points equal to 0.05 when their distance reaches 0.2.

As n increases, the plot tends to flatten, as suggested by the assumption.

Some additional explanation is also possible from a theoretical viewpoint. Since P_0 and P_1 are equivalent, Corollary 3.1 of Stein (1990b) implies that $E_1(e_{i,1} - e_{i,0})^2/E_1 e_{i,1}^2 \to 0$ as $n, i \to \infty$. By the stationarity and the symmetry of the Matérn model, $e_{i,j}$ is distributed as the error of the least square estimate of $y_0 := y(0)$, given observations $y_{(i)} = (y_1, \ldots, y_i)^\top$. Hence, $e_{i,j}$ can be realized as $y_0 - E_j(y_0 \,|\, y_{(i)})$. Similarly, $\tilde{e}_{i,j}$ can be realized as $y_0 - E_j(y_0 \,|\, S_{(i-1)})$. Now, consider the infinitely sampled locations $\{i\delta : i \ge 0\}$, and extend the finite sample $Y_n := (y_0, \ldots, y_n)^\top$ to $Y := (y_0, y_1, \ldots)^\top$, with $y_i := y(i\delta)$. The sampled locations of Y form an infinite grid on $[0, \infty)$, and $\delta = 1/n$ is determined based on the sample size of Y_n. Let $e_{\infty,j}$ be the error $y_0 - E_j(y_0 \,|\, y_1, \ldots)$ for the infinite sequence Y. For f_0 (resp., f_1), the spectral density under P_0 (resp., P_1), it is easily seen that $f_0, f_1 \sim C\sigma_0^2\phi_0^\nu u^{-2\nu-1}$ as $u \to \infty$, and $(f_1 - f_0)/f_0 \asymp u^{-2}$. Therefore, by Theorem 2 of Stein (1999b),

$$\frac{E_1(e_{\infty,1} - e_{\infty,0})^2}{E_1 e_{\infty,1}^2} = O(\delta^{\min(2\nu+1,4)} \log(\delta^{-1})^{1(\nu=3/2)})\,. \tag{3.2}$$

Intuitively, $e_{i,j} \approx e_{\infty,j}$ for large i. It is not unreasonable to speculate a stronger result in which (3.2) still holds by replacing $e_{\infty,j}$ with $e_{i,j}$ for large i, which would imply (3.1). As indicated on p.138 of Stein (1999a), obtaining the rate of $E_1(e_{i,1} - e_{i,0})^2/E_1 e_{i,1}^2$ for any bounded domain $\mathcal{D}$ is a highly nontrivial task. The only known results are obtained by Stein (1990a, 1999b) for $\nu = 1/2, 3/2, \ldots$, which also imply (3.1). In fact, we need that $E_1 e_{i,1}^2 = E_1 e_{\infty,1}^2(1 + \mathcal{O}(i^{-\kappa}))$, for any $\kappa > 0$. Hence, the only missing piece in these heuristics is an estimate of $E_1 e_{i,0}^2/E_1 e_{\infty,0}^2$, which we do not explore further here. Our result is stated as follows.

Theorem 2. *Let $\mathcal{D} = [0,1]$, $\chi = \{i\delta : 0 \le i \le n\}$, and $S_{(i)} = S_{(i)}[n^\epsilon]$, for $\epsilon \in (0,1)$. If* (3.1) *holds, then* (2.6) *also holds. Consequently,* (2.7) *holds for the Matérn model ($\nu > 0$).*

We make a few remarks before presenting a proof. Theorem 2 states that the asymptotic normality of the microergodic parameter $\sigma^2\phi^{2\nu}$ still holds under Vecchia's approximation in a neighborhood of size at most $k = n^\epsilon \ll n$ (sample size), where the computation of $\hat{\sigma}^2_{n,vecch}$ is much cheaper. This justifies the validity of Vecchia's approximation for Matérn models from a fixed-domain perspective. The range n^ϵ may not be optimal, and it might be possible to improve to $k = \mathcal{O}(\log n)$. However, we do not pursue this direction here from a theoretical standpoint. A simulation study is provided for the case of $k = \mathcal{O}(\log n)$ in Section 4.

An interesting situation arises with $\nu = 1/2$, where the process reduces to the Ornstein–Uhlenbeck process, and $p(y_i \,|\, y_{(i-1)}) = p(y_i \,|\, y_{i-1})$. Therefore, (2.7) holds trivially for Vecc hia's approximation with a neighborhood of size $k = 1 = \mathcal{O}(1)$. Therefore, it is natural to enquire whether the asymptotic normality of (2.7) holds under Vecchia's approximation within a range $k = \mathcal{O}(1)$. If this is true, computational efforts can be reduced further. Unfortunately, this need not be the case. For $\nu < 1/4$ and $k = \mathcal{O}(1)$, $n^{2\nu}(\hat{\sigma}^2_{n,vecch}\phi_1^{2\nu} - \sigma_0^2\phi_0^{2\nu})$ converges to a non-Gaussian distribution (Bachoc and Lagnoux (2020)). The cases for $\nu \ge 1/4$ remain unresolved.

Now, we turn to the proof of Theorem 2. The key to this analysis is the following proposition, which relies on a result on the bound of $e_{\infty,j} - e_{i,j}$, that is, the difference between the errors of the finite and the infinite least square estimates (Baxter (1962)). The study dates back to the work of Kolmogorov (1941); see also Grenander and Szegö (1958); Ibragimov (1964); Dym and McKean (1970, 1976) and Ginovian (1999) for related discussions.

Proposition 1. *Let $\kappa > 0$. There exist $C_0, C_\kappa > 0$ such that for $\delta < 1$ and $j = 0, 1$,*

$$E_j e^2_{\infty,j} \sim C_0 \delta^{2\nu}, \tag{3.3}$$

$$E_j(e_{\infty,j} - e_{i,j})^2 \le C_\kappa \delta^{2\nu} i^{-\kappa}. \tag{3.4}$$

Proof. From the discussion below (3.1) that $e_{i,j}$ and $e_{\infty,j}$ can be realized as $e_{i,j} = y_0 - E_j(y_0 \,|\, y_1, \ldots, y_i)$ and $e_{\infty,j} = y_0 - E_j(y_0 \,|\, y_1, y_2, \ldots)$, respectively, where $y_i := y(i\delta)$ is indexed by nonnegative integers, we know from Stein (1999a, p.77) that the spectral density of Y under P_j is

$$\overline{f}_j^\delta(u) = \frac{1}{\delta} \sum_{\ell=-\infty}^{\infty} f_j\left(\frac{u + 2\pi\ell}{\delta}\right) \quad \text{for } u \in (-\pi, \pi], \quad j = 0, 1\ ,$$

where f_j is the spectral density defined by (1.2) corresponding to P_j. For $j = 0, 1$,

$f_j(u) \sim C\sigma_0^2\phi_0^{2\nu}u^{-2\nu-1}$ as $u \to \infty$. From Stein (1999a, p.80, (17)), we obtain

$$E_je_{\infty,j}^2 \sim 2\pi C\sigma_0^2\phi_0^{2\nu}\delta^{2\nu}\exp\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\left(\sum_{\ell=-\infty}^{\infty}|u+2\pi\ell|^{-2\nu-1}\right)du\right),$$

which implies (3.3), with

$$C_0 = 2\pi C\sigma_0^2\phi_0^{2\nu}\exp\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\left(\sum_{\ell=-\infty}^{\infty}|u+2\pi\ell|^{-2\nu-1}\right)du\right).$$

Turning to (3.4), we know from Baxter (1962, p.142, (15)) that

$$E_j(e_{i,j}-e_{\infty,j})^2 = E_je_{\infty,j}^2\,E_je_{i,j}^2\sum_{m=i}^{\infty}|\phi_{m,j}(0)|^2, \tag{3.5}$$

where $\phi_{m,j}(\cdot)$ are the Szegö polynomials associated with the spectral $\overline{f}_j^{\delta}$ (see Section 2.1 of Grenander and Szegö (1958), for background). Note that $E_je_{\infty,j}^2 \sim C_0\delta^{2\nu}$ and $E_je_{i,j}^2 \le E_je_{1,j}^2 \to 0$ as $\delta \to 0$. It is sufficient to establish

$$\sum_{m=0}^{\infty} m^{\kappa}|\phi_{m,j}(0)| \le D_{\kappa}\,, \quad \text{for some } D_{\kappa} > 0, \tag{3.6}$$

in which case, the identity (3.5) implies (3.4). The key observation of Baxter (1962) (Theorem 2.3) is that (3.6) holds if the κ^{th} moment of the Fourier coefficients associated with $\overline{f}_j^{\delta}$ is bounded from above by D'_{κ}, for some $D'_{\kappa} > 0$, that is,

$$\sum_{m=0}^{\infty} m^{\kappa}|c_{m,j}| < D'_{\kappa}, \quad \text{where } c_{m,j} := \frac{1}{2\pi}\int_{-\pi}^{\pi}\overline{f}_j^{\delta}(u)e^{-inu}du.$$

A sufficient condition for the latter to hold is that the κ^{th} derivative of $\overline{f}_j^{\delta}$ is integrable, and $\int_{-\pi}^{\pi}|d^{\kappa}\overline{f}_j^{\delta}(u)/du^{\kappa}|du \le D''_{\kappa}$, for some $D''_{\kappa} > 0$, which does not depend on $\delta < 1$. Breaking the sum of $\overline{f}_j^{\delta}$ according to $\ell = 0$ and $\ell \neq 0$ produces

$$\int_{-\pi}^{\pi}\left|\frac{d^{\kappa}}{du^{\kappa}}\overline{f}_j^{\delta}(u)\right|du \le A\int_{-\infty}^{\infty}(1+u^{2\nu+1+\kappa})^{-1}du + A'\delta^{2\nu}\sum_{\ell\neq 0}l^{-2\nu-1-\kappa}\,, \tag{3.7}$$

where $A, A' > 0$ are numerical constants. Hence, the right-hand side of (3.7) is bounded by $D''_{\kappa} = A\int_{-\infty}^{\infty}(1+u^{2\nu+1+\kappa})^{-1}du + A'\sum_{\ell\neq 0}l^{-2\nu-1-\kappa}$, which depends only on κ.

Proof of Theorem 2. We fix $\kappa = 2/\epsilon$, and use $C_{\kappa}^{(1)}, C_{\kappa}^{(2)}, \ldots$ to denote constants depending only on κ. Note that $E_0e_{i,0}^2 = E_0e_{\infty,0}^2 + E_0(e_{\infty,0}-e_{i,0})^2$ because $e_{\infty,0}$ and $e_{\infty,0} - e_{i,0}$ are independent under P_0. By Proposition 1, we get $E_0e_{i,0}^2 \sim$

$C_0\delta^{2\nu}(1 + C_\kappa^{(1)} i^{-\kappa})$. Similarly, $E_1\tilde{e}_{i,1}^2 = E_1 e_{\infty,1}^2 + E_1(\tilde{e}_{i,1} - e_{\infty,1})^2 \sim C_0\delta^{2\nu}(1 + C_\kappa^{(2)} \min(i, n^\epsilon)^{-\kappa})$, because $\tilde{e}_{i,1}$ is realized as $y_0 - E_1\left(y_0 \,|\, S_{(i-1)}[k]\right)$, with $k = \min(i, n^\epsilon)$. Therefore,

$$\sum_{i=1}^{n}\left(\frac{E_1\tilde{e}_{i,1}^2}{E_0 e_{i,0}^2} - 1\right)^2 \leq \sum_{i\geq n^\epsilon}\left(\frac{C_\kappa^{(3)} n^{-\epsilon\kappa}}{1 + C_\kappa^{(1)} i^{-\kappa}}\right)^2 + \sum_{i=1}^{n}\left(\frac{C_\kappa^{(4)} i^{-\kappa}}{1 + C_\kappa^{(1)} i^{-\kappa}}\right)^2$$
$$\leq \frac{C_\kappa^{(5)}}{n^3} + C_\kappa^{(6)}\sum_{i=1}^{n} i^{-2\kappa} \leq C_\kappa^{(7)}.$$

Moreover, we have

$$\frac{E_1(\tilde{e}_{i,1} - e_{i,0})^2}{E_0 e_{i,0}^2} \leq 3\frac{E_1(\tilde{e}_{i,1} - e_{\infty,1})^2}{E_0 e_{i,0}^2} + 3\frac{E_1(e_{i,1} - e_{\infty,1})^2}{E_0 e_{i,0}^2} + 3\frac{E_1(e_{i,1} - e_{i,0})^2}{E_1 e_{i,1}^2}\frac{E_1 e_{i,1}^2}{E_0 e_{i,0}^2}.$$

By the same argument as above, for the first two terms,

$$\sum_{i=1}^{n}\frac{E_1(\tilde{e}_{i,1} - e_{\infty,1})^2}{E_0 e_{i,0}^2} < C_\kappa^{(8)} \quad \text{and} \quad \sum_{i=1}^{n}\frac{E_1(e_{i,1} - e_{\infty,1})^2}{E_0 e_{i,0}^2} < C_\kappa^{(9)}.$$

For the last term,

$$\sum_{i=1}^{n}\frac{E_1(e_{i,1} - e_{i,0})^2}{E_1 e_{i,1}^2}\frac{E_1 e_{i,1}^2}{E_0 e_{i,0}^2} \leq C_\kappa^{(10)}\sum_{i=1}^{n}\frac{E_1(e_{i,1} - e_{i,0})^2}{E_1 e_{i,1}^2},$$

which converges because of (3.1). This establishes (2.6) and, hence, (2.7) follows.

4. Simulations

Based on (2.8) and (2.11) provided in Theorem 1, Theorem 2 proves that

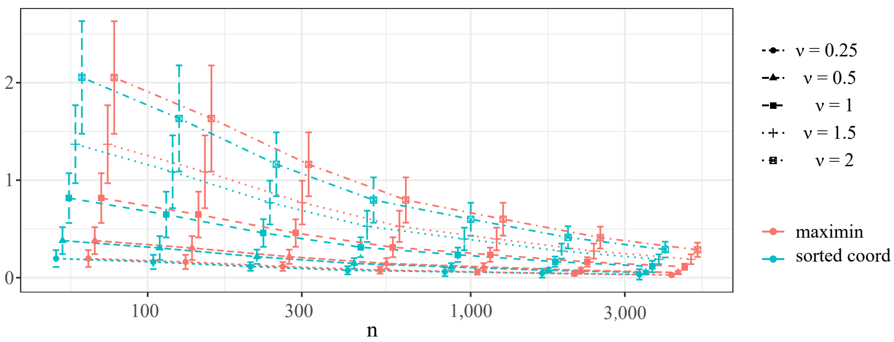
$$c_n(\phi_1, \phi_0, k) = \frac{1}{\sqrt{n}}\left[\sum_{i=1}^{n}\frac{\tilde{e}_{i,1}^2}{E_1\tilde{e}_{i,1}^2} - \sum_{i=1}^{n}\frac{e_{i,0}^2}{E_0 e_{i,0}^2}\right] = o(1) \tag{4.1}$$

when $k = n^\epsilon$, for $\epsilon \in (0, 1)$. The equation (4.1) induces the critical condition (2.10), resulting in the convergence in law in (2.7). Considering the more challenging case $k = \mathcal{O}(\log(n))$, we extend the discussion in Theorem 2 by investigating the behavior of $c_n(\phi_1, \phi_0, k)$ in (4.1) for a sequence of data sets with increasing sample sizes. Our experiments involve two data generation schemes. The first scheme considers the study domains $\mathcal{D}_1 = [0, 1]$ with n observations on the grid $\chi_1 = \{i/(n-1) : 0 \leq i \leq n-1\}$ and $\mathcal{D}_2 = [0, 1]^2$ with $n = n_s^2$ observations on the grid $\chi_2 = \{(i/(n_s - 1), j/(n_s - 1)) : 0 \leq i, \leq n_s - 1, 0 \leq j \leq n_s - 1\}$. With this scheme, we generate a sequence of data sets with increasing sample size on increasingly finer grids on the study domains. The second scheme generates data on a "disturbed grid". On $\mathcal{D}_1 = [0, 1]$ with n observations,

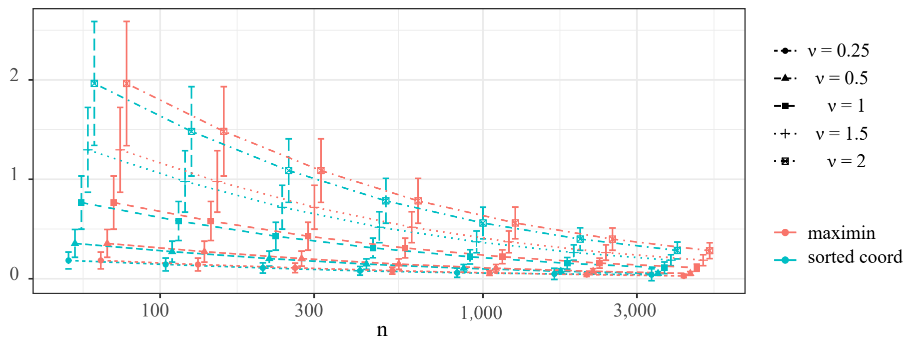
χ_1 comprises locations randomly sampled by $\mathrm{N}(i/(n+2), 0.15/(n+2))$, for $i = 1, \ldots, n$. On $\mathcal{D}_2 = [0,1]^2$ with $n = n_s^2$ observations, locations in χ_2 are generated by $\{N(i/(n_s+2), 0.15/(n_s+2)), N(i/(n_s+2), 0.15/(n_s+2))\}$, for $i, j = 1, \ldots, n_s$. With this scheme, we first generate simulations with the largest sample size, and then randomly select successively larger subsets from the same data set to examine the tendency of $c_n(\phi_1, \phi_0, k)$ with an increasing n. The first scheme matches the setup of our proofs in the preceding sections, and the second scheme serves as a more directly informative regime for simulation studies about asymptotics. In practice, estimations using Vecchia's approximation (1.4) are complicated by the fixed ordering of locations. Guinness (2018) provides excellent practical insights into this issue that can considerably improve finite-sample behavior in certain settings. In this study, we test two orderings of locations, namely, maximin ordering and sorted coordinate ordering. The sorted coordinate ordering first orders locations on χ_2 based on the second coordinate, and then breaks ties based on the associated first coordinate. We take $S_{(i)}[k]$ as the, at most, k nearest neighbors of y_{i+1}. In both studies on $\mathcal{D}_1$ and $\mathcal{D}_2$, we fix $\sigma^2 = 1.0$ and consider five smoothness values $\nu \in \{0.25, 0.5, 1.0, 1.5, 2\}$. We choose different decay parameters ϕ_0 for ν so that $K_\theta(h) = 0.05$ when $h = 0.2$ and 0.5 for the study on $\mathcal{D}_1$ and $\mathcal{D}_2$, respectively.

For each fixed value of $\theta = \{\sigma^2, \phi, \nu\}$, we generate 100 data sets with Y_n being the realization from $y(s) \sim GP(0, K_\theta(\cdot))$, and calculate $c_n(\phi_1, \phi_0, k)$, with k being the closest integer to $3\log(n)$, and $\phi_1 = 1.2\phi_0$ and $1.1\phi_0$ for $\mathcal{D}_1$ and $\mathcal{D}_2$, respectively. Then, we record the mean and standard deviation of the 100 values of $c_n(\phi_1, \phi_0, k)$. We repeat this process for different values of n ranging from $2^6 = 64$ to $2^{12} = 4096$ in the study on $\mathcal{D}_1$. The study on $\mathcal{D}_2$ follows the study on $\mathcal{D}_1$, with n_s ranging from 9 to 81. The code for this simulation study is available from `https://github.com/LuZhangstat/vecchia_consistency`. Figures 2a and 2b summarize the study results on $\mathcal{D}_1$ under the two data generation schemes. Each figure presents 10 graphs, one for each value of ν and each ordering, showing the mean and standard deviation of $c_n(\phi_1, \phi_0, k)$ for different values of n.

The value of $c_n(\phi_1, \phi_0, k)$, as shown in Figure 2, decreases rapidly as the sample size increases, supporting the main conclusion in Theorem 2. We do not observe a strong effect of the ordering and the data generation scheme on the results. The corresponding graphs for the study on $\mathcal{D}_2$ are presented in Figure 3. These graphs also reveal decreasing trends, but with more gentle slopes than those in Figure 2. The results under the second data-generation scheme are slightly better than those under the first scheme. When the smoothness ν is small, the standard deviation decreases faster with maximin ordering than it does with sorted coordinate ordering. At the same time, the standard deviation does not decrease significantly with an increase of n when ν is large. To explore further, we reproduce the study on $\mathcal{D}_2$, with k being the closest integer to $\sqrt{n}$, and show the results in Figure 4. We observe that, in all cases, the standard deviation



(a) Increasingly finer grids (Data generation scheme 1)



(b) Successively increasing data sets (Data generation scheme 2)

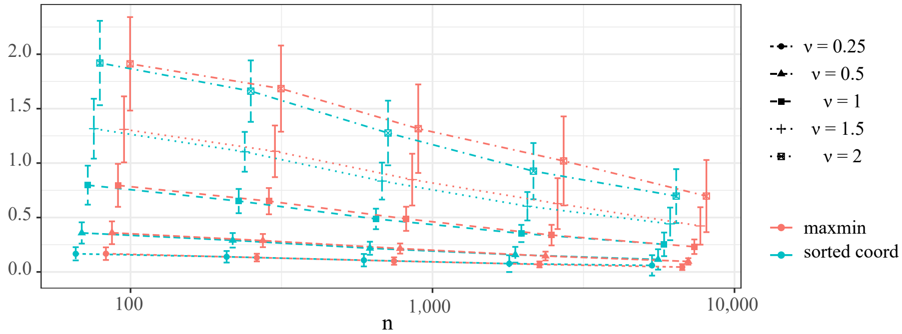
Figure 2. The mean of $c_n(\phi_1, \phi_0, k)$ of 100 simulations on $\mathcal{D}_1 = [0, 1]$. The error bars represent one standard deviation. The sample size n takes values in $64, 128, 256, 512, 1024, 2048$, and 4096. The graph describing the results obtained using maximin ordering is dodged to the right of the graph describing the results obtained using sorted coordinate ordering. The error bars in both graphs are connected using distinct line styles: solid lines for maximin ordering and dashed lines for sorted coordinate ordering.

decreases rapidly as n increases.

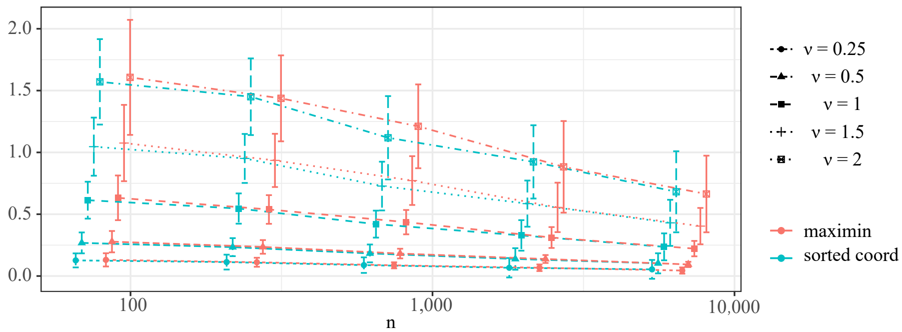
Recall that, from the proof in Theorem 1, $c_n(k, \phi_1, \phi_0) = \sqrt{n}(\hat{\sigma}^2_{n,vecch}/\sigma_1^2 - \hat{\sigma}^2_{0,n}/\sigma_0^2)$, where $\hat{\sigma}^2_{0,n} = \text{argmax}_{\sigma^2}\{p(y; \phi_0, \sigma^2),\ \sigma^2 \in \mathbb{R}^+\}$ is the maximum likelihood estimator from (1.3) when fixing $\phi_1 = \phi_0$. Hence, $c_n(k, \phi_1, \phi_0)$ also measures the discrepancy between $\hat{\sigma}^2_{n,vecch}/\sigma_1^2$ and $\hat{\sigma}^2_{0,n}/\sigma_0^2$, and the decreasing trend of $c_n(\phi_0, \phi_1, k)$ indicates that the inference based on Vecchia's approximation approaches that based on the full likelihood as the sample size increases. This phenomenon reveals that Vecchia's approximation is still efficient when the neighborhood size k is substantially smaller than the sample size.

5. Conclusion

We have developed insights into inferences based on GP likelihood approximations by Vecchia (1988) under fixed-domain asymptotics for geostatistical



(a) Increasingly finer grids (Data generation scheme 1)

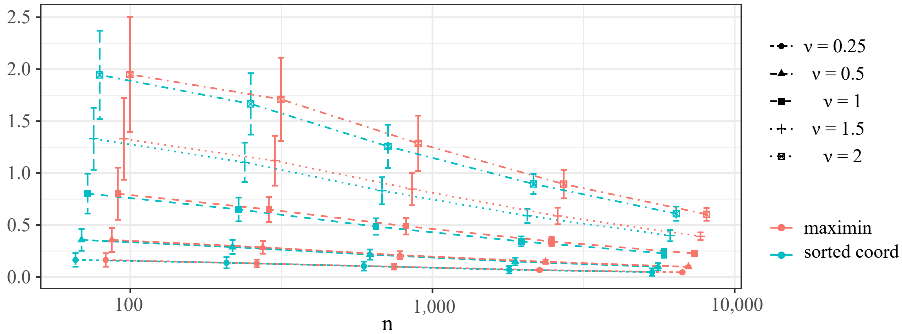


(b) Successively increasing data sets (Data generation scheme 2)

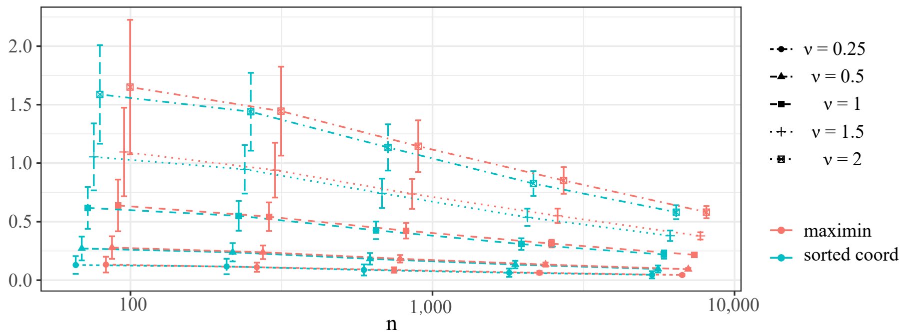
Figure 3. The mean of $c_n(\phi_1, \phi_0, k)$ of 100 simulations on $\mathcal{D}_2 = [0,1]^2$. The error bars represent one standard deviation. The sample size n takes values in $81, 256, 729, 2209$, and 6561. The graph describing the results obtained using maximin ordering is situated to the right of the graph describing the results obtained using sorted coordinate ordering. The error bars in both graphs are connected using distinct line styles: solid lines for maximin ordering and dashed lines for sorted coordinate ordering.

data analysis. We have formally established sufficient conditions for such approximations to have the same asymptotic efficiency as that of a a full GP likelihood when estimating parameters in a Matérn covariance function. The insights obtained here enhance our understanding of the identifiability of the process parameters, and can be used to develop priors for the microergodic parameters in Bayesian modeling. The results derived here also offer insights into formally establishing posterior consistency of process parameters for a number of Bayesian models that have emerged from (1.4) (Datta et al. (2016a,b); Katzfuss and Guinness (2021); Peruzzi, Banerjee and Finley (2022)).

We anticipate that our findings will motivate further research in variants of geostatistical models, such as asymptotic investigations of covariance-tapered models (see, e.g., Wang and Loh (2011)), and in adapting results, such as Theorems 2 and 3 in Kaufman and Shaby (2013), where ϕ is estimated, to the approximate likelihoods presented here. Another direction for future research can



(a) Increasingly finer grids (Data generation scheme 1)



(b) Successively increasing data sets (Data generation scheme 2)

Figure 4. The mean of $c_n(\phi_1, \phi_0, k)$ of 100 simulations on $\mathcal{D}_2 = [0,1]^2$. The error bars represent one standard deviation. The sample size n takes values in $81, 256, 729, 2209$, and 6561. The graph describing the results obtained using maximin ordering is dodged to the right of the graph describing the results obtained using sorted coordinate ordering. The error bars in both graphs are connected using distinct line styles: solid lines for maximin ordering and dashed lines for sorted coordinate ordering.

lead to formal developments about the inferential consistency of the "nugget" or the variance of the measurement error when the spatial process has a discontinuity at the origin arising from white noise (Tang, Zhang and Banerjee (2021)). Finally, there is scope to investigate fixed-domain inference for other likelihood approximations that extend or generalize (1.4) (see, e.g., Katzfuss and Guinness (2021); Peruzzi, Banerjee and Finley (2022)).

Acknowledgements

The work of the first and third author was supported, in part, by funding from the National Science Foundation grants NSF/DMS 1916349 and NSF/IIS 1562303, and by the National Institute of Environmental Health Sciences (NIEHS) under grants R01ES030210 and 5R01ES027027.

References

Abramowitz, M. and Stegun, A. (1965). *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables.* Dover.

Bachoc, F. and Lagnoux, A. (2020). Fixed-domain asymptotic properties of maximum composite likelihood estimators for Gaussian processes. *Journal of Statistical Planning and Inference* **209**, 62–75.

Banerjee, S. (2017). High-dimensional Bayesian geostatistics. *Bayesian Analysis* **12**, 583–614.

Baxter, G. (1962). An asymptotic result for the finite predictor. *Mathematica Scandinavica* **10**, 137–144.

Datta, A., Banerjee, S., Finley, A. O. and Gelfand, A. E. (2016a). Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. *Journal of the American Statistical Association* **111**, 800–812.

Datta, A., Banerjee, S., Finley, A. O., Hamm, N. A. S. and Schaap, M. (2016b). Non-separable dynamic nearest-neighbor Gaussian process models for large spatio-temporal data with an application to particulate matter analysis. *Annals of Applied Statistics* **10**, 1286–1316.

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

Dym, H. and McKean, H. P. (1970). Extrapolation and interpolation of stationary Gaussian processes. *The Annals of Mathematical Statistics* **41**, 1817–1844.

Dym, H. and McKean, H. P. (1976). *Gaussian Processes, Function Theory, and the Inverse Spectral Problem.* Academic Press, New York-London.

Eidsvik, J., Shaby, B. A., Reich, B. J., Wheeler, M. and Niemi, J. (2014). Estimation and prediction in spatial models with block composite likelihoods. *Journal of Computational and Graphical Statistics* **23**, 295–315.

Finley, A. O., Datta, A., Cook, B. C., Morton, D. C., Andersen, H. E. and Banerjee, S. (2019). Efficient algorithms for Bayesian nearest neighbor Gaussian processes. *Journal of Computational and Graphical Statistics* **28**, 401–414.

Ginovian, M. S. (1999). Asymptotic behavior of the prediction error for stationary random sequences. *Natsionalnaya Akademiya Nauk Armenii. Izvestiya. Matematika* **34**, 18–36.

Grenander, U. and Szegö, G. (1958). *Toeplitz Forms and Their Applications.* California Monographs in Mathematical Sciences. University of California Press, Berkeley-Los Angeles.

Guinness, J. (2018). Permutation and grouping methods for sharpening Gaussian process approximations. *Technometrics* **60**, 415–429.

Heaton, M. J., Datta, A., Finley, A. O., Furrer, R., Guinness, J., Guhaniyogi, R. et al. (2019). A case study competition among methods for analyzing large spatial data. *Journal of Agricultural, Biological and Environmental Statistics* **24**, 398–425.

Ibragimov, I. A. (1964). On the asymptotic behavior of the prediction error. *Theory of Probability & Its Applications* **9**, 627–634.

Ibragimov, I. A. and Rozanov, Y. A. (1978). *Gaussian Random Processes.* Springer-Verlag, New York-Berlin.

Katzfuss, M. and Guinness, J. (2021). A general framework for Vecchia approximations of Gaussian processes. *Statistical Science* **36**, 124–141.

Katzfuss, M., Guinness, J., Gong, W. and Zilber, D. (2020). Vecchia approximations of Gaussian-process predictions. *Journal of Agricultural, Biological and Environmental Statistics* **25**, 383–414.

Kaufman, C. G. and Shaby, B. A. (2013). The role of the range parameter for estimation and prediction in geostatistics. *Biometrika* **100**, 473–484.

Kolmogorov, A. (1941). Interpolation und Extrapolation von stationären zufälligen Folgen. *Izvestiya Akademii Nauk SSSR, Seriya Matematicheskaya* **5**, 3–14.

Matérn, B. (1986). *Spatial Variation*. Springer-Verlag.

Peruzzi, M., Banerjee, S. and Finley, A. O. (2022). Highly scalable Bayesian geostatistical modeling via meshed Gaussian processes on partitioned domains. *Journal of the American Statistical Association* **117**, 969–982.

Shiryaev, A. N. (1996). *Probability*. 2nd Edition. Springer-Verlag, New York.

Stein, M. L. (1990a). Bounds on the efficiency of linear predictions using an incorrect covariance function. *The Annals of Statistics* **18**, 1116–1138.

Stein, M. L. (1990b). Uniform asymptotic optimality of linear predictions of a random field using an incorrect second-order structure. *The Annals of Statistics* **18**, 850–872.

Stein, M. L. (1999a). *Interpolation of Spatial Data: Some Theory for Kriging*. Springer-Verlag.

Stein, M. L. (1999b). Predicting random fields with increasing dense observations. *The Annals of Applied Probability* **9**, 242–273.

Stein, M. L., Chi, Z. and Welty, L. J. (2004). Approximating likelihoods for large spatial data sets. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **66**, 275–296.

Sun, Y., Li, B. and Genton, M. G. (2012). Geostatistics for large datasets. In *Advances and Challenges in Space-Time Modelling of Natural Events*, 55–77. Springer.

Tang, W., Zhang, L. and Banerjee, S. (2021). On identifiability and consistency of the nugget in Gaussian spatial process models. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **83**, 1044–1070.

Vecchia, A. V. (1988). Estimation and model identification for continuous spatial processes. *Journal of the Royal Statistical Society. Series B (Methodological)* **50**, 297–312.

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.

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

Zhang, H. (2012). Asymptotics and computation for spatial statistics. In *Advances and Challenges in Space-Time Modelling of Natural Events*, 239–252. Springer.

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

Zhang, L., Datta, A. and Banerjee, S. (2019). Practical Bayesian modeling and inference for massive spatial datasets on modest computing environments. *Statistical Analysis and Data Mining: The ASA Data Science Journal* **12**, 197–209.

Lu Zhang

Department of Population and Public health Sciences, Division of Biostatistics, University of Southern California, Los Angeles, CA 90089, USA.

E-mail: lzhang63@usc.edu

Wenpin Tang

Department of Industrial Engineering and Operations Research, Columbia University, New York, NY 10027, USA.

E-mail: wt2319@columbia.edu

Sudipto Banerjee
Department of Biostatistics, University of California, Los Angeles, CA 90095, USA.
E-mail: sudipto@ucla.edu

(Received December 2021; accepted January 2023)