The Annals of Statistics
2022, Vol. 50, No. 6, 3334–3363
https://doi.org/10.1214/22-AOS2230

BAYESIAN FIXED-DOMAIN ASYMPTOTICS FOR COVARIANCE PARAMETERS IN A GAUSSIAN PROCESS MODEL

BY CHENG LI[a]

Department of Statistics and Data Science, National University of Singapore, [a]*stalic@nus.edu.sg*

Gaussian process models typically contain finite-dimensional parameters in the covariance function that need to be estimated from the data. We study the Bayesian fixed-domain asymptotics for the covariance parameters in a universal kriging model with an isotropic Matérn covariance function, which has many applications in spatial statistics. We show that when the dimension of domain is less than or equal to three, the joint posterior distribution of the microergodic parameter and the range parameter can be factored independently into the product of their marginal posteriors under fixed-domain asymptotics. The posterior of the microergodic parameter is asymptotically close in total variation distance to a normal distribution with shrinking variance, while the posterior distribution of the range parameter does not converge to any point mass distribution in general. Our theory allows an unbounded prior support for the range parameter and flexible designs of sampling points. We further study the asymptotic efficiency and convergence rates in posterior prediction for the Bayesian kriging predictor with covariance parameters randomly drawn from their posterior distribution. In the special case of one-dimensional Ornstein–Uhlenbeck process, we derive explicitly the limiting posterior of the range parameter and the posterior convergence rate for asymptotic efficiency in posterior prediction. We verify these asymptotic results in numerical experiments.

1. Introduction. Gaussian processes (GP) have been widely used in spatial statistics, computer experiments, machine learning and many other fields. In this paper, we consider the observation from the following spatial Gaussian process regression model, known as the *universal kriging model* (Chapter 3, Section 3.4, [17]):

$$Y(s_i) = \mathrm{m}(s_i)^\top \beta + X(s_i) \quad \text{for } i = 1, \dots, n. \tag{1}$$

In the model (1), $\mathcal{S}_n = \{s_1, \dots, s_n\}$ is a sequence of distinct sampling points in the fixed domain $\mathcal{S} = [0, T]^d$, and $0 < T < \infty$ is a known constant and the dimension $d \in \{1, 2, 3\}$. Such a dimension d is of primary interest in spatial statistics. Here, $\mathrm{m}(\cdot) = (\mathrm{m}_1(\cdot), \dots, \mathrm{m}_p(\cdot))^\top$ is a p-dimensional vector of linearly independent and known deterministic functions defined on $\mathcal{S}$, and $\beta \in \mathbb{R}^p$ is the regression coefficient vector. In applications, $\mathrm{m}_1, \dots, \mathrm{m}_p$ can include the constant function 1, and hence β can include an intercept term. In the model (1), $X(\cdot)$ is a mean-zero Gaussian stochastic process $X = \{X(s) : s \in \mathcal{S}\}$. We assume that the covariance function of X is the isotropic Matérn covariance function given by

$$\mathrm{Cov}\big(X(s), X(t)\big) = \sigma^2 K_{\alpha,\nu}(s - t) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \big(\alpha \|s - t\|\big)^\nu \mathcal{K}_\nu\big(\alpha \|s - t\|\big), \tag{2}$$

for any $s, t \in \mathcal{S}$, where $\nu > 0$ is the smoothness parameter, $\sigma^2 > 0$ is the variance (or partial sill) parameter, and $\alpha > 0$ is the inverse range (or length-scale) parameter, $\mathcal{K}_\nu(\cdot)$ is the modified Bessel function of the second kind [40] and $\|\cdot\|$ is the Euclidean norm. The Matérn

Received October 2021; revised May 2022.
MSC2020 subject classifications. 62F15, 62H11, 62E20.
Key words and phrases. Fixed-domain asymptotics, limiting posterior distribution, Matérn covariance function, asymptotic efficiency in posterior prediction.

covariance function is popular in applications of spatial statistics and computer experiments because the smoothness parameter ν provides flexibility in controlling the smoothness of sample paths [64]. The observed data from the model (1) are $Y_n = (Y(s_1), \ldots, Y(s_n))^\top$. Parameter estimation and prediction of $Y(\cdot)$ at a new spatial location (known as kriging) is based on Y_n. For simplicity, we call α the range parameter in the rest of the paper.

In Bayesian inference on GP models [20, 30], it is common practice to assign prior distributions on the regression coefficient β and the covariance parameters (σ^2, α), and the prediction of $Y(s^*)$ at a new location s^* is based on the posterior distribution of $(\beta, \sigma^2, \alpha)$. There is abundant literature in Bayesian spatial statistics on speeding up the costly GP posterior computation for spatial data sets with a large sample size n ([5, 19, 27, 31, 54], etc.) However, there is a clear lack of theoretical understanding of the asymptotic properties of the Bayesian posterior distributions of covariance parameters (σ^2, α). This theory is important because in Bayesian inference, instead of taken as fixed values, the covariance parameters (σ^2, α) are randomly drawn from their posterior using sampling algorithms such as Markov chain Monte Carlo (MCMC), which eventually affect the posterior prediction performance of the GP model.

To illustrate our motivation, we fit a Bayesian universal kriging model in (1) to the sea surface temperature (SST) data. The data set is obtained from National Oceanographic Data Centres (NODC) World Ocean Database (https://www.ncei.noaa.gov/products/world-ocean-database) and corresponds to sea surface temperature measured by remote sensing satellites on August 16, 2016. The data we test come from the Pacific Ocean between 45°–48° north latitudes and 150°–153° west longitudes. The original data set is high resolution on a $0.025° \times 0.025°$ fine grid. We choose subsets of size $\{400, 800, 1200, 1600, 2000\}$ on equispaced grids. For the regressors $\mathrm{m}(\cdot)$, we include all $p = 10$ monomials of the latitude and longitude up to degree 3, since on average SST is lower at higher latitudes. We set $\nu = 1/2$, and assign a flat prior $\pi(\beta) \propto 1$ on β, an inverse gamma prior with shape and rate parameters both equal to 2 on σ^2 and an independent Uniform$(0.01, 300)$ prior on α. The marginal posterior densities of α and $\theta = \sigma^2\alpha$ are shown in Figure 1 below. As the sample size increases, the marginal posterior density of the parameter $\theta = \sigma^2\alpha$ seems to contract faster with n than that of the range parameter α. Even with sample size $n = 2000$, the posterior of α still has a relatively large uncertainty. It is natural to ask the following questions: (i) Do the posteriors of (σ^2, α) (or (θ, α)) converge, and if so, at what rates? (ii) How does the posterior uncertainty in (σ^2, α) affect the posterior prediction of the response $Y(\cdot)$ at a new location?

We provide an answer to (i) by studying the limiting posterior distributions of the covariance parameters (σ^2, α) in the Matérn covariance function in (2), under the *fixed-domain asymptotics (or infill asymptotics)* framework [59, 64, 85]. We further answer (ii) and show that the randomness in (σ^2, α) in general does not affect the posterior prediction performance. To the best of our knowledge, this paper is the first theoretical work on the fixed-domain asymptotics for the Bayesian posterior distribution of the finite-dimensional parameters in Gaussian process covariance functions. In the following, we explain the reasons we adopt the fixed-domain asymptotics regime and the main technical challenges.

1.1. *Why fixed-domain asymptotics?* In the fixed-domain asymptotics regime, the domain $\mathcal{S}$ remains fixed and bounded regardless of the increasing sampling size n. This implies that as n goes to infinity, the sampling points $\mathcal{S}_n$ become increasingly dense in the domain $\mathcal{S}$, leading to increasingly stronger dependence between adjacent observations in Y_n. Besides the fixed-domain asymptotics regime, there are also increasing-domain asymptotics [48] and mixed-domain asymptotics [13], in which the domain is assumed to increase as n goes to infinity and, therefore, the minimum distance between two adjacent sampling points is either not decreasing or decreasing slowly with n.

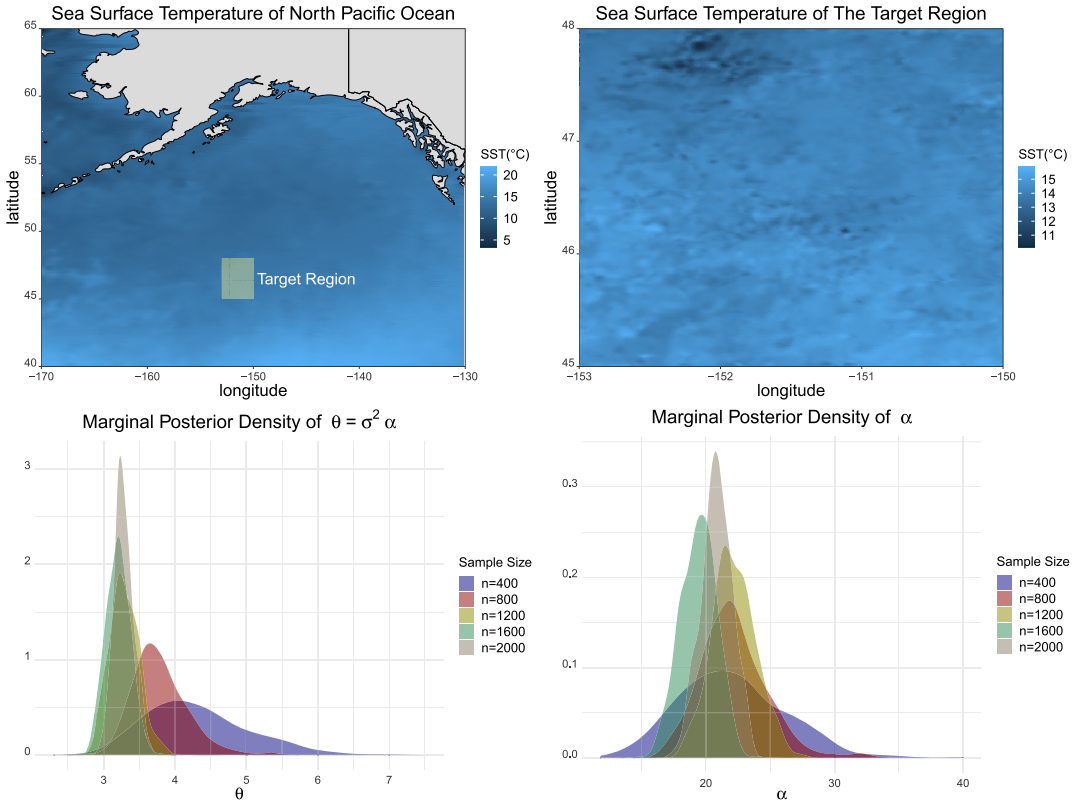

FIG. 1. *Example of the Sea Surface Temperature (SST) data. Top left: The SST data in North Pacific Ocean and the target region of our sampled data. Top right: The SST data in the target region. Bottom left: The marginal posterior densities of* $\theta = \sigma^2\alpha$ *for sample sizes* $n = 400, 800, 1200, 1600, 2000$. *Bottom right: The marginal posterior densities of* α *for sample sizes* $n = 400, 800, 1200, 1600, 2000$. *The posterior densities are based on* 2000 *MCMC draws.*

Compared to these alternatives, the fixed-domain setup has several advantages. First and foremost, a fixed domain matches up with the reality in many spatial applications. The advances in remote sensing technology make it possible to collected spatial data in larger volume and higher resolution in a given region [66]. The motivating example above of the SST data from NODC has millions of observations with high resolution on the $0.025° \times 0.025°$ fine grid (about $2 \sim 4$km range). Second, since the model (1) has a stationary Matérn covariance function, this stationarity assumption of GP is more likely to hold on a fixed domain rather than an expanding domain. Therefore, the fixed-domain asymptotics regime is more suitable for interpolation of spatial processes; see Section 3.3 of [64] for a cogent argument. Third, [86] has shown that the fixed-domain asymptotics has better parameter estimation performance than the increasing-domain asymptotics.

1.2. *What are the main difficulties in Bayesian fixed-domain asymptotics?* Theoretically, the increasingly stronger spatial dependence among the observed data Y_n in fixed-domain asymptotics leads to a lack of consistent estimation for the covariance parameters (σ^2, α) [85] and, therefore, poses significant challenges to theory development. When the dimension of sampling points $d = 1, 2, 3$, a well-known fixed-domain asymptotics result [85] says that it is only possible to consistently estimate the *microergodic parameter* $\theta = \sigma^2\alpha^{2\nu}$ in an isotropic Matérn covariance function, but not the individual variance parameter σ^2 and the range parameter α. The microergodic parameter is defined to be the parameter that uniquely

determines the Gaussian measure induced by a Gaussian process, such that different values of microergodic parameter will lead to mutually orthogonal Gaussian measures; see Section 6.2 of [64] for a detailed explanation on this definition. On the other hand, both the variance and range parameters (σ^2, α) can be consistently estimated if $d \geq 5$, with the case of $d = 4$ still open [1]. Nevertheless, the cases with $d = 1, 2, 3$ are of primary interest in spatial and spatiotemporal applications and will be our main focus.

The standard Bayesian asymptotic theory consists of results such as posterior consistency, posterior convergence rates, and the Bernstein–von Mises (BvM) theorem [23]. For parametric models, the BvM theorem typically relies on the local asymptotic normality (LAN) condition and the existence of uniformly consistent tests; see, for example, Chapter 10 in [71]. Since no consistent frequentist estimator exists for (σ^2, α) under fixed-domain asymptotics, one cannot expect to establish posterior consistency for (σ^2, α). Instead, we will consider the microergodic parameter $\theta = \sigma^2\alpha^{2\nu}$, which can be consistently estimated, and reparametrize the covariance function (2) by (θ, α). Crowder [18] is an early work on the asymptotic normality of maximum likelihood estimator (MLE) in the presence of dependent observations and nuisance parameters. We will establish the LAN condition for the microergodic parameter θ, *uniformly* over a wide range of values of the "nuisance" range parameter α. Such a uniform LAN condition based on data with increasingly stronger dependence is new in the literature and differs significantly from the LAN in classic parametric models with independent or weakly dependent data. The asymptotic normality for microergodic parameter θ is crucial and guarantees the posterior prediction performance of $Y(\cdot)$ at a new location.

For Bayesian inference on the GP covariance parameters, the only theoretical work we are aware of is [56], who have worked under the increasing-domain asymptotics regime and have established that the joint posterior of all parameters in the tapered covariance functions converges to a limiting normal distribution. This is similar to the classic BvM theorem since the dependence among data does not get stronger under increasing-domain asymptotics. A key assumption in [56] is that the observed covariance matrix have lower and upper bounded eigenvalues, which no longer holds under fixed-domain asymptotics.

We define some universal notation. Let $\mathbb{R}^+ = (0, +\infty)$. For two positive sequences a_n and b_n, we use $a_n \preceq b_n$ and $b_n \succeq a_n$ to denote the relation $\limsup_{n\to\infty} a_n/b_n < +\infty$, and $a_n \asymp b_n$ to denote the relation $a_n \preceq b_n$ and $a_n \succeq b_n$. For any integers k, m, we let I_k be the $k \times k$ identity matrix, 0_k and 1_k be the k-dimensional column vectors of all zeros and all ones, $0_{k\times m}$ be the $k \times m$ zero matrix. For any generic matrix A, cA denotes the matrix of A with all entries multiplied by the number c, and $|A|$ denotes the determinant of A. If A is positive semidefinite, then $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote the smallest and largest eigenvalues of A. Let $\mathcal{N}(\mu, \Sigma)$ be the normal distribution with mean μ and covariance matrix Σ. Sometimes to highlight the random variable $Z \sim \mathcal{N}(\mu, \Sigma)$, we also write the normal measure as $\mathcal{N}(\mathrm{d}z|\mu, \Sigma)$.

The remainder of the paper is organized as follows. In Section 2, we introduce the basic model setup and present the main theorems on limiting posterior distribution of covariance parameters under fixed-domain asymptotics. Section 3 presents the theory on asymptotic efficiency in posterior prediction. Section 4 presents some empirical results from simulation study to verify the main theory. Section 5 includes some discussion on further extensions. The technical proofs of all theorems, propositions, lemmas, corollaries and additional simulation results are in the Supplementary Material [42].

2. Limiting posterior distribution for covariance parameters.

2.1. *Bayesian model setup.* We consider the Bayesian estimation of $(\beta, \sigma^2, \alpha)$ in the model (1) based on the observed data Y_n. Throughout the paper, we assume that the domain

dimension satisfies $d \in \{1, 2, 3\}$, and that the smoothness parameter $\nu > 0$ is fixed and known. Estimation of the smoothness parameter ν is an important research topic with some recent developments in frequentist literature [45, 46]), but is beyond the scope of the current paper. We let the true parameter values in the Matérn covariance function that generates X be (σ_0^2, α_0) and let the true regression coefficient vector be β_0. We use the notation $X \sim \mathrm{GP}(0, \sigma_0^2 K_{\alpha_0,\nu})$, and hence $Y \sim \mathrm{GP}(\mathrm{m}^\top \beta_0, \sigma_0^2 K_{\alpha_0,\nu})$.

Let $Y_n = (Y(s_1), \ldots, Y(s_n))^\top$. Let M_n be the $n \times p$ matrix by stacking the row vectors $\mathrm{m}(s_i)^\top$ for $i = 1, \ldots, n$. Throughout the paper, we assume that M_n is a rank-p matrix without loss of generality, since all our results are asymptotic with $n \to \infty$. Let R_α be the implied $n \times n$ Matérn correlation matrix on $\mathcal{S}_n$ indexed by α, whose (i, j)th entry is $R_{\alpha,ij} = K_{\alpha,\nu}(s_i - s_j)$, for $i, j \in \{1, \ldots, n\}$. We omit the dependence of R_α on ν. The covariance matrix of X_n is then $\sigma^2 R_\alpha$. Therefore, the model (1) can be equivalently written as $Y_n = M_n\beta + X_n$. The log-likelihood function based on Y_n is

$$\mathcal{L}_n(\beta, \sigma^2, \alpha) = -\frac{n}{2}\log\sigma^2 - \frac{1}{2}\log|R_\alpha| - \frac{1}{2\sigma^2}(Y_n - M_n\beta)^\top R_\alpha^{-1}(Y_n - M_n\beta). \tag{3}$$

We study the Bayesian posterior distribution based on the log-likelihood (3). We follow the common practice in Bayesian spatial modeling literature [5, 19, 27, 31, 50, 54], etc.) and assign the conjugate normal prior on β, given by

$$\beta|\sigma^2, \alpha \sim \mathcal{N}(0_p, \sigma^2\Omega_\beta^{-1}), \tag{4}$$

which uses a rescaling with σ^2, and the prior precision matrix $\Omega_\beta \in \mathbb{R}^{p\times p}$ is assumed to be symmetric positive semidefinite. Here, we can set the prior mean to be 0_p without any loss of generality. This is because if the prior is $\beta|\sigma^2, \alpha \sim \mathcal{N}(\mu_\beta, \sigma^2\Omega_\beta^{-1})$ and the prior mean is $\mu_\beta \neq 0_p$, we can always define a new response variable $Y'(s) = Y(s) - \mathrm{m}(s)^\top\mu_\beta$, the new regression coefficient vector $\beta' = \beta - \mu_\beta$, and rewrite the original model (1) as $Y'(s) = \mathrm{m}(s)^\top\beta' + X(s)$ for $s \in \mathcal{S}$, where $Y'(s)$ is still fully observable on $\mathcal{S}_n$ given that $\mathrm{m}(\cdot)$ is observable and μ_β is known. Furthermore, we allow the precision matrix Ω_β to be arbitrarily small, leading to a prior of β with arbitrarily large variance. In particular, all our later theory covers the extreme case of improper noninformative prior $\pi(\beta|\sigma^2, \alpha) \propto 1$ [6, 26], which corresponds to $\Omega_\beta = 0_{p\times p}$. The joint posterior density of $(\beta, \sigma^2, \alpha)$ is then $\pi(\beta, \sigma^2, \alpha|Y_n) \propto \exp\{\mathcal{L}_n(\beta, \sigma^2, \alpha)\}\pi(\beta|\sigma^2, \alpha)\pi(\sigma^2, \alpha)$. Since $\pi(\beta|\sigma^2, \alpha)$ is the normal prior density, it is straightforward to obtain the conditional posterior of β:

$$\beta|\sigma^2, \alpha, Y_n \sim \mathcal{N}(\widetilde{\beta}_\alpha, \sigma^2(M_n^\top R_\alpha^{-1}M_n + \Omega_\beta)^{-1}), \tag{5}$$

where $\widetilde{\beta}_\alpha = (M_n^\top R_\alpha^{-1}M_n + \Omega_\beta)^{-1}M_n^\top R_\alpha^{-1}Y_n$ and the subscript is to highlight its dependence on α but not σ^2. We can further integrate out β and obtain the marginal posterior density of the covariance parameters (σ^2, α). We write $\pi(\sigma^2, \alpha|Y_n) \propto \exp\{\mathcal{L}_n(\sigma^2, \alpha)\}\pi(\sigma^2, \alpha)$, where the *restricted log-likelihood* $\mathcal{L}_n(\sigma^2, \alpha)$ is given by

$$\begin{aligned}\mathcal{L}_n(\sigma^2, \alpha) = &-\frac{1}{2\sigma^2}Y_n^\top\big[R_\alpha^{-1} - R_\alpha^{-1}M_n(M_n^\top R_\alpha^{-1}M_n + \Omega_\beta)^{-1}M_n^\top R_\alpha^{-1}\big]Y_n \\ &-\frac{n-p}{2}\log\sigma^2 - \frac{1}{2}\log|R_\alpha| - \frac{1}{2}\log|M_n^\top R_\alpha^{-1}M_n + \Omega_\beta|.\end{aligned} \tag{6}$$

In spatial statistical theory, it is well known [85] that the parameters (σ^2, α) cannot be consistently estimated under fixed-domain asymptotics. The main reason is that for two Gaussian measures $\mathrm{GP}(0, \sigma_j^2 K_{\alpha_j,\nu})$ $(j = 1, 2)$ on the space of sample paths on the domain $\mathcal{S} = [0, T]^d$ and $d \in \{1, 2, 3\}$, they are equivalent (or mutually absolutely continuous) as long as $\sigma_1^2\alpha_1^{2\nu} = \sigma_2^2\alpha_2^{2\nu}$, and they are orthogonal otherwise. As a result, one cannot tell from a

finite sample which parameter values (σ_j^2, α_j) $(j = 1, 2)$ are correct. Empirically, this phenomenon has been also observed [1, 22]. Despite the lack of consistent estimator for (σ^2, α), the microergodic parameter $\theta = \sigma^2\alpha^{2\nu}$ can still be consistently estimated [85]. For a fixed $\alpha > 0$, we maximize $\mathcal{L}_n(\sigma^2, \alpha)$ with respect to σ^2 (and so θ) to derive the *restricted maximum likelihood estimator (REML)*, given by

$$\widetilde{\sigma}_\alpha^2 = \frac{1}{n-p} Y_n^\top \big[R_\alpha^{-1} - R_\alpha^{-1} M_n \big(M_n^\top R_\alpha^{-1} M_n + \Omega_\beta\big)^{-1} M_n^\top R_\alpha^{-1}\big] Y_n, \qquad \widetilde{\theta}_\alpha = \alpha^{2\nu}\widetilde{\sigma}_\alpha^2. \tag{7}$$

In (7), we have slightly extended the meaning of REML such that we can account for general prior precision matrix Ω_β, including the special case of $\Omega_\beta = 0_{p\times p}$ where β can be viewed as normal random effects of $\mathrm{m}(\cdot)$, such that $\widetilde{\sigma}_\alpha^2$ (and $\widetilde{\theta}_\alpha$) can be viewed as the conventional REML of σ^2 (and θ) in random effects models. We can plug in $\widetilde{\theta}_\alpha$ in (3) to obtain the *profile restricted log-likelihood of* α (up to an additive constant), which plays an important role in our theory:

$$\begin{aligned}\widetilde{\mathcal{L}}_n(\alpha) &\equiv \mathcal{L}_n\big(\alpha^{-2\nu}\widetilde{\theta}_\alpha, \alpha\big) \\ &= -\frac{n-p}{2}\log\left\{\frac{1}{n-p} Y_n^\top \big[R_\alpha^{-1} - R_\alpha^{-1} M_n \big(M_n^\top R_\alpha^{-1} M_n + \Omega_\beta\big)^{-1} M_n^\top R_\alpha^{-1}\big] Y_n\right\} \\ &\quad - \frac{1}{2}\log|R_\alpha| - \frac{1}{2}\log\big|M_n^\top R_\alpha^{-1} M_n + \Omega_\beta\big| - \frac{n-p}{2}.\end{aligned} \tag{8}$$

The frequentist asymptotic normality for the MLE of θ has been studied for the model (1) without the regression term, that is, $Y(\cdot) \equiv X(\cdot) \sim \mathrm{GP}(0, \sigma^2 K_{\alpha,\nu})$. For this simplified model, [83] first studied the special case of $d = 1$ and $\nu = 1/2$, followed by [21, 76, 85] and [37] for a general $\nu > 0$. If $\alpha \in [\alpha_1, \alpha_2]$ for some constants $0 < \alpha_1 < \alpha_2 < \infty$, the MLE of θ, denoted by $\widehat{\theta}$, satisfies that $\sqrt{n}(\widehat{\theta} - \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 2\theta_0^2)$ as $n \to \infty$ under fixed-domain asymptotics, where $\theta_0 = \sigma_0^2\alpha_0^{2\nu}$ is the true value, and $\xrightarrow{\mathcal{D}}$ is the convergence in distribution.

We study the fixed-domain asymptotic limit for the Bayesian posterior distribution of (σ^2, α) based on the log likelihood (6). We reparametrize the model using (θ, α), with $\theta = \sigma^2\alpha^{2\nu}$ being the microergodic parameter. This reparametrization has been suggested in [64] (p. 175) and also used in recent Bayesian GP works such as [22]. For the consistency of notation, we will still maintain the parametrization of (σ^2, α) for the log-likelihood functions and quantities related to the probability distributions, such as $P_{(\beta,\sigma^2,\alpha)}$ for the probability distribution of $\mathrm{GP}(m^\top\beta, \sigma^2 K_{\alpha,\nu})$. The change of variable from σ^2 to $\theta = \sigma^2\alpha^{2\nu}$ is often clear from the context. We assign prior distributions on (θ, α) and write the joint prior density as $\pi(\theta, \alpha) = \pi(\theta|\alpha)\pi(\alpha)$. The joint posterior density of (θ, α) is given by

$$\pi(\theta, \alpha|Y_n) = \frac{\exp\{\mathcal{L}_n(\theta/\alpha^{2\nu}, \alpha)\}\pi(\theta|\alpha)\pi(\alpha)}{\int_0^\infty\int_0^\infty \exp\{\mathcal{L}_n(\theta'/\alpha'^{2\nu}, \alpha')\}\pi(\theta'|\alpha')\pi(\alpha')\,\mathrm{d}\alpha'\,\mathrm{d}\theta'}. \tag{9}$$

We will use $\Pi(\mathrm{d}\theta, \mathrm{d}\alpha|Y_n)$ to denote the posterior probability measure with the density in (9).

2.2. *Main results.* We first present the limiting posterior distribution of θ conditional on a fixed $\alpha > 0$. Let $L_2(\mathcal{S})$ be the space of square integrable functions on $\mathcal{S}$ and $\|f\|_2$ be the $L_2(\mathcal{S})$ norm of f for any $f \in L_2(\mathcal{S})$. Let $\mathsf{j} = (j_1, \ldots, j_d)$ with $j_1, \ldots, j_d \in \mathbb{N}$, $|j| = \sum_{i=1}^d j_i$, and D^{j} be the partial differentiation operator of order j. For $k > 0$, define the Sobolev space $\mathcal{W}_2^k(\mathcal{S}) = \{f \in L_2(\mathcal{S}) : \|f\|_{\mathcal{W}_2^k(\mathcal{S})}^2 = \sum_{\mathsf{j}\in\mathbb{N}^d:|\mathsf{j}|\le k} \|\mathsf{D}^{\mathsf{j}} f\|_2^2 < \infty\}$. We make the following assumptions.

(A.1) $\mathrm{m}_j \in \mathcal{W}_2^{\nu+d/2}(\mathcal{S})$ for each $j = 1, \ldots, p$. M_n is a rank-p matrix for all $\mathcal{S}_n$ with $n \ge p$.

(A.2) The prior of β given (σ^2, α) is $\mathcal{N}(0_p, \sigma^2\Omega_\beta^{-1})$ for a symmetric positive semidefinite matrix Ω_β. The conditional prior density of θ given α, $\pi(\theta|\alpha)$, is a proper prior density that is continuously differentiable in θ, continuous in α, and finite everywhere for all $\theta \in \mathbb{R}^+$ and $\alpha \in \mathbb{R}^+$. $\pi(\theta|\alpha)$ does not depend on n. $\pi(\theta_0|\alpha) > 0$ for all $\alpha > 0$.

Assumption (A.1) is the regularity assumption on the regression functions $\mathrm{m}_1, \ldots, \mathrm{m}_p$. By Theorem 10.35 of [78], $\mathcal{W}_2^{\nu+d/2}(\mathcal{S})$ is norm equivalent to the reproducing kernel Hilbert space (RKHS) associated with the Matérn kernel $\sigma_0^2 K_{\alpha_0,\nu}$. As a result, Assumption (A.1) implies that $\mathrm{m}_1(\cdot), \ldots, \mathrm{m}_p(\cdot)$ are smoother functions than the sample paths from $\mathrm{GP}(0, \sigma^2 K_{\alpha,\nu})$ for any $(\sigma^2, \alpha) \in \mathbb{R}^+ \times \mathbb{R}^+$; see, for example, Corollary 4.15 of [35]. Such a smoothness assumption is necessary. Otherwise, if $\mathrm{m}_1, \ldots, \mathrm{m}_p$ are rougher functions than the sample path of X, their roughness will overwhelm the information contained in the smoother $\mathrm{GP}(0, \sigma_0^2 K_{\alpha_0,\nu})$, and one cannot expect to estimate any covariance parameter consistently, including θ. As argued in page 12 of [64], $\mathrm{m}_1(\cdot), \ldots, \mathrm{m}_p(\cdot)$ in applications are often highly regular functions such as monomials, which are infinitely differentiable on $\mathcal{S}$ and, therefore, satisfy Assumption (A.1). Assumption (A.2) on $\pi(\theta|\alpha)$ is mild and satisfied in most applications.

For two probability measures P_1, P_2, let $\|P_1(\cdot) - P_2(\cdot)\|_{\mathrm{TV}} = \sup_{\mathcal{A}} |P_1(\mathcal{A}) - P_2(\mathcal{A})|$, where the supremum is taken over all measurable sets $\mathcal{A}$.

THEOREM 2.1 (Limiting Distribution for Conditional Posterior). *Suppose that $\alpha > 0$ is fixed and does not depend on n. Under Assumptions (A.1) and (A.2), the REML $\widetilde{\theta}_\alpha$ defined in (7) is asymptotically normal, with $\sqrt{n}(\widetilde{\theta}_\alpha - \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 2\theta_0^2)$ as $n \to \infty$. Furthermore, the conditional posterior distribution of θ given $\alpha > 0$ satisfies that*

$$\|\Pi(\mathrm{d}\theta|Y_n, \alpha) - \mathcal{N}(\mathrm{d}\theta|\widetilde{\theta}_\alpha, 2\theta_0^2/n)\|_{\mathrm{TV}} \preceq n^{-1/2}\log^3 n \to 0, \tag{10}$$

as $n \to \infty$ almost surely $P_{(\beta_0,\sigma_0^2,\alpha_0)}$, where $\widetilde{\theta}_\alpha$ is given in (7), and $\Pi(\cdot|Y_n, \alpha)$ is the conditional posterior probability measure of θ given a fixed $\alpha > 0$ with the density

$$\pi(\theta|Y_n, \alpha) = \frac{\exp\{\mathcal{L}_n(\theta/\alpha^{2\nu}, \alpha)\}\pi(\theta|\alpha)}{\int_0^\infty \exp\{\mathcal{L}_n(\theta'/\alpha^{2\nu}, \alpha)\}\pi(\theta'|\alpha)\,\mathrm{d}\theta'}. \tag{11}$$

Theorem 2.1 shows that under fixed-domain asymptotics, the REML $\widetilde{\theta}_\alpha$ is asymptotically normal, and the conditional posterior $\pi(\theta|Y_n, \alpha)$ is asymptotically close the normal distribution $\mathcal{N}(\widetilde{\theta}_\alpha, 2\theta_0^2/n)$ in total variation distance. Some comments are in order.

First, to the best of our knowledge, Theorem 2.1 is the first in the literature to establish both frequentist and Bayeisan asymptotic normality for the microergodic parameter θ for any $\nu > 0$ and $d \in \{1, 2, 3\}$ in the universal kriging model (1) with regression terms $\mathrm{m}(\cdot)^\top\beta$. Most of the existing frequentist fixed-domain asymptotic theory has considered either only the GP model with mean zero and no regression terms [1, 4, 21, 37, 76, 85], or only for some particular values of ν (such as $\nu = 1/2$ in [3, 12, 14, 74, 83, 84] and $\nu = 3/2$ in [44]). Theorem 3 of [83] has shown the asymptotic normality for the MLE of θ in the GP model with regression terms, but only for the special case of $\nu = 1/2$ and $d = 1$, and their proof techniques cannot be generalized to any $\nu > 0$ and $d > 1$. Our proof is based on the general RKHS theory and spectral analysis of isotropic Matérn covariance functions; see Section S1 of the Supplementary Material [42]. Theorem 3 of [83] almost needs that $\mathrm{m}_1, \ldots, \mathrm{m}_p \in \mathcal{W}_2^1([0, 1])$ in the special case of $\nu = 1/2$ and $d = 1$, that is, they are bounded functions with square integrable derivatives (following the comments after their Theorem 3), which is exactly the same as the space $\mathcal{W}_2^{\nu+d/2}(\mathcal{S})$ assumed in our Assumption (A.1).

Second, if the model (1) does not have regression terms $\mathrm{m}(\cdot)^\top\beta$, that is, if $p=0$ and we observe $Y_n = X_n$ directly from $\mathrm{GP}(0, \sigma_0^2 K_{\alpha_0,\nu})$, then the REML $\widetilde{\theta}_\alpha$ in (7) coincides with the MLE of θ, and $2\theta_0^2$ is also the asymptotic variance of this MLE [37, 76].

Third, the posterior convergence of (10) Theorem 2.1 has a similar format to the classic BvM theorem in regular parametric models for independent data, such as Theorem 8.2 in [41] and Theorem 10.1 in [71], where the limiting normal distribution is centered at the MLE with variance equal to the asymptotic variance of MLE. However, the classic BvM theorem usually relies on the LAN condition and the existence of uniformly consistent tests (Theorem 10.1 in [71]), which can be readily verified for models with independent and weakly dependent data. The main technical challenge for proving Theorem 2.1 is to establish the LAN condition for data with increasingly stronger dependence under fixed-domain asymptotics. We need the asymptotic normality of the REML $\widetilde{\theta}_\alpha$ at a given range parameter $\alpha > 0$, which can be different from the true α_0. Our proof leverages the spectral analysis of Matérn covariance functions (see Section S1.4 in the Supplementary Material [42], which has also been used in the previous works for the MLE of θ for GP with mean zero ([21, 76] and [37]), though they have not considered the model with regression terms as ours. Finally, we provide an explicit convergence rate $n^{-1/2}\log^3 n$ for the convergence in total variation distance. The $\log^3 n$ term is mainly used to ensure the strong mode of almost sure convergence.

In most spatial applications, the range parameter α is unknown and assigned a prior $\pi(\alpha)$. Next, we present a much stronger theorem for the limit of the joint posterior distribution of $(\theta, \alpha) \in \mathbb{R}^+ \times \mathbb{R}^+$. The consistency of the REML of θ and the nonexistence of consistent frequentist estimator for α indicates that the posterior of θ should converge to a normal limit, while the posterior of α does not necessarily converge to any fixed value under fixed-domain asymptotics. We prove this idea rigorously.

We define two small positive constants $\underline{\kappa}$ and $\overline{\kappa}$ that depend on the smoothness $\nu > 0$ and the dimension d ($d \in \{1, 2, 3\}$), together with two deterministic sequences $\underline{\alpha}_n$ and $\overline{\alpha}_n$:

$$\begin{aligned} \underline{\kappa} &= \frac{1}{2}\min\left\{\frac{0.9}{(2d+0.94)(8\nu+3d-0.9)}, \frac{1}{4(3\nu+d)}, 0.01\right\}, \qquad \underline{\alpha}_n = n^{-\underline{\kappa}}, \\ \overline{\kappa} &= \frac{1}{2}\min\left\{\frac{0.9}{(2d+0.94)(8\nu+5d+0.9)}, \frac{1}{2(2\nu+d)}, 0.01\right\}, \qquad \overline{\alpha}_n = n^{\overline{\kappa}}. \end{aligned} \tag{12}$$

The choices of $\overline{\kappa}$ and $\underline{\kappa}$ in (12) are not unique and can be replaced by other sufficiently small positive numbers; see Lemma S.20 in the Supplementary Material [42]. By definition, $\underline{\alpha}_n \to 0$ and $\overline{\alpha}_n \to +\infty$ as $n \to \infty$, and both are in slow polynomial rates. A key result below is that uniformly for all α in the slowly expanding interval $[\underline{\alpha}_n, \overline{\alpha}_n]$, the difference between $\widetilde{\theta}_\alpha$ and $\widetilde{\theta}_{\alpha_0}$ converges to zero at a faster rate than $n^{-1/2}$.

LEMMA 2.2 (Monotonicity and Uniform Convergence of $\widetilde{\theta}_\alpha$). *Suppose that Assumption* (*A*.1) *holds. Then for the REML* $\widetilde{\theta}_\alpha$ *defined in* (7):

(i) $\widetilde{\theta}_\alpha$ *is a nondecreasing function of* α *for all* $\alpha \in \mathbb{R}^+$;

(ii) *There exists a large integer* N_1 *and a positive constant* $\tau \in (0, 1/2)$ *that only depend on* $\nu, d, T, \beta_0, \theta_0, \alpha_0$ *and the* $\mathcal{W}_2^{\nu+d/2}(\mathcal{S})$ *norms of* $\mathrm{m}_1(\cdot), \ldots, \mathrm{m}_p(\cdot)$, *such that for all* $n > N_1$,

$$\Pr\Big(\sup_{\alpha\in[\underline{\alpha}_n,\overline{\alpha}_n]} \sqrt{n}|\widetilde{\theta}_\alpha - \widetilde{\theta}_{\alpha_0}| \le \theta_0 n^{-\tau}\Big) \ge 1 - \exp(-2\log^2 n),$$

where $\Pr(\cdot)$ *denotes the probability under the true probability measure* $P_{(\beta_0,\sigma_0^2,\alpha_0)}$.

Lemma 2.2 involves a new discovery in part (i) that the REML $\widetilde{\theta}_\alpha$ is monotone in α for the universal kriging model (1). The monotonicity of $\widetilde{\theta}_\alpha$ for the universal kriging model (1)

has significantly extended the previous work of [37], which only considered the MLE of θ for GP with mean zero. Previously [76] has shown that $|\widetilde{\theta}_\alpha - \widetilde{\theta}_{\alpha_0}|$ can be small, but only for a *fixed and known* value of range parameter α and only for GP with mean zero. In part (ii) of Lemma 2.2, we make a novel utilization of the monotonicity of $\widetilde{\theta}_\alpha$ in α, and prove in Lemma 2.2 that the difference $|\widetilde{\theta}_\alpha - \widetilde{\theta}_{\alpha_0}|$ can be uniformly small over an expanding interval $[\underline{\alpha}_n, \overline{\alpha}_n]$ for the more general model (1) with regression terms. Even though the REML $\widetilde{\theta}_\alpha$ defined in (7) is in fact a stochastic process indexed by α, our techniques using the monotonicity property of $\widetilde{\theta}_\alpha$ have the advantage of completely circumventing any empirical process argument. Our proof of part (ii) also develops a much strengthened concentration inequality for $\widetilde{\theta}_\alpha$ using more detailed spectral analysis of Matérn covariance functions than [76].

The nondecreasing property of $\widetilde{\theta}_\alpha$ in (7) is crucial for both establishing the uniform convergence of $\widetilde{\theta}_\alpha$ on the interval $[\underline{\alpha}_n, \overline{\alpha}_n]$ and understanding the asymptotic behavior of the joint posterior $\pi(\theta, \alpha|Y_n)$. Based on the uniform convergence in Lemma 2.2, a heuristic argument to extend the limiting conditional posterior in Theorem 2.1 to the joint posterior $\pi(\theta, \alpha|Y_n)$ is as follows: For each $\alpha \in [\underline{\alpha}_n, \overline{\alpha}_n]$, the conditional posterior $\pi(\theta|Y_n, \alpha)$ can be approximated by the normal distribution $\mathcal{N}(\widetilde{\theta}_\alpha, 2\theta_0^2/n)$. Since the center $\widetilde{\theta}_\alpha$ only differs from $\widetilde{\theta}_{\alpha_0}$ by a higher order term $O(n^{-1/2-\tau})$, this normal distribution can be further approximated by $\mathcal{N}(\widetilde{\theta}_{\alpha_0}, 2\theta_0^2/n)$, whose mean parameter only depends on the data Y_n but not α. Hence, the limiting distribution of θ is approximately independent of α.

To solidify this idea, we need additional prior conditions to ensure that the posterior probabilities outside the interval $[\underline{\alpha}_n, \overline{\alpha}_n]$ can be made small, such that the convergence to the normal limit inside $[\underline{\alpha}_n, \overline{\alpha}_n]$ is dominant in driving the asymptotics of the joint posterior $\pi(\theta, \alpha|Y_n)$. We specify the following general assumptions on the prior densities $\pi(\theta|\alpha)$ and $\pi(\alpha)$.

(A.3) There exist positive constants $C_{\pi,1}$, $C_{\pi,2}$, and $C_{\pi,3}$ that can depend on ν, d, T, α_0, θ_0, such that $0 < C_{\pi,1} + C_{\pi,2} < 1/2$, $0 < C_{\pi,3} < 1$ and for $\underline{\alpha}_n$ and $\overline{\alpha}_n$ defined in (12), for all sufficiently large n,

$$\sup_{\alpha\in[\underline{\alpha}_n,\overline{\alpha}_n]} \sup_{\theta\in(\theta_0/2,2\theta_0)} \left| \frac{\partial \log \pi(\theta|\alpha)}{\partial \theta} \right| \le n^{C_{\pi,1}}, \tag{13}$$

$$\sup_{\alpha\in[\underline{\alpha}_n,\overline{\alpha}_n]} \sup_{\theta\in(\theta_0/2,2\theta_0)} \frac{\pi(\theta|\alpha)}{\pi(\theta_0|\alpha)} \le n^{C_{\pi,2}}, \tag{14}$$

$$\inf_{\alpha\in[\underline{\alpha}_n,\overline{\alpha}_n]} \log \pi(\theta_0|\alpha) \ge -n^{C_{\pi,3}}. \tag{15}$$

(A.4) The marginal prior $\pi(\alpha)$ is a proper and continuous density function on $\mathbb{R}^+$. $\pi(\alpha)$ does not depend on n. $\pi(\alpha_0) > 0$. $\int_0^\infty \pi(\theta_0|\alpha)\pi(\alpha)\,d\alpha < \infty$. There exist positive constants $\underline{c_\pi} < (\nu + d/2)\underline{\kappa}$ and $\overline{c_\pi} < (\nu + d/2)\overline{\kappa}$ for $\underline{\kappa}$ and $\overline{\kappa}$ defined in (12), such that for $\underline{\alpha}_n$ and $\overline{\alpha}_n$ defined in (12), and for all sufficiently large n,

$$\begin{aligned} &\max\left\{ \int_0^{\underline{\alpha}_n} \alpha^{-n(\nu+d/2)} \pi(\alpha)\,d\alpha, \int_0^{\underline{\alpha}_n} \alpha^{-n(\nu+d/2)} \pi(\theta_0|\alpha)\pi(\alpha)\,d\alpha \right\} \\ &\quad \le \exp(\underline{c_\pi} n \log n), \end{aligned} \tag{16}$$

$$\begin{aligned} &\max\left\{ \int_{\overline{\alpha}_n}^\infty \alpha^{n(\nu+d/2)} \pi(\alpha)\,d\alpha, \int_{\overline{\alpha}_n}^\infty \alpha^{n(\nu+d/2)} \pi(\theta_0|\alpha)\pi(\alpha)\,d\alpha \right\} \\ &\quad \le \exp(\overline{c_\pi} n \log n). \end{aligned} \tag{17}$$

We will discuss these two assumptions in greater detail after presenting our main theorem for the joint posterior of (θ, α).

THEOREM 2.3 (Limiting Distributions for Joint and Marginal Posteriors). *Under Assumptions* (*A*.1), (*A*.2), (*A*.3) *and* (*A*.4), *the posterior distributions of* θ *and* α *are asymptotically independent*, *in the sense that the joint posterior distribution of* (θ, α) *satisfies*

$$\left\| \Pi(\mathrm{d}\theta, \mathrm{d}\alpha | Y_n) - \mathcal{N}\big(\mathrm{d}\theta | \widetilde{\theta}_{\alpha_0}, 2\theta_0^2/n\big) \times \widetilde{\Pi}(\mathrm{d}\alpha | Y_n) \right\|_{\mathrm{TV}} \to 0, \tag{18}$$

as $n \to \infty$ *almost surely* $P_{(\beta_0, \sigma_0^2, \alpha_0)}$, *where* $\widetilde{\Pi}(\mathrm{d}\alpha | Y_n)$ *is the profile posterior distribution with density* $\widetilde{\pi}(\alpha | Y_n)$ *given by*

$$\widetilde{\pi}(\alpha | Y_n) = \frac{\exp\{\widetilde{\mathcal{L}}_n(\alpha)\}\pi(\alpha | \theta_0)}{\int_0^\infty \exp\{\widetilde{\mathcal{L}}_n(\alpha')\}\pi(\alpha' | \theta_0)\, \mathrm{d}\alpha'}, \tag{19}$$

where the profile restricted log-likelihood $\widetilde{\mathcal{L}}_n(\alpha)$ *is given in* (8) *and* $\pi(\alpha | \theta_0)$ *is the conditional prior density of* α *given* $\theta = \theta_0$. *Furthermore*, *this profile posterior density* $\widetilde{\pi}(\alpha | Y_n)$ *is well defined for any given* $n \geq p$ *almost surely* $P_{(\beta_0, \sigma_0^2, \alpha_0)}$. *As a result*, *the total variation distance between* $\Pi(\mathrm{d}\theta | Y_n)$ *and* $\mathcal{N}(\mathrm{d}\theta | \widetilde{\theta}_{\alpha_0}, 2\theta_0^2/n)$ *converges to zero*, *and the total variation distance between* $\Pi(\mathrm{d}\alpha | Y_n)$ *and* $\widetilde{\Pi}(\mathrm{d}\alpha | Y_n)$ *converges to zero*, *as* $n \to \infty$ *almost surely* $P_{(\beta_0, \sigma_0^2, \alpha_0)}$.

Theorem 2.3 provides a clear description of the limiting behavior of the joint posterior of (θ, α) in the universal kriging model (1). Under fixed-domain asymptotics, the microergodic parameter θ and the range parameter α have *asymptotically independent* posterior distributions. The posterior of θ is centered at the REML $\widetilde{\theta}_{\alpha_0}$ and the variance is the same $2\theta_0^2/n$ as the asymptotic variance of REML $\widetilde{\theta}_{\alpha_0}$ in Theorem 2.1. In fact, according to part (ii) of Lemma 2.2, the center $\widetilde{\theta}_{\alpha_0}$ can be replaced by $\widetilde{\theta}_{\alpha_1}$ for any fixed $\alpha_1 > 0$, since α_1 will be eventually covered by the slowly expanding interval $[\underline{\alpha}_n, \overline{\alpha}_n]$, and the difference between $\widetilde{\theta}_{\alpha_0}$ and $\widetilde{\theta}_{\alpha_1}$ is negligible compared to the limiting normal standard deviation $\sqrt{2\theta_0^2/n}$.

The posterior convergence of microergodic parameter θ with a varying range parameter α shows that we can consistently estimate the equivalent class of Gaussian measures using the Bayesian procedure even if the range parameter α has possibly large posterior uncertainty. An important consequence is that based on a random draw of parameters (θ, α) from the posterior, the predictive variance at a new location is asymptotically close to the predictive variance based on the true parameters (θ_0, α_0). We will elaborate this in Section 3.

Theorem 2.3 has three advantages in its generality. First, the theorem works for the universal kriging model with regression terms $\mathrm{m}(\cdot)^\top \beta$. Second, it allows an unbounded prior support for α, which is not available in previous frequentist fixed-domain asymptotics literature. Third, the theorem does not require any assumption on the design points $\mathcal{S}_n$. In other words, the asymptotic factorization and normality works for *arbitrary* design of the sampling points $\mathcal{S}_n$, not even requiring $\mathcal{S}_n$ to be dense in the domain $\mathcal{S}$. Theorem 2.3 also shows that the marginal posterior density of α can be approximated by the more abstract profile posterior with density $\widetilde{\pi}(\alpha | Y_n)$, which is based on the profile restricted likelihood of α. Using the result in [26], we can show that this profile posterior is always well defined. On the other hand, without further assumptions on $\mathcal{S}_n$, it is not likely that the form of the profile posterior density $\widetilde{\pi}(\alpha | Y_n)$ can be simplified. In general, this profile posterior of α does not necessarily converge to any point mass. In Theorem 2.6 below, for a special case of 1-dimensional Ornstein–Uhlenbeck process (Matérn with $\nu = 1/2$) observed on an equispaced grid without regression terms, we approximate $\widetilde{\pi}(\alpha | Y_n)$ using an explicit density of α that asymptotically does not contract to any fixed value with high probability. Such nonconverging property of $\pi(\alpha | Y_n)$ explains the seemingly slow convergence of posterior of α in our SST data example in Section 1. We also demonstrate this phenomenon using simulation examples in Section 4.

The difficulty in the estimation of range parameter α is a well-known problem in the GP literature [38]. Gaussian processes with different values of α but the same microergodic parameter θ in the Matérn covariance function (2) can have similar sample paths [22], making

it difficult to infer an appropriate value for α from the data. Zhang [85] and many others have observed that for a fixed value of $\theta > 0$, $\mathcal{L}_n(\theta/\alpha^{2\nu}, \alpha)$ has a long right tail in $1/\alpha$ that creates problem for finding the MLE of α. The sampling distribution of the MLE of α does not show any sign of convergence as $n \to \infty$. For Bayesian inference, [26] identifies prior conditions using the objective priors in [6] for robust estimation of $1/\alpha$ in finite samples. Though we do not study point estimation of α, our technical proofs have derived some new properties for the profile posterior $\widetilde{\pi}(\alpha|Y_n)$, which could be of independent interest for Matérn covariance functions; see Section S2 of the Supplementary Material [42] for details.

Theorem 2.3 works for the domain dimension $d \in \{1, 2, 3\}$. For completeness, we also derive a similar theorem for the limiting joint posterior distribution when $d \geq 5$ under additional assumptions; see Section S3.4 of the Supplementary Material [42].

2.3. *On the prior assumptions*. We discuss the two technical prior assumptions (A.3) and (A.4). The inequalities (13) and (14) in (A.3) require that the conditional prior $\pi(\theta|\alpha)$ does not vary too dramatically in a neighborhood of θ_0 and in the slowly expanding interval $[\underline{\alpha}_n, \overline{\alpha}_n]$. The interval $(\theta_0/2, 2\theta_0)$ in principle can be replaced by any neighborhood of the true parameter θ_0, such as $(\theta_0 - \delta_0, \theta_0 + \delta_0)$ for some $0 < \delta_0 < \theta_0$. The inequality (15) in (A.3) requires that the prior assigns a minimum of $\exp(-n^{C_{\pi,3}})$ prior mass on the true parameter θ_0 uniformly over all $\alpha \in [\underline{\alpha}_n, \overline{\alpha}_n]$. Such minimal prior mass assumption is often necessary for achieving the basic posterior consistency in Bayesian models [23]. In particular, we can verify Assumption (A.3) for the following examples of the prior $\pi(\theta|\alpha)$, some of which are commonly used in applications.

PROPOSITION 2.4. *Suppose that the prior $\pi(\theta|\alpha)$ does not depend on the sample size n. Then Assumption* (A.3) *holds in either one of the following cases*:

(i) $\pi(\theta|\alpha) = \pi(\theta)$ *is independent of* α. $\pi(\theta)$ *has continuous first derivative on* $\mathbb{R}^+$ *and* $\pi(\theta) > 0$ *for all* $\theta \in \mathbb{R}^+$.

(ii) $\pi(\alpha)$ *is supported on a compact interval* $[\alpha_1, \alpha_2]$, *with constant lower and upper bounds* $0 < \alpha_1 < \alpha_2 < \infty$. $\pi(\theta|\alpha)$ *is positive for all* $(\theta, \alpha) \in \mathbb{R}^+ \times \mathbb{R}^+$, *continuous in* $\alpha \in \mathbb{R}^+$, *and has continuous first derivative with respect to* θ *on* $\mathbb{R}^+$ *for all* $\alpha \in \mathbb{R}^+$.

(iii) *The prior of* σ^2 *is independent of* α *and belongs to the broad distribution family of the generalized beta of the second kind* (*or the Feller–Pareto family*, [2], *with the density* $\pi(\sigma^2) = \frac{\Gamma(\gamma_1+\gamma_2)}{\Gamma(\gamma_1)\Gamma(\gamma_2)} \frac{(\sigma^2/b)^{\gamma_2/\gamma-1}}{b\gamma[1+(\sigma^2/b)^{1/\gamma}]^{\gamma_1+\gamma_2}}$ *with parameters* $b > 0$, $\gamma > 0$, $\gamma_1 > 0$, $\gamma_2 > 0$.

Proposition 2.4 shows that Assumption (A.3) about $\pi(\theta|\alpha)$ is satisfied by a wide range of prior distributions on θ with continuously differentiable densities. Case (i) says that (A.3) holds as long as the priors of θ and α are independent. Case (ii) says that (A.3) holds as long as the support of the prior of α is bounded away from zero and infinity. Compactly supported priors for the range parameter α have been widely used in Bayesian spatial statistics literature; see, for example, [5, 19, 27, 55], etc. Case (iii) provides the example in which an independent prior is assigned on the variance parameter σ^2 instead of on θ. The generalized beta of the second kind (or Feller–Pareto family, [2, 10]) has polynomially decaying tails at both $\sigma^2 \to 0+$ and $\sigma^2 \to +\infty$. This family covers a wide range of continuous distributions on $(0, +\infty)$ including the half-Student's t distributions, the F distributions, the log-logistic distributions, the Burr distributions and many others [2]. Case (iii) mainly illustrates that if $\pi(\alpha)$ has a full support on $[0, +\infty)$, then $\pi(\theta|\alpha)$ cannot decay too fast in the two tails. For example, if $\pi(\theta|\alpha)$ has exponentially decaying tails at either $\theta \to 0+$ and $\theta \to +\infty$, then (A.3) is not satisfied when $\pi(\alpha)$ has a full support on $[0, +\infty)$. Fortunately, most spatial applications use a compactly supported prior for α, and (A.3) is satisfied as in Case (ii).

Next, we discuss Assumption (A.4), which imposes some technical conditions on the tail behavior of $\pi(\alpha)$ as $\alpha \to 0+$ and $\alpha \to +\infty$.

PROPOSITION 2.5. *Let* $\underline{\kappa}, \overline{\kappa}, \underline{\alpha}_n, \overline{\alpha}_n$ *be defined in* (12). *If a nonnegative function* $p(\alpha)$ *for* $\alpha > 0$ *satisfies either one of the following conditions*:

(i) $p(\alpha) \le \exp(-\alpha^{\delta_1})$ *for all* $\alpha > \overline{\alpha}_n$, *for some constant* $\delta_1 > 1/\overline{\kappa}$ *and for all sufficiently large* n;

(ii) $p(\alpha) \le n^{\delta_3} \exp(-n^{\delta_2}\alpha)$ *for all* $\alpha > \overline{\alpha}_n$, *for some constant* $1 - \overline{\kappa} < \delta_2 \le \delta_3 < \infty$ *and all sufficiently large* n;

then there exists a constant $0 < \overline{c_\pi} < (\nu + d/2)\overline{\kappa}$ *such that for all sufficiently large* n,

$$\int_{\overline{\alpha}_n}^{\infty} \alpha^{n(\nu+d/2)} p(\alpha)\, \mathrm{d}\alpha \le \exp(\overline{c_\pi} n \log n). \tag{20}$$

Similarly, if a nonnegative function $p(\alpha)$ *for* $\alpha > 0$ *satisfies either one of the following conditions*:

(i) $p(\alpha) \le \exp(-\alpha^{-\delta_1})$ *for all* $0 < \alpha < \underline{\alpha}_n$, *for some constant* $\delta_1 > 1/\underline{\kappa}$ *and for all sufficiently large* n;

(ii) $p(\alpha) \le n^{\delta_3} \exp(-n^{\delta_2}/\alpha)$ *for all* $0 < \alpha < \underline{\alpha}_n$, *for some constant* $1 - \underline{\kappa} < \delta_2 \le \delta_3 < \infty$ *and all sufficiently large* n;

then there exists a constant $0 < \underline{c_\pi} < (\nu + d/2)\underline{\kappa}$ *such that for all sufficiently large* n,

$$\int_{0}^{\underline{\alpha}_n} \alpha^{-n(\nu+d/2)} p(\alpha)\, \mathrm{d}\alpha \le \exp(\underline{c_\pi} n \log n). \tag{21}$$

Whilst having formulated Proposition (2.5) for a generic function $p(\alpha)$, we have in mind to apply it to the priors $\pi(\alpha)$ and $\pi(\theta_0|\alpha)\pi(\alpha)$ in (16) and (17) in Assumption (A.4). Since $\int_0^\infty \pi(\theta_0|\alpha)\pi(\alpha)\,\mathrm{d}\alpha < \infty$ as in (A.4), the tail conditions on $\pi(\theta_0|\alpha)\pi(\alpha)$ are the same as the tail conditions on $\pi(\alpha|\theta_0)$. Two types of tail decaying conditions are given in Proposition 2.5. In the first case, the tail of $\pi(\alpha)$ or $\pi(\alpha|\theta_0)$ decays at the exponential power rate $\exp(-\alpha^{\delta_1})$ in the right tail (or $\exp(-\alpha^{-\delta_1})$ in the left tail), with some lower conditions on δ_1 depending on the values of $\overline{\kappa}$ (or $\underline{\kappa}$). This condition requires that $\pi(\alpha)$ and $\pi(\alpha|\theta_0)$ decay very fast in the right (or left) tail. One example of $\pi(\alpha)$ is that $\alpha^{1/\min(\underline{\kappa},\overline{\kappa})}$ follows the inverse Gaussian distribution, since the inverse Gaussian distribution has exponentially decaying tails at zero and infinity. In the second case of Proposition 2.5, we allow the tails of $\pi(\alpha)$ and $\pi(\alpha|\theta_0)$ to be upper bounded by some exponential rate in α that depends on n. These tail decaying conditions in Proposition 2.5 and Assumption (A.4) can ensure that the convergence to a normal limit will be dominant in the joint posterior of (θ, α).

We remark that the tail conditions in (A.4) are often stronger than necessary in practice. This is partly because we have made *no assumption* on the design of the sampling points $\mathcal{S}_n$. Even when $\mathcal{S}_n$ is highly unevenly distributed in $\mathcal{S}$ or is not dense in the full space of $\mathcal{S}$, Theorem 2.3 still holds true under (A.4), which allows the prior $\pi(\alpha)$ to have a full support in $[0, +\infty)$. If one is willing to impose more assumptions on $\mathcal{S}_n$, for example, the maximum distance between two adjacent points decreases at a certain rate to zero, then it is possible to relax the tail conditions in (A.4). Furthermore, such assumptions on the sampling design $\mathcal{S}_n$ may also improve how fast the total variation distance between the joint posterior distribution $\Pi(\mathrm{d}\theta, \mathrm{d}\alpha|Y_n)$ and its limiting distribution in Theorem 2.3 converges to zero. For a general smoothness parameter ν, analyzing the effect of design $\mathcal{S}_n$ inevitably requires more sophisticated matrix theory for the properties of the Matérn correlation matrix R_α and the related quantities $Y_n^\top R_\alpha^{-1} Y_n$ and $|R_\alpha|$ as $\alpha \to 0+$ and $\alpha \to +\infty$, since these two terms determine the properties of the profile restricted log-likelihood function (8). We will see in Theorem 2.6 below that in a special case when the sampling points are from an equispaced grid, the tail conditions in (A.4) can be significantly weakened and the conclusion of Theorem 2.3 can hold for a broader class of priors on α.

2.4. *Limiting posterior distribution for* 1*-dimensional Ornstein–Uhlenbeck process* For a concrete example of Theorem 2.3, we consider the special case of $d = 1$, $\mathcal{S} = [0, 1]$, and $\nu = 1/2$ in the Matérn covariance function. The covariance function becomes $\operatorname{Cov}(X(s), X(t)) = \sigma^2 \exp(-\alpha|s - t|)$ for $s, t \in [0, 1]$, which is also known as the exponential covariance function. The resulted stochastic process X is the 1-dimensional Ornstein–Uhlenbeck process [52]. We assume that the sampling points in $\mathcal{S}_n$ are on the equispaced grid with $s_i = i/n$ for $i = 1, \ldots, n$. For the regression terms, we consider two different cases:

(i) Model (1) without the regression term $\mathrm{m}(\cdot)^\top \beta$, that is, $p = 0$, $Y(s) = X(s)$ for any $s \in [0, 1]$, which implies that $Y_n \sim \mathcal{N}(0, \sigma_0^2 R_{\alpha_0})$;

(ii) Model (1) with a constant regression term, that is, $p = 1$, $\mathrm{m}_1(\cdot) \equiv 1$, $\beta \in \mathbb{R}$, $Y(s) = \beta + X(s)$ for any $s \in [0, 1]$, which implies that $Y_n \sim \mathcal{N}(1_n \beta_0, \sigma_0^2 R_{\alpha_0})$, where 1_n denotes the n-dimensional column vector of all 1's.

For Case (i), we derive an explicit formula for the limiting posterior of α and relax the condition on the tail of $\pi(\alpha)$ in the new Assumption (A.4'). For Case (ii), we show that the posterior of β does not converge to the true parameter β_0 as $n \to \infty$.

For the model in Case (i), the frequentist MLE of (θ, α) under fixed-domain asymptotics has been extensively studied in [14, 21, 83, 84], etc. Since $s_i = i/n$ for $i = 1, \ldots, n$, the inverse matrix R_α^{-1} is given by

$$\big(R_\alpha^{-1}\big)_{ii} = \begin{cases} \big(1 - \mathrm{e}^{-2\alpha/n}\big)^{-1}, & i = 1, n, \\ \big(1 + \mathrm{e}^{-2\alpha/n}\big)/\big(1 - \mathrm{e}^{-2\alpha/n}\big), & i = 2, \ldots, n-1, \end{cases}$$

$$\big(R_\alpha^{-1}\big)_{i,i+1} = \big(R_\alpha^{-1}\big)_{i+1,i} = -\mathrm{e}^{-\alpha/n}\big(1 - \mathrm{e}^{-2\alpha/n}\big)^{-1}, \quad i = 1, \ldots, n-1,$$

and all other entries of R_α are zero. Furthermore, the determinant of R_α is $|R_\alpha| = (1 - \mathrm{e}^{-2\alpha/n})^{n-1}$. Since the model does not contain β, the profile restricted log-likelihood in (8) has the explicit form

$$\widetilde{\mathcal{L}}_n(\alpha) = -\frac{n}{2}\log\big(A_1 \mathrm{e}^{-2\alpha/n} - 2A_2 \mathrm{e}^{-\alpha/n} + A_3\big) + \frac{1}{2}\log\big(1 - \mathrm{e}^{-2\alpha/n}\big), \tag{22}$$

$$\text{where } A_1 = \sum_{i=2}^{n-1} Y(s_i)^2,\ A_2 = \sum_{i=1}^{n-1} Y(s_i)Y(s_{i+1}),\ A_3 = \sum_{i=1}^{n} Y(s_i)^2. \tag{23}$$

For the prior of α, instead of Assumption (A.4), we use a weaker alternative assumption.

(A.4') The marginal prior $\pi(\alpha)$ is a proper and continuous density on $\mathbb{R}^+$. $\pi(\alpha)$ does not depend on n. $\pi(\alpha_0) > 0$. $\int_0^\infty \pi(\theta_0|\alpha)\pi(\alpha)\,\mathrm{d}\alpha < \infty$. $\int_0^\infty \sqrt{\alpha}\pi(\theta_0|\alpha)\pi(\alpha)\,\mathrm{d}\alpha < \infty$. $\int_0^\infty \sqrt{\alpha}\pi(\alpha)\,\mathrm{d}\alpha < \infty$. Furthermore, for $\underline{\alpha}_n$ and $\overline{\alpha}_n$ defined in (12), the following relations hold as $n \to \infty$:

$$\sqrt{n}\int_0^{\underline{\alpha}_n} \sqrt{\alpha}\pi(\alpha)\,\mathrm{d}\alpha \to 0, \qquad \sqrt{n}\int_{\overline{\alpha}_n}^\infty \sqrt{\alpha}\pi(\alpha)\,\mathrm{d}\alpha \to 0. \tag{24}$$

Assumption (A.4') is considerably weaker than Assumption (A.4). Assumption (A.4') only requires that $\pi(\alpha)$ and $\pi(\theta_0|\alpha)\pi(\alpha)$ (or equivalently, $\pi(\alpha|\theta_0)$) to have polynomially decaying tails at zero and infinity, compared to the exponential power tails as in Proposition 2.5. With appropriate choice of hyperparameters, $\pi(\alpha)$ in (A.4') can be taken as gamma, inverse gamma, inverse Gaussian, or the family of generalized beta of the second kind defined in Proposition 2.4; see the beginning of Section S5 in the Supplementary Material [42] for detailed discussion on the choice of hyperparameters.

THEOREM 2.6. *Consider the model* (1) *with* $p = 0$, $d = 1$, $\mathcal{S} = [0, 1]$, $\nu = 1/2$, *and observations* Y_n *on the equispaced grid* $s_i = i/n$ *for* $i = 1, \ldots, n$. *Suppose that Assumptions* (*A*.2), (*A*.3), *and* (*A*.4') *hold. Then*

$$\|\Pi(\mathrm{d}\theta, \mathrm{d}\alpha|Y_n) - \mathcal{N}(\mathrm{d}\theta|\widetilde{\theta}_{\alpha_0}, 2\theta_0^2/n) \times \widetilde{\Pi}(\mathrm{d}\alpha|Y_n)\|_{\mathrm{TV}} \to 0, \tag{25}$$

$$\|\Pi(\mathrm{d}\theta, \mathrm{d}\alpha|Y_n) - \mathcal{N}(\mathrm{d}\theta|\widetilde{\theta}_{\alpha_0}, 2\theta_0^2/n) \times \Pi_*(\mathrm{d}\alpha|Y_n)\|_{\mathrm{TV}} \to 0, \tag{26}$$

as $n \to \infty$ *in* $P_{(\sigma_0^2,\alpha_0)}$*-probability, where* $\widetilde{\theta}_{\alpha_0} = n^{-1}\alpha_0^{2\nu} Y_n^\top R_{\alpha_0}^{-1} Y_n$, *the profile posterior distribution* $\widetilde{\Pi}(\mathrm{d}\alpha|Y_n)$ *has the density* $\widetilde{\pi}(\alpha|Y_n) \propto \exp\{\widetilde{\mathcal{L}}_n(\alpha)\} \cdot \pi(\alpha|\theta_0)$ *with* $\widetilde{\mathcal{L}}_n(\alpha)$ *given in* (22), *and the distribution* $\Pi_*(\mathrm{d}\alpha|Y_n)$ *has the density*

$$\pi_*(\alpha|Y_n) \propto \sqrt{\alpha}\exp\left\{-\frac{(\alpha - u_*)^2}{2v_*}\right\} \cdot \pi(\alpha|\theta_0) \quad \textit{for all } \alpha \in \mathbb{R}^+,$$

$$\textit{where } u_* = \frac{n(A_1 - A_2)}{A_1}, v_* = \frac{n(A_1 - 2A_2 + A_3)}{A_1},$$

and A_1, A_2, A_3 *are defined in* (23). *Furthermore,* $|u_*| \preceq 1$, $v_* > 0$ *and* $v_* \asymp 1$ *as* $n \to \infty$ *in* $P_{(\sigma_0^2,\alpha_0)}$*-probability. Therefore,* $\pi(\alpha|Y_n)$ *does not converge to any point mass distribution as* $n \to \infty$ *in* $P_{(\sigma_0^2,\alpha_0)}$*-probability.*

Theorem 2.6 provides a concrete form for the limiting joint posterior distribution of (θ, α) in the 1-dimensional Ornstein–Uhlenbeck process under fixed-domain asymptotics. Since the model does not contain β, we write $P_{(\sigma_0^2,\alpha_0)}$ instead of $P_{(\beta_0,\sigma_0^2,\alpha_0)}$ in Theorem 2.6. Compared to Theorem 2.3, Theorem 2.6 shows the same limiting distribution under the weaker (A.4'). Furthermore, Theorem 2.6 simplifies the profile posterior density $\widetilde{\pi}(\alpha)$ to a more explicit form $\pi_*(\alpha|Y_n)$, which is a *polynomially tilted normal density* [9] times the conditional prior density $\pi(\alpha|\theta_0)$. The "normal" part of $\pi_*(\alpha|Y_n)$ is centered at u_* with scale v_*. Both center u_* and the scale v_* are of constant order in $P_{(\sigma_0^2,\alpha_0)}$-probability. Moreover, (A.2) and (A.4') ensure that $\pi(\alpha|\theta_0)$ is positive for all $\alpha \in \mathbb{R}^+$. Therefore, the limiting distribution $\pi_*(\alpha|Y_n)$ has a continuous and positive density with a nonshrinking variance on $\mathbb{R}^+$. If $\pi(\alpha|\theta_0)$ does not depend on n, then as a result of the convergence in total variation distance in (26), the marginal posterior $\pi(\alpha|Y_n)$ also cannot converge to any point mass distribution as $n \to \infty$. Therefore, the posterior of α does not converge to the true parameter α_0. This Bayesian asymptotic result matches with the frequentist theory in [85] that there exists no consistent estimator for α under fixed-domain asymptotics.

Next we consider Case (ii). To simplify the expressions, we assume the noninformative prior $\pi(\beta|\sigma^2, \alpha) \propto 1$ which corresponds to $\Omega_\beta = 0_{p\times p}$ in Assumption (A.2). We notice that in Case (ii), $m_1(\cdot) \equiv 1$ and it is infinitely differentiable on $[0, 1]$ with all derivatives equal to zero. Hence it lies in $\mathcal{W}_2^{\nu+d/2}([0, 1])$ for any $\nu > 0$ and $d \in \{1, 2, 3\}$, and Assumption (A.1) is satisfied. We have the following corollary from Theorem 2.3.

COROLLARY 2.7. *Consider the model* (1) *with* $p = 1, \mathrm{m}_1(\cdot) \equiv 1, \pi(\beta|\sigma^2, \alpha) \propto 1, d = 1$, $\mathcal{S} = [0, 1]$, $\nu = 1/2$, *and observations* Y_n *on the equispaced grid* $s_i = i/n$ *for* $i = 1, \ldots, n$. *Suppose that Assumptions* (*A*.2), (*A*.3) *and* (*A*.4) *hold. Then*

$$\beta|Y_n, \theta, \alpha \sim \mathcal{N}\left(\frac{B_2 - B_1 \mathrm{e}^{-\alpha/n}}{(n-2)(1-\mathrm{e}^{-\alpha/n}) + 2}, \frac{\theta(1+\mathrm{e}^{-\alpha/n})}{[(n-2)(1-\mathrm{e}^{-\alpha/n}) + 2]\alpha}\right), \tag{27}$$

$$\|\Pi(\mathrm{d}\theta, \mathrm{d}\alpha|Y_n) - \mathcal{N}(\mathrm{d}\theta|\widetilde{\theta}_{\alpha_0}, 2\theta_0^2/n) \times \widetilde{\Pi}(\mathrm{d}\alpha|Y_n)\|_{\mathrm{TV}} \to 0, , \tag{28}$$

as $n \to \infty$ almost surely $P_{(\beta_0,\sigma_0^2,\alpha_0)}$, where the profile posterior distribution $\widetilde{\Pi}(\mathrm{d}\alpha|Y_n)$ has the density $\widetilde{\pi}(\alpha|Y_n) \propto \exp\{\widetilde{\mathcal{L}}_n(\alpha)\} \cdot \pi(\alpha|\theta_0)$, and the formulas of $\widetilde{\theta}_\alpha$ and $\widetilde{\mathcal{L}}_n(\alpha)$ are given by

$$\widetilde{\theta}_\alpha = \frac{\alpha(1-\mathrm{e}^{-2\alpha/n})^{-1}}{n-1}\left\{(A_1\mathrm{e}^{-2\alpha/n} - 2A_2\mathrm{e}^{-\alpha/n} + A_3) - \frac{(1-\mathrm{e}^{-\alpha/n})(B_2 - B_1\mathrm{e}^{-\alpha/n})^2}{(n-2)(1-\mathrm{e}^{-\alpha/n})+2}\right\},$$

$$\begin{aligned}\widetilde{\mathcal{L}}_n(\alpha) &= -\frac{n-1}{2}\log\left\{(A_1\mathrm{e}^{-2\alpha/n} - 2A_2\mathrm{e}^{-\alpha/n} + A_3) - \frac{(1-\mathrm{e}^{-\alpha/n})(B_2 - B_1\mathrm{e}^{-\alpha/n})^2}{(n-2)(1-\mathrm{e}^{-\alpha/n})+2}\right\} \\ &\quad + \frac{1}{2}\log\frac{1+\mathrm{e}^{-\alpha/n}}{(n-2)(1-\mathrm{e}^{-\alpha/n})+2},\end{aligned}$$

where $B_1 = \sum_{i=2}^{n-1} Y(s_i)$, $B_2 = \sum_{i=1}^{n} Y(s_i)$, and A_1, A_2, A_3 are as defined in (23).

Furthermore, for any $\eta \in (0, 1/4)$, there exists constants $\epsilon_0 > 0$, $\delta_0 \in (0,1)$ and a large integer N_2, such that $\Pr(\Pi(|\beta - \beta_0| > \epsilon_0|Y_n) > \delta_0) > 1 - \eta$ for all $n > N_2$. Therefore, the posterior distribution of β is inconsistent for the true parameter β_0.

Corollary 2.7 provides a concrete example that the posterior of β is not consistent under fixed-domain asymptotics. In fact, this can be seen from the conditional posterior variance of β given in (27). For a fixed α, this variance is close to $2\theta_0/[\alpha(\alpha+2)]$ as $n \to \infty$ since θ drawn from the posterior is close to θ_0. Therefore, the posterior variance of β does not vanish as $n \to \infty$. We expect that this is also true for general $\mathrm{m}(\cdot)$ functions, since one cannot expect to consistently estimate the regression coefficients β only based on a single sample path $Y(\cdot)$. This echoes the frequentist result that the MLE of β is inconsistent under fixed-domain asymptotics; see, for example, Lemma 5 of [25].

2.5. *Relation to previous Bayesian results. Relation to previous BvM results.* In the presence of nuisance parameters, [57] and [8] have developed general machinery for proving BvM results in the presence of possibly nonparametric nuisance parameters. They assume that the model depends on an identifiable parameter and a nuisance parameter. Bickel and Kleijn [8] first establish a LAN result for each value of the identifiable parameter inside a neighborhood of the "least-favorable submodel," which is a contracting neighborhood of the nuisance parameter around the minimizer of the Kullback–Leibler divergence. Then their Theorem 4.2 gives the integral LAN property with integration over the nuisance parameter. They further proposes a rate free BvM theorem in their Corollary 5.2 that allows a noncontracting posterior for the nuisance parameter, which can be related to the posterior distribution of α in our GP model.

Despite the similarity, we adopt a more direct proof technique for the GP model with isotropic Matérn covariance function, instead of checking the condition on Hellinger distance in [8] for uniform tests. There are several additional challenges. First, the likelihood function in our GP model cannot be written in an independent product form. The design of the sampling points $\mathcal{S}_n$ is arbitrary, making R_α^{-1} and $|R_\alpha|$ completely intractable. This determines that the LAN condition in our model is fundamentally different from that for independent or weakly dependent data considered in [8]. We instead use the tools of RKHS theory and spectral analysis to establish the LAN condition for θ. We integrate out θ for each given α and obtain the profile posterior distribution of α as in (19). Second, our LAN condition holds uniformly over all $\alpha \in [\underline{\alpha}_n, \overline{\alpha}_n]$, but we still need to handle those α outside $[\underline{\alpha}_n, \overline{\alpha}_n]$. We derive sufficient tail conditions on $\pi(\alpha)$ such that the posterior probability outside $[\underline{\alpha}_n, \overline{\alpha}_n]$ vanishes as $n \to \infty$. This involves detailed analysis on the properties of the profile posterior distribution in (19); see Section S2 of the Supplementary Material [42].

In the broader sense, our work contributes a new example to the literature of limiting posterior distribution for nonregular models; see, for example, [9, 15, 16, 34, 39], etc.

Relation to partially identified models. Our theorems for the covariance parameters can also be related to the Bayesian literature of *partially identified models*. Such models have been studied extensively in statistics and econometrics literature, but only for independent and weakly dependent data [29, 47, 67]. In partially identified models, the probability distribution of the data is compatible with a set of different parameter values. This parameter set is referred to as the *identification region*. As a result, consistent point estimator for the true parameter does not exist, though one can still consistently estimate the identification region. The asymptotic property of posterior distributions in partially identified models have been studied in [15, 28, 32, 33, 49], etc. However, the Bayesian theory from these works only applies to independent data and weakly dependent data, and does not apply to our GP model. Depending on the assumptions, the limiting posterior of the nuisance parameter can either only depend on the prior [49], or depend on the prior and some asymptotically deterministic function of the identifiable part of the parameter vector [32].

Our paper contributes a new example to the Bayesian partial identification literature. Consider the model (1) with isotropic Matérn covariance function $\sigma^2 K_{\alpha,\nu}$ and without regression terms, that is, $Y(\cdot) = X(\cdot)$. Under fixed-domain asymptotics, the distribution of Y_n is asymptotically compatible with any parameters on the curve $\Gamma_{\theta_0} = \{(\sigma^2, \alpha) \in \mathbb{R}^+ \times \mathbb{R}^+ : \sigma^2\alpha^{2\nu} = \theta_0\}$, which is the identification region in our problem. Different from [49], our Theorem 2.3 shows that both the prior and the data Y_n play important roles in the posterior of α. The data Y_n influences the posterior through the profile restricted likelihood function. Different from [32], Theorem 2.6 shows that the influence from Y_n is always stochastic instead of asymptotically deterministic, as the polynomially tilted normal distribution $\pi_*(\alpha|Y_n)$ has a scale v_* dependent on Y_n and not converging to any point limit asymptotically.

3. Asymptotic efficiency and convergence rate of posterior prediction. The limiting theorems in Section 2 shows that the posterior of the microergodic parameter θ in the Matérn covariance function satisfies the same $n^{-1/2}$-convergence to a normal limit. This result has an important implication for the Bayesian GP (or kriging) prediction with covarinace parameters randomly drawn from the posterior distribution at a new location $s^* \in \mathcal{S}\backslash\mathcal{S}_n$, that is, s^* is an arbitrary point in $\mathcal{S}$ but different from the sampling points $\mathcal{S}_n$. We first show that for the general model (1), the Bayesian GP predictive variance is almost equal to the one with a known θ_0. Then we discuss the detailed posterior asymptotic efficiency for the model without regression terms and the convergence rates for the model with regression terms. We also present results both for a fixed α and for a range of α values.

Consider the linear prediction (or kriging) of $Y(s^*)$ using the data Y_n. Let $r_\alpha(s^*) = (K_{\alpha,\nu}(s_1 - s^*), \ldots, K_{\alpha,\nu}(s_n - s^*))^\top$ be the correlation vector between s^* and $\{s_1, \ldots, s_n\}$. Then under a possibly misspecified model $Y \sim \mathrm{GP}(\mathrm{m}^\top\beta, \sigma^2 K_{\alpha,\nu})$, the best linear unbiased predictor (BLUP) for $Y(s^*)$ using (Section 1.5 of [64]) is

$$\widehat{Y}(s^*; \beta, \alpha) = \mathrm{m}(s^*)^\top\beta + r_\alpha(s^*)^\top R_\alpha^{-1}(Y_n - M_n\beta). \tag{29}$$

This kriging predictor only depends on (β, α) but not σ^2. Now under the Bayesian setup, we randomly draw $(\beta, \sigma^2, \alpha)$ from the posterior $\Pi(\cdot|Y_n)$ to predict $Y(s^*)$. We denote the predicted variable as $\widetilde{Y}(s^*)$. Using the Gaussian process predictive distribution, we have

$$\widetilde{Y}(s^*)|Y_n, \beta, \sigma^2, \alpha \sim \mathcal{N}\big(\widehat{Y}(s^*; \beta, \alpha), \sigma^2\{1 - r_\alpha(s^*)^\top R_\alpha^{-1} r_\alpha(s^*)\}\big).$$

We can integrate out β using (5) to derive that

$$
\begin{aligned}
&\widetilde{Y}(s^*)|Y_n,\sigma^2,\alpha\sim\mathcal{N}\big(\widehat{Y}(s^*;\alpha),\mathrm{v}_n(s^*;\sigma^2,\alpha)\big),\\
&\quad\text{where }\widehat{Y}(s^*;\alpha)=r_\alpha(s^*)^\top R_\alpha^{-1}Y_n+b_\alpha(s^*)^\top\big(M_n^\top R_\alpha^{-1}M_n+\Omega_\beta\big)^{-1}M_n^\top R_\alpha^{-1}Y_n,\\
&\mathrm{v}_n(s^*;\sigma^2,\alpha)\\
&\quad=\sigma^2\big\{1-r_\alpha(s^*)^\top R_\alpha^{-1}r_\alpha(s^*)\big\}+\sigma^2 b_\alpha(s^*)^\top\big(M_n^\top R_\alpha^{-1}M_n+\Omega_\beta\big)^{-1}b_\alpha(s^*),\\
&\quad\text{and }b_\alpha(s^*)=\mathrm{m}(s^*)-M_n^\top R_\alpha^{-1}r_\alpha(s^*),\ \text{for any } s^*\in\mathcal{S}.
\end{aligned}
\tag{30}
$$

The detailed derivation of (30) is in Section S6.1 of the Supplementary Material [42]. This normal predictive distribution is the same as in equation (2.4) of [30], which is for the special case of $\Omega_\beta=0_{p\times p}$. The predictive variance of $\widetilde{Y}(s^*)$, $\mathrm{v}_n(s^*;\sigma^2,\alpha)$ in (30), is the main focus of this section, because it directly quantifies the Bayesian uncertainty of GP prediction.

We first show that if (σ^2,α) is randomly drawn from the posterior $\Pi(\cdot|Y_n)$, then the GP predictive variance $\mathrm{v}_n(s^*;\sigma^2,\alpha)$ is almost equal to $\mathrm{v}_n(s^*;\theta_0/\alpha^{2\nu},\alpha)$, that is, as if the true microergodic parameter θ_0 were known. We notice that $\mathrm{v}_n(s^*;\sigma^2,\alpha)$ is random due to the randomness in the posterior distribution of (σ^2,α).

THEOREM 3.1 (Posterior asymptotic efficiency compared to the half oracle model).

(i) *Under Assumptions* (A.1) *and* (A.2), *for any fixed* $\alpha>0$, *as* $n\to\infty$, *almost surely* $P_{(\beta_0,\sigma_0^2,\alpha_0)}$,

$$
\Pi\Bigg[\sup_{s^*\in\mathcal{S}\setminus\mathcal{S}_n}\bigg|\frac{\mathrm{v}_n(s^*;\sigma^2,\alpha)}{\mathrm{v}_n(s^*;\theta_0/\alpha^{2\nu},\alpha)}-1\bigg|>7n^{-1/2}\log n\Big|Y_n,\alpha\Bigg]\to 0.
$$

(ii) *Under Assumptions* (A.1), (A.2), (A.3) *and* (A.4), *as* $n\to\infty$, *almost surely* $P_{(\beta_0,\sigma_0^2,\alpha_0)}$,

$$
\Pi\Bigg[\sup_{s^*\in\mathcal{S}\setminus\mathcal{S}_n}\bigg|\frac{\mathrm{v}_n(s^*;\sigma^2,\alpha)}{\mathrm{v}_n(s^*;\theta_0/\alpha^{2\nu},\alpha)}-1\bigg|>7n^{-1/2}\log n\Big|Y_n\Bigg]\to 0.
$$

Theorem 3.1 shows that the GP predictive variance at an arbitrary new location s^* evaluated under the measure $P_{(\beta,\sigma^2,\alpha)}$ is asymptotically equal to the predictive MSE evaluated under the measure $P_{(\beta_0,\theta_0/\alpha^{2\nu},\alpha)}$. Part (i) and part (ii) are the direct consequence of Theorem 2.1 for the posterior of θ given α and Theorem 2.3 for the joint posterior of (θ,α), respectively. We also give the explicit convergence rate $n^{-1/2}\log n$, in which the $\log n$ factor is to ensure the almost sure convergence. Theorem 3.1 shows that the prediction performance from a random draw of (θ,α) from the posterior is as good as the "half oracle" model with the true microergodic parameter θ_0 and the same range parameter α. It is half oracle because Theorem 3.1 has not yet set the range parameter at the true α_0 and compared with $\mathrm{v}_n(s^*;\theta_0/\alpha_0^{2\nu},\alpha_0)$. On the other hand, Theorem 3.1 only requires the same conditions as Theorem 2.3.

In the following, we will compare $\mathrm{v}_n(s^*;\sigma^2,\alpha)$ with $\mathrm{v}_n(s^*;\theta_0/\alpha_0^{2\nu},\alpha_0)$, the predictive variance from the full oracle model where both θ and α are set at their true values. We first study a simplified model without regression terms and prove the asymptotic efficiency in posterior prediction with respect to the full oracle model, and then consider the general model (1) and show the same optimal posterior convergence rates as the full oracle model.

3.1. *Posterior asymptotic efficiency without regression terms.* In this subsection, we consider a special case of the model (1) where the regression term $m(\cdot)^\top\beta$ is absent and the model simplifies to

$$Y(s) = X(s) \text{ for any } s \in \mathcal{S},\ X \sim \mathrm{GP}\big(0, \sigma^2 K_{\alpha,\nu}\big). \tag{31}$$

We observe $Y_n \sim \mathcal{N}(0, \sigma_0^2 R_{\alpha_0})$ at the sampling points $\mathcal{S}_n$. This is equivalent to setting $p = 0$. For this model, we prove the strong result that $\mathrm{v}_n(s^*; \sigma^2, \alpha)$ with (σ^2, α) randomly drawn from the posterior is asymptotically equal to $\mathrm{v}_n(s^*; \theta_0/\alpha_0^{2\nu}, \alpha_0)$ and quantify the convergence rate. We need the following dense assumption.

(A.5) The sequence of $\mathcal{S}_n = \{s_1, \ldots, s_n\}$ is getting dense in $\mathcal{S} = [0, T]^d$ as $n \to \infty$, in the sense that $\sup_{s^* \in \mathcal{S}} \min_{1 \le i \le n} \|s^* - s_i\| \to 0$ as $n \to \infty$.

The sets $\mathcal{S}_1, \mathcal{S}_2, \ldots$ are increasingly dense in the fixed domain $\mathcal{S}$, so that we can predict at any new location accurately. But we do not require the sequence $\mathcal{S}_1, \mathcal{S}_2, \ldots$ to be nested.

In the model (31), the BLUP of $Y(s^*)$ is $\widehat{Y}(s^*; \alpha) = r_\alpha(s^*)^\top R_\alpha^{-1} Y_n$, and $\mathrm{v}_n(s^*; \sigma^2, \alpha) = \sigma^2\{1 - r_\alpha(s^*)^\top R_\alpha^{-1} r_\alpha(s^*)\}$ from (30). We notice that in this case, another interpretation of $\mathrm{v}_n(s^*; \sigma^2, \alpha)$ is the GP prediction mean squared error of the BLUP $\widehat{Y}(s^*; \alpha)$ [37]. That is, if we let $e_n(s^*; \alpha) = \widehat{Y}(s^*; \alpha) - Y(s^*)$, then $\mathrm{v}_n(s^*; \sigma^2, \alpha) = \mathrm{E}_{(\sigma^2,\alpha)}\{e_n(s^*; \alpha)^2\}$. The optimal "oracle" predictive MSE using the true parameters is $\mathrm{v}_n(s^*; \sigma_0^2, \alpha_0) = \mathrm{E}_{(\sigma_0^2,\alpha_0)}\{e_n(s^*; \alpha_0)^2\}$. Under the true model $\mathrm{GP}(0, \sigma_0^2 K_{\alpha_0,\nu})$, the predictive MSE based on a misspecified α is

$$\mathrm{E}_{(\sigma_0^2,\alpha_0)}\big\{e_n\big(s^*; \alpha\big)^2\big\} = \sigma_0^2\big\{1 - 2r_\alpha\big(s^*\big)^\top R_\alpha^{-1} r_{\alpha_0}\big(s^*\big) + r_\alpha\big(s^*\big)^\top R_\alpha^{-1} R_{\alpha_0} R_\alpha^{-1} r_\alpha\big(s^*\big)\big\}.$$

We are interested in whether $\mathrm{E}_{(\sigma^2,\alpha)}\{e_n(s^*; \alpha)^2\}$, the predictive MSE under the true measure $\mathrm{E}_{(\sigma_0^2,\alpha_0)}\{e_n(s^*; \alpha)^2\}$, and the oracle predictive MSE $\mathrm{E}_{(\sigma_0^2,\alpha_0)}\{e_n(s^*; \alpha)^2\}$ are close to each other. In a series of works [58–60, 62, 63] and [65], Stein has systematically studied the GP prediction problem and shown that if an incorrect Gaussian process model is used for prediction, the predictive variance at s^* is asymptotically equal to the predictive variance at s^* using the incorrect model but evaluated under the true Gaussian process model, as long as the two Gaussian measures are compatible (or mutually absolutely continuous). For our GP model with mean-zero and isotropic Matérn covariance function with $d \in \{1, 2, 3\}$, the compatibility of the incorrect model $\mathrm{GP}(0, \sigma^2 K_{\alpha,\nu})$ and the true model $\mathrm{GP}(0, \sigma_0^2 K_{\alpha_0,\nu})$ simplifies to the equivalence condition $\sigma^2\alpha^{2\nu} = \theta_0 = \sigma_0^2\alpha_0^{2\nu}$, that is, they have the same microergodic parameter θ_0. If the equivalence condition holds, then [58, 59] and [60] have shown that for the model without regression terms (31), as $n \to \infty$,

$$\sup_{s^* \in \mathcal{S} \setminus \mathcal{S}_n} \left| \frac{\mathrm{E}_{(\sigma^2,\alpha)}\{e_n(s^*; \alpha)^2\}}{\mathrm{E}_{(\sigma_0^2,\alpha_0)}\{e_n(s^*; \alpha)^2\}} - 1 \right| \to 0, \qquad \sup_{s^* \in \mathcal{S} \setminus \mathcal{S}_n} \left| \frac{\mathrm{E}_{(\sigma^2,\alpha)}\{e_n(s^*; \alpha)^2\}}{\mathrm{E}_{(\sigma_0^2,\alpha_0)}\{e_n(s^*; \alpha_0)^2\}} - 1 \right| \to 0, \tag{32}$$

which is called *asymptotic efficiency in linear prediction*. The first convergence shows that for the BLUP (29), the predictive MSEs are almost the same under either the incorrect Gaussian measure $P_{(\sigma^2,\alpha)}$ or the true Gaussian measure $P_{(\sigma_0^2,\alpha_0)}$. The second convergence shows that the predictive MSEs obtained from the incorrect model $\mathrm{GP}(0, \sigma^2 K_{\alpha,\nu})$ is asymptotically equal to the optimal predictive MSE from the true model $\mathrm{GP}(0, \sigma_0^2 K_{\alpha_0,\nu})$.

Using the weakened conditions in [62], Theorem 4 of [37] shows that in the model (31), for a given $\alpha > 0$, the prediction based on the MLE of σ^2 for a *fixed* $\alpha > 0$ satisfies that

$$\sup_{s^* \in \mathcal{S} \setminus \mathcal{S}_n} \left| \frac{\mathrm{E}_{(\widetilde{\sigma}_\alpha^2,\alpha)}\{e_n(s^*; \alpha)^2\}}{\mathrm{E}_{(\sigma_0^2,\alpha_0)}\{e_n(s^*; \alpha)^2\}} - 1 \right| \to 0,$$

as $n \to \infty$ almost surely $P_{(\sigma_0^2,\alpha_0)}$, where $\widetilde{\sigma}_\alpha^2 = n^{-1} Y_n^\top R_\alpha^{-1} Y_n$ is the MLE of σ^2.

Motivated by these works, we establish the Bayesian version of (32), called *asymptotic efficiency in posterior prediction*, which is the posterior asymptotic efficiency compared to the full oracle model. In Bayesian inference, we randomly draw (σ^2, α) from the joint posterior distribution, and compute the predictive MSE at a new location $s^* \in \mathcal{S}\backslash\mathcal{S}_n$ using the Gaussian measure $P_{(\sigma^2,\alpha)}$.

For a given $\alpha > 0$, we define the following sequence $\varsigma_n(\alpha)$, which will be useful:

$$\varsigma_n(\alpha) = \max\Bigg\{ \sup_{s^*\in\mathcal{S}\backslash\mathcal{S}_n} \left| \frac{\mathrm{E}_{(\theta_0/\alpha^{2\nu},\alpha)}\{e_n(s^*;\alpha)^2\}}{\mathrm{E}_{(\sigma_0^2,\alpha_0)}\{e_n(s^*;\alpha)^2\}} - 1 \right|, \tag{33}$$
$$\sup_{s^*\in\mathcal{S}\backslash\mathcal{S}_n} \left| \frac{\mathrm{E}_{(\theta_0/\alpha^{2\nu},\alpha)}\{e_n(s^*;\alpha)^2\}}{\mathrm{E}_{(\sigma_0^2,\alpha_0)}\{e_n(s^*;\alpha_0)^2\}} - 1 \right| \Bigg\}.$$

For a given $\alpha > 0$, as $n \to \infty$, Theorem 3.1 of [58] shows that the first rate in $\varsigma_n(\alpha)$ in (33) converges to zero, and Theorem 1 of [60] further implies that the second rate in $\varsigma_n(\alpha)$ in (33) converges to zero. To handle a random range parameter α, we need the following uniform convergence condition.

(A.6) There exists a positive deterministic sequence $\varsigma_n \to 0$ as $n \to \infty$, such that $\sup_{\alpha\in[\underline{\alpha}_n,\overline{\alpha}_n]} \varsigma_n(\alpha) \le \varsigma_n$ for the sequence $\varsigma_n(\alpha)$ defined in (33).

We have the following theorem for the prediction MSE in the model (31).

THEOREM 3.2 (Posterior asymptotic efficiency compared to the full oracle model under (31)).

(i) (*For a fixed* α) *Under Assumptions* (*A*.2) *and* (*A*.5), *as* $n \to \infty$, *almost surely* $P_{(\sigma_0^2,\alpha_0)}$,

$$\Pi\left[\sup_{s^*\in\mathcal{S}\backslash\mathcal{S}_n} \left| \frac{\mathrm{E}_{(\sigma^2,\alpha)}\{e_n(s^*;\alpha)^2\}}{\mathrm{E}_{(\sigma_0^2,\alpha_0)}\{e_n(s^*;\alpha)^2\}} - 1 \right| > \max\{16n^{-1/2}\log n, 2\varsigma_n(\alpha)\} \Big| Y_n, \alpha \right] \to 0,$$

$$\Pi\left[\sup_{s^*\in\mathcal{S}\backslash\mathcal{S}_n} \left| \frac{\mathrm{E}_{(\sigma^2,\alpha)}\{e_n(s^*;\alpha)^2\}}{\mathrm{E}_{(\sigma_0^2,\alpha_0)}\{e_n(s^*;\alpha_0)^2\}} - 1 \right| > \max\{16n^{-1/2}\log n, 2\varsigma_n(\alpha)\} \Big| Y_n, \alpha \right] \to 0,$$

where $\varsigma_n(\alpha)$ *is given in* (33);

(ii) (*For random* α) *Under Assumptions* (*A*.2), (*A*.3), (*A*.4), (*A*.5) *and* (*A*.6), *as* $n \to \infty$, *almost surely* $P_{(\sigma_0^2,\alpha_0)}$,

$$\Pi\left[\sup_{s^*\in\mathcal{S}\backslash\mathcal{S}_n} \left| \frac{\mathrm{E}_{(\sigma^2,\alpha)}\{e_n(s^*;\alpha)^2\}}{\mathrm{E}_{(\sigma_0^2,\alpha_0)}\{e_n(s^*;\alpha)^2\}} - 1 \right| > \max(16n^{-1/2}\log n, 2\varsigma_n) \Big| Y_n \right] \to 0, \tag{34}$$
$$\Pi\left[\sup_{s^*\in\mathcal{S}\backslash\mathcal{S}_n} \left| \frac{\mathrm{E}_{(\sigma^2,\alpha)}\{e_n(s^*;\alpha)^2\}}{\mathrm{E}_{(\sigma_0^2,\alpha_0)}\{e_n(s^*;\alpha_0)^2\}} - 1 \right| > \max(16n^{-1/2}\log n, 2\varsigma_n) \Big| Y_n \right] \to 0,$$

where ς_n *is given in Assumption* (*A*.6).

We emphasize again that $\mathrm{E}_{(\sigma^2,\alpha)}\{e_n(s^*;\alpha)^2\} = \mathrm{v}_n(s^*;\sigma^2,\alpha)$ and $\mathrm{E}_{(\sigma_0^2,\alpha_0)}\{e_n(s^*;\alpha_0)^2\} = \mathrm{v}_n(s^*;\sigma_0^2,\alpha_0)$ for the model (31) without regression terms. Part (i) of Theorem 3.2 establishes two posterior convergence results. The first convergence is about the ratio of the predictive MSEs using a misspecified range parameter α evaluated under the measure $P_{(\sigma^2,\alpha)}$ and the true measure $P_{(\sigma_0^2,\alpha_0)}$, which implies that these two predictive MSEs are asymptotically equal. The second convergence is about the ratio of the predictive MSEs using the incorrect model

$P_{(\sigma^2,\alpha)}$ and the full oracle optimal predictive MSE using the true model $P_{(\sigma_0^2,\alpha_0)}$. This implies that the predictive MSE computed with random parameters (θ,α) drawn from the posterior can asymptotically recover the exact full oracle optimal predictive MSE. Both convergence rates depend on two parts: one is the posterior convergence rate of θ to θ_0, which is as fast as $n^{-1/2}\log n$; the other is the convergence rate from the convergence of the two ratios in the definition of $\varsigma_n(\alpha)$ in (33), which has been shown before by [58] and [60].

Part (ii) of Theorem 3.2 is similar to part (i) with the same interpretation of asymptotic efficiency, except that α is also random and (σ^2,α) is drawn from their joint posterior. Furthermore, Assumption (A.6) is used to guarantee the uniform convergence over the majority of α values in the interval $[\underline{\alpha}_n,\overline{\alpha}_n]$. Part (ii) shows that the predictive MSE computed from randomly drawn (σ^2,α) from the posterior is asymptotically equal to the oracle optimal predictive MSE with the true parameters.

We emphasize that the posterior asymptotic efficiency in Theorem 3.2 automatically implies that $\mathrm{E}_{(\sigma^2,\alpha)}\{e_n(s^*;\alpha)^2\}$ with (σ^2,α) randomly drawn from the posterior must always converge at exactly the same rate to zero as $\mathrm{E}_{(\sigma_0^2,\alpha_0)}\{e_n(s^*;\alpha_0)^2\}$, regardless of how fast $\mathrm{E}_{(\sigma_0^2,\alpha_0)}\{e_n(s^*;\alpha_0)^2\}$ converges to zero. Therefore, the posterior asymptotic efficiency is stronger than posterior convergence rate results.

To clarify the rate ς_n in Assumption (A.6), we revisit the 1-dimensional Ornstein–Uhlenbeck process in Case (i) in Section 2.4 and derive an explicit form for ς_n.

THEOREM 3.3. *For the case of $d=1$, $\nu=1/2$, $\mathcal{S}=[0,1]$ and equispaced grid $s_i=i/n$, for $i=1,\dots,n$, Assumption (A.6) is satisfied with $\varsigma_n=3n^{-1/2+(\overline{\kappa}+\underline{\kappa}/2)}$, where $\overline{\kappa}$ and $\underline{\kappa}$ are defined in (12). As a result, under Assumptions (A.2), (A.3), (A.4), (A.5), as $n\to\infty$, almost surely $P_{(\sigma_0^2,\alpha_0)}$,*

$$\Pi\Bigg[\sup_{s^*\in\mathcal{S}\setminus\mathcal{S}_n}\bigg|\frac{\mathrm{E}_{(\sigma^2,\alpha)}\{e_n(s^*;\alpha)^2\}}{\mathrm{E}_{(\sigma_0^2,\alpha_0)}\{e_n(s^*;\alpha)^2\}}-1\bigg|>6n^{-1/2+(\overline{\kappa}+\underline{\kappa}/2)}\Big|Y_n\Bigg]\to 0,$$

$$\Pi\Bigg[\sup_{s^*\in\mathcal{S}\setminus\mathcal{S}_n}\bigg|\frac{\mathrm{E}_{(\sigma^2,\alpha)}\{e_n(s^*;\alpha)^2\}}{\mathrm{E}_{(\sigma_0^2,\alpha_0)}\{e_n(s^*;\alpha_0)^2\}}-1\bigg|>6n^{-1/2+(\overline{\kappa}+\underline{\kappa}/2)}\Big|Y_n\Bigg]\to 0.$$

To prove Theorem 3.3, we use the result in [60] and relate the rate ς_n in Assumption (A.6) to the convergence rate of the finite sample version of the symmetrized Kullback–Leibler divergence between two equivalent Gaussian measures toward its limit. Since $\overline{\kappa}$ and $\underline{\kappa}$ are both small positive numbers as given in (12), the two posterior convergence rates for asymptotic efficiency in Theorem 3.3 are both close to the rate $n^{-1/2}$.

3.2. *Optimal rates for GP predictive variance with regression terms.* We now consider the general universal kriging model (1) with the regression term $\mathrm{m}(\cdot)^\top\beta$. Like Theorem 3.2, we also need a similar assumption to Assumption (A.6).

(A.6') There exists a positive deterministic sequence $\tilde{\varsigma}_n\to 0$ as $n\to\infty$, such that

$$\sup_{\alpha\in[\underline{\alpha}_n,\overline{\alpha}_n]}\sup_{s^*\in\mathcal{S}\setminus\mathcal{S}_n}\bigg|\frac{(\theta_0/\alpha^{2\nu})[1-r_\alpha(s^*)^\top R_\alpha^{-1}r_\alpha(s^*)]}{\sigma_0^2[1-r_{\alpha_0}(s^*)^\top R_{\alpha_0}^{-1}r_{\alpha_0}(s^*)]}-1\bigg|\le\tilde{\varsigma}_n. \tag{35}$$

Because the relative error in (35) is exactly the second relative error in the definition of $\varsigma_n(\alpha)$ in (33), Assumption (A.6') is weaker than and implied by Assumption (A.6). Therefore, by Theorem 3.3, we can take $\tilde{\varsigma}_n=3n^{-1/2+(\overline{\kappa}+\underline{\kappa}/2)}$ for 1-dimensional Ornstein–Uhlenbeck process in Assumption (A.6').

To quantify the convergence rate of $\mathrm{v}_n(s^*;\sigma^2,\alpha)$, we follow the literature on kriging and define the *fill distance* given a set of design points $\mathcal{S}_n=\{s_1,\ldots,s_n\}$ as

$$h_{\mathcal{S}_n}=\sup_{s\in\mathcal{S}}\min_{s_i\in\mathcal{S}_n}\|s-s_i\|. \tag{36}$$

The fill distance quantifies the space-filling properties of $\mathcal{S}_n$. The convergence rates of kriging in Model (1) can often be expressed as a function of $h_{\mathcal{S}_n}$ [69, 77, 78, 80]. Then we have the following theorem on the posterior convergence rate of Bayesian GP predictive variance.

THEOREM 3.4. *Suppose that Assumptions* (A.1), (A.2), (A.3), (A.4), (A.5) *and* (A.6') *hold. Let* $C_{\mathrm{m}}=\sum_{j=1}^p\|\mathrm{m}_j\|^2_{\mathcal{W}_2^{\nu+d/2}(\mathcal{S})}$. *For an index set* $\mathcal{I}\subseteq\{1,\ldots,n\}$, *let* $|\mathcal{I}|$ *be its cardinality and* $M_{\mathcal{I}}$ *be the submatrix of* M_n *with row indexes in* $\mathcal{I}$. *Assume that for each* $\mathcal{S}_n$, $\underline{\lambda}(M_n,p)=\max_{\mathcal{I}\subseteq\{1,\ldots,n\},|\mathcal{I}|=p}\lambda_{\min}(M_{\mathcal{I}}^\top M_{\mathcal{I}})/p>0$. *Then for any* $\eta,\delta\in(0,1)$, *there exist large constants* $C_{\mathrm{v},1}>0, C_{\mathrm{v},2}>0$ *that depend on* $\sigma_0^2,\alpha_0,\nu,d,T$, *and a large constant* $C_{\mathrm{v},3}>0$ *and large integer* N_3 *that depend on* $\eta,\delta,\sigma_0^2,\alpha_0,\nu,d,T$, *such that for all* $n>N_3$,

$$\sup_{s^*\in\mathcal{S}}\mathrm{v}_n(s^*;\sigma_0^2,\alpha_0)\le C_{\mathrm{v},1}\big[C_{\mathrm{m}}\sigma_0^2\underline{\lambda}(M_n,p)^{-1}+1\big]h_{\mathcal{S}_n}^{2\nu}\quad\text{and}$$

$$\Pr\Big(\Pi\Big[\sup_{s^*\in\mathcal{S}}\mathrm{v}_n(s^*;\sigma^2,\alpha)\le C_{\mathrm{v},2}\big[C_{\mathrm{v},3}C_{\mathrm{m}}\underline{\lambda}(M_n,p)^{-1}+1\big]h_{\mathcal{S}_n}^{2\nu}\Big|Y_n\Big]>1-\delta\Big)>1-\eta.$$

Theorem 3.4 essentially shows that with (σ^2,α) randomly drawn from the posterior distribution $\Pi(\cdot|Y_n)$, the Bayesian GP predictive variance $\mathrm{v}_n(s^*;\sigma^2,\alpha)$ converges to zero at almost the same rate as the oracle predictive variance $\mathrm{v}_n(s^*;\sigma_0^2,\alpha_0)$ using the true parameters (σ_0^2,α_0), as $n\to\infty$ in $P_{(\sigma_0^2,\alpha_0)}$-probability. Given that the posterior support of (σ^2,α) is unbounded, $\sup_{s^*\in\mathcal{S}}\mathrm{v}_n(s^*;\sigma^2,\alpha)$ with $\mathrm{v}_n(s^*;\sigma^2,\alpha)$ defined in (30) could be potentially very large if σ^2 is large. However, our Theorem 3.4 shows that the posterior convergence rate can still be controlled even with (σ^2,α) randomly drawn from the posterior with unbounded support. The proof of Theorem 3.4 crucially depends on the limiting posterior distribution of (θ,α) proved in Theorem 2.3.

The convergence rates of GP predictive error have been extensively studied in the frequentist literature ([58, 69, 77, 81], etc.) Wu and Schaback [79] has shown that the squared L_2 kriging prediction error for the GP with a Matérn covariance function, fixed covariance parameters, and no regression terms is $O(h_{\mathcal{S}_n}^{2\nu})$ for sufficiently small $h_{\mathcal{S}_n}$. Ritter [53] and [69] have proved that for the GP with isotropic Matérn $\sigma_0^2K_{\alpha_0,\nu}$ and no regression terms, the *optimal* convergence rate of squared L_2 kriging prediction error is $n^{-2\nu/d}$, which is also a lower bound and not improvable. This optimal rate $n^{-2\nu/d}$ can be attained when $\mathcal{S}_n$ has the quasi-uniform design, such as a regular grid in $\mathcal{S}$, such that $h_{\mathcal{S}_n}\asymp n^{-1/d}$; see Table 1 of [69]. If $C_{\mathrm{m}}\underline{\lambda}(M_n,p)^{-1}$ in Theorem 3.4 is of constant order, then Theorem 3.4 provides the upper bound of the order $h_{\mathcal{S}_n}^{2\nu/d}\asymp n^{-2\nu/d}$ for the Bayesian GP predictive variance $\mathrm{v}_n(s^*;\sigma^2,\alpha)$ with a quasi-uniform design $\mathcal{S}_n$, which matches up with the optimal rate of squared L_2 kriging prediction error.

The multiplicative factor $C_{\mathrm{m}}\underline{\lambda}(M_n,p)^{-1}$ in the upper bounds in Theorem 3.4 is due to the regression terms $\mathrm{m}(\cdot)^\top\beta$. The same factor also appears in the frequentist kriging error bound in Theorem 2 of [77] under a fixed covariance functions. By Assumption (A.1), C_{m} is already a constant. In many applications, the term $\underline{\lambda}(M_n,p)$ is bounded from below by constant for fixed p as $n\to\infty$, for example, when $\mathcal{S}_n$ is either some regular grid in $\mathcal{S}$ or drawn from some sampling distribution [77]. Then Theorem 3.4 leads to the optimal convergence rate for the posterior predictive variance with randomly drawn (σ^2,α).

In the special case of $p = 1$, $\mathrm{m}_1(\cdot) \equiv 1$, and Matérn with $d = 1$ and $\nu = 1/2$, [51] has shown the stronger frequentist asymptotic efficiency in linear prediction. Therefore, one can possibly establish the Bayesian posterior asymptotic efficiency similar to Theorem 3.2 for this special case. However, posterior asymptotic efficiency for the general universal kriging model (1) with $p > 1$ regression functions, a general smoothness parameter $\nu > 0$ and $d \in \{1, 2, 3\}$ is technically very challenging and likely to involve more demanding assumptions on the functions $\mathrm{m}_1(\cdot), \ldots, \mathrm{m}_p(\cdot)$ and the sampling design of $\mathcal{S}_n$. While we leave this general problem for future research, we provide some empirical evidence of this posterior asymptotic efficiency in the simulation study in Section S7 of the Supplementary Material [42].

Our results on convergence rates are not directly comparable with the previous literature on Bayesian Gaussian process regression, such as [70, 72, 73, 82], etc., since our model assumes a random sample path $Y(\cdot)$ from a GP instead of a deterministic true function, and our model does not contain the additional measurement error as in these works.

4. Simulation study. We verify our limiting theorems and posterior asymptotic efficiency using several numerical examples. In this section, we consider the 1 and 2-dimensional Ornstein–Uhlenbeck process with $\nu = 1/2$ in the isotropic Matérn covariance function without the regression terms $\mathrm{m}(\cdot)^\top \beta$. We provide additional simulation results for the model with regression terms $\mathrm{m}(\cdot)^\top \beta$ for $\nu = 1/2, 1/4, 3/2$ and dimension $d = 1, 2$ in Section S7 of the Supplementary Material [42].

In the model without regression terms, we have $Y(s) = X(s)$ for $s \in \mathcal{S}$, $d = 1, 2$ and $X(\cdot) \sim \mathrm{GP}(0, \sigma_0^2 K_{\alpha_0, 1/2})$. The main purpose is to verify Theorems 2.3 and 2.6. The true covariance parameters are $\sigma_0^2 = 2$, $\alpha_0 = 1$, and $\theta_0 = 2$. We assign independent gamma priors to θ and α, with the same shape parameter 1.1 and rate parameter 0.1. This prior satisfies Assumptions (A.2), (A.3) and the right tail condition (the second relation of (24)) in (A.4'), but does not satisfy the left tail condition (the first relation of (24)) in (A.4'); see Proposition 2.4. We will see that empirically this prior still yields convergent results.

We consider two cases with dimensions $d = 1$ and $d = 2$. For the $d = 1$ case, we set $\mathcal{S} = [0, 1]$ and the sampling points of $\mathcal{S}_n$ to be the grid $s_i = \frac{2i-1}{2n}$ $(i = 1, \ldots, n)$, for $n = 25, 50, 100, 200, 400$. For the $d = 2$ case, we set $\mathcal{S} = [0, 1]^2$ and the sampling points of $\mathcal{S}_n$ to be the regular grid $(\frac{2i-1}{2m}, \frac{2j-1}{2m})$ $(i, j = 1, \ldots, m)$, for $m = 10, 20, 30$ and $n = m^2$. Then we draw Y_n from the mean zero Gaussian process with the $\nu = 1/2$ Matérn covariance function observed on $\mathcal{S}_n$. We use the random walk Metropolis algorithm (RWM) to draw 5000 samples after 1000 burn-ins from the joint posterior $\Pi(\mathrm{d}\theta, \mathrm{d}\alpha|Y_n)$ and the limiting posterior $\mathcal{N}(\mathrm{d}\theta|\widetilde{\theta}_{\alpha_0}, 2\theta_0^2/n) \times \widetilde{\Pi}(\mathrm{d}\alpha|Y_n)$ in Theorem 2.3, respectively. For the $d = 1$ case, we further use RWM to draw 5000 samples from the limiting posterior $\mathcal{N}(\mathrm{d}\theta|\widetilde{\theta}_{\alpha_0}, 2\theta_0^2/n) \times \Pi_*(\mathrm{d}\alpha|Y_n)$ in Theorem 2.6.

We compare the true posterior distribution with the limiting posterior distributions using two criteria: (a) the closeness of our limiting distributions in Theorems 2.3 and 2.6 to the true posterior, and (b) the convergence of the two asymptotic efficiency measures in (32) with (θ, α) drawn from the joint posterior. For (a), since it is difficult to evaluate the total variation distance between two 2-dimensional posterior distributions based on finite posterior samples, we instead compute the Wasserstein-2 (W_2) distance between the marginal posteriors for θ and α, respectively. The W_2 distance between two 1-dimensional distributions F_1 and F_2 has the simple expression $W_2(F_1, F_2)^2 = \int_0^1 [F_1^{-1}(u) - F_2^{-1}(u)]^2 \,\mathrm{d}u$, where F_1^{-1} and F_2^{-1} are the corresponding quantile functions. With finite samples from F_1 and F_2, $W_2(F_1, F_2)$ can be accurately estimated by replacing F_1^{-1} and F_2^{-1} with the empirical quantile functions [43]. In our simulation study, we replace F_1 and F_2 with $\Pi(\mathrm{d}\theta|Y_n)$ and $\mathcal{N}(\mathrm{d}\theta|\widetilde{\theta}_{\alpha_0}, 2\theta_0^2/n)$ for θ, and $\Pi(\mathrm{d}\alpha|Y_n)$ and $\widetilde{\Pi}(\mathrm{d}\alpha|Y_n)$ for α, respectively. For the $d = 1$ case, we also compute the W_2 distance between $\Pi(\mathrm{d}\alpha|Y_n)$ and $\Pi_*(\mathrm{d}\alpha|Y_n)$. The convergence in W_2 distance is equivalent to

TABLE 1

Parameter estimation and Wasserstein-2 distances between the true posterior and the limiting posteriors in Theorems 2.3 and 2.6 for the model with $\nu = 1/2$, $d = 1$ and without regression terms. $\mathrm{E}(\cdot|Y_n)$, $\mathrm{Var}(\cdot|Y_n)$, $\widetilde{\mathrm{E}}(\cdot|Y_n)$, $\widetilde{\mathrm{Var}}(\cdot|Y_n)$, $\mathrm{E}_*(\cdot|Y_n)$, *and* $\mathrm{Var}_*(\cdot|Y_n)$ *are the posterior means and variances under the true posterior, the limiting posterior in Theorem 2.3, and the limiting posterior in Theorem 2.6. The true parameter values are* $\theta_0 = 2$ *and* $\alpha_0 = 1$. *All numbers are averaged over* 100 *macro replications. The standard errors are in the parentheses*

$d = 1$	$n = 25$	$n = 50$	$n = 100$	$n = 200$	$n = 400$
$\mathrm{E}(\theta\|Y_n)$	2.6795 (0.0763)	2.1932 (0.0434)	2.1467 (0.0269)	2.0740 (0.0202)	2.0320 (0.0139)
$\mathrm{Var}(\theta\|Y_n)$	0.9825 (0.0557)	0.2441 (0.0096)	0.1031 (0.0026)	0.0455 (0.0010)	0.0212 (0.0003)
$\widetilde{\mathrm{E}}(\theta\|Y_n)$	2.0404 (0.0592)	1.9357 (0.0391)	2.0214 (0.0193)	2.0130 (0.0251)	2.0028 (0.0136)
$\widetilde{\mathrm{Var}}(\theta\|Y_n)$	0.3197 (0.0007)	0.1599 (0.0003)	0.0798 (0.0002)	0.0399 (0.0001)	0.0200 (0.0000)
$\mathrm{E}(\alpha\|Y_n)$	3.1924 (0.2459)	2.9803 (0.2527)	2.7392 (0.2049)	2.9947 (0.2819)	2.5075 (0.2044)
$\mathrm{Var}(\alpha\|Y_n)$	5.3673 (0.8032)	4.0441 (0.6657)	2.9987 (0.4264)	3.7074 (0.6484)	2.5080 (0.3876)
$\widetilde{\mathrm{E}}(\alpha\|Y_n)$	2.9717 (0.2246)	2.8767 (0.2389)	2.6941 (0.2001)	2.9534 (0.2791)	2.5012 (0.2044)
$\widetilde{\mathrm{Var}}(\alpha\|Y_n)$	4.5474 (0.6732)	3.7045 (0.5762)	2.9094 (0.4093)	3.6840 (0.6396)	2.4664 (0.3818)
$\mathrm{E}_*(\alpha\|Y_n)$	2.5267 (0.1789)	2.6534 (0.2135)	2.5873 (0.1874)	2.9105 (0.2723)	2.4933 (0.2044)
$\mathrm{Var}_*(\alpha\|Y_n)$	2.5207 (0.3018)	2.7894 (0.3862)	2.5783 (0.3414)	3.3733 (0.5548)	2.4291 (0.3660)
$W_2(\Pi(\mathrm{d}\theta\|Y_n),$	0.8051	0.3000	0.1449	0.0706	0.0335
$\mathcal{N}(\mathrm{d}\theta\|\widetilde{\theta}_{\alpha_0}, \frac{2\theta_0^2}{n}))$	(0.0326)	(0.0101)	(0.0042)	(0.0024)	(0.0010)
$W_2(\Pi(\mathrm{d}\alpha\|Y_n),$	0.3175	0.1807	0.1260	0.1303	0.1073
$\widetilde{\Pi}(\mathrm{d}\alpha\|Y_n))$	(0.0290)	(0.0183)	(0.0086)	(0.0099)	(0.0077)
$W_2(\Pi(\mathrm{d}\alpha\|Y_n),$	0.8972	0.4259	0.2131	0.1583	0.1095
$\Pi_*(\mathrm{d}\alpha\|Y_n))$	(0.0874)	(0.0504)	(0.0211)	(0.0160)	(0.0075)

the weak convergence plus the convergence in the second moment [75]. Therefore, it provides useful empirical evidence for convergence in the posterior means and variances of θ and α. Theoretically, [11] has shown that the Wasserstein distance provides an upper bound for the total variation distance between two kernel smoothed densities from discrete draws.

For the $d = 1$ case, Table 1 reports the estimated posterior means under the true posterior $\Pi(\cdot|Y_n)$, the limiting posterior $\widetilde{\Pi}(\cdot|Y_n)$ in Theorem 2.3, the limiting posterior $\Pi_*(\cdot|Y_n)$ in Theorem 2.6, and the W_2 distances between the marginal posteriors. The posterior mean estimates of the microergodic θ are accurate for the true value $\theta_0 = 2$ and the posterior variance decreases as n increases. As expected, the posterior mean estimates of α are not consistent for the true $\alpha_0 = 1$, and show no sign of convergence for all three distributions. For the approximation accuracy, we can see that the W_2 distance between the true marginal posterior of θ and the normal limit in our theorem decreases quickly to zero as n increases. Furthermore, the W_2 distances between the true marginal posterior of α and the two approximations, the profile posterior $\widetilde{\Pi}(\mathrm{d}\alpha|Y_n)$ and the polynomially tilted normal distribution $\Pi_*(\mathrm{d}\alpha|Y_n)$ in Theorem 2.6 also show clear decreasing trends toward zero as n increases. These empirical observations have verified our limiting distributions in Theorems 2.3 and 2.6 for the 1-dimensional Ornstein–Uhlenbeck process.

For the $d = 2$ case, the results are summarized in Table 2, showing similar trends to those from the $d = 1$ case. The posterior mean estimates of θ are accurate with standard errors decreasing with n. The posterior mean estimates of α happen to be close to $\alpha_0 = 1$, though both the true posterior variance and the asymptotic posterior variance remain above 0.4 as n increases. The W_2 distance between the true marginal posteriors and the limiting posteriors in Theorem 2.3 converges to zero as n increases. This has verified the limiting distribution in Theorem 2.3 for the 2-dimensional process.

TABLE 2

Parameter estimation and Wasserstein-2 distances between the true posterior and the limiting posteriors in Theorem 2.3 for the model with $\nu = 1/2$, $d = 2$ and without regression terms. $\mathrm{E}(\cdot|Y_n)$, $\mathrm{Var}(\cdot|Y_n)$, $\widetilde{\mathrm{E}}(\cdot|Y_n)$ and $\widetilde{\mathrm{Var}}(\cdot|Y_n)$ are the posterior means and variances under the true posterior and the limiting posterior in Theorem 2.3. The true parameter values are $\theta_0 = 2$ and $\alpha_0 = 1$. All numbers are averaged over 100 macro replications. The standard errors are in the parentheses

$d = 2$	$n = 10^2$	$n = 20^2$	$n = 30^2$
$\mathrm{E}(\theta\|Y_n)$	2.0211 (0.0258)	2.0152 (0.0135)	1.9959 (0.0097)
$\mathrm{Var}(\theta\|Y_n)$	0.0835 (0.0022)	0.0203 (0.0003)	0.0089 (0.0001)
$\widetilde{\mathrm{E}}(\theta\|Y_n)$	2.0150 (0.0262)	2.0110 (0.0134)	1.9939 (0.0096)
$\widetilde{\mathrm{Var}}(\theta\|Y_n)$	0.0798 (0.0002)	0.0200 (0.0000)	0.0089 (0.0001)
$\mathrm{E}(\alpha\|Y_n)$	1.0936 (0.0479)	1.1317 (0.0456)	1.0909 (0.0397)
$\mathrm{Var}(\alpha\|Y_n)$	0.5054 (0.0392)	0.4864 (0.0352)	0.4500 (0.0266)
$\widetilde{\mathrm{E}}(\alpha\|Y_n)$	1.1094 (0.0486)	1.1392 (0.0459)	1.0941 (0.0397)
$\widetilde{\mathrm{Var}}(\alpha\|Y_n)$	0.5131 (0.0406)	0.4796 (0.0348)	0.4385 (0.0261)
$W_2(\Pi(\mathrm{d}\theta\|Y_n), \mathcal{N}(\mathrm{d}\theta\|\widetilde{\theta}_{\alpha_0}, \frac{2\theta_0^2}{n}))$	0.0652 (0.0024)	0.0185 (0.0008)	0.0090 (0.0003)
$W_2(\Pi(\mathrm{d}\alpha\|Y_n), \widetilde{\Pi}(\mathrm{d}\alpha\|Y_n))$	0.0547 (0.0030)	0.0514 (0.0024)	0.0505 (0.0021)

Figure 2 illustrates the convergence of posterior densities for the $d = 1$ case. With $n = 50$, there exists noticeable difference between the true posterior and the limiting posteriors. But their difference gradually disappears as n increases. Furthermore, as n increases, the posterior shrinks along the θ direction, but remains spread out in the α direction. The "ridge" of the joint posterior is the REML $\widetilde{\theta}_\alpha$, which increases with α as proved in Lemma 2.2, but becomes flatter as n increases, indicating the convergence from $\widetilde{\theta}_\alpha$ to $\theta_0 = 2$ over all values of α.

For the posterior asymptotic efficiency in (b), we compute the two asymptotic efficiency measures in (32) and Theorems 3.2 and 3.3 empirically, using the posterior samples of (θ, α). To approximate the supremums, we take the maximum of the ratios that depend on the random (σ^2, α) drawn from the posterior:

$$\mathrm{r}_{1n}(s^*) = \left| \frac{\mathrm{E}_{(\sigma^2,\alpha)}\{e_n(s^*;\alpha)^2\}}{\mathrm{E}_{(\sigma_0^2,\alpha_0)}\{e_n(s^*;\alpha)^2\}} - 1 \right| \quad \text{and} \quad \mathrm{r}_{2n}(s^*) = \left| \frac{\mathrm{E}_{(\sigma^2,\alpha)}\{e_n(s^*;\alpha)^2\}}{\mathrm{E}_{(\sigma_0^2,\alpha_0)}\{e_n(s^*;\alpha_0)^2\}} - 1 \right| \tag{37}$$

over a large number of testing points s^* from the Latin hypercube design. We use 1000 testing points in $\mathcal{S} = [0, 1]$ for the $d = 1$ case, and 2500 testing points in $\mathcal{S} = [0, 1]^2$ for the $d = 2$ case. Let the testing set be $\mathcal{S}^*$. We report the estimated posterior mean $\mathrm{E}[\max_{s^*\in\mathcal{S}^*} \mathrm{r}_{1n}(s^*)|Y_n]$ and $\mathrm{E}[\max_{s^*\in\mathcal{S}^*} \mathrm{r}_{2n}(s^*)|Y_n]$. The results are summarized in Table 3. The simulation results show that the posterior means of the two ratios in (37) decrease as n increases, and their standard errors also decrease. This is observed for both 1 and 2-dimensional domains.

5. Discussion Our theory has answered the two questions from the SST example in Section 1. For question (i), Theorems 2.3 and 2.6 in Section 2 show that the posterior of the microergodic parameter θ converges to a normal limit at the parametric rate, while the posterior of the range parameter α does not converge to any point mass in general. For question (ii), Theorems 3.1, 3.2, 3.3 and 3.4 in Section 3 show that the predictive performance based on the covariance parameters randomly drawn from their posterior distribution is asymptotically as good as the oracle predictive performance based on the true covariance parameters.

We discuss several future directions based on the current work. In many spatial applications, one may also add a measurement error term to the model, such that $Y(s_i) = \mathrm{m}(\cdot)^\top \beta + X(s_i) + \varepsilon(s_i)$ for $i = 1, \ldots, n$ with a noise process $\{\varepsilon(s) : s \in \mathcal{S}\}$ that is independent of X. Often it is assumed that $\varepsilon(s) \sim \mathcal{N}(0, \tau^2)$ for all $s \in \mathcal{S}$. The parameter τ^2 is

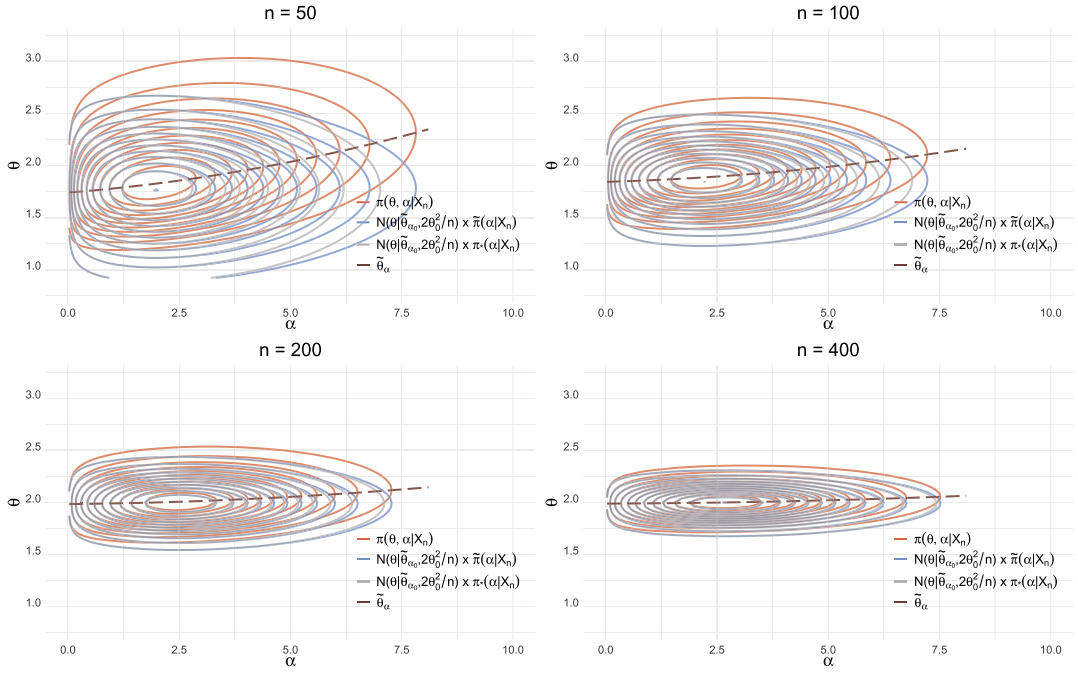

FIG. 2. *Contour plots of the true joint posterior density* $\pi(\theta, \alpha|Y_n)$ *(in red), the limiting posterior density* $\mathcal{N}(\theta|\widetilde{\theta}_{\alpha_0}, 2\theta_0^2/n) \times \widetilde{\pi}(\alpha|Y_n)$ *in Theorem 2.6, equation (25) (in blue), and the limiting posterior density* $\mathcal{N}(\theta|\widetilde{\theta}_{\alpha_0}, 2\theta_0^2/n) \times \pi_*(\alpha|Y_n)$ *in Theorem 2.6, equation (26) (in grey), for the 1-d Ornstein–Uhlenbeck process with sample size* $n = 50, 100, 200, 400$ *in the model without regression terms. The dashed line is the "ridge" REML* $\widetilde{\theta}_\alpha$ *given in (7). The true parameter values are* $\theta_0 = 2$ *and* $\alpha_0 = 1$.

the *nugget* parameter [17]. From the frequentist fixed-domain asymptotic theory, it is already known [61] that the presence of nugget parameter τ^2 will significantly change the convergence rate of the microergodic parameter θ, due to the convolution with Gaussian noise. For example, as shown in [14] for the 1-dimensional Ornstein–Uhlenbeck process ($\nu = 1/2$) on an equispaced grid, the convergence rate of the MLE of θ deteriorates from $n^{-1/2}$ to $n^{-1/4}$, though both θ and the nugget τ^2 can still be consistently estimated; see also the recent de-

TABLE 3

The posterior means of the two ratios of predictive MSEs defined in (37) maximized over 2500 testing points s^* *for the model with* $\nu = 1/2$ *and without regression terms, averaged over 100 macro replications. The standard errors are in the parentheses*

$d = 1$	$n = 25$	$n = 50$	$n = 100$	$n = 200$	$n = 400$
$\mathrm{E}[\max_{s^* \in \mathcal{S}^*} r_{1n}(s^*)\|Y_n]$	0.5129	0.2804	0.1796	0.1232	0.0823
	(0.0442)	(0.0197)	(0.0125)	(0.0082)	(0.0055)
$\mathrm{E}[\max_{s^* \in \mathcal{S}^*} r_{2n}(s^*)\|Y_n]$	0.4958	0.2626	0.1741	0.1188	0.0804
	(0.0447)	(0.0198)	(0.0126)	(0.0082)	(0.0055)
$d = 2$	$n = 10^2$	$n = 20^2$	$n = 30^2$		
$\mathrm{E}[\max_{s^* \in \mathcal{S}^*} r_{1n}(s^*)\|Y_n]$	0.1887	0.0736	0.0702		
	(0.0104)	(0.0051)	(0.0041)		
$\mathrm{E}[\max_{s^* \in \mathcal{S}^*} r_{2n}(s^*)\|Y_n]$	0.1827	0.0718	0.0705		
	(0.0101)	(0.0050)	(0.0041)		

velopment in [68]. Therefore, in the Bayesian setting, we expect that the limiting posterior distribution of (θ, α, τ^2) will be dramatically different from those in Theorems 2.3 and 2.6.

In the proof of Lemma 2.2 and Theorem 2.3, we have derived many useful properties of the spectral density of Matérn covariance functions. These derivations can be possibly extended to the tapered Matérn covariance functions [21, 76] and the generalized Wendland (GW) covariance functions [24], whose spectral densities also have polynomially decaying tails [7, 36]. As shown in Lemma 1 of [7], for the GP model with mean zero, the MLE of the GW microergodic parameter also has the monotonicity property. Therefore, with suitable modification, we expect that our technical proofs can be generalized to a broader class of covariance functions whose spectral densities share similar tail behavior to Matérn.

We have only considered the isotropic Matérn covariance functions. For anisotropic Matérn covariance functions, the existing fixed-domain asymptotic theory is very limited. Only a few special cases such as $\nu = 1/2$ [84], $\nu = 3/2$ [44] and $d > 4$ [1] have been studied, while the theory for the anisotropic Matérn with a general $\nu > 0$ and $d = 1, 2, 3$ remains unknown. We leave these directions for future research.

Acknowledgments The author sincerely thanks the Associate Editor and two anonymous referees for valuable comments that have significantly improved the paper. The author thanks Michael L. Stein, Wei-Liem Loh, Wenxin Jiang, Sanvesh Srivastava and Yichen Zhu for helpful discussion.

Funding The author was supported by the Singapore Ministry of Education Academic Research Funds Tier 1 Grants R-155-000-201-114 and A-0004822-00-00.

SUPPLEMENTARY MATERIAL

Supplementary Material to "Bayesian fixed-domain asymptotics for covariance parameters in a Gaussian process model" (DOI: 10.1214/22-AOS2230SUPPA; .pdf). Technical proofs of all the theorems, lemmas, propositions and corollaries in the main text, as well as additional simulation results for the universal kriging model with regression terms.

Codes for "Bayesian fixed-domain asymptotics for covariance parameters in a Gaussian process model" (DOI: 10.1214/22-AOS2230SUPPB; .zip). A zipped folder of R codes for the simulation studies in both the main text and the Supplementary Material.

REFERENCES

[1] ANDERES, E. (2010). On the consistent separation of scale and variance for Gaussian random fields. *Ann. Statist.* **38** 870–893. MR2604700 https://doi.org/10.1214/09-AOS725

[2] ARNOLD, B. C. (2015). *Pareto Distributions*, 2nd ed. *Monographs on Statistics and Applied Probability* **140**. CRC Press, Boca Raton, FL. MR3618736

[3] BACHOC, F., BEVILACQUA, M. and VELANDIA, D. (2019). Composite likelihood estimation for a Gaussian process under fixed domain asymptotics. *J. Multivariate Anal.* **174** 104534. MR3985049 https://doi.org/10.1016/j.jmva.2019.104534

[4] BACHOC, F. and LAGNOUX, A. (2020). Fixed-domain asymptotic properties of maximum composite likelihood estimators for Gaussian processes. *J. Statist. Plann. Inference* **209** 62–75. MR4096254 https://doi.org/10.1016/j.jspi.2020.02.008

[5] BANERJEE, S., GELFAND, A. E., FINLEY, A. O. and SANG, H. (2008). Gaussian predictive process models for large spatial data sets. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 825–848. MR2523906 https://doi.org/10.1111/j.1467-9868.2008.00663.x

[6] BERGER, J. O., DE OLIVEIRA, V. and SANSÓ, B. (2001). Objective Bayesian analysis of spatially correlated data. *J. Amer. Statist. Assoc.* **96** 1361–1374. MR1946582 https://doi.org/10.1198/016214501753382282

[7] BEVILACQUA, M., FAOUZI, T., FURRER, R. and PORCU, E. (2019). Estimation and prediction using generalized Wendland covariance functions under fixed domain asymptotics. *Ann. Statist.* **47** 828–856. MR3909952 https://doi.org/10.1214/17-AOS1652

[8] BICKEL, P. J. and KLEIJN, B. J. K. (2012). The semiparametric Bernstein–von Mises theorem. *Ann. Statist.* **40** 206–237. MR3013185 https://doi.org/10.1214/11-AOS921

[9] BOCHKINA, N. A. and GREEN, P. J. (2014). The Bernstein–von Mises theorem and nonregular models. *Ann. Statist.* **42** 1850–1878. MR3262470 https://doi.org/10.1214/14-AOS1239

[10] BRAZAUSKAS, V. (2002). Fisher information matrix for the Feller–Pareto distribution. *Statist. Probab. Lett.* **59** 159–167. MR1927527 https://doi.org/10.1016/S0167-7152(02)00143-8

[11] CHAE, M. and WALKER, S. G. (2020). Wasserstein upper bounds of the total variation for smooth densities. *Statist. Probab. Lett.* **163** 108771. MR4083852 https://doi.org/10.1016/j.spl.2020.108771

[12] CHANG, C.-H., HUANG, H.-C. and ING, C.-K. (2014). Asymptotic theory of generalized information criterion for geostatistical regression model selection. *Ann. Statist.* **42** 2441–2468. MR3269985 https://doi.org/10.1214/14-AOS1258

[13] CHANG, C.-H., HUANG, H.-C. and ING, C.-K. (2017). Mixed domain asymptotics for a stochastic process model with time trend and measurement error. *Bernoulli* **23** 159–190. MR3556770 https://doi.org/10.3150/15-BEJ740

[14] CHEN, H.-S., SIMPSON, D. G. and YING, Z. (2000). Infill asymptotics for a stochastic process model with measurement error. *Statist. Sinica* **10** 141–156. MR1742105

[15] CHEN, X., CHRISTENSEN, T. M. and TAMER, E. (2018). Monte Carlo confidence sets for identified sets. *Econometrica* **86** 1965–2018. MR3901244 https://doi.org/10.3982/ECTA14525

[16] CHERNOZHUKOV, V. and HONG, H. (2004). Likelihood estimation and inference in a class of nonregular econometric models. *Econometrica* **72** 1445–1480. MR2077489 https://doi.org/10.1111/j.1468-0262.2004.00540.x

[17] CRESSIE, N. A. C. (1993). *Statistics for Spatial Data*. *Wiley Series in Probability and Mathematical Statistics*: *Applied Probability and Statistics*. Wiley, New York. MR1239641 https://doi.org/10.1002/9781119115151

[18] CROWDER, M. J. (1976). Maximum likelihood estimation for dependent observations. *J. Roy. Statist. Soc. Ser. B* **38** 45–53. MR0403035 https://doi.org/10.1111/j.2517-6161.1976.tb01565.x

[19] DATTA, A., BANERJEE, S., FINLEY, A. O. and GELFAND, A. E. (2016). Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. *J. Amer. Statist. Assoc.* **111** 800–812. MR3538706 https://doi.org/10.1080/01621459.2015.1044091

[20] DE OLIVEIRA, V., KEDEM, B. and SHORT, D. A. (1997). Bayesian prediction of transformed Gaussian random fields. *J. Amer. Statist. Assoc.* **92** 1422–1433. MR1615252 https://doi.org/10.2307/2965412

[21] DU, J., ZHANG, H. and MANDREKAR, V. S. (2009). Fixed-domain asymptotic properties of tapered maximum likelihood estimators. *Ann. Statist.* **37** 3330–3361. MR2549562 https://doi.org/10.1214/08-AOS676

[22] FUGLSTAD, G.-A., SIMPSON, D., LINDGREN, F. and RUE, H. (2019). Constructing priors that penalize the complexity of Gaussian random fields. *J. Amer. Statist. Assoc.* **114** 445–452. MR3941267 https://doi.org/10.1080/01621459.2017.1415907

[23] GHOSAL, S. and VAN DER VAART, A. (2017). *Fundamentals of Nonparametric Bayesian Inference*. *Cambridge Series in Statistical and Probabilistic Mathematics* **44**. Cambridge Univ. Press, Cambridge. MR3587782 https://doi.org/10.1017/9781139029834

[24] GNEITING, T. (2002). Compactly supported correlation functions. *J. Multivariate Anal.* **83** 493–508. MR1945966 https://doi.org/10.1006/jmva.2001.2056

[25] GU, M. and ANDERSON, K. (2018). Calibration of imperfect mathematical models by multiple sources of data with measurement bias. ArXiv Preprint. Available at arXiv:1810.11664.

[26] GU, M., WANG, X. and BERGER, J. O. (2018). Robust Gaussian stochastic process emulation. *Ann. Statist.* **46** 3038–3066. MR3851764 https://doi.org/10.1214/17-AOS1648

[27] GUHANIYOGI, R., LI, C., SAVITSKY, T. D. and SRIVASTAVA, S. (2022). Distributed Bayesian inference in massive spatial data. *Statist. Sci.* To appear.

[28] GUSTAFSON, P. (2014). Bayesian inference in partially identified models: Is the shape of the posterior distribution useful? *Electron. J. Stat.* **8** 476–496. MR3205730 https://doi.org/10.1214/14-EJS891

[29] GUSTAFSON, P. (2015). *Bayesian Inference for Partially Identified Models*: *Exploring the Limits of Limited Data*. *Monographs on Statistics and Applied Probability* **141**. CRC Press, Boca Raton, FL. MR3642458

[30] HANDCOCK, M. S. and STEIN, M. L. (1993). A Bayesian analysis of kriging. *Technometrics* **35** 403–410. https://doi.org/10.1080/00401706.1993.10485354

[31] HEATON, J. H., DATTA, A., FINLEY, A. O., FURRER, R., GUINNESS, J., GUHANIYOGI, R., GERBER, F., GRAMACY, R. B., HAMMERLING, D. et al. (2019). A case study competition among methods for analyzing large spatial data. *J. Agric. Biol. Environ. Stat.* **24** 398–825. https://doi.org/10.1007/s13253-018-00348-w

[32] JIANG, W. (2017). On limiting distribution of quasi-posteriors under partial identification. *Econom. Stat.* **3** 60–72. MR3666242 https://doi.org/10.1016/j.ecosta.2017.03.006

[33] JIANG, W. and LI, C. (2019). On Bayesian oracle properties. *Bayesian Anal.* **14** 235–260. MR3910045 https://doi.org/10.1214/18-BA1097

[34] JUN, S. J., PINKSE, J. and WAN, Y. (2015). Classical Laplace estimation for $\sqrt[3]{n}$-consistent estimators: Improved convergence rates and rate-adaptive inference. *J. Econometrics* **187** 201–216. MR3347303 https://doi.org/10.1016/j.jeconom.2015.01.005

[35] KANAGAWA, M., HENNIG, P., SEJDINOVIC, D. and SRIPERUMBUDUR, B. K. (2018). Gaussian processes and kernel methods: A review on connections and equivalences. ArXiv Preprint. Available at arXiv:1807.02582.

[36] KAUFMAN, C. G., SCHERVISH, M. J. and NYCHKA, D. W. (2008). Covariance tapering for likelihood-based estimation in large spatial data sets. *J. Amer. Statist. Assoc.* **103** 1545–1555. MR2504203 https://doi.org/10.1198/016214508000000959

[37] KAUFMAN, C. G. and SHABY, B. A. (2013). The role of the range parameter for estimation and prediction in geostatistics. *Biometrika* **100** 473–484. MR3068447 https://doi.org/10.1093/biomet/ass079

[38] KENNEDY, M. C. and O'HAGAN, A. (2001). Bayesian calibration of computer models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **63** 425–464. MR1858398 https://doi.org/10.1111/1467-9868.00294

[39] KLEIJN, B. J. K. and KNAPIK, B. (2012). Semiparametric posterior limits under local asymptotic exponentiality. ArXiv Preprint. Available at arXiv:1210.6204v3.

[40] KREH, M. (2012). *Bessel Functions. Lecture Notes, Penn State—Göttingen Summer School on Number Theory*.

[41] LEHMANN, E. L. and CASELLA, G. (1998). *Theory of Point Estimation*, 2nd ed. *Springer Texts in Statistics*. Springer, New York. MR1639875 https://doi.org/10.1007/b98854

[42] LI, C. (2022). Supplement to "Bayesian fixed-domain asymptotics for covariance parameters in a Gaussian process model." https://doi.org/10.1214/22-AOS2230SUPPA, https://doi.org/10.1214/22-AOS2230SUPPB

[43] LI, C., SRIVASTAVA, S. and DUNSON, D. B. (2017). Simple, scalable and accurate posterior interval estimation. *Biometrika* **104** 665–680. MR3694589 https://doi.org/10.1093/biomet/asx033

[44] LOH, W.-L. (2005). Fixed-domain asymptotics for a subclass of Matérn-type Gaussian random fields. *Ann. Statist.* **33** 2344–2394. MR2211089 https://doi.org/10.1214/009053605000000516

[45] LOH, W.-L. (2015). Estimating the smoothness of a Gaussian random field from irregularly spaced data via higher-order quadratic variations. *Ann. Statist.* **43** 2766–2794. MR3405611 https://doi.org/10.1214/15-AOS1365

[46] LOH, W.-L., SUN, S. and WEN, J. (2021). On fixed-domain asymptotics, parameter estimation and isotropic Gaussian random fields with Matérn covariance functions. *Ann. Statist.* **49** 3127–3152. MR4352525 https://doi.org/10.1214/21-aos2077

[47] MANSKI, C. F. (2003). *Partial Identification of Probability Distributions. Springer Series in Statistics*. Springer, New York. MR2151380 https://doi.org/10.1007/b97478

[48] MARDIA, K. V. and MARSHALL, R. J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika* **71** 135–146. MR0738334 https://doi.org/10.1093/biomet/71.1.135

[49] MOON, H. R. and SCHORFHEIDE, F. (2012). Bayesian and frequentist inference in partially identified models. *Econometrica* **80** 755–782. MR2951948 https://doi.org/10.3982/ECTA8360

[50] PERUZZI, M., BANERJEE, S. and FINLEY, A. O. (2022). Highly scalable Bayesian geostatistical modeling via meshed Gaussian processes on partitioned domains. *J. Amer. Statist. Assoc.* **117** 969–982. MR4436326 https://doi.org/10.1080/01621459.2020.1833889

[51] PUTTER, H. and YOUNG, G. A. (2001). On the effect of covariance function estimation on the accuracy of kriging predictors. *Bernoulli* **7** 421–438. MR1836738 https://doi.org/10.2307/3318494

[52] RASMUSSEN, C. E. and WILLIAMS, C. K. I. (2006). *Gaussian Processes for Machine Learning. Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR2514435

[53] RITTER, K. (2000). *Average-Case Analysis of Numerical Problems. Lecture Notes in Math.* **1733**. Springer, Berlin. MR1763973 https://doi.org/10.1007/BFb0103934

[54] SANG, H. and HUANG, J. Z. (2012). A full scale approximation of covariance functions for large spatial data sets. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** 111–132. MR2885842 https://doi.org/10.1111/j.1467-9868.2011.01007.x

[55] SANG, H., JUN, M. and HUANG, J. Z. (2011). Covariance approximation for large multivariate spatial data sets with an application to multiple climate model errors. *Ann. Appl. Stat.* **5** 2519–2548. MR2907125 https://doi.org/10.1214/11-AOAS478

[56] SHABY, B. and RUPPERT, D. (2012). Tapered covariance: Bayesian estimation and asymptotics. *J. Comput. Graph. Statist.* **21** 433–452. MR2945475 https://doi.org/10.1080/10618600.2012.680819

[57] SHEN, X. (2002). Asymptotic normality of semiparametric and nonparametric posterior distributions. *J. Amer. Statist. Assoc.* **97** 222–235. MR1947282 https://doi.org/10.1198/016214502753479365

[58] STEIN, M. (1990). Uniform asymptotic optimality of linear predictions of a random field using an incorrect second-order structure. *Ann. Statist.* **18** 850–872. MR1056340 https://doi.org/10.1214/aos/1176347629

[59] STEIN, M. L. (1988). Asymptotically efficient prediction of a random field with a misspecified covariance function. *Ann. Statist.* **16** 55–63. MR0924856 https://doi.org/10.1214/aos/1176350690

[60] STEIN, M. L. (1990). Bounds on the efficiency of linear predictions using an incorrect covariance function. *Ann. Statist.* **18** 1116–1138. MR1062701 https://doi.org/10.1214/aos/1176347742

[61] STEIN, M. L. (1990). A comparison of generalized cross validation and modified maximum likelihood for estimating the parameters of a stochastic process. *Ann. Statist.* **18** 1139–1157. MR1062702 https://doi.org/10.1214/aos/1176347743

[62] STEIN, M. L. (1993). A simple condition for asymptotic optimality of linear predictions of random fields. *Statist. Probab. Lett.* **17** 399–404. MR1237787 https://doi.org/10.1016/0167-7152(93)90261-G

[63] STEIN, M. L. (1997). Efficiency of linear predictors for periodic processes using an incorrect covariance function. *J. Statist. Plann. Inference* **58** 321–331. MR1450019 https://doi.org/10.1016/S0378-3758(96)00088-2

[64] STEIN, M. L. (1999). *Interpolation of Spatial Data*: *Some Theory for Kriging*. *Springer Series in Statistics*. Springer, New York. MR1697409 https://doi.org/10.1007/978-1-4612-1494-6

[65] STEIN, M. L. (1999). Predicting random fields with increasing dense observations. *Ann. Appl. Probab.* **9** 242–273. MR1682572 https://doi.org/10.1214/aoap/1029962604

[66] SUN, Q., MIAO, C., DUAN, Q., ASHOURI, H., SOROOSHIAN, S. and HSU, K. L. (2018). A review of global precipitation data sets: Data sources, estimation, and intercomparisons. *Rev. Geophys.* **56** 79–107. https://doi.org/10.1002/2017RG000574

[67] TAMER, E. (2010). Partial identification in econometrics. *Ann. Rev. Econ.* **3** 167–195. https://doi.org/10.1146/annurev.economics.050708.143401

[68] TANG, W., ZHANG, L. and BANERJEE, S. (2021). On identifiability and consistency of the nugget in Gaussian spatial process models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **83** 1044–1070. MR4349127 https://doi.org/10.1111/rssb.12472

[69] TUO, R. and WANG, W. (2020). Kriging prediction with isotropic Matérn correlations: Robustness and experimental designs. *J. Mach. Learn. Res.* **21** Paper No. 187. MR4209473

[70] VAN DER VAART, A. and VAN ZANTEN, H. (2011). Information rates of nonparametric Gaussian process methods. *J. Mach. Learn. Res.* **12** 2095–2119. MR2819028

[71] VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247 https://doi.org/10.1017/CBO9780511802256

[72] VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2008). Rates of contraction of posterior distributions based on Gaussian process priors. *Ann. Statist.* **36** 1435–1463. MR2418663 https://doi.org/10.1214/009053607000000613

[73] VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2009). Adaptive Bayesian estimation using a Gaussian random field with inverse gamma bandwidth. *Ann. Statist.* **37** 2655–2675. MR2541442 https://doi.org/10.1214/08-AOS678

[74] VELANDIA, D., BACHOC, F., BEVILACQUA, M., GENDRE, X. and LOUBES, J.-M. (2017). Maximum likelihood estimation for a bivariate Gaussian process under fixed domain asymptotics. *Electron. J. Stat.* **11** 2978–3007. MR3694574 https://doi.org/10.1214/17-EJS1298

[75] VILLANI, C. (2009). *Optimal Transport*: *Old and New*. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **338**. Springer, Berlin. MR2459454 https://doi.org/10.1007/978-3-540-71050-9

[76] WANG, D. and LOH, W.-L. (2011). On fixed-domain asymptotics and covariance tapering in Gaussian random field models. *Electron. J. Stat.* **5** 238–269. MR2792553 https://doi.org/10.1214/11-EJS607

[77] WANG, W., TUO, R. and WU, C. F. J. (2020). On prediction properties of kriging: Uniform error bounds and robustness. *J. Amer. Statist. Assoc.* **115** 920–930. MR4107689 https://doi.org/10.1080/01621459.2019.1598868

[78] WENDLAND, H. (2005). *Scattered Data Approximation*. *Cambridge Monographs on Applied and Computational Mathematics* **17**. Cambridge Univ. Press, Cambridge. MR2131724 https://doi.org/10.1017/CBO9780511617539

[79] WU, Z. M. and SCHABACK, R. (1993). Local error estimates for radial basis function interpolation of scattered data. *IMA J. Numer. Anal.* **13** 13–27. MR1199027 https://doi.org/10.1093/imanum/13.1.13

[80] WYNNE, G., BRIOL, F. X. and GIROLAMI, M. (2021). Convergence guarantees for Gaussian process approximations under several observation models. *J. Mach. Learn. Res.* **123** 1–40.

[81] YAKOWITZ, S. J. and SZIDAROVSZKY, F. (1985). A comparison of kriging with nonparametric regression methods. *J. Multivariate Anal.* **16** 21–53. MR0778488 https://doi.org/10.1016/0047-259X(85)90050-8

[82] YANG, Y. and TOKDAR, S. T. (2015). Minimax-optimal nonparametric regression in high dimensions. *Ann. Statist.* **43** 652–674. MR3319139 https://doi.org/10.1214/14-AOS1289

[83] YING, Z. (1991). Asymptotic properties of a maximum likelihood estimator with data from a Gaussian process. *J. Multivariate Anal.* **36** 280–296. MR1096671 https://doi.org/10.1016/0047-259X(91)90062-7

[84] YING, Z. (1993). Maximum likelihood estimation of parameters under a spatial sampling scheme. *Ann. Statist.* **21** 1567–1590. MR1241279 https://doi.org/10.1214/aos/1176349272

[85] ZHANG, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *J. Amer. Statist. Assoc.* **99** 250–261. MR2054303 https://doi.org/10.1198/016214504000000241

[86] ZHANG, H. and ZIMMERMAN, D. L. (2005). Towards reconciling two asymptotic frameworks in spatial statistics. *Biometrika* **92** 921–936. MR2234195 https://doi.org/10.1093/biomet/92.4.921