Supplementary Materials: Beyond Matérn: On A Class of Interpretable Confluent Hypergeometric Covariance Functions

Pulong Ma*
School of Mathematical and Statistical Sciences, Clemson University
and
Anindya Bhadra
Department of Statistics, Purdue University

This supplement contains seven sections. Section S.1 gives an illustration to compare the timing to evaluate Bessel function and confluent hypergeometric function. Section S.2 shows 1-dimensional process realizations under different parameter values for the Matérn class and the CH class. Section S.3 shows some theoretical results that are used to prove main theorems in the main text. Section S.4 contains technical proofs that are omitted in the main text. Section S.5 contains simulation results that verify the theoretical results in Section 3. Section S.6 contains additional simulation results referenced in Section 4. Section S.7 contains parameter estimation results and figures referenced in Section 5.

*Address of Correspondence: Pulong Ma, School of Mathematical and Statistical Sciences, Clemson University, 220 Parkway Drive, Clemson, SC 29634. Email: plma@clemson.edu.

S.1 Illustration of Timing to Evaluate Bessel Function and Confluent Hypergeometric Function

When the confluent hypergeometric function is evaluated by calling the GNU scientific library GSL via `Rcpp` package and the modified Bessel function of second kind is evaluated with the `base` package in R, with 10000 times repeated evaluations, on average, the confluent hypergeometric function takes about 10.7 microseconds for each evaluation, while the Bessel function takes about 7.8 microseconds for each evaluation. The timings are recorded using the R package `microbenchmark` with results shown in Figure S.1.

S.2 1-D Process Realizations

In Figure S.2, we show the realizations from zero mean Gaussian processes with the CH class and the Matérn class under different parameter settings. When the distance is within the effective range, the Matérn covariance function results in more large correlations than the CH covariance function. This makes the process realizations from the Matérn class smoother even though the smoothness parameter is fixed at the same value for both the Matérn class and the CH class. For the CH class, if α has a smaller value, the corresponding correlation function has more small values within the effective range. This makes the process realizations under the CH class look rougher. As we expect, when the effective range and the tail decay parameter are fixed, the process realizations under the CH class look smoother for a larger value of the smoothness parameter.

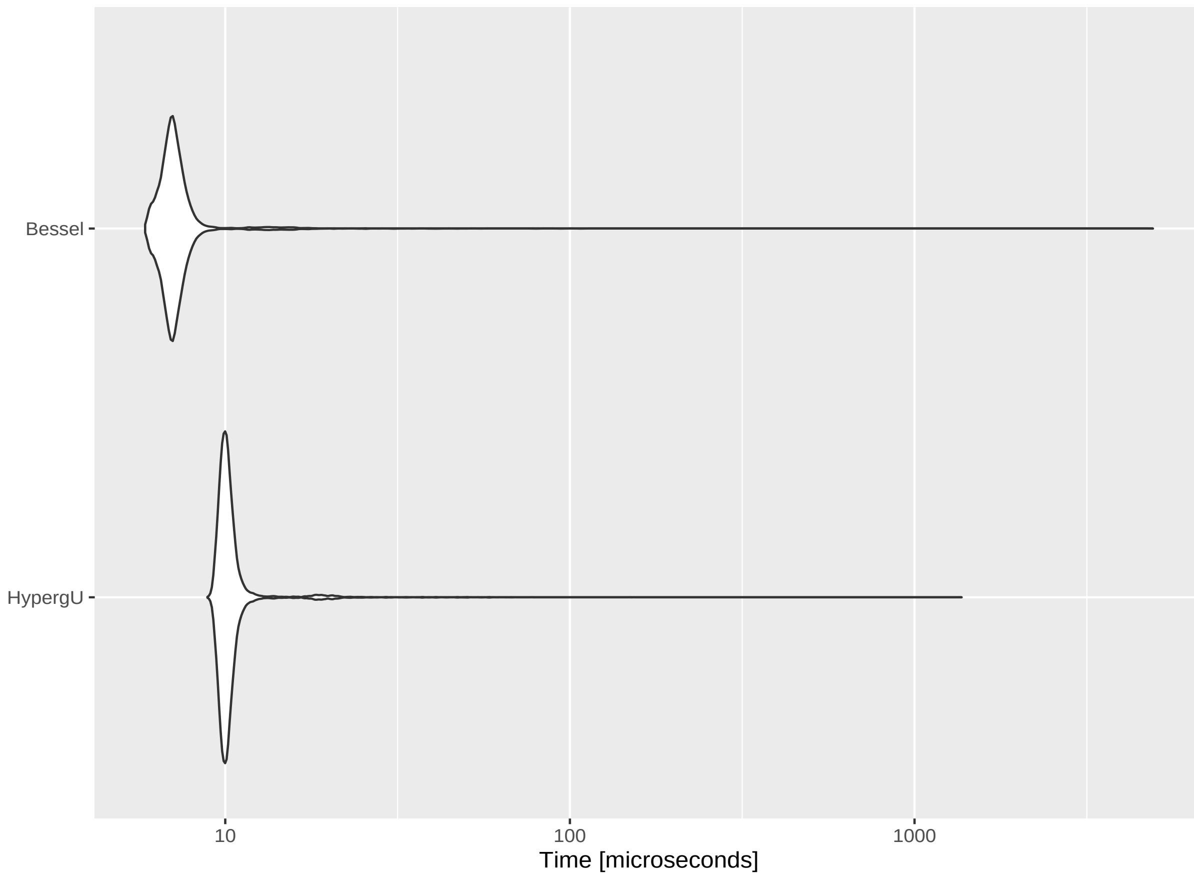
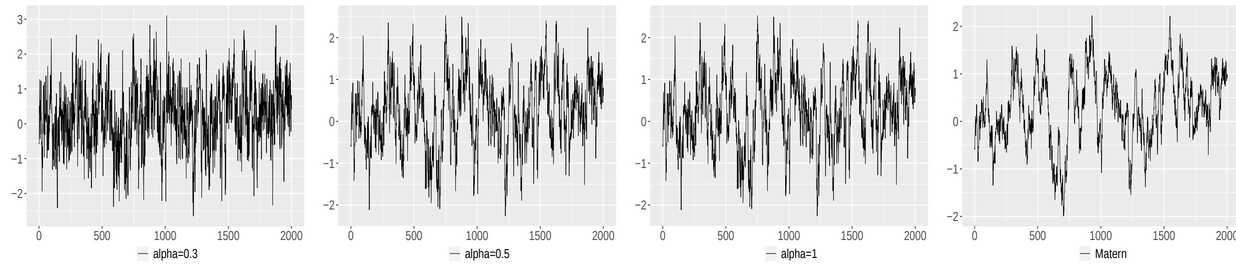
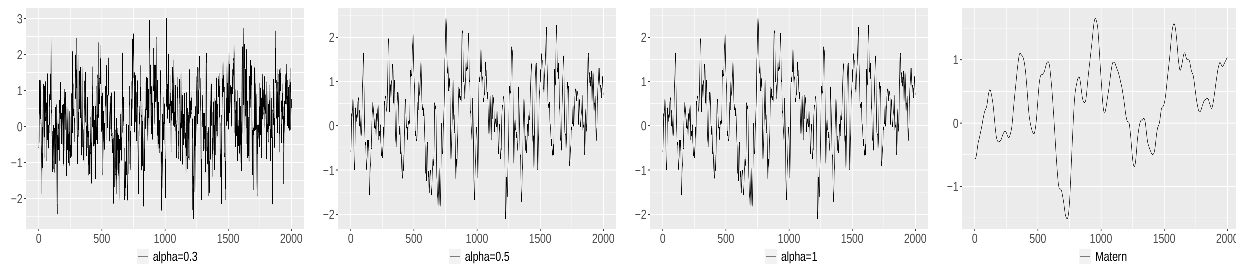


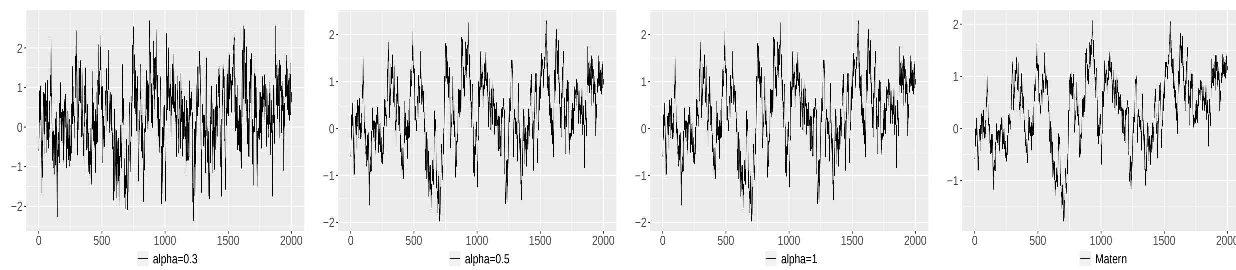
Fig. S.1. Benchmark of the computing time to evaluate the confluent hypergeometric function and the Bessel function in R. "Bessel" refers to the timing for evaluating the modified Bessel function of the second kind and "HypergU" refers to the timing for evaluating the confluent hypergeometric function of the second kind.



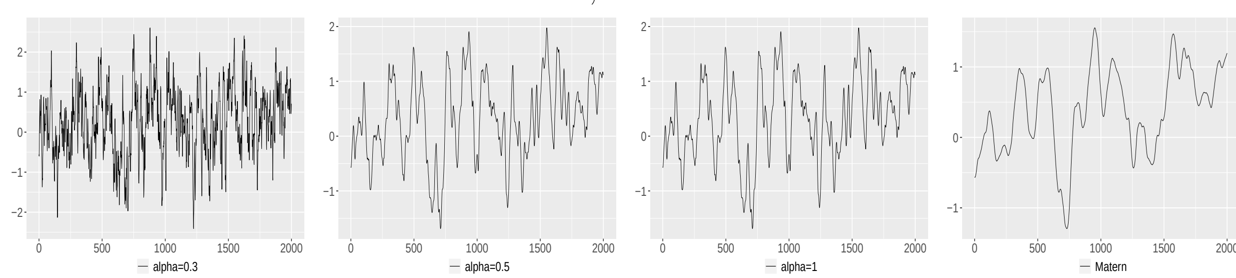
$\nu = 0.5, ER = 200$



$\nu = 2.5, ER = 200$



$\nu = 0.5, ER = 500$



$\nu = 2.5, ER = 500$

Fig. S.2. Realizations over 2000 regular grid points in the domain $[0, 2000]$ from zero mean Gaussian processes with the CH covariance model and the Matérn covariance model under different parameter settings. The realizations from the CH covariance are shown in the first three columns and those from the Matérn covariance are shown in the last column. For the first two rows, the effective range (ER) is fixed at 200. For the last two rows, the effective range is fixed at 500. ER is defined as the distance at which correlation is approximately 0.05.

S.3 Ancillary Results

To show the asymptotic behavior of the MLE of the microergodic parameter, we need some results in terms of spectral densities of covariance functions. More precisely, the tail behavior of the spectral densities can be used to check the equivalence of probability measures generated by stationary Gaussian random fields. Equivalence of Gaussian measures defined by Gaussian processes has been studied in probability and statistics with sufficient conditions given in Theorem 17 of Chapter III of Ibragimov and Rozanov (1978) for $d = 1$ and given on page 156 of Yadrenko (1983) and page 120 of Stein (1999) for $d > 1$. In particular, the following sufficient conditions can be used to check the equivalence of Gaussian probability measures defined by covariance functions. If for some $\lambda > 0$ and for some finite $c \in \mathbb{R}$, one has

$$0 < f_1(\boldsymbol{\omega})|\boldsymbol{\omega}|^{\lambda} < \infty \quad \text{as} \quad |\boldsymbol{\omega}| \to \infty, \quad \text{and} \tag{S.1}$$

$$\int_{|\boldsymbol{\omega}|>c} \left\{ \frac{f_1(\boldsymbol{\omega}) - f_2(\boldsymbol{\omega})}{f_1(\boldsymbol{\omega})} \right\}^2 d\boldsymbol{\omega} < \infty, \tag{S.2}$$

then the two corresponding Gaussian measures $\mathcal{P}_1$ and $\mathcal{P}_2$ are equivalent. For isotropic Gaussian random fields, the condition (S.2) can be expressed as

$$\int_c^{\infty} \omega^{d-1} \left\{ \frac{f_1(\omega) - f_2(\omega)}{f_1(\omega)} \right\}^2 d\omega < \infty, \tag{S.3}$$

where $\omega := |\boldsymbol{\omega}|$ with $|\cdot|$ denoting the Euclidean norm. The detailed discussion on equivalence of Gaussian measures and the condition for equivalence can be found in Chapter 4 of Stein (1999) and references (e.g., Stein, 1988, 1993; Stein and Handcock, 1989).

In what follows, we will introduce a few useful lemmas. Lemma 1 is used to diagonalize two covariance matrices and it is needed in Lemma 2, which gives an important result on the behavior of eigenvalues of a correlation matrix constructed from the CH correlation function.

Lemma 1. *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be two* $n \times n$ *symmetric positive definite matrices. Then there exists a non-singular matrix* $\mathbf{U}$ *such that* $\mathbf{U}^\top \mathbf{A} \mathbf{U} = \mathbf{I}_{n\times n}$ *and* $\mathbf{U}^\top \mathbf{B} \mathbf{U} = \mathbf{D}$*, where* $\mathbf{D}$ *is an* $n \times n$ *diagonal matrix with positive diagonal entries.*

Proof. For the symmetric matrix $\mathbf{A}$, it follows from the Schur Decomposition Theorem (e.g., Magnus and Neudecker, 1999, p. 17) that there exists an orthogonal $n \times n$ matrix $\mathbf{S}$ consisting of eigenvectors of $\mathbf{A}$ and a diagonal matrix $\mathbf{\Lambda} := \text{diag}\{\lambda_1, \ldots, \lambda_n\}$ such that $\mathbf{S}^\top \mathbf{A} \mathbf{S} = \mathbf{\Lambda}$. Since $\mathbf{A}$ is positive definite, the diagonal entries of $\mathbf{\Lambda}$ are all positive. Let $\mathbf{\Lambda}^{1/2} := \text{diag}\{\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n}\}$ be the "square root" of $\mathbf{\Lambda}$. Then we call $\mathbf{A}^{1/2} := \mathbf{\Lambda}^{1/2}\mathbf{S}$ a square root of $\mathbf{A}$ satisfying $\mathbf{A} = (\mathbf{A}^\top)^{1/2}\mathbf{A}^{1/2}$. As the matrix $\mathbf{A}^{1/2}$ is invertible, the symmetric matrix $(\mathbf{A}^\top)^{-1/2}\mathbf{B}\mathbf{A}^{-1/2}$ is well-defined. Note that $(\mathbf{A}^\top)^{-1/2}\mathbf{B}\mathbf{A}^{-1/2}$ is positive definite since for all $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{x}^\top(\mathbf{A}^\top)^{-1/2}\mathbf{B}\mathbf{A}^{-1/2}\mathbf{x} = \|\mathbf{B}^{1/2}\mathbf{A}^{-1/2}\mathbf{x}\|^2 \geq 0$ with the inequality becoming an equality only if $\mathbf{x} = \mathbf{0}$. Hence $(\mathbf{A}^\top)^{-1/2}\mathbf{B}\mathbf{A}^{-1/2}$ is also a symmetric and positive definite matrix. According to the Schur Decomposition Theorem, there exists an orthogonal matrix $\mathbf{O}$ and diagonal matrix $\mathbf{D}$ with positive diagonal entries such that $\mathbf{O}^\top(\mathbf{A}^\top)^{-1/2}\mathbf{B}\mathbf{A}^{-1/2}\mathbf{O} = \mathbf{D}$. Now we define the non-singular matrix $\mathbf{U} := \mathbf{A}^{-1/2}\mathbf{O}$, which satisfies $\mathbf{U}^\top \mathbf{A} \mathbf{U} = \mathbf{I}_{n\times n}$, as to be established. $\square$

Lemma 2. *Suppose that* $\nu > 0$ *is fixed. Given a set of* n *observation locations in a bounded domain* $\mathcal{D}$*, let* $\sigma_0^2\mathbf{R}_n(\boldsymbol{\theta}_0)$ *be the* $n\times n$ *covariance matrix defined by the CH covariance function* $C(h, \nu, \alpha_0, \beta_0, \sigma_0^2)$ *with* $\boldsymbol{\theta}_0 := \{\alpha_0, \beta_0\}$ *and* $\sigma^2\mathbf{R}_n(\boldsymbol{\theta})$ *be the* $n \times n$ *covariance matrix defined by the CH covariance function* $C(h, \nu, \alpha, \beta, \sigma^2)$ *with* $\boldsymbol{\theta} := \{\alpha, \beta\}$*. Assume that* $\alpha_0, \alpha > d/2$*. Let* $\mathbf{\Lambda} := \mathit{diag}\{\lambda_{1,n}, \ldots, \lambda_{n,n}\}$ *be an* $n \times n$ *diagonal matrix with diagonal elements* $\lambda_{k,n} > 0$ *for* $k = 1, \ldots, n$ *such that* $\mathbf{U}^\top\sigma_0^2\mathbf{R}_n(\boldsymbol{\theta}_0)\mathbf{U} = \mathbf{I}_n$ *and* $\mathbf{U}^\top\sigma^2\mathbf{R}_n(\boldsymbol{\theta})\mathbf{U} = \mathbf{\Lambda}$ *for some non-singular* *matrix* $\mathbf{U}$. *Then it can be established that for any* $\epsilon > 0$, *as* $n \to \infty$,

$$\frac{1}{\epsilon\sqrt{n}} \max_{1 \le i \le n} \sum_{k=1}^{n} \{\lambda_{i,n}^{-1}\} |\lambda_{k,n} - 1| \to 0.$$

Proof. Note that the existence of the matrix $\mathbf{U}$ is true according to Lemma 1 and thus $\lambda_{k,n}, k = 1, \ldots, n$ are well-defined.

Let $\xi_0 : \mathbb{R}^d \to \mathbb{R}$ be a function of the form $\xi_0(\boldsymbol{\omega}) = \int_{\mathbb{R}^d} \exp\{-i\mathbf{x}^\top \boldsymbol{\omega}\} c_0(\mathbf{x}) d\mathbf{x}$ for any $\boldsymbol{\omega} \in \mathbb{R}^d$, where $c_0(\mathbf{x}) = |\mathbf{x}|^{\kappa - d} I(|\mathbf{x}| \le 1)$ for any $\mathbf{x} \in \mathbb{R}^d$, $\kappa = (\nu + d/2)/(2m)$ with $|\cdot|$ denoting the Euclidean norm, and $m = \lfloor \nu + d/2 \rfloor + 1$ with $\lfloor x \rfloor$ denoting the largest integer less than or equal to x. As $d \in \{1, 2, 3\}$, $\kappa \in (0, 1/2)$, it follows from Lemma 2.3 of Wang (2010) and proof of Theorem 8 of Bevilacqua et al. (2019) that the function ξ_0 is a continuous, isotropic, and strictly positive function with $\xi_0(\boldsymbol{\omega}) \asymp |\boldsymbol{\omega}|^{-\kappa}$ when $|\boldsymbol{\omega}| \to \infty$.

Let $c_1 = c_0 * \ldots * c_0$ denote the $2m$-fold convolution of the function c_0 with itself, and let $\xi_1(\boldsymbol{\omega}) = \int_{\mathbb{R}^d} \exp\{-i\mathbf{x}^\top \boldsymbol{\omega}\} c_1(\mathbf{x}) d\mathbf{x}$. Then $\xi_1(\boldsymbol{\omega}) = \xi_0(\boldsymbol{\omega})^{2m}$ for all $\boldsymbol{\omega} \in \mathbb{R}^d$. This implies that ξ_1 is also a continuous, isotropic, and strictly positive function. By Proposition 1, the spectral density $f(|\boldsymbol{\omega}|)$ of the CH covariance function satisfies $f(|\boldsymbol{\omega}|) \asymp |\boldsymbol{\omega}|^{-(2\nu+d)}$, and hence we have $f(|\boldsymbol{\omega}|)/\xi_1(\boldsymbol{\omega}) \asymp 1$ as $|\boldsymbol{\omega}| \to \infty$. Note that this ratio (as a function of $|\boldsymbol{\omega}|$) is a well-defined and continuous function on arbitrary compact interval of the positive real line with $\xi_1 > 0$. Thus, there exist two positive constants (not depending on $\boldsymbol{\omega}$) such that

$$c_{\xi_1} \le \frac{f(|\boldsymbol{\omega}|)}{\xi_1(\boldsymbol{\omega})} \le C_{\xi_1}, \quad \text{as } |\boldsymbol{\omega}| \to \infty. \tag{S.4}$$

For any fixed $\nu > 0$, let $f_{\sigma,\alpha,\beta}(|\boldsymbol{\omega}|)$ denote the spectral density of the CH covariance $C(h; \nu, \alpha, \beta, \sigma^2)$ and let $f_{\sigma_0,\alpha_0,\beta_0}(|\boldsymbol{\omega}|)$ denote the spectral density of the CH covariance $C(h; \nu, \alpha_0, \beta_0, \sigma_0^2)$. Then we define

$$\eta(\boldsymbol{\omega}) := \frac{f_{\sigma,\alpha,\beta}(|\boldsymbol{\omega}|) - f_{\sigma_0,\alpha_0,\beta_0}(|\boldsymbol{\omega}|)}{\xi_1(\boldsymbol{\omega})}.$$

It follows from direct calculation that for a constant $C_\eta > 0$,

$$\begin{aligned}\int_{\mathbb{R}^d} \eta(\boldsymbol{\omega})^2 d\boldsymbol{\omega} = \frac{2\pi^{d/2}}{\Gamma(d/2)} \Bigg\{ \int_0^{C_\eta} r^{d-1} \left(\frac{f_{\sigma,\alpha,\beta}(r) - f_{\sigma_0,\alpha_0,\beta_0}(r)}{\xi_1(\mathbf{r})} \right)^2 dr \\ + \int_{C_\eta}^{\infty} r^{d-1} \left(\frac{f_{\sigma,\alpha,\beta}(r) - f_{\sigma_0,\alpha_0,\beta_0}(r)}{\xi_1(\mathbf{r})} \right)^2 dr \Bigg\},\end{aligned} \tag{S.5}$$

where $r = |\mathbf{r}|$ and $\mathbf{r} \in \mathbb{R}^d$.

As shown in Theorem 3, for any fixed $\nu > 0$, the spectral density of the CH class satisfies the conditions (S.1) and (S.2) when $\frac{\sigma^2\Gamma(\nu+\alpha)}{\beta^{2\nu}\Gamma(\alpha)} = \frac{\sigma_0^2\Gamma(\nu+\alpha_0)}{\beta_0^{2\nu}\Gamma(\alpha_0)}$. This implies that there exists a constant C_η^0 (not depending on $\boldsymbol{\omega}$) such that

$$|\eta(\boldsymbol{\omega})| \leq \frac{C_\eta^0}{(1 + |\boldsymbol{\omega}|^2)}, \quad \forall \boldsymbol{\omega} \in \mathbb{R}^d.$$

It follows immediately that the two integrals in the righ-hand side of Equation (S.5) are hence finite for $d = 1, 2, 3$. Thus, η is square integrable, i.e., $\eta \in L^2(\mathbb{R}^d)$. From classic Fourier theory (see Chapter 1 of Stein and Weiss (1971)), an immediate consequence of the square integrability of η is that there exists a square-integrable function $g : \mathbb{R}^d \to \mathbb{C}$ such that

$$\int_{\mathbb{R}^d} |\eta(\boldsymbol{\omega}) - \hat{g}_k(\boldsymbol{\omega})|^2 d\boldsymbol{\omega}, \text{ as } k \to \infty,$$

where $\hat{g}_k(\boldsymbol{\omega}) = \int_{\mathbb{R}^d} \exp\{-i\mathbf{x}^\top\boldsymbol{\omega}\} g(\mathbf{x}) I(|\mathbf{x}|_{\max} \leq k) d\mathbf{x}$ for all $\boldsymbol{\omega} \in \mathbb{R}^d$ and $|\mathbf{x}|_{\max} = \max_{1\leq j\leq d} |x_j|$ for $\mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$.

Let $a > 0$, $m_a := \lfloor a + d/2 \rfloor + 1, a_0 := (a + d/2)/(2m_a)$. Define

$$\begin{aligned}\tilde{c}_0(\mathbf{x}) &:= |\mathbf{x}|^{a_0 - d} I(|\mathbf{x}| \leq 1), \quad \forall \mathbf{x} \in \mathbb{R}^d, \\ \tilde{\xi}_0(\boldsymbol{\omega}) &:= \int_{\mathbb{R}^d} \exp\{-i\mathbf{x}^\top\boldsymbol{\omega}\} \tilde{c}_0(\mathbf{x}) d\mathbf{x}, \quad \forall \boldsymbol{\omega} \in \mathbb{R}^d.\end{aligned}$$

Let $\tilde{c}_1 := \tilde{c}_0 * \cdots * \tilde{c}_0$ denote the $2m_a$-fold convolution of $\tilde{c}_0$ with itself. Let $\{\epsilon_n : n = 1, 2, \ldots\}$ be a sequence of real numbers such that $\epsilon_n \in (0, 1)$ for all n and $\lim_{n\to\infty} \epsilon_n = 0$. Then we define

$$e_n(\mathbf{x}) := \frac{1}{C_e \epsilon_n^d} \tilde{c}_1 \left(\frac{\mathbf{x}}{\epsilon_n} \right),$$

where $C_e := \int_{\mathbb{R}^d} \tilde{c}_1(\mathbf{x}) d\mathbf{x}$. Then we obtain the Fourier transform of e_n:

$$\hat{e}_n(\mathbf{x}) = \int_{\mathbb{R}^d} \exp\{-i\mathbf{x}^\top \boldsymbol{\omega}\} e_n(\mathbf{x}) d\mathbf{x} = \frac{\tilde{\xi}_1(\epsilon_n \boldsymbol{\omega})}{C_e},$$

where $\tilde{\xi}_1(\boldsymbol{\omega}) := \int_{\mathbb{R}^d} \exp\{-i\mathbf{x}^\top \boldsymbol{\omega}\} \tilde{c}_1(\mathbf{x}) d\mathbf{x} = \tilde{\xi}_0^{2m_a}(\boldsymbol{\omega})$. This implies that there exists a constant $C_{\hat{e}}$ (not depending on $\boldsymbol{\omega}$ and n) such that

$$|\hat{e}_n(\boldsymbol{\omega})| \leq \frac{C_{\hat{e}}}{(1 + \epsilon_n |\boldsymbol{\omega}|)^{a+d/2}}, \quad \forall \boldsymbol{\omega} \in \mathbb{R}^d. \tag{S.6}$$

Note that it follows from Plancherel's theorem that

$$\begin{aligned}
\int_{\mathbb{R}^d} |g(\mathbf{x} - \mathbf{y}) - g(\mathbf{y})|^2 d\mathbf{x} &= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} |(\exp\{-i\mathbf{w}^\top \mathbf{y}\} - 1)\eta(\boldsymbol{\omega})|^2 d\boldsymbol{\omega} \\
&\leq \frac{2^{2-\ell_0} |\mathbf{y}|^{\ell_0}}{(2\pi)^d} \int_{\mathbb{R}^d} |\boldsymbol{\omega}|^{\ell_0} |\eta(\boldsymbol{\omega})|^2 d\boldsymbol{\omega},
\end{aligned}$$

and it follows from Minkowski's equality that

$$\begin{aligned}
\left\{ \int_{\mathbb{R}^d} |e_n * g(\mathbf{x}) - g(\mathbf{x})|^2 d\mathbf{x} \right\}^{1/2} &= \left\{ \int_{\mathbb{R}^d} \int_{|\mathbf{y}| \leq 2m_a \epsilon_n} |(g(\mathbf{x} - \mathbf{y}) - g(\mathbf{x})) e_n(\mathbf{y}) d\mathbf{y}|^2 d\mathbf{x} \right\}^{1/2} \\
&\leq \frac{2^{1-\ell_0/2} (2m_a \epsilon_n)^{\ell_0/2}}{(2\pi)^{d/2}} \left\{ \int_{\mathbb{R}^d} |\boldsymbol{\omega}|^{\ell_0} |\eta(\boldsymbol{\omega})|^2 d\boldsymbol{\omega} \right\}^{1/2} \\
&\leq \frac{2^{1-\ell_0/2} (2m_a \epsilon_n)^{\ell_0/2}}{(2\pi)^{d/2}} C_\eta^0 \left\{ \int_{\mathbb{R}^d} \frac{|\boldsymbol{\omega}|^{\ell_0}}{(1 + |\boldsymbol{\omega}|^2)^2} d\boldsymbol{\omega} \right\}^{1/2},
\end{aligned}$$

where the integral $\int_{\mathbb{R}^d} |\boldsymbol{\omega}|^{\ell_0}(1+|\boldsymbol{\omega}|^2)^{-2}d\boldsymbol{\omega}$ is finite for $\ell_0 < \min\{2, 4-d\}$. Thus, there exists a constant C_{ℓ_0} such that

$$\int_{\mathbb{R}^d} |e_n * g(\mathbf{x}) - g(\mathbf{x})|^2 d\mathbf{x} \leq C_{\ell_0}\epsilon_n^{\ell_0} \tag{S.7}$$

for $\ell_0 < \min\{2, 4-d\}$.

Next, we will show some useful bounds on eigenvalues of covariance matrices based on results from spectral theory.

Let $b(\mathbf{s}, \mathbf{u}) := E_{f_{\sigma,\alpha,\beta}}[Z(\mathbf{s})Z(\mathbf{u})] - E_{f_{\sigma_0,\alpha_0,\beta_0}}[Z(\mathbf{s})Z(\mathbf{u})]$ for all $\mathbf{s}, \mathbf{u} \in \mathcal{D} = [0, L]^d$. It follows from Equation (2.24) of Wang (2010) and the fact that $\text{supp}(c_1) \subset [-2m, 2m]^d$ that for all $\mathbf{s}, \mathbf{u} \in \mathcal{D}$,

$$\begin{aligned}
b(\mathbf{s}, \mathbf{u}) &= \int_{\mathbb{R}^d} \exp\{-i(\mathbf{s}-\mathbf{u})^\top\boldsymbol{\omega}\}\{f_{\sigma,\alpha,\beta}(|\boldsymbol{\omega}|) - f_{\sigma_0,\alpha_0,\beta_0}(|\boldsymbol{\omega}|)\}d\boldsymbol{\omega} \\
&= \int_{\mathbb{R}^d} \exp\{-i(\mathbf{s}-\mathbf{u})^\top\boldsymbol{\omega}\}\eta(\boldsymbol{\omega})\xi_1(\boldsymbol{\omega})d\boldsymbol{\omega} \\
&= (2\pi)^d \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} g(\mathbf{x}-\mathbf{y})c_1(\mathbf{s}-\mathbf{x})c_1(\mathbf{u}-\mathbf{y})d\mathbf{x}d\mathbf{y} \\
&= (2\pi)^d \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} e_n * g(\mathbf{x}-\mathbf{y})c_1(\mathbf{s}-\mathbf{x})c_1(\mathbf{u}-\mathbf{y})d\mathbf{x}d\mathbf{y} \\
&\quad + (2\pi)^d \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} h_n^*(\mathbf{x},\mathbf{y})c_1(\mathbf{s}-\mathbf{x})c_1(\mathbf{u}-\mathbf{y})d\mathbf{x}d\mathbf{y},
\end{aligned} \tag{S.8}$$

where $h_n^*(\mathbf{x}, \mathbf{y}) = \{g(\mathbf{x}-\mathbf{y}) - e_n * g(\mathbf{x}-\mathbf{y})\}I(|\mathbf{x}+\mathbf{y}|_{\max} \leq 4m + 2L)$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ and h_n^* is square integrable.

Define

$$h_n^{**}(\mathbf{x}, \mathbf{y}) := \int_{|\mathbf{u}|_{\max} \leq 2m+2m_a+L} e_n(\mathbf{x}-\mathbf{u})g(\mathbf{u}-\mathbf{y})d\mathbf{u}, \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^d.$$

The function $h_n^{**} : \mathbb{R}^{2d} \to \mathbb{C}$ is again square integrable. Direct calculation yields that the first part of Equation (S.8) can be re-expressed as

$$\begin{aligned}
&(2\pi)^d \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} e_n * g(\mathbf{x}-\mathbf{y}) c_1(\mathbf{s}-\mathbf{x}) c_2(\mathbf{u}-\mathbf{y}) d\mathbf{x} d\mathbf{y} \\
&= (2\pi)^d \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} h_n^{**}(\mathbf{x},\mathbf{y}) c_1(\mathbf{s}-\mathbf{x}) c_1(\mathbf{u}-\mathbf{y}) d\mathbf{x} d\mathbf{y} \\
&= (2\pi)^{-d} \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \exp\{i(\boldsymbol{\omega}^\top \mathbf{s} - \mathbf{v}^\top \mathbf{u})\} \xi_1(\boldsymbol{\omega}) \xi_1(\mathbf{v}) \\
&\quad \times \left\{ \int_{|\mathbf{t}|_{\max} \le 2m + 2m_a + L} \exp\{-i(\boldsymbol{\omega}^\top \mathbf{t} - \mathbf{v}^\top \mathbf{t})\} \hat{e}_n(\boldsymbol{\omega}) \eta(\mathbf{v}) d\mathbf{t} \right\} d\mathbf{v} d\boldsymbol{\omega}.
\end{aligned} \tag{S.9}$$

Let $\eta_n^* : \mathbb{R}^d \to \mathbb{C}$ be the Fourier transform of $g - e_n * g$ and define

$$\hat{g}_{n,k}(\boldsymbol{\omega}) := \int_{\mathbb{R}^d} \exp\{-i\boldsymbol{\omega}^\top \mathbf{x}\}[g(\mathbf{x}) - e_n * g(\mathbf{x})] I(|\mathbf{x}|_{\max} \le k) d\mathbf{x}.$$

This implies that

$$\int_{\mathbb{R}^d} |\eta_n^*(\boldsymbol{\omega}) - \hat{g}_{n,k}(\boldsymbol{\omega})|^2 d\boldsymbol{\omega} \to 0, \text{ as } k \to \infty. \tag{S.10}$$

Let $\theta(\boldsymbol{\omega}) = 2^{-d} \int_{\mathbb{R}^d} \exp\{-i\mathbf{t}^\top \boldsymbol{\omega}\} I(|\mathbf{t}|_{\max} \le 4m + 2L) d\mathbf{t}$. Then θ is continuous and square integrable with

$$\int_{\mathbb{R}^d} \theta(\boldsymbol{\omega})^2 d\boldsymbol{\omega} < \infty. \tag{S.11}$$

Direct calculation yields that the second part of Equation (S.8) can be re-expressed as

$$\begin{aligned}
&(2\pi)^d \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} h_n^{*}(\mathbf{x},\mathbf{y}) c_1(\mathbf{s}-\mathbf{x}) c_1(\mathbf{u}-\mathbf{y}) d\mathbf{x} d\mathbf{y} \\
&= (2\pi)^{-d} \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \exp\{i(\boldsymbol{\omega}^\top \mathbf{s} - \mathbf{v}^\top \mathbf{u})\} \eta_n^* \left(\frac{\boldsymbol{\omega}+\mathbf{v}}{2}\right) \theta\left(\frac{\boldsymbol{\omega}-\mathbf{v}}{2}\right) \xi_1(\boldsymbol{\omega}) \xi_1(\mathbf{v}) d\boldsymbol{\omega} d\mathbf{v}.
\end{aligned} \tag{S.12}$$

Combining Equations (S.9) and (S.12) allows us to write $b(\mathbf{s},\mathbf{u})$ as a sum of two parts:

$$
\begin{aligned}
b(\mathbf{s},\mathbf{u}) = (2\pi)^{-d} & \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \exp\{i(\boldsymbol{\omega}^\top \mathbf{s} - \mathbf{v}^\top \mathbf{u})\}\xi_{\mathbf{1}}(\boldsymbol{\omega})\xi_{\mathbf{1}}(\mathbf{v}) \\
& \times \left\{ \int_{|\mathbf{t}|_{\max}\le 2m+2m_a+L} \exp\{-i(\boldsymbol{\omega}^\top \mathbf{t} - \mathbf{v}^\top \mathbf{t})\}\hat{e}_n(\boldsymbol{\omega})\eta(\mathbf{v})d\mathbf{t} \right\} d\mathbf{v}d\boldsymbol{\omega} \\
& + (2\pi)^{-d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \exp\{i(\boldsymbol{\omega}^\top \mathbf{s} - \mathbf{v}^\top \mathbf{u})\}\eta_n^*\left(\frac{\boldsymbol{\omega}+\mathbf{v}}{2}\right)\theta\left(\frac{\boldsymbol{\omega}-\mathbf{v}}{2}\right)\xi_1(\boldsymbol{\omega})\xi_1(\mathbf{v})d\boldsymbol{\omega}d\mathbf{v}.
\end{aligned}
$$

In the rest of the proof, we will relate $b(\mathbf{s},\mathbf{u})$ with the eigenvalues of the CH covariance matrix and give bounds on these eigenvalues. Let $\{\psi_1,\ldots,\psi_n\}$ be as in Equation (2.15) of Wang (2010). Then it follows from Equations (2.16) and (2.60) of Wang (2010) that

$$
\langle \psi_k, \psi_k\rangle_{f_{\sigma,\alpha,\beta}} - \langle \psi_k, \psi_k\rangle_{f_{\sigma_0,\alpha_0,\beta_0}} = \lambda_{k,n} - 1 =: \tilde{\nu}_{k,n}^* + \tilde{\nu}_{k,n}^\dagger,
$$

where

$$
\begin{aligned}
\tilde{\nu}_{k,n}^* &:= \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \psi_k(\boldsymbol{\omega})\overline{\psi_k(\mathbf{v})}\eta_n^*\left(\frac{\boldsymbol{\omega}+\mathbf{v}}{2}\right)\theta\left(\frac{\boldsymbol{\omega}-\mathbf{v}}{2}\right)\xi_1(\boldsymbol{\omega})\xi_1(\mathbf{v})d\boldsymbol{\omega}d\mathbf{v}, \\
\tilde{\nu}_{k,n}^\dagger &:= \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \psi_k(\boldsymbol{\omega})\overline{\psi_k(\mathbf{v})}\xi_1(\boldsymbol{\omega})\xi_1(\mathbf{v}) \\
&\quad \times \left\{ \int_{|\mathbf{t}|_{\max}\le 2m+2m_a+L} \exp\{-i(\boldsymbol{\omega}^\top \mathbf{t} - \mathbf{v}^\top \mathbf{t})\}\hat{e}_n(\boldsymbol{\omega})\eta(\mathbf{v})d\mathbf{t} \right\} d\mathbf{v}d\boldsymbol{\omega},
\end{aligned}
$$

with $\bar{x}$ denote the complex conjugate of x. It then follows from Bessel's inequality that

$$
\begin{aligned}
\sum_{k=1}^n |\tilde{\nu}_{k,n}^*|^2 &\le 2^{-d-1}\pi^{-d}\left\{\sup_{\mathbf{s}\in\mathbb{R}^d}\frac{\xi_1(\mathbf{s})^2}{f_{\sigma_0,\alpha_0,\beta_0}(\mathbf{s})}\right\}\int_{\mathbb{R}^d}|\eta_n^*(\boldsymbol{\omega})|^2 d\boldsymbol{\omega}\int_{\mathbb{R}^d}|\boldsymbol{\theta}(\mathbf{v})|^2 d\mathbf{v}, \\
\sum_{k=1}^n |\tilde{\nu}_{k,n}^\dagger| &\le 2^{-d-1}\pi^{-d}\left\{\sup_{\mathbf{s}\in\mathbb{R}^d}\frac{\xi_1(\mathbf{s})^2}{f_{\sigma_0,\alpha_0,\beta_0}(\mathbf{s})}\right\}\int_{|\mathbf{t}|_{\max}\le 2m+2m_a+L} d\mathbf{t} \\
&\quad \times \left\{\int_{\mathbb{R}^d}|\hat{e}_n(\boldsymbol{\omega})|^2 d\boldsymbol{\omega} + \int_{\mathbb{R}^d}\eta(\mathbf{v})^2 d\mathbf{v}\right\}.
\end{aligned}
$$

Combining the above inequalities with Equations (S.4), (S.6), (S.7), (S.10), and (S.11), we observe that there exist constants C, C_1, C_2 (not depending on n) such that

$$\sum_{k=1}^{n} |\tilde{\nu}_{k,n}^{*}|^2 \leq C\epsilon_n^{\ell_0},$$

$$\sum_{k=1}^{n} |\tilde{\nu}_{k,n}^{*}| \leq \left\{ n \sum_{k=1}^{n} |\tilde{\nu}_{k,n}^{*}|^2 \right\}^{1/2} \leq \sqrt{Cn\epsilon_n^{\ell_0}},$$

$$\sum_{k=1}^{n} |\tilde{\nu}_{k,n}^{\dagger}| \leq \left(\frac{C_1}{\epsilon_n^d} + C_2 E \right),$$

where $E := \int_{\mathbb{R}^d} \eta(\mathbf{v})^2 d\mathbf{v}$ is finite. Thus it follows that

$$\sum_{k=1}^{n} |\lambda_{k,n} - 1| \leq \sqrt{Cn\epsilon_n^{\ell_0}} + \frac{C_1}{\epsilon_n^d} + C_2 E. \tag{S.13}$$

When $\alpha, \alpha_0 > d/2$, the spectral density of the CH covariance function is well-defined and is finite for any frequency. Thus, the ratio

$$\frac{f_{\sigma,\alpha,\beta}(|\boldsymbol{\omega}|)}{f_{\sigma_0,\alpha_0,\beta_0}(|\boldsymbol{\omega}|)}$$

is well-defined for all $\boldsymbol{\omega} \in \mathbb{R}^d$. We also observe that there exist constants $\tilde{c}_f > 0$ and $\tilde{C}_f > 0$ (not depending on $\boldsymbol{\omega}$) such that

$$\tilde{c}_f \leq \frac{f_{\sigma,\alpha,\beta}(|\boldsymbol{\omega}|)}{f_{\sigma_0,\alpha_0,\beta_0}(|\boldsymbol{\omega}|)} \leq \tilde{C}_f, \quad \forall \boldsymbol{\omega} \in \mathbb{R}^d.$$

It follows immediately that $\tilde{c}_f \leq \lambda_{k,n} \leq \tilde{C}_f$ for all $k = 1, \ldots, n$.

Finally, using (S.13) yields that for any $\epsilon > 0$,

$$\frac{1}{\epsilon\sqrt{n}} \left\{ \max_{1\le i\le n} \{\lambda_{i,n}^{-1}\} \right\} \sum_{k=1}^{n} |\lambda_{k,n} - 1| \le \frac{C^{1/2}\epsilon_n^{\ell_0/2}}{\tilde{c}_f \epsilon} + \frac{1}{\tilde{c}_f \epsilon n^{1/2}} \left(\frac{C_1}{\epsilon_n^d} + C_2 E \right). \tag{S.14}$$

Therefore, the right-hand side of Equation (S.14) tends to 0 as $n \to \infty$, as desired.

□

Based on Lemma 1 and Lemma 2, we can study the asymptotic behavior of the MLE for the microergodic parameter of the CH covariance function.

Lemma 3. *Let $\{\mathcal{D}_n\}_{n\ge1}$ be an increasing sequence of subsets of a bounded domain $\mathcal{D}$ such that $\cup_{n=1}^{\infty}\mathcal{D}_n$ is bounded. Assume that ν is fixed. Let $\mathcal{P}_0$ be the Gaussian probability measure defined under $C(h;\nu,\alpha_0,\beta_0,\sigma_0^2)$. Let $c(\boldsymbol{\theta}_0) := \sigma_0^2\beta_0^{-2\nu}\Gamma(\nu+\alpha_0)/\Gamma(\alpha_0)$. Assume that $\alpha_0 > d/2, \beta_0 > 0, \sigma_0^2 > 0$. The following results can be established.*

(a) As $n \to \infty$, $\hat{c}_n(\boldsymbol{\theta}) \xrightarrow{a.s.} c(\boldsymbol{\theta}_0)$ under measure $\mathcal{P}_0$ for any fixed $\alpha > d/2$ and $\beta > 0$;

(b) As $n \to \infty$, $\sqrt{n}\{\hat{c}_n(\boldsymbol{\theta}) - c(\boldsymbol{\theta}_0)\} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 2[c(\boldsymbol{\theta}_0)]^2)$ for any fixed $\alpha > d/2$ and $\beta > 0$.

Proof. The proof of Part (a) follows from the same arguments as in the proof of Theorem 3 in Zhang (2004) and is omitted. For the proof of Part (b), we follow the arguments in Wang (2010); Wang and Loh (2011) and Bevilacqua et al. (2019). Without loss of generality, we assume $\mathcal{D} = [0, L]^d, 0 < L < \infty$ is a bounded subset of $\mathbb{R}^d$ with $d = 1, 2, 3$. Let σ^2 be a positive constant such that $\sigma^2\beta^{-2\nu}\Gamma(\nu+\alpha)/\Gamma(\alpha) = \sigma_0^2\beta_0^{-2\nu}\Gamma(\nu+\alpha_0)/\Gamma(\alpha_0)$. Let $c(\boldsymbol{\theta}) = \sigma^2\beta^{-2\nu}\Gamma(\nu+\alpha)/\Gamma(\alpha)$ and $\hat{c}_n(\boldsymbol{\theta}) = \hat{\sigma}_n^2\beta^{-2\nu}\Gamma(\nu+\alpha)/\Gamma(\alpha)$. Then we have

$$\begin{aligned}\sqrt{n}\{\hat{c}_n(\boldsymbol{\theta}) - c(\boldsymbol{\theta}_0)\} = \frac{c(\boldsymbol{\theta}_0)}{\sqrt{n}} \left\{ \frac{1}{\sigma^2}\mathbf{Z}_n^\top\mathbf{R}_n^{-1}(\boldsymbol{\theta})\mathbf{Z}_n - \frac{1}{\sigma_0^2}\mathbf{Z}_n^\top\mathbf{R}_n^{-1}(\boldsymbol{\theta}_0)\mathbf{Z}_n \right\} \\ + \frac{c(\boldsymbol{\theta}_0)}{\sqrt{n}} \left\{ \frac{1}{\sigma_0^2}\mathbf{Z}_n^\top\mathbf{R}_n^{-1}(\boldsymbol{\theta}_0)\mathbf{Z}_n - n \right\}.\end{aligned}$$

Under Gaussian measure $\mathcal{P}_0$ defined by the covariance function $C(h; \nu, \alpha_0, \beta_0, \sigma_0^2)$, we have

$$\mathbf{Z}_n^\top \mathbf{R}_n^{-1}(\boldsymbol{\theta}_0)\mathbf{Z}_n/\sigma_0^2 \sim \chi_n^2 \quad \text{and} \quad \frac{c(\boldsymbol{\theta}_0)}{\sqrt{n}} \left\{ \frac{1}{\sigma_0^2} \mathbf{Z}_n^\top \mathbf{R}_n^{-1}(\boldsymbol{\theta}_0)\mathbf{Z}_n - n \right\} \xrightarrow{\mathcal{L}} \mathcal{N}(0,\, 2[c(\boldsymbol{\theta}_0)]^2),$$

as $n \to \infty$. To prove the result, it suffices to show that

$$\frac{1}{\sqrt{n}} \left\{ \frac{1}{\sigma^2} \mathbf{Z}_n^\top \mathbf{R}_n^{-1}(\boldsymbol{\theta})\mathbf{Z}_n - \frac{1}{\sigma_0^2} \mathbf{Z}_n^\top \mathbf{R}_n^{-1}(\boldsymbol{\theta}_0)\mathbf{Z}_n \right\} \xrightarrow{\mathcal{P}_0} 0, \quad \text{as} \quad n \to \infty, \tag{S.15}$$

under Gaussian measure $\mathcal{P}_0$.

According to Lemma 1, there exists an $n \times n$ non-singular matrix $\mathbf{U}$ such that

$$\sigma_0^2 \mathbf{U}^\top \mathbf{R}_n(\boldsymbol{\theta}_0)\mathbf{U} = \mathbf{I}_n, \quad \sigma^2 \mathbf{U}^\top \mathbf{R}_n(\boldsymbol{\theta})\mathbf{U} = \boldsymbol{\Lambda},$$

where $\boldsymbol{\Lambda} := \mathrm{diag}\{\lambda_{1,n}, \ldots, \lambda_{n,n}\}$ is an $n \times n$ diagonal matrix with diagonal elements satisfying $\lambda_{k,n} > 0$ for $k = 1, \ldots, n$. Now we define the random vector $\mathbf{Y} := (Y_1, \ldots, Y_n)^\top = \mathbf{U}^\top \mathbf{Z}_n$. It is easy to check that $\mathbf{Y} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I}_n)$ for $\mathbf{Z}_n$ generated under the measure $\mathcal{P}_0$. Thus, the assertion (S.15) is true if for any $\epsilon > 0$,

$$\begin{aligned} &\mathcal{P}_0 \left(\frac{1}{\sqrt{n}} \left| \frac{1}{\sigma^2} \mathbf{Z}_n^\top \mathbf{R}_n^{-1}(\boldsymbol{\theta})\mathbf{Z}_n - \frac{1}{\sigma_0^2} \mathbf{Z}_n^\top \mathbf{R}_n^{-1}(\boldsymbol{\theta}_0)\mathbf{Z}_n \right| > \epsilon \right) \\ &= \mathcal{P}_0 \left(\frac{1}{\sqrt{n}} \left| \sum_{k=1}^n (\lambda_{k,n}^{-1} - 1) Y_k^2 \right| > \epsilon \right) \to 0, \quad \text{as} \quad n \to \infty. \end{aligned} \tag{S.16}$$

By Markov's inequality, the probability in the assertion (S.16) can be bounded as

$$\mathcal{P}_0 \left(\frac{1}{\sqrt{n}} \left| \sum_{k=1}^n (\lambda_{k,n}^{-1} - 1) Y_k^2 \right| > \epsilon \right) \leq \frac{1}{\epsilon \sqrt{n}} \sum_{k=1}^n |\lambda_{k,n}^{-1} - 1| \leq \frac{1}{\epsilon \sqrt{n}} \left\{ \max_{1 \leq i \leq n} \{\lambda_{i,n}^{-1}\} \right\} \sum_{k=1}^n |\lambda_{k,n} - 1|.$$

The rest of the proof is to show that for any $\epsilon > 0$, the term

$$\frac{1}{\epsilon\sqrt{n}}\left\{\max_{1\le i\le n}\{\lambda_{i,n}^{-1}\}\right\}\sum_{k=1}^{n}|\lambda_{k,n}-1|$$

goes to 0 as $n \to \infty$. This is true according to Lemma 2. □

Lemma 3 implies that the estimator $\hat{c}_n(\boldsymbol{\theta})$ of the microergodic parameter converges to the true microergodic parameter, almost surely, when the number of observations tends to infinity in a fixed and bounded domain. This result holds true for any value of $\boldsymbol{\theta}$. As will be shown, if one replaces $\boldsymbol{\theta}$ with its maximum likelihood estimator in $\hat{c}_n(\boldsymbol{\theta})$, this conclusion is true as well. The second statement of Lemma 3 indicates that $\hat{c}_n(\boldsymbol{\theta})$ converges to a normal distribution.

A key fact is that the above lemma holds true for arbitrarily fixed $\boldsymbol{\theta}$. A more practical situation is to estimate $\boldsymbol{\theta}$ and σ^2 by maximizing the log-likelihood (7). The following lemma is needed to prove the asymptotic behavior of $\hat{c}_n(\alpha, \hat{\beta}_n)$ and $\hat{c}_n(\hat{\alpha}_n, \beta)$ under infill asymptotics.

Lemma 4. *Suppose that d is the dimension of the domain $\mathcal{D}$ and $\mathbf{Z}_n$ is a vector of n observations in $\mathcal{D}$. For any α_1, α_2 such that $d/2 < \alpha_1 < \alpha_2$ and any β_1, β_2 such that $0 < \beta_1 < \beta_2$, we have the following results:*

(a) $\hat{c}_n(\alpha, \beta_1) \le \hat{c}_n(\alpha, \beta_2)$ for any fixed $\alpha > d/2$.

(b) $\hat{c}_n(\alpha_1, \beta) \ge \hat{c}_n(\alpha_2, \beta)$ for any fixed $\beta > 0$.

Proof. The difference

$$\hat{c}_n(\boldsymbol{\theta}_1) - \hat{c}_n(\boldsymbol{\theta}_2) = \mathbf{Z}_n^\top\left\{\frac{\Gamma(\nu+\alpha_1)}{\beta_1^{2\nu}\Gamma(\alpha_1)}\mathbf{R}_n^{-1}(\boldsymbol{\theta}_1) - \frac{\Gamma(\nu+\alpha_2)}{\beta_2^{2\nu}\Gamma(\alpha_2)}\mathbf{R}_n^{-1}(\boldsymbol{\theta}_2)\right\}\mathbf{Z}_n/n$$

is nonnegative for any $\mathbf{Z}_n$ if the matrix $\mathbf{A} := \frac{\Gamma(\nu+\alpha_1)}{\beta_1^{2\nu}\Gamma(\alpha_1)}\mathbf{R}_n^{-1}(\boldsymbol{\theta}_1) - \frac{\Gamma(\nu+\alpha_2)}{\beta_2^{2\nu}\Gamma(\alpha_2)}\mathbf{R}_n^{-1}(\boldsymbol{\theta}_2)$ is positive semidefinite. Notice that $\mathbf{A}$ is positive semidefinite if and only if $\mathbf{B} := \frac{\beta_2^{2\nu}\Gamma(\alpha_2)}{\Gamma(\nu+\alpha_2)}\mathbf{R}_n(\boldsymbol{\theta}_2) -$

$\frac{\beta_1^{2\nu}\Gamma(\alpha_1)}{\Gamma(\nu+\alpha_1)}\mathbf{R}_n(\boldsymbol{\theta}_1)$ is positive semidefinite. The entries of $\mathbf{B}$ can be expressed in terms of a function $K_B : \mathbb{R}^d \to \mathbb{R}$, with

$$B_{ij} = K_B(\mathbf{s}_i - \mathbf{s}_j) = \frac{\beta_2^{2\nu}\Gamma(\alpha_2)}{\Gamma(\nu+\alpha_2)}R(|\mathbf{s}_i - \mathbf{s}_j|;\alpha_2,\beta_2,\nu) - \frac{\beta_1^{2\nu}\Gamma(\alpha_1)}{\Gamma(\nu+\alpha_1)}R(|\mathbf{s}_i - \mathbf{s}_j|;\alpha_1,\beta_1,\nu),$$

where $|\cdot|$ denotes the Euclidean norm. The matrix $\mathbf{B}$ is positive semidefinite if K_B is a positive definite function. Define the Fourier transform of K_B by

$$\begin{aligned}
f_B(\boldsymbol{\omega}) &:= \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\exp\{-i\boldsymbol{\omega}^\top\mathbf{x}\}K_B(\mathbf{x})d\mathbf{x} \\
&= \frac{\beta_2^{2\nu}\Gamma(\alpha_2)}{\Gamma(\nu+\alpha_2)}\left\{\frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\exp\{-i\boldsymbol{\omega}^\top\mathbf{x}\}R(|\mathbf{x}|;\alpha_2,\beta_2,\nu)d\mathbf{x}\right\} \\
&\quad - \frac{\beta_1^{2\nu}\Gamma(\alpha_1)}{\Gamma(\nu+\alpha_1)}\left\{\frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\exp\{-i\boldsymbol{\omega}^\top\mathbf{x}\}R(|\mathbf{x}|;\alpha_1,\beta_1,\nu)d\mathbf{x}\right\}.
\end{aligned}$$

The integrals in $f_B(\boldsymbol{\omega})$ are finite for $\alpha_1,\alpha_2 > d/2$. Let $g(\boldsymbol{\omega})$ be the spectral density of the CH correlation function with parameters α,β,ν:

$$\begin{aligned}
g(\boldsymbol{\omega}) &:= \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\exp\{-i\boldsymbol{\omega}^\top\mathbf{x}\}R(|\mathbf{x}|;\alpha,\beta,\nu)d\mathbf{x} \\
&= \frac{2^{2\nu}\nu^\nu}{\pi^{d/2}\beta^{2\nu}\Gamma(\alpha)}\int_0^\infty\{4\nu/(\beta^2 t) + |\boldsymbol{\omega}|^2\}^{-(\nu+d/2)}t^{-(\nu+\alpha+1)}\exp\{-1/t\}dt.
\end{aligned}$$

Thus, K_B is positive definite if f_B is positive for all $\boldsymbol{\omega}\in\mathbb{R}^d$. Notice that f_B is given by

$$\begin{aligned}
f_B(\boldsymbol{\omega}) = &\frac{(4\nu)^\nu}{\pi^{d/2}\Gamma(\nu+\alpha_2)}\int_0^\infty\{4\nu/(\beta_2^2 t) + |\boldsymbol{\omega}|^2\}^{-(\nu+d/2)}t^{-(\nu+\alpha_2+1)}\exp\{-1/t\}dt \\
&- \frac{(4\nu)^\nu}{\pi^{d/2}\Gamma(\nu+\alpha_1)}\int_0^\infty\{4\nu/(\beta_1^2 t) + |\boldsymbol{\omega}|^2\}^{-(\nu+d/2)}t^{-(\nu+\alpha_1+1)}\exp\{-1/t\}dt.
\end{aligned}$$

It is straightforward to check that when $\alpha := \alpha_1 = \alpha_2 > d/2$,

$$\beta_1 < \beta_2 \implies f_B(\boldsymbol{\omega}) > 0, \quad \forall\boldsymbol{\omega}\in\mathbb{R}^d.$$

Thus, if $\beta_1 < \beta_2$, then $\hat{c}_n(\alpha, \beta_1) \leq \hat{c}_n(\alpha, \beta_2)$, as claimed in Part (a).

The proof of Part (b) is as follows. Note that $f_B(\boldsymbol{\omega})$ can be expressed as $f_B(\boldsymbol{\omega}) = \{(4\nu)^\nu/\pi^{d/2}\}(I(\alpha_2) - I(\alpha_1))$, where

$$
\begin{aligned}
I(\alpha) &:= \int_0^\infty \frac{1}{\Gamma(\nu+\alpha)} \{4\nu/(\beta^2 t) + |\boldsymbol{\omega}|^2\}^{-(\nu+d/2)} t^{-(\nu+\alpha+1)} \exp\{-1/t\} dt \\
&= \int_0^\infty \frac{u^{\nu+\alpha-1}}{\Gamma(\nu+\alpha)} \exp\{-u\} \left(4\nu u/\beta^2 + |\boldsymbol{\omega}|^2\right)^{-(\nu+d/2)} du \\
&= E_U \left(4\nu U/\beta^2 + |\boldsymbol{\omega}|^2\right)^{-(\nu+d/2)},
\end{aligned}
$$

with $U \sim \text{Gamma}(\nu+\alpha, 1)$. This expectation is finite if $\alpha > d/2$. Suppose that $\alpha_1 < \alpha_2$ and $\beta := \beta_1 = \beta_2$. To show $f_B(\boldsymbol{\omega})$ is negative for all $\boldsymbol{\omega} \in \mathbb{R}^d$, it suffices to show that $I(\alpha_2) - I(\alpha_1) \leq 0$. Let $U_1 \sim \text{Gamma}(\nu+\alpha_1, 1)$ and $U_2 \sim \text{Gamma}(\nu+\alpha_2, 1)$. Then $U_2 \overset{\mathcal{L}}{=} U_1 + U_0$, where $U_0 \sim \text{Gamma}(\alpha_2 - \alpha_1, 1)$ and U_0 is independent of U_1. Thus, the quantify $I(\alpha_2)$ can be upper bounded by $I(\alpha_1)$, since,

$$
\begin{aligned}
I(\alpha_2) &= E_{U_1,U_0} \left\{ \frac{4\nu}{\beta^2}(U_1 + U_0) + |\boldsymbol{\omega}|^2 \right\}^{-(\nu+d/2)} \\
&= E_{U_1,U_0} \left\{ \frac{4\nu}{\beta^2}U_1 + |\boldsymbol{\omega}|^2 + \frac{4\nu}{\beta^2}U_0 \right\}^{-(\nu+d/2)} \\
&\leq E_{U_1} \left\{ \frac{4\nu}{\beta^2}U_1 + |\boldsymbol{\omega}|^2 \right\}^{-(\nu+d/2)} = I(\alpha_1).
\end{aligned}
$$

$\square$

This lemma indicates that the MLE of the microergodic parameter is monotone when one of its parameters is fixed. This property is used to prove the asymptotics of the MLE for the microergodic parameter. Based on Lemma 3 and Lemma 4 , one can show that $\hat{c}_n(\hat{\boldsymbol{\theta}}_n)$ has the same asymptotic properties as $\hat{c}_n(\boldsymbol{\theta})$ for any fixed $\boldsymbol{\theta}$.

S.4 Technical Proofs

This section contains all the proofs that are not given in the main text. For notational convenience, we drop the parameters of the covariance function when there is no scope for ambiguity.

S.4.1 Proof of Theorem 1

Proof. As $C(0) = \sigma^2 > 0$, it remains to verify the positive definiteness of the function $C(\cdot)$. For any n, all sequences $\{a_i \in \mathbb{R} : i = 1, \dots, n\}$ and all sequences of spatial locations $\{\mathbf{s}_i \in \mathbb{R}^d : i = 1, \dots, n\}$, it follows that

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j C(h_{ij}; \nu, \alpha, \beta, \sigma^2) &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \int_0^\infty \mathcal{M}(h_{ij}; \nu, \phi, \sigma^2)\pi(\phi^2; \alpha, \beta) d\phi^2 \\ &= \int_0^\infty \mathbf{a}^\top \mathbf{A}\mathbf{a}\pi(\phi^2; \alpha, \beta) d\phi^2 \geq 0,\end{aligned}$$

where $h_{ij} = |\mathbf{s}_i - \mathbf{s}_j|$ with $|\cdot|$ denoting the Euclidean norm and $\mathbf{a} := (a_1, \dots, a_n)^\top$. The matrix $\mathbf{A} := [\mathcal{M}(h_{ij})]_{i,j=1,\dots,n}$ is a covariance matrix constructed via a Matérn covariance function that is positive definite in $\mathbb{R}^d$ for all d, and hence $\mathbf{A}$ is a positive definite matrix, which yields that $\mathbf{a}^\top \mathbf{A}\mathbf{a} \geq 0$ for any $\mathbf{a}$. This implies that the resultant integral is nonnegative for any $\mathbf{a}$, and it is strictly positive for $\mathbf{a} \neq \mathbf{0}$. Thus, the function $C(h)$ is positive definite in $\mathbb{R}^d$ for any all d.

To derive the form of Equation (3), we start with the gamma mixture representation in Equation (2), and substitute for $\pi(\phi^2)$ the required inverse gamma density.

$$\begin{aligned}C(h; \nu, \alpha, \beta, \sigma^2) &= \frac{\sigma^2}{2^\nu \Gamma(\nu)} \int_0^\infty x^{(\nu-1)} \left[\int_0^\infty \phi^{-2\nu} \exp\{-x/(2\phi^2)\}\pi(\phi^2) d\phi^2\right] \exp\left(-\nu h^2/x\right) dx \\ &= \frac{\sigma^2 \beta^{2\alpha}}{2^{\nu+\alpha}\Gamma(\nu)\Gamma(\alpha)} \int_0^\infty x^{(\nu-1)} \left[\int_0^\infty \phi^{-2\nu} \exp\{-x/(2\phi^2)\}\phi^{-2(\alpha+1)}\right.\end{aligned}$$

$$
\begin{aligned}
&\times \exp\{-\beta^2/(2\phi^2)\}d\phi^2\big] \exp\left(-\nu h^2/x\right)dx \\
&= \frac{\sigma^2\beta^{2\alpha}}{2^{\nu+\alpha}\Gamma(\nu)\Gamma(\alpha)} \int_0^\infty x^{(\nu-1)} \left[\int_0^\infty \phi^{-2(\nu+\alpha+1)} \exp\{-(\beta^2+x)/(2\phi^2)\}d\phi^2\right] \\
&\times \exp\{-\nu h^2/x\}dx \\
&= \frac{\sigma^2\beta^{2\alpha}\Gamma(\nu+\alpha)}{\Gamma(\nu)\Gamma(\alpha)} \int_0^\infty x^{(\nu-1)}(x+\beta^2)^{-(\nu+\alpha)} \exp\left(-\nu h^2/x\right)dx.
\end{aligned}
$$

□

S.4.2 Proof of Theorem 2

Proof. (a) Using the property of modified Bessel function (see Abramowitz and Stegun, 1965, p. 375), as $|h| \to 0$, we can express the Matérn covariance function as

$$
\mathcal{M}(h) = \begin{cases} a_1(h) + a_2(\phi,\nu,\sigma^2)|h|^{2\nu}\log|h| + O(|h|^{2\nu}); & \text{when} \quad \nu = 0,1,2,\ldots, \\ a_3(h) + a_4(\phi,\nu,\sigma^2)|h|^{2\nu} + O(|h|^{2\lceil\nu\rceil}); & \text{otherwise}, \end{cases}
$$

where $a_i(h), i = 1,3$ are of the form $\sum_{k=0}^{\lfloor\nu\rfloor} c_k(\phi,\nu,\sigma^2)h^{2k}$ with $c_k(\phi,\nu,\sigma^2)$ being the coefficients that depend on parameters ϕ,ν,σ^2. The terms $a_2(\phi,\nu,\sigma^2) = \frac{(-1)^{\nu+1}\sigma^2}{2^{\nu-1}\Gamma(\nu)\Gamma(\nu+1)\phi^{2\nu}}$ and $a_4(\phi,\nu,\sigma^2) = \frac{-\pi\sigma^2}{2^\nu\sin(\nu\pi)\Gamma(\nu)\Gamma(\nu+1)\phi^{2\nu}}$. The terms $a_2(\phi,\nu,\sigma^2)|h|^{2\nu}\log|h|$ and $a_4(\phi,\nu,\sigma^2)|h|^{2\nu}$ are called *principal irregular* terms that determine the differentiability of a random field (see Stein, 1999, p. 32). This implies that the Matérn covariance function is $2m$ times differentiable if and only if $\nu > m$ for an integer m. By mixing the parameter ϕ^2 over an inverse gamma distribution $\mathcal{IG}(\alpha,\beta^2/2)$, when $h \to 0$, the covariance function $C(h)$ can be written as

$$
C(h) = \begin{cases} \int_0^\infty a_1(h)\pi(\phi^2)d\phi^2 + \tilde{a}_2(\nu,\sigma^2)|h|^{2\nu}\log|h| + O(|h|^{2\nu}); & \text{when} \quad \nu = 0,1,2,\ldots, \\ \int_0^\infty a_3(h)\pi(\phi^2)d\phi^2 + \tilde{a}_4(\nu,\sigma^2)|h|^{2\nu} + O(|h|^{2\lceil\nu\rceil}); & \text{otherwise}, \end{cases}
$$

where

$$\tilde{a}_2(\nu,\sigma^2) := \int_0^\infty a_2(\phi,\nu,\sigma^2)\pi(\phi^2)d\phi^2$$
$$= \frac{2(-1)^{\nu+1}\sigma^2}{\Gamma(\nu)\Gamma(\nu+1)}\frac{\Gamma(\nu+\alpha)}{\beta^{2\nu}\Gamma(\alpha)},$$

and

$$\tilde{a}_4(\nu,\sigma^2) := \int_0^\infty a_4(\phi,\nu,\sigma^2)\pi(\phi^2)d\phi^2$$
$$= \frac{-\pi\sigma^2}{\sin(\nu\pi)\Gamma(\nu)\Gamma(\nu+1)}\frac{\Gamma(\nu+\alpha)}{\beta^{2\nu}\Gamma(\alpha)}$$
$$= \frac{-\sigma^2\Gamma(1-\nu)}{\Gamma(\nu+1)}\frac{\Gamma(\nu+\alpha)}{\beta^{2\nu}\Gamma(\alpha)}.$$

Note that $\tilde{a}_2(\nu,\sigma^2)$ is finite for any positive integer ν and any fixed $\alpha>0,\beta>0$, and $\tilde{a}_4(\nu,\sigma^2)$ is finite for $\nu\in(0,\infty)\setminus\mathbb{Z}$ and $\alpha>0,\beta>0$. Thus, the covariance $C(h)$ has the same differentiability as the Matérn covariance.

(b) It follows from Theorem 1 that

$$C(h;\nu,\alpha,\beta,\sigma^2) = \frac{\sigma^2\beta^{2\alpha}\Gamma(\nu+\alpha)}{\Gamma(\nu)\Gamma(\alpha)}\int_0^\infty\left(\frac{x}{x+\beta^2}\right)^{\nu+\alpha}x^{-\alpha-1}\exp\left(-\nu h^2/x\right)dx$$
$$\overset{t=x/(2\nu)}{=\!=\!=} \frac{\sigma^2\Gamma(\nu+\alpha)}{(2\nu/\beta^2)^\alpha\Gamma(\nu)\Gamma(\alpha)}\int_0^\infty t^{\nu-1}(t+\beta^2/(2\nu))^{-(\nu+\alpha)}$$
$$\times\exp\{-h^2/(2t)\}\,dt$$
$$= \frac{\sigma^2\sqrt{2\pi}\Gamma(\nu+\alpha)}{(2\nu/\beta^2)^\alpha\Gamma(\nu)\Gamma(\alpha)}\int_0^\infty\left(\frac{t}{t+\beta^2/(2\nu)}\right)^{\nu+\alpha}t^{-\alpha-1/2}$$
$$\times\frac{1}{\sqrt{2\pi t}}\exp\{-h^2/(2t)\}\,dt.$$

Let $L(x)=\left(\frac{x}{x+\beta^2/(2\nu)}\right)^{\nu+\alpha}$. Then $L(x)$ is a slowly varying function at ∞. Viewed as a function of h, the above integral is a Gaussian scale mixture with respect to t. Thus, an application of Theorem 6.1 of Barndorff-Nielsen et al. (1982) yields

$$\begin{aligned} C(h;\nu,\alpha,\beta,\sigma^2) &\sim \frac{\sigma^2\sqrt{2\pi}\Gamma(\nu+\alpha)}{(2\nu/\beta^2)^\alpha\Gamma(\nu)\Gamma(\alpha)}(2\pi)^{-1/2}2^\alpha\Gamma(\alpha)|h|^{-2\alpha}L(h^2), \quad \text{as} \quad h\to\infty, \\ &\sim \frac{\sigma^2 2^\alpha\Gamma(\nu+\alpha)}{(2\nu/\beta^2)^\alpha\Gamma(\nu)}|h|^{-2\alpha}L(h^2), \quad \text{as} \quad h\to\infty \\ &\sim \frac{\sigma^2\beta^{2\alpha}\Gamma(\nu+\alpha)}{\nu^\alpha\Gamma(\nu)}|h|^{-2\alpha}L(h^2), \quad \text{as} \quad h\to\infty. \end{aligned}$$

Thus, the tail decays as $|h|^{-2\alpha}L(h^2)$ when $\alpha>0$.

□

S.4.3 Proof of Proposition 1

Proof. Let Φ_d denote the family of the continuous functions from $[0,\infty)$ to $\mathbb{R}$ that represent correlation functions of stationary and isotropic random processes on $\mathbb{R}^d$. Then the family Φ_d is nested satisfying $\Phi_1\supset\Phi_2\supset\cdots\supset\Phi_\infty$, where $\Phi_\infty:=\cap_{d\geq1}\Phi_d$ is the family of radial functions that are positive definite on any number of dimensions in Euclidean space.

The proof consists of two parts. We first show that the CH correlation function belongs to Φ_d, from $[0,\infty)$ to $\mathbb{R}$. Then we use Theorem 6.1 of Barndorff-Nielsen et al. (1982) to derive the tail behavior of the spectral density.

Note that Schoenberg (1938) shows that any member ψ that is in the family Φ_d can be written as a scale mixture with a probability measure F on $[0,\infty)$:

$$\psi(h)=\int_0^\infty h^{-(d-2)/2}\mathcal{J}_{(d-2)/2}(\omega h)dF(\omega), \quad h\geq 0,$$

where $\mathcal{J}_\nu(\cdot)$ is the ordinary Bessel function (see 9.1.20 of Abramowitz and Stegun, 1965). It is well-known (see Chapter 2 of Matérn, 1960) that the Matérn correlation is positive definite in any number of dimensions in Euclidean space and it is a member of Φ_∞ for any positive values of ϕ and ν. The CH correlation function as a scale mixture of the Matérn correlation is also a member of Φ_∞ (see the proof in Theorem 1). The Fourier transform of $f \in \Phi_d$, denoted by $\mathcal{F}(f)$, is available in a convenient form (Yaglom, 1987) with

$$\mathcal{F}(f)(\omega) = (2\pi)^{-d/2} \int_0^\infty (u\omega)^{-(d-2)/2} \mathcal{J}_{(d-2)/2}(u\omega) u^{d-1} f(u) du, \quad \omega \geq 0.$$

Notice that the Matérn covariance function (1) has spectral density

$$\begin{aligned} f_{\mathcal{M}}(\omega) &= (2\pi)^{-d/2} \int_0^\infty (\omega h)^{-(d-2)/2} \mathcal{J}_{(d-2)/2}(\omega h) h^{d-1} \mathcal{M}(h)\, dh, \\ &= \frac{\sigma^2 (\sqrt{2\nu}/\phi)^{2\nu}}{\pi^{d/2}((\sqrt{2\nu}/\phi)^2 + \omega^2)^{\nu+d/2}}. \end{aligned}$$

Thus, the spectral density of the covariance function $C(h)$ is

$$\begin{aligned} f(\omega) &= (2\pi)^{-d/2} \int_0^\infty (\omega h)^{-(d-2)/2} \mathcal{J}_{(d-2)/2}(\omega h) h^{d-1} \int_0^\infty \mathcal{M}(h; \nu, \phi, \sigma^2) \pi(\phi^2) d\phi^2 dh \\ &= \frac{\sigma^2 2^\nu \nu^\nu (\beta^2/2)^\alpha}{\Gamma(\alpha)} \int_0^\infty \frac{\phi^{-2\nu}}{\pi^{d/2}(2\nu\phi^{-2} + \omega^2)^{\nu+d/2}} \phi^{-2(\alpha+1)} \exp\{-\beta^2/(2\phi^2)\} d\phi^2 \\ &= \frac{\sigma^2 2^{\nu-\alpha} \nu^\nu \beta^{2\alpha}}{\pi^{d/2}\Gamma(\alpha)} \int_0^\infty (2\nu\phi^{-2} + \omega^2)^{-\nu-d/2} \phi^{-2(\nu+\alpha+1)} \exp\{-\beta^2/(2\phi^2)\} d\phi^2. \end{aligned}$$

where the above spectral density is finite for $\alpha > d/2$ and is infinite for $\alpha \in (0, d/2]$.

To derive the tail behavior, we make the change of variable $\phi^2 = \beta^2 t/\omega^2$. The spectral density above can be expressed as

$$\begin{aligned} f(\omega) &= \frac{\sigma^2 2^{\nu-\alpha} \nu^\nu \beta^{2\alpha}}{\pi^{d/2}\Gamma(\alpha)} \omega^{2\alpha-d} \int_0^\infty ((2\nu/\beta^2)t^{-1} + 1)^{-(\nu+d/2)} t^{-(\nu+\alpha+1)} \exp\{-\omega^2/(2t)\} dt \\ &= \frac{\sigma^2 2^{\nu-\alpha} \nu^\nu}{\pi^{d/2}\beta^{2\nu}\Gamma(\alpha)} (2\pi)^{1/2} \omega^{2\alpha-d} \int_0^\infty \left(\frac{t}{2\nu/\beta^2 + t}\right)^{(\nu+d/2)} t^{(-\nu-\alpha+1/2)-1} \frac{1}{\sqrt{2\pi t}} \\ &\quad \times \exp\{-\omega^2/(2t)\} dt. \end{aligned}$$

We define

$$L(x) := \left\{ \frac{x}{x + \beta^2/(2\nu)} \right\}^{\nu+d/2}.$$

Then $L(x)$ is a slowly varying function at ∞. The above integral is also a Gaussian scale mixture. Thus, an application of Theorem 6.1 of Barndorff-Nielsen et al. (1982) yields that as $|\omega| \to \infty$,

$$\begin{aligned} f(\omega) &\sim \frac{\sigma^2 2^{\nu-\alpha}\nu^\nu}{\pi^{d/2}\beta^{2\nu}\Gamma(\alpha)}(2\pi)^{1/2}\omega^{2\alpha-d}(2\pi)^{-1/2}2^{1/2+(\nu+\alpha-1/2))}|\omega|^{-2(\nu+\alpha-1/2)-1}L(\omega^2) \\ &\sim \frac{\sigma^2 2^{2\nu}\nu^\nu\Gamma(\nu+\alpha)}{\pi^{d/2}\beta^{2\nu}\Gamma(\alpha)}\omega^{-(2\nu+d)}L(\omega^2). \end{aligned}$$

□

S.4.4 Proof of Theorem 3

Proof. Let $f_i(\omega), i = 1, 2$ be the spectral densities with parameters $\{\sigma_i^2, \beta_i, \alpha_i, \nu\}$ for two covariance functions $C_1(\cdot), C_2(\cdot)$. The condition (S.1) says the spectral density $f_i(\omega)$ is bounded at zero and ∞ when $\omega \to \infty$. In fact, the boundness of f_i near zero follows from the assumption that $\alpha_i > d/2$. Let $\lambda = 2\nu + d$. Then, one can show that

$$\lim_{\omega\to\infty} f_1(\omega)|\omega|^{2\nu+d} = \frac{\sigma_1^2(\beta_1^2/2)^{-\nu}(2\nu)^\nu\Gamma(\nu+\alpha_1)}{\pi^{d/2}\Gamma(\alpha_1)}.$$

Thus, the condition (S.1) is satisfied.

We first show the sufficiency. Assume that the condition in Equation (5) holds. To prove the equivalence of two measures, it suffices to show that the condition (S.2) is satisfied. Notice that as $\omega \to \infty$,

$$\left|\frac{f_1(\omega) - f_2(\omega)}{f_1(\omega)}\right| = \left|\frac{\{\omega^2 + \beta_2^2/(2\nu)\}^{-(\nu+d/2)}}{\{\omega^2 + \beta_1^2/(2\nu)\}^{-(\nu+d/2)}} - 1\right|$$

$$
\begin{aligned}
&\leq \omega^{-(2\nu+d)} \left|\{\omega^2 + \beta_2^2/(2\nu)\}^{\nu+d/2} - \{\omega^2 + \beta_1^2/(2\nu)\}^{\nu+d/2}\right| \\
&\leq \left|\{1 + (\beta_2^2/2\nu)\omega^{-2}\}^{\nu+d/2} - \{1 + (\beta_1^2/2\nu)\omega^{-2}\}^{\nu+d/2}\right| \\
&\leq \big|\{1 + (\nu + d/2)(\beta_2^2/2\nu)\omega^{-2} + O(\omega^{-4})\} \\
&\quad - \{1 + (\nu + d/2)(\beta_1^2/2\nu)\omega^{-2} + O(\omega^{-4})\}\big| \\
&\leq |\beta_1^2 - \beta_2^2|(\nu + d/2)/(2\nu)\omega^{-2} + O(\omega^{-4}).
\end{aligned}
$$

The integral in (S.2) is finite for $d = 1, 2, 3$. Therefore, the two measures are equivalent.

It remains to show the necessary condition. Suppose that

$$\frac{\sigma_1^2\beta_1^{-2\nu}\Gamma(\nu+\alpha_1)}{\Gamma(\alpha_1)} \neq \frac{\sigma_2^2\beta_2^{-2\nu}\Gamma(\nu+\alpha_2)}{\Gamma(\alpha_2)}.$$

Let

$$\sigma_0^2 := \sigma_2^2 \frac{\beta_2^{-2\nu}\Gamma(\alpha_1)\Gamma(\nu+\alpha_2)}{\beta_1^{-2\nu}\Gamma(\alpha_2)\Gamma(\nu+\alpha_1)}.$$

Then

$$\frac{\sigma_0^2\beta_1^{-2\nu}\Gamma(\nu+\alpha_1)}{\Gamma(\alpha_1)} = \frac{\sigma_2^2\beta_2^{-2\nu}\Gamma(\nu+\alpha_2)}{\Gamma(\alpha_2)}.$$

Thus, the two covariances $C(h;\nu,\alpha_1,\beta_1,\sigma_0^2)$ and $C(h;\nu,\alpha_1,\beta_1,\sigma_1^2)$ define two equivalent measures. It remains to show that $C(h;\nu,\alpha_1,\beta_1,\sigma_0^2)$ and $C(h;\nu,\alpha_2,\beta_2,\sigma_2^2)$ define two equivalent Gaussian measures, which follows from the proof in Theorem 2 of Zhang (2004). □

S.4.5 Proof of Theorem 4

Proof. Let $k_1 = \sigma_1^2 \frac{2^{2\nu}\nu^\nu\Gamma(\nu+\alpha)}{\pi^{d/2}\beta^{2\nu}\Gamma(\alpha)}$ and $k_2 = \sigma_2^2(2\nu)^\nu\phi^{-2\nu}/\pi^{d/2}$. Then the condition in Equation (6) implies that $k_1 = k_2$. It follows that as $|\omega| \to \infty$,

$$\left|\frac{f_1(\omega) - f_2(\omega)}{f_1(\omega)}\right| = \left|\frac{k_2}{k_1}(\omega^2 + 2\nu/\phi^2)^{-(\nu+d/2)}(\omega^2 + \beta^2/(2\nu))^{(\nu+d/2)} - 1\right|$$

$$
\begin{aligned}
&= (\omega^2 + 2\nu/\phi^2)^{-(\nu+d/2)} \times \big| k_2/k_1 (\omega^2 + 2\nu/\phi^2)^{\nu+d/2} \\
&\quad -(\omega^2 + \beta^2/(2\nu))^{(\nu+d/2)} \big| \\
&\le \omega^{-(2\nu+d)} \times \big| (\omega^2 + 2\nu/\phi^2)^{\nu+d/2} - (\omega^2 + \beta^2/(2\nu))^{(\nu+d/2)} \big| \\
&\le \big| \{1 + (2\nu/\phi^2)\omega^{-2}\}^{-(\nu+d/2)} - \{1 + \beta^2/(2\nu)\omega^{-2}\}^{(\nu+d/2)} \big| \\
&\le \big| \{1 + (2\nu/\phi^2)(\nu + d/2)\omega^{-2} + O(\omega^{-4})\} - \{1 + (\beta^2/(2\nu))(\nu + d/2)\omega^{-2} \\
&\quad + O(\omega^{-4})\} \big| . \\
&\le |2\nu/\phi^2 - \beta^2/(2\nu)|(\nu + d/2)\omega^{-2} + O(\omega^{-4}).
\end{aligned}
$$

The integral in (S.2) is finite for $d = 1, 2, 3$. Therefore, these two measures are equivalent. □

S.4.6 Proof of Theorem 5

Proof. Note that the CH covariance function $C(h; \nu, \alpha, \beta, \sigma^2)$ is a continuous function of the covariance parameters α, β, σ^2 over their natural parameter space $\{(\sigma^2, \alpha, \beta) : \sigma^2 > 0, \alpha > 0, \beta > 0\}$, and hence the likelihood function is also a continuous function over this natural parameter space.

For case (a), it follows from the continuity of the likelihood function and the assumption in case (a) that $\hat{\beta}_n \in [\beta_L, \beta_U]$ for all n. Applying Lemma 4 yields that $\hat{c}_n(\alpha, \beta_L) \le \hat{c}_n(\alpha, \hat{\beta}_n) \le \hat{c}_n(\alpha, \hat{\beta}_U)$. The result thus follows from Lemma 3 immediately.

For case (b), it follows from the continuity of the likelihood function and the assumption in case (b) that $\hat{\alpha}_n \in [\alpha_L, \alpha_U]$ for all n. Applying Lemma 4 yields that $\hat{c}_n(\alpha_U, \beta) \le \hat{c}_n(\hat{\alpha}_n, \beta) \le \hat{c}_n(\alpha_U, \beta)$. The result thus follows from Lemma 3 immediately.

For case (c), it follows from the continuity of the likelihood function and the assumption in case (c) that $\hat{\alpha}_n \in [\alpha_L, \alpha_U]$ and $\hat{\beta}_n \in [\beta_L, \beta_U]$ for all n. According to Lemma 4, $\hat{c}_n(\alpha_U, \beta_L) \le \hat{c}_n(\alpha_U, \hat{\beta}_n) \le \hat{c}_n(\hat{\alpha}_n, \hat{\beta}_n)$ and $\hat{c}_n(\hat{\alpha}_n,\ \hat{\beta}_n) \le \hat{c}_n(\alpha_L, \hat{\beta}_n) \le \hat{c}_n(\alpha_L, \beta_U)$. The result thus follows from Lemma 3 immediately. □

S.4.7 Proof of Theorem 6

Proof. Part (a) and Part (b) can be proven by applying Theorem 1 and Theorem 2 of Stein (1993). Let $f_i(\omega)$ be the spectral density of the CH class $C(h;\nu,\alpha_i,\beta_i,\sigma_i^2)$ with $i=1,2$. Note that $\lim_{\omega\to\infty} f_i(\omega)|\omega|^{2\nu+d}$ is finite. If the condition in Equation (5) is satisfied, then,

$$\lim_{\omega\to\infty}\frac{f_2(\omega)}{f_1(\omega)}=\lim_{\omega\to\infty}\frac{f_2(\omega)|\omega|^{2\nu+d}}{f_1(\omega)|\omega|^{2\nu+d}}=1.$$

The proof of Part (c) is analogous to the proof of Theorem 4 in Kaufman and Shaby (2013). Let

$$\sigma_1^2:=\sigma_0^2(\beta_1/\beta_0)^{2\nu}\frac{\Gamma(\nu+\alpha_0)\Gamma(\alpha_1)}{\Gamma(\nu+\alpha_1)\Gamma(\alpha_0)}.$$

Then $\mathcal{P}_0$ and $\mathcal{P}_1$ define two equivalent measures. We write

$$\frac{\mathrm{Var}_{\nu,\boldsymbol{\theta}_1,\hat{\sigma}_n^2}\{\hat{Z}_n(\boldsymbol{\theta}_1)-Z(\mathbf{s}_0)\}}{\mathrm{Var}_{\nu,\boldsymbol{\theta}_0,\sigma_0^2}\{\hat{Z}_n(\boldsymbol{\theta}_1)-Z(\mathbf{s}_0)\}}=\frac{\mathrm{Var}_{\nu,\boldsymbol{\theta}_1,\hat{\sigma}_n^2}\{\hat{Z}_n(\boldsymbol{\theta}_1)-Z(\mathbf{s}_0)\}}{\mathrm{Var}_{\nu,\boldsymbol{\theta}_1,\sigma_1^2}\{\hat{Z}_n(\boldsymbol{\theta}_1)-Z(\mathbf{s}_0)\}}\frac{\mathrm{Var}_{\nu,\boldsymbol{\theta}_1,\sigma_1^2}\{\hat{Z}_n(\boldsymbol{\theta}_1)-Z(\mathbf{s}_0)\}}{\mathrm{Var}_{\nu,\boldsymbol{\theta}_0,\sigma_0^2}\{\hat{Z}_n(\boldsymbol{\theta}_1)-Z(\mathbf{s}_0)\}}.$$

According to Part (b) of Theorem 6, it suffices to show that almost surely under $\mathcal{P}_1$,

$$\frac{\mathrm{Var}_{\nu,\boldsymbol{\theta}_1,\hat{\sigma}_n^2}\{\hat{Z}_n(\boldsymbol{\theta}_1)-Z(\mathbf{s}_0)\}}{\mathrm{Var}_{\nu,\boldsymbol{\theta}_1,\sigma_1^2}\{\hat{Z}_n(\boldsymbol{\theta}_1)-Z(\mathbf{s}_0)\}}\to 1.$$

By Equation (9),

$$\frac{\mathrm{Var}_{\nu,\boldsymbol{\theta}_1,\hat{\sigma}_n^2}\{\hat{Z}_n(\boldsymbol{\theta}_1)-Z(\mathbf{s}_0)\}}{\mathrm{Var}_{\nu,\boldsymbol{\theta}_1,\sigma_1^2}\{\hat{Z}_n(\boldsymbol{\theta}_1)-Z(\mathbf{s}_0)\}}=\frac{\hat{\sigma}_n^2}{\sigma_1^2}.$$

Note that under $\mathcal{P}_1$, we have $\hat{\sigma}_n^2\sim(\sigma_0^2/n)\chi_n$, and hence $\hat{\sigma}_n^2$ converges almost surely to σ_0^2 as $n\to\infty$. As $\mathcal{P}_0$ is equivalent to $\mathcal{P}_1$, It follows from Lemma 3 that $\hat{\sigma}_n^2\to\sigma_1^2$, almost surely under $\mathcal{P}_0$. ☐

S.4.8 Proof of Theorem 7

Proof. Let $f_0(\omega)$ be the spectral density of the Matérn covariance function $\mathcal{M}(h;\nu,\phi,\sigma_0^2)$ and $f_1(\omega)$ be the spectral density of the covariance function $C(h;\nu,\alpha,\beta,\sigma_1^2)$. Notice that the spectral density of the Matérn covariance satisfies the condition (S.1). It suffices to show that $\lim_{\omega\to\infty} f_1(\omega)/f_0(\omega) = 1$. Let $k_0 = \sigma_0^2\phi^{-2\nu}$ and $k_1 = \sigma_1^2\ (\beta^2/2)^{-\nu}\Gamma(\nu+\alpha)/\Gamma(\alpha)$. If $k_0 = k_1$, it follows that

$$\lim_{\omega\to\infty}\frac{f_1(\omega)}{f_0(\omega)} = \lim_{\omega\to\infty}\frac{f_1(\omega)|\omega|^{2\nu+d}}{f_0(\omega)|\omega|^{2\nu+d}} = \lim_{\omega\to\infty}\frac{k_1}{k_0}\left(2\nu\phi^{-2}\omega^{-2}+1\right)^{\nu+d/2} = k_1/k_0 = 1.$$

Thus, the covariance function $C(h;\nu,\alpha,\beta,\sigma_1^2)$ yields an asymptotically equivalent BLP as the Matérn covariance $\mathcal{M}(h;\nu,\phi,\sigma_0^2)$. □

S.5 Examples to Illustrate Asymptotic Normality

As shown in Section 3.2, each individual parameter in the CH model cannot be estimated consistently, however, the microergodic parameter can be estimated consistently.

To study the finite sample performance of the asymptotic properties of MLE for the microergodic parameter, we simulate 1000 realizations from a zero-mean Gaussian process with the CH class over 100-by-100 regular grid in the unit domain $\mathcal{D} = [0,1] \times [0,1]$. As there are no clear guidelines to pick the sample sizes such that the finite sample performances can appropriately reflect the asymptotic results, we randomly select $n = 4000, 5000, 6000$ locations from these 10,000 grid points. The variance parameter is fixed at 1 for all realizations. We consider two different values for the smoothness parameter ν at 0.5 and 1.5, three different values for the tail decay parameter α at 0.5, 2 and 5. The scale parameter β is chosen such that the effective range is 0.6 or 0.9. Although all the theoretical results in Section 3 are valid for $\alpha > d/2$, we also run the simulation setting with $\alpha = 0.5$ to see whether there is any interesting numerical results compared to cases where $\alpha > d/2$.

Let $C(h; \nu, \alpha_0, \beta_0, \sigma_0^2)$ be the true covariance. We use $\hat{c}_n(\boldsymbol{\theta})$ to denote the maximum likelihood estimator of the microergodic parameter $c(\boldsymbol{\theta}_0) = \sigma_0^2 \beta_0^{-2\nu} \Gamma(\nu + \alpha_0)/\Gamma(\alpha_0)$ for any $\boldsymbol{\theta}$. Then the 95% confidence interval for $c(\boldsymbol{\theta}_0)$ is given by $\hat{c}_n(\boldsymbol{\theta}) \pm 1.96\sqrt{2\hat{c}_n(\boldsymbol{\theta})^2/n}$. Lemma 3 and Theorem 5 show that this interval is asymptotically valid when n is large and $\alpha > d/2$ for (1) arbitrarily fixed $\boldsymbol{\theta}$, (2) $\boldsymbol{\theta} = (\alpha, \hat{\beta}_n)$, (3) $\boldsymbol{\theta} = (\hat{\alpha}_n, \beta)$ and (4) $\boldsymbol{\theta} = (\hat{\alpha}_n, \hat{\beta}_n)$. In this simulation study, we primarily focus on the finite sample performance of $\hat{c}_n(\boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\alpha_0, \sqrt{0.5}\beta_0)$, $\boldsymbol{\theta} = (\alpha_0, \beta_0)$, $\boldsymbol{\theta} = (\alpha_0, \sqrt{2}\beta_0)$, $\boldsymbol{\theta} = (\alpha_0, \hat{\beta}_n)$, and $\boldsymbol{\theta} = (\hat{\alpha}_n, \hat{\beta}_n)$. Exhaustive simulations with all other settings of $\boldsymbol{\theta}$ is considered future work. Let

$$\xi := \frac{\sqrt{n}\{\hat{c}_n(\boldsymbol{\theta}) - c(\boldsymbol{\theta}_0)\}}{\sqrt{2}c(\boldsymbol{\theta}_0)}.$$

Then ξ should asymptotically follow the standard normal distribution. Based on these 1000 realizations, we compute the empirical coverage probability of the 95% percentile confidence interval, bias and root-mean-square error (RMSE) for $c(\boldsymbol{\theta}_0)$ and compare the quantiles of ξ with the standard normal quantiles.

The results are reported in Table S.1, Table S.2 and Table S.3 of the Supplementary Material. They can be summarized as follows. When the true parameters are used, i.e., $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, as expected, the sampling distribution of $\hat{c}_n(\boldsymbol{\theta}_0)$ gives the best normal approximation and converges to the asymptotic distribution in Lemma 3 when n increases. The sampling distribution of $\hat{c}_n(\boldsymbol{\theta})$ can be highly biased and approach to the truth can be very slow with increase in n. Fixing β at a larger value gives better empirical results than fixing β at a small value. When the scale parameter is chosen to be its maximum likelihood estimator, i.e., $\beta = \hat{\beta}_n$, the sampling distribution of $\hat{c}_n(\alpha, \hat{\beta}_n)$ converges to the asymptotic distribution given in Theorem 5 as n increases. When α is small, e.g., $\alpha = 0.5$, the sampling distributions of $\hat{c}_n(\boldsymbol{\theta})$, with $(\alpha_0, \sqrt{0.5}\beta_0)$, $(\alpha_0, \sqrt{2}\beta_0)$, $(\alpha_0, \hat{\beta}_n)$ and $(\hat{\alpha}_n, \hat{\beta}_n)$ substituted for $\boldsymbol{\theta}$, has noticeable biases. As the tail decay parameter or the effective range increases, the sampling distributions of $\hat{c}_n(\boldsymbol{\theta})$ have smaller biases. As ν becomes smaller, the sampling distributions of $\hat{c}_n(\boldsymbol{\theta})$ approaches the truth better with increase in n. When $\nu = 0.5$ and $\alpha \in \{2, 5\}$, these sampling distributions have negligible biases as n increases. When both α and β are substituted by their maximum likelihood estimator, the sampling distribution of $\hat{c}_n(\boldsymbol{\theta})$ has smaller bias and gives better approximation to the true asymptotic distribution given in Theorem 5 as n increases for $\alpha > d/2 = 1$.

When α is fixed at its true value and β is estimated by maximum likelihood method, the MLE of the microergodic parameter, $\hat{c}_n(\alpha, \hat{\beta}_n)$, gives better finite sample performance than the cases where β is misspecified. When both α and β are estimated by maximum likelihood method, the MLE of the microergodic parameter, $\hat{c}_n(\hat{\alpha}_n, \hat{\beta}_n)$, also gives better finite sample performance than the cases where β is misspecified and α is fixed at its true value. One would also expect that this is true when either α or β is misspecified at incorrect values. In general, the MLE of the microergodic parameter has better finite sampler performance than those with any individual parameter fixed at an incorrect value in the microergodic parameter. Theorem 5 requires $\alpha > d/2$ in order to derive asymptotic results for $\hat{c}_n(\hat{\alpha}_n, \hat{\beta}_n)$. However, it is interesting to observe from these simulation results that $\hat{c}_n(\boldsymbol{\theta})$ seems to converge to a normal distribution even when $\alpha < d/2$, i.e., when $\alpha = 0.5$. It is an open problem to determine the exact distribution that the maximum likelihood estimator $\hat{c}_n(\hat{\alpha}_n, \hat{\beta}_n)$ of the microergodic parameter converges to asymptotically when α and β are substituted with their maximum likelihood estimators for true $\alpha_0 \in (0, d/2]$.

Table S.1. Percentiles of ξ and CVG, bias, and RMSE of $\hat{c}_n(\boldsymbol{\theta})$ when $\alpha_0 = 0.5$.

Settings		5%	25%	50%	75%	95%	CVG	bias	RMSE
$\mathcal{N}(0,1)$		-1.6449	-0.6749	0	0.6749	1.6449	0.95	0	
				$ER = 0.6, \nu = 0.5$					
$\boldsymbol{\theta}$									
$\alpha = \alpha_0, \beta = \beta_0$	$n = 4000$	-1.449	-0.542	0.009	0.686	1.767	0.955	0.020	0.327
	$n = 5000$	-1.469	-0.665	-0.077	0.696	1.573	0.965	-0.003	0.289
	$n = 6000$	-1.705	-0.618	0.056	0.662	1.847	0.929	0.010	0.280
$\alpha = \alpha_0, \beta = \sqrt{0.5}\beta_0$	$n = 4000$	2.113	3.098	3.730	4.424	5.549	0.044	1.259	1.308
	$n = 5000$	2.129	2.930	3.578	4.439	5.347	0.040	1.097	1.140
	$n = 6000$	1.798	3.015	3.705	4.397	5.606	0.071	1.005	1.049
$\alpha = \alpha_0, \beta = \sqrt{2}\beta_0$	$n = 4000$	-3.471	-2.608	-2.073	-1.420	-0.394	0.415	-0.676	0.746
	$n = 5000$	-3.480	-2.693	-2.114	-1.395	-0.519	0.404	-0.611	0.676
	$n = 6000$	-3.688	-2.623	-1.967	-1.376	-0.214	0.462	-0.543	0.606
$\alpha = \alpha_0, \beta = \hat{\beta}_n$	$n = 4000$	-1.871	-0.711	0.095	1.000	2.244	0.889	0.047	0.428
	$n = 5000$	-1.912	-0.767	0.022	0.881	2.134	0.881	0.013	0.371
	$n = 6000$	-2.016	-0.760	0.096	0.862	2.097	0.879	0.019	0.343
$\alpha = \hat{\alpha}_n, \beta = \hat{\beta}_n$	$n = 4000$	-1.778	-0.875	0.000	0.925	2.382	0.887	0.030	0.446
	$n = 5000$	-2.129	-0.816	0.026	0.893	2.227	0.870	0.019	0.395
	$n = 6000$	-2.268	-0.911	-0.015	0.865	2.117	0.875	-0.006	0.363
				$ER = 0.6, \nu = 1.5$					
$\boldsymbol{\theta}$									
$\alpha = \alpha_0, \beta = \beta_0$	$n = 4000$	-1.654	-0.604	-0.014	0.701	1.776	0.949	12.650	370.7
	$n = 5000$	-1.430	-0.687	-0.046	0.672	1.576	0.969	0.283	312.0
	$n = 6000$	-1.731	-0.649	0.070	0.710	1.740	0.929	7.182	307.5
$\alpha = \alpha_0, \beta = \sqrt{0.5}\beta_0$	$n = 4000$	26.20	27.77	28.82	30.03	31.76	0.000	10495	10513
	$n = 5000$	27.09	28.35	29.37	30.55	31.99	0.000	9567	9581
	$n = 6000$	27.22	28.79	29.85	30.89	32.68	0.000	8860	8874
$\alpha = \alpha_0, \beta = \sqrt{2}\beta_0$	$n = 4000$	-13.70	-12.93	-12.48	-11.99	-11.21	0.000	-4526	4534
	$n = 5000$	-13.99	-13.35	-12.90	-12.36	-11.62	0.000	-4177	4184
	$n = 6000$	-14.47	-13.61	-13.11	-12.58	-11.80	0.000	-3886	3893
$\alpha = \alpha_0, \beta = \hat{\beta}_n$	$n = 4000$	-2.993	-1.121	0.172	1.515	3.505	0.670	72.52	732.5
	$n = 5000$	-2.823	-1.155	0.146	1.452	3.398	0.700	49.49	624.8
	$n = 6000$	-3.068	-1.090	0.235	1.433	3.093	0.733	44.59	543.1
$\alpha = \hat{\alpha}_n, \beta = \hat{\beta}_n$	$n = 4000$	-3.887	-1.656	0.061	1.681	4.142	0.565	9.059	895.5
	$n = 5000$	-3.607	-1.643	0.055	1.497	4.142	0.592	16.27	772.1
	$n = 6000$	-4.107	-1.678	-0.206	1.488	3.774	0.592	-70.59	832.3
				$ER = 0.9, \nu = 0.5$					
$\boldsymbol{\theta}$									
$\alpha = \alpha_0, \beta = \beta_0$	$n = 4000$	-1.589	-0.557	-0.013	0.669	1.748	0.955	0.007	0.220
	$n = 5000$	-1.429	-0.654	0.065	0.759	1.683	0.978	0.012	0.190
	$n = 6000$	-1.512	-0.591	0.004	0.702	1.768	0.943	0.009	0.179
$\alpha = \alpha_0, \beta = \sqrt{0.5}\beta_0$	$n = 4000$	0.628	1.679	2.278	2.967	4.052	0.399	0.513	0.563
	$n = 5000$	0.727	1.546	2.306	2.994	3.958	0.420	0.454	0.496
	$n = 6000$	0.548	1.543	2.200	2.888	4.013	0.445	0.403	0.444
$\alpha = \alpha_0, \beta = \sqrt{2}\beta_0$	$n = 4000$	-2.812	-1.808	-1.259	-0.595	0.440	0.769	-0.272	0.346
	$n = 5000$	-2.620	-1.846	-1.170	-0.478	0.379	0.760	-0.229	0.295
	$n = 6000$	-2.635	-1.756	-1.187	-0.477	0.586	0.799	-0.205	0.270
$\alpha = \alpha_0, \beta = \hat{\beta}_n$	$n = 4000$	-1.856	-0.688	0.087	0.911	1.946	0.905	0.021	0.262
	$n = 5000$	-1.587	-0.696	0.062	0.822	1.930	0.926	0.018	0.220
	$n = 6000$	-1.646	-0.589	0.045	0.833	2.008	0.918	0.020	0.202
$\alpha = \hat{\alpha}_n, \beta = \hat{\beta}_n$	$n = 4000$	-1.876	-0.748	0.082	0.882	2.157	0.887	0.016	0.276
	$n = 5000$	-1.865	-0.692	0.023	0.853	1.994	0.902	0.014	0.233
	$n = 6000$	-1.884	-0.744	0.006	0.901	1.978	0.904	0.008	0.213
				$ER = 0.9, \nu = 1.5$					
$\boldsymbol{\theta}$									
$\alpha = \alpha_0, \beta = \beta_0$	$n = 4000$	-1.598	-0.618	-0.015	0.663	1.747	0.958	2.284	106.8
	$n = 5000$	-1.426	-0.647	0.063	0.774	1.711	0.977	6.598	92.45
	$n = 6000$	-1.660	-0.623	0.059	0.685	1.743	0.945	2.287	87.41
$\alpha = \alpha_0, \beta = \sqrt{0.5}\beta_0$	$n = 4000$	17.15	18.51	19.42	20.44	21.94	0.000	2096	2102
	$n = 5000$	17.34	18.50	19.50	20.42	21.83	0.000	1877	1881
	$n = 6000$	17.06	18.45	19.36	20.24	21.70	0.000	1699	1703
$\alpha = \alpha_0, \beta = \sqrt{2}\beta_0$	$n = 4000$	-10.02	-9.207	-8.722	-8.151	-7.284	0.000	-934.2	938.3
	$n = 5000$	-9.964	-9.263	-8.691	-8.092	-7.331	0.000	-8.835	839.5
	$n = 6000$	-10.13	-9.280	-8.717	-8.159	-7.26	0.000	-766.2	769.8
$\alpha = \alpha_0, \beta = \hat{\beta}_n$	$n = 4000$	-2.455	-0.993	0.029	1.282	2.953	0.771	15.62	180.4
	$n = 5000$	-2.224	-1.038	-0.031	1.140	2.706	0.789	8.661	151.8
	$n = 6000$	-2.259	-0.918	0.082	1.152	2.586	0.803	11.20	131.2
$\alpha = \hat{\alpha}_n, \beta = \hat{\beta}_n$	$n = 4000$	-3.178	-1.316	-0.002	1.215	3.289	0.691	1.887	208.8
	$n = 5000$	-3.055	-1.211	-0.003	1.229	3.129	0.708	2.440	178.3
	$n = 6000$	-2.820	-1.285	0.006	1.236	3.051	0.710	0.875	156.7

Table S.2. Percentiles of ξ and CVG, bias, and RMSE of $\hat{c}_n(\boldsymbol{\theta})$ when $\alpha_0 = 2$.

Settings		5%	25%	50%	75%	95%	CVG	bias	RMSE
$\mathcal{N}(0,1)$		-1.6449	-0.6749	0	0.6749	1.6449	0.95	0	
$ER = 0.6, \nu = 0.5$									
$\boldsymbol{\theta}$									
$\alpha = \alpha_0, \beta = \beta_0$	$n = 4000$	-1.557	-0.604	-0.013	0.691	1.759	0.954	0.004	0.099
	$n = 5000$	-1.442	-0.614	0.003	0.723	1.575	0.962	0.002	0.086
	$n = 6000$	-1.689	-0.462	0.093	0.728	1.970	0.947	0.003	0.084
$\alpha = \alpha_0, \beta = \sqrt{0.5}\beta_0$	$n = 4000$	-1.072	-0.128	0.486	1.179	2.264	0.921	0.052	0.113
	$n = 5000$	-0.999	-0.179	0.493	1.183	2.047	0.939	0.043	0.097
	$n = 6000$	-1.315	-0.010	0.556	1.184	2.396	0.929	0.041	0.094
$\alpha = \alpha_0, \beta = \sqrt{2}\beta_0$	$n = 4000$	-1.801	-0.860	-0.258	0.440	1.505	0.949	-0.021	0.101
	$n = 5000$	-1.680	-0.840	-0.232	0.479	1.347	0.946	-0.018	0.088
	$n = 6000$	-1.880	-0.681	-0.136	0.517	1.756	0.931	-0.012	0.084
$\alpha = \alpha_0, \beta = \hat{\beta}_n$	$n = 4000$	-1.616	-0.600	0.040	0.758	1.796	0.954	0.008	0.103
	$n = 5000$	-1.443	-0.583	0.070	0.752	1.705	0.962	0.006	0.088
	$n = 6000$	-1.564	-0.505	0.171	0.774	1.941	0.938	0.008	0.087
$\alpha = \hat{\alpha}_n, \beta = \hat{\beta}_n$	$n = 4000$	-1.576	-0.546	0.140	0.798	1.880	0.944	0.014	0.104
	$n = 5000$	-1.426	-0.565	0.094	0.785	1.747	0.956	0.009	0.089
	$n = 6000$	-1.595	-0.614	0.079	0.764	1.882	0.953	0.007	0.085
$ER = 0.6, \nu = 1.5$									
$\boldsymbol{\theta}$									
$\alpha = \alpha_0, \beta = \beta_0$	$n = 4000$	-1.567	-0.624	-0.010	0.689	1.764	0.952	0.103	2.513
	$n = 5000$	-1.469	-0.633	0.005	0.734	1.620	0.958	0.083	2.200
	$n = 6000$	-1.729	-0.614	0.016	0.592	1.646	0.953	-0.013	2.027
$\alpha = \alpha_0, \beta = \sqrt{0.5}\beta_0$	$n = 4000$	1.226	2.227	2.885	3.622	4.748	0.215	7.351	7.840
	$n = 5000$	1.056	2.005	2.772	3.469	4.427	0.257	6.145	6.586
	$n = 6000$	0.725	1.861	2.602	3.180	4.350	0.296	5.225	5.657
$\alpha = \alpha_0, \beta = \sqrt{2}\beta_0$	$n = 4000$	-2.801	-1.876	-1.283	-0.593	0.433	0.749	-3.103	3.948
	$n = 5000$	-2.590	-1.829	-1.185	-0.497	0.385	0.765	-2.612	3.379
	$n = 6000$	-2.807	-1.771	-1.128	-0.565	0.493	0.779	-2.350	3.073
$\alpha = \alpha_0, \beta = \hat{\beta}_n$	$n = 4000$	-1.613	-0.649	0.093	0.790	1.923	0.928	0.223	2.701
	$n = 5000$	-1.417	-0.648	0.046	0.818	1.821	0.957	0.197	2.332
	$n = 6000$	-1.623	-0.655	0.056	0.691	1.793	0.954	0.071	2.111
$\alpha = \hat{\alpha}_n, \beta = \hat{\beta}_n$	$n = 4000$	-1.558	-0.598	0.095	0.811	1.977	0.921	0.301	2.759
	$n = 5000$	-1.543	-0.594	0.040	0.798	1.827	0.950	0.186	2.342
	$n = 6000$	-1.604	-0.569	0.090	0.763	1.827	0.946	0.198	2.132
$ER = 0.9, \nu = 0.5$									
$\boldsymbol{\theta}$									
$\alpha = \alpha_0, \beta = \beta_0$	$n = 4000$	-1.574	-0.594	-0.026	0.666	1.776	0.952	0.002	0.066
	$n = 5000$	-1.458	-0.664	-0.052	0.638	1.547	0.962	-0.001	0.057
	$n = 6000$	-1.742	-0.548	0.145	0.809	1.784	0.942	0.005	0.055
$\alpha = \alpha_0, \beta = \sqrt{0.5}\beta_0$	$n = 4000$	-1.319	-0.321	0.254	0.930	2.042	0.938	0.020	0.069
	$n = 5000$	-1.220	-0.411	0.185	0.904	1.817	0.962	0.014	0.059
	$n = 6000$	-1.567	-0.366	0.385	1.022	2.052	0.925	0.017	0.058
$\alpha = \alpha_0, \beta = \sqrt{2}\beta_0$	$n = 4000$	-1.704	-0.723	-0.158	0.524	1.635	0.950	-0.007	0.066
	$n = 5000$	-1.579	-0.786	-0.173	0.517	1.424	0.963	-0.007	0.057
	$n = 6000$	-1.862	-0.653	0.024	0.700	1.646	0.950	-0.001	0.054
$\alpha = \alpha_0, \beta = \hat{\beta}_n$	$n = 4000$	-1.586	-0.578	0.022	0.695	1.799	0.948	0.005	0.068
	$n = 5000$	-1.440	-0.615	-0.006	0.696	1.648	0.959	0.001	0.058
	$n = 6000$	-1.653	-0.450	0.244	0.880	1.646	0.930	0.009	0.055
$\alpha = \hat{\alpha}_n, \beta = \hat{\beta}_n$	$n = 4000$	-1.572	-0.550	0.123	0.789	1.862	0.950	0.008	0.068
	$n = 5000$	-1.378	-0.541	0.097	0.798	1.782	0.957	0.007	0.059
	$n = 6000$	-1.659	-0.595	0.055	0.727	1.786	0.954	0.003	0.055
$ER = 0.9, \nu = 1.5$									
$\boldsymbol{\theta}$									
$\alpha = \alpha_0, \beta = \beta_0$	$n = 4000$	-1.589	-0.595	-0.015	0.700	1.748	0.955	0.028	0.744
	$n = 5000$	-1.454	-0.668	-0.029	0.673	1.531	0.966	-0.006	0.638
	$n = 6000$	-1.701	-0.671	-0.098	0.572	1.747	0.940	-0.031	0.652
$\alpha = \alpha_0, \beta = \sqrt{0.5}\beta_0$	$n = 4000$	-0.026	0.961	1.579	2.307	3.369	0.675	1.205	1.434
	$n = 5000$	-0.117	0.805	1.443	2.165	3.003	3.705	0.978	1.183
	$n = 6000$	-0.371	0.705	1.327	1.957	3.189	0.764	0.813	1.037
$\alpha = \alpha_0, \beta = \sqrt{2}\beta_0$	$n = 4000$	-2.288	-1.288	-0.722	-0.002	1.009	0.886	-0.498	0.886
	$n = 5000$	-2.098	-1.348	-0.706	-0.010	0.858	0.916	-0.447	0.771
	$n = 6000$	-2.312	-1.309	-0.728	-0.050	1.071	0.887	-0.411	0.742
$\alpha = \alpha_0, \beta = \hat{\beta}_n$	$n = 4000$	-1.684	-0.599	0.092	0.751	1.805	0.934	0.058	0.780
	$n = 5000$	-1.468	-0.698	-0.003	0.696	1.641	0.966	0.007	0.658
	$n = 6000$	-1.670	-0.725	-0.068	0.660	1.775	0.931	-0.023	0.644
$\alpha = \hat{\alpha}_n, \beta = \hat{\beta}_n$	$n = 4000$	-1.532	-0.611	0.100	0.820	1.903	0.934	0.080	0.781
	$n = 5000$	-1.422	-0.584	0.050	0.746	1.745	0.959	0.049	0.664
	$n = 6000$	-1.498	-0.544	0.068	0.810	1.843	0.950	0.042	0.618

Table S.3. Percentiles of ξ and CVG, bias, and RMSE of $\hat{c}_n(\boldsymbol{\theta})$ when $\alpha_0 = 5$.

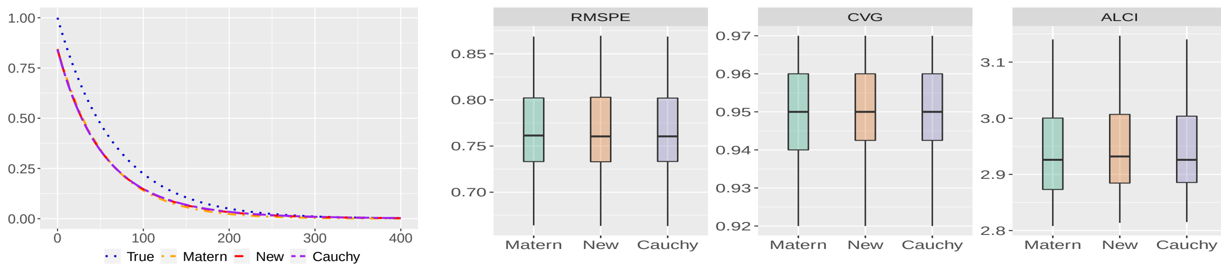
Settings		5%	25%	50%	75%	95%	CVG	bias	RMSE
$\mathcal{N}(0,1)$		-1.6449	-0.6749	0	0.6749	1.6449	0.95	0	
$ER = 0.6, \nu = 0.5$									
$\boldsymbol{\theta}$									
$\alpha = \alpha_0, \beta = \beta_0$	$n = 4000$	-1.612	-0.659	-0.049	0.655	1.661	0.954	0.000	0.085
	$n = 5000$	-1.468	-0.636	-0.027	0.685	1.683	0.961	0.002	0.074
	$n = 6000$	-1.633	-0.560	0.079	0.723	1.751	0.940	0.003	0.070
$\alpha = \alpha_0, \beta = \sqrt{0.5}\beta_0$	$n = 4000$	-1.273	-0.289	0.319	1.010	2.038	0.942	0.030	0.091
	$n = 5000$	-1.183	-0.311	0.298	1.033	2.054	0.942	0.027	0.079
	$n = 6000$	-1.261	-0.245	0.401	1.023	2.070	0.932	0.025	0.074
$\alpha = \alpha_0, \beta = \sqrt{2}\beta_0$	$n = 4000$	-1.757	-0.829	-0.222	0.479	1.477	0.945	-0.015	0.086
	$n = 5000$	-1.616	-0.793	-0.191	-0.500	-1.520	0.958	-0.011	0.074
	$n = 6000$	-1.787	-0.709	-0.071	0.556	1.587	0.936	-0.007	0.070
$\alpha = \alpha_0, \beta = \hat{\beta}_n$	$n = 4000$	-1.607	-0.633	0.000	0.690	1.705	0.951	0.003	0.087
	$n = 5000$	-1.434	-0.591	0.030	0.719	1.766	0.953	0.005	0.075
	$n = 6000$	-1.609	-0.564	0.094	0.758	1.822	0.930	0.006	0.071
$\alpha = \hat{\alpha}_n, \beta = \hat{\beta}_n$	$n = 4000$	-1.556	-0.538	0.142	0.813	1.884	0.948	0.012	0.089
	$n = 5000$	-1.378	-0.514	0.116	0.835	1.762	0.958	0.010	0.076
	$n = 6000$	-1.530	-0.557	0.100	0.758	1.832	0.950	0.008	0.070
$ER = 0.6, \nu = 1.5$									
$\boldsymbol{\theta}$									
$\alpha = \alpha_0, \beta = \beta_0$	$n = 4000$	-1.611	-0.651	-0.015	0.676	1.748	0.957	0.026	1.179
	$n = 5000$	-1.409	-0.606	0.052	0.727	1.705	0.958	0.069	1.031
	$n = 6000$	-1.346	-0.495	0.146	0.763	1.660	0.950	0.114	0.930
$\alpha = \alpha_0, \beta = \sqrt{0.5}\beta_0$	$n = 4000$	0.153	1.146	1.783	2.469	3.601	0.618	2.132	2.463
	$n = 5000$	0.109	1.070	1.720	2.432	3.475	0.628	1.818	2.113
	$n = 6000$	0.053	1.018	1.675	2.311	3.351	0.653	1.623	1.886
$\alpha = \alpha_0, \beta = \sqrt{2}\beta_0$	$n = 4000$	-2.308	-1.383	-0.802	-0.088	0.934	0.875	-0.877	1.456
	$n = 5000$	-2.093	-1.323	-0.675	0.004	0.964	0.909	-0.685	1.224
	$n = 6000$	-1.985	-1.186	-0.521	0.083	0.993	0.929	-0.538	1.059
$\alpha = \alpha_0, \beta = \hat{\beta}_n$	$n = 4000$	-1.664	-0.641	0.069	0.757	1.802	0.951	0.070	1.239
	$n = 5000$	-1.362	-0.615	0.080	0.792	1.788	0.956	0.105	1.074
	$n = 6000$	-1.368	-0.588	0.177	0.809	1.812	0.950	0.129	0.966
$\alpha = \hat{\alpha}_n, \beta = \hat{\beta}_n$	$n = 4000$	-1.528	-0569	0.138	0.845	1.937	0.929	0.181	1.263
	$n = 5000$	-1.342	-0.522	0.105	0.838	1.822	0.956	0.150	1.065
	$n = 6000$	-1.468	-0.518	0.102	0.842	1.920	0.937	0.142	1.003
$ER = 0.9, \nu = 0.5$									
$\boldsymbol{\theta}$									
$\alpha = \alpha_0, \beta = \beta_0$	$n = 4000$	-1.567	-0.602	-0.013	0.681	1.759	0.955	0.002	0.057
	$n = 5000$	-1.482	-0.664	-0.034	0.643	1.563	0.961	0.000	0.049
	$n = 6000$	-1.513	-0.615	0.077	0.637	1.745	0.953	0.002	0.045
$\alpha = \alpha_0, \beta = \sqrt{0.5}\beta_0$	$n = 4000$	-1.377	-0.423	0.179	0.871	1.956	0.947	0.013	0.058
	$n = 5000$	-1.333	-0.478	0.153	0.835	1.743	0.961	0.009	0.050
	$n = 6000$	-1.357	-0.468	0.233	0.800	1.894	0.947	0.009	0.046
$\alpha = \alpha_0, \beta = \sqrt{2}\beta_0$	$n = 4000$	-1.663	-0.708	-0.109	0.587	1.646	0.957	-0.004	0.057
	$n = 5000$	-1.554	-0.765	-0.121	0.556	1.465	0.958	-0.005	0.049
	$n = 6000$	-1.605	-0.692	-0.006	0.559	1.655	0.942	-0.002	0.045
$\alpha = \alpha_0, \beta = \hat{\beta}_n$	$n = 4000$	-1.571	-0.593	0.016	0.716	1.744	0.951	0.004	0.058
	$n = 5000$	-1.424	-0.653	-0.007	0.669	1.638	0.962	0.001	0.049
	$n = 6000$	-1.462	-0.572	0.095	0.704	1.803	0.949	0.003	0.045
$\alpha = \hat{\alpha}_n, \beta = \hat{\beta}_n$	$n = 4000$	-1.574	-0.539	0.106	0.803	1.853	0.952	0.007	0.059
	$n = 5000$	-1.327	-0.537	0.134	0.810	1.778	0.960	0.007	0.049
	$n = 6000$	-1.606	-0.601	0.058	0.713	1.756	0.956	0.003	0.046
$ER = 0.9, \nu = 1.5$									
$\boldsymbol{\theta}$									
$\alpha = \alpha_0, \beta = \beta_0$	$n = 4000$	-1.574	-0.616	-0.015	0.661	1.764	0.955	0.009	0.345
	$n = 5000$	-1.414	-0.617	0.015	0.741	1.612	0.974	0.017	0.298
	$n = 6000$	-1.743	-0.591	0.066	0.677	1.795	0.933	0.008	0.292
$\alpha = \alpha_0, \beta = \sqrt{0.5}\beta_0$	$n = 4000$	-0.614	0.325	0.969	1.631	2.762	0.837	0.351	0.498
	$n = 5000$	-0.587	0.274	0.936	1.652	2.593	0.857	0.299	0.428
	$n = 6000$	-0.978	0.198	0.889	1.566	2.643	0.870	0.248	0.389
$\alpha = \alpha_0, \beta = \sqrt{2}\beta_0$	$n = 4000$	-1.988	-1.042	-0.448	0.234	1.285	0.933	-0.139	0.369
	$n = 5000$	-1.784	-1.013	-0.373	0.330	1.207	0.956	-0.105	0.313
	$n = 6000$	-2.094	-0.968	-0.301	0.312	1.420	0.925	-0.097	0.306
$\alpha = \alpha_0, \beta = \hat{\beta}_n$	$n = 4000$	-1.590	-0.609	0.029	0.720	1.791	0.943	0.021	0.356
	$n = 5000$	-1.364	-0.605	0.055	0.779	1.677	0.967	0.025	0.303
	$n = 6000$	-1.648	-0.608	0.107	0.709	1.823	0.947	0.015	0.296
$\alpha = \hat{\alpha}_n, \beta = \hat{\beta}_n$	$n = 4000$	-1.533	-0.549	0.113	0.849	1.836	0.940	0.047	0.369
	$n = 5000$	-1.387	-0.526	0.111	0.798	1.733	0.961	0.039	0.307
	$n = 6000$	-1.352	-0.534	0.111	0.765	1.755	0.956	0.039	0.281

S.6 Additional Simulation Examples

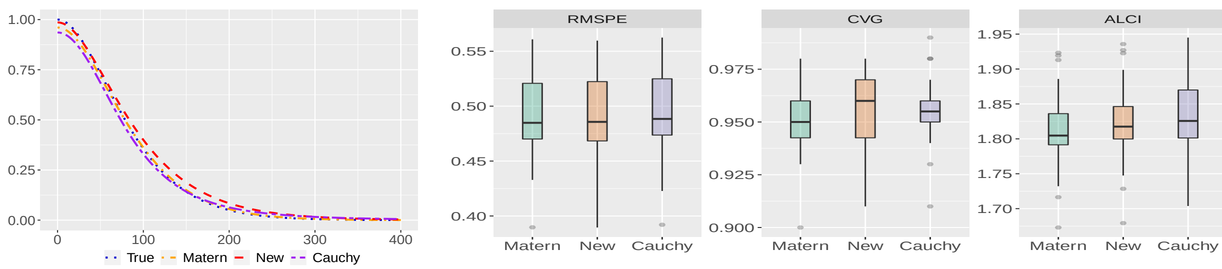
The results in the three cases in Section 4 of the main manuscript are based $n = 2000$ observations. In Section S.6.1 of the Supplementary Material, we provide results on exactly the same simulation settings with $n = 500$ and 1000. Similar conclusions can be drawn there. In addition, we also investigate the predictive performance when the covariance of the underlying true process is a product of individual covariance functions in Section S.6.2 of the Supplementary Material. The examples there show significant improvement of the CH class over the Matérn class and the GC class. In all these simulation examples, we found that the CH class is quite flexible in terms of capturing both the smoothness and the tail behavior. No matter which covariance structure (the Matérn class or the GC class) the true underlying process is generated from, the CH class is able to capture the underlying true covariance structure with satisfactory performance as implied by our theoretical developments. In contrast, the Matérn class is not able to capture the underlying true covariance structure with polynomially decaying dependence and the GC class is not able to capture the underlying true covariance structure with different degrees of smoothness behaviors. Below are the detailed results.

S.6.1 Predictive Performance with Different Sample Sizes

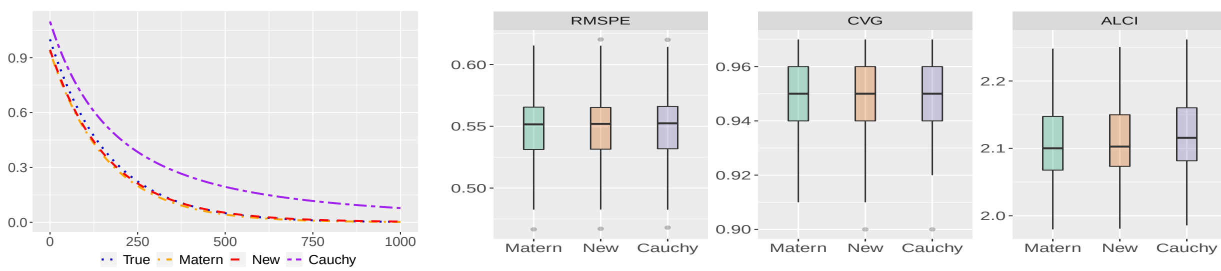
In this section, we use the same simulation settings as in Section 4 but with $n = 500$ and 1000 observations for parameter estimation. The simulation setup here is the same as the one considered in Section 4. For $n = 500$ observations, the results are shown in Figure S.3 for Case 1, Figure S.4 for Case 2, and Figure S.5 for Case 3. For $n = 1000$ observations, the results are shown in Figure S.6 for Case 1, Figure S.7 for Case 2, and Figure S.8 for Case 3. To conclude, the CH class is very flexible since it can allow different smoothness behaviors in the same way as the Matérn class and can allow different degrees of tail behaviors that can capture the one in the GC class.



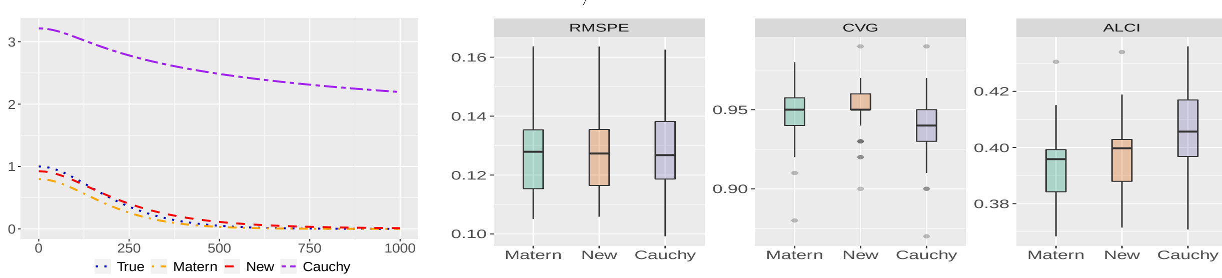
$\nu = 0.5, ER = 200$



$\nu = 2.5, ER = 200$

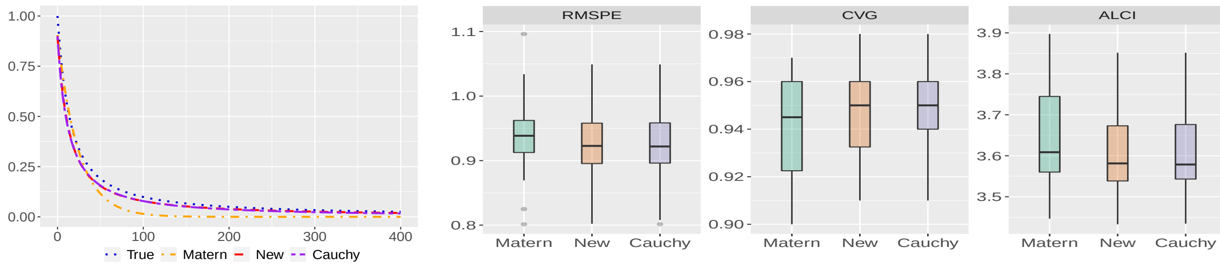


$\nu = 0.5, ER = 500$

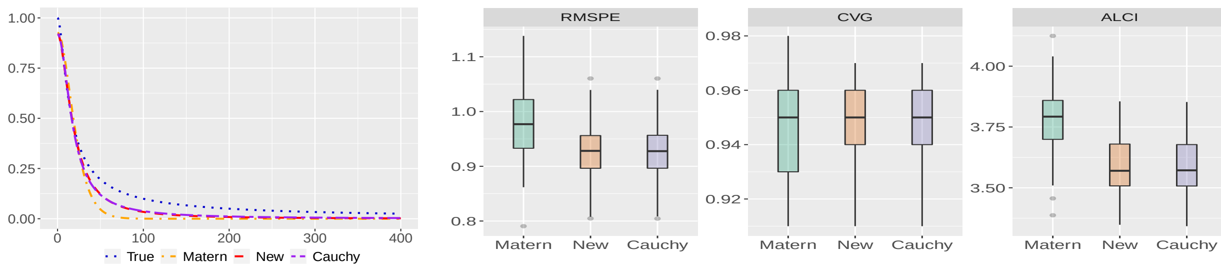


$\nu = 2.5, ER = 500$

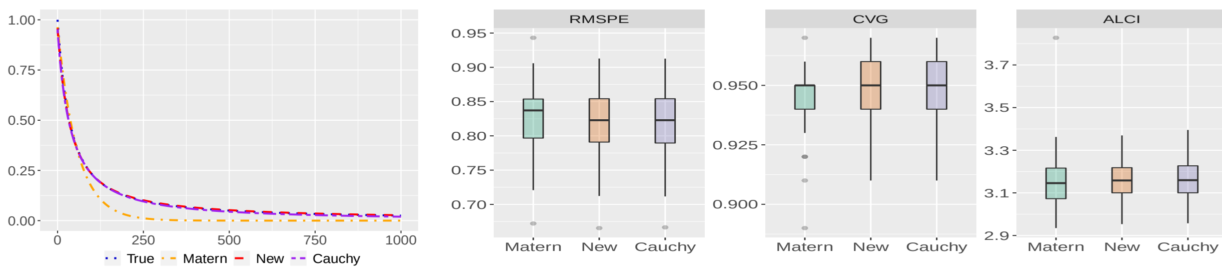
Fig. S.3. Case 1: Comparison of predictive performance and estimated covariance structures when the true covariance is the Matérn class with 500 observations. The predictive performance is evaluated at 10-by-10 regular grids in the square domain. These figures summarize the predictive measures based on RMSPE, CVG and ALCI under 30 simulated realizations.



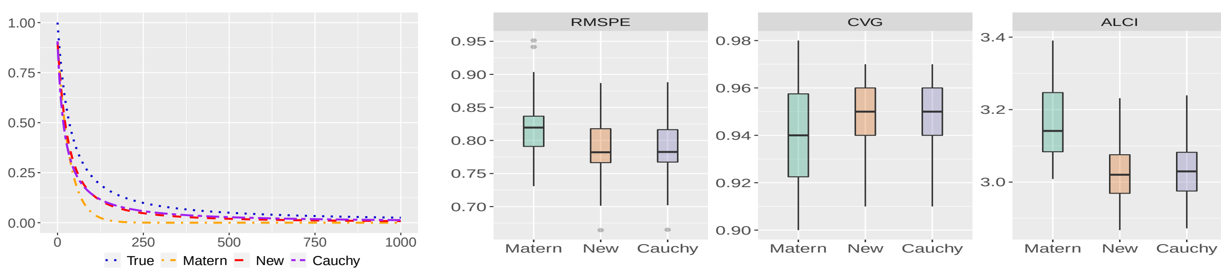
$\nu = 0.5, ER = 200$



$\nu = 2.5, ER = 200$

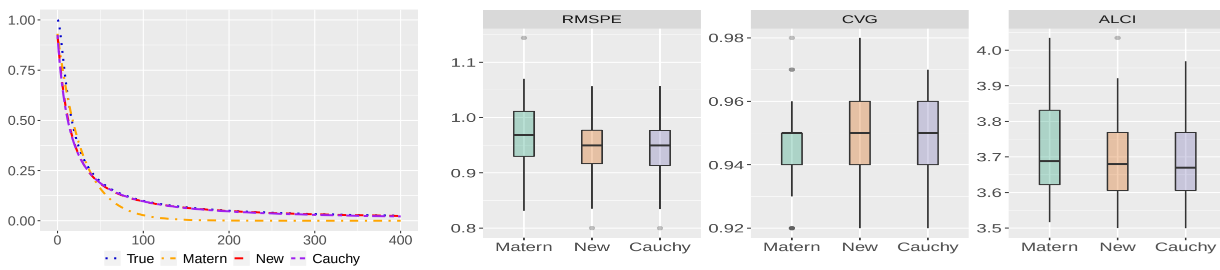


$\nu = 0.5, ER = 500$

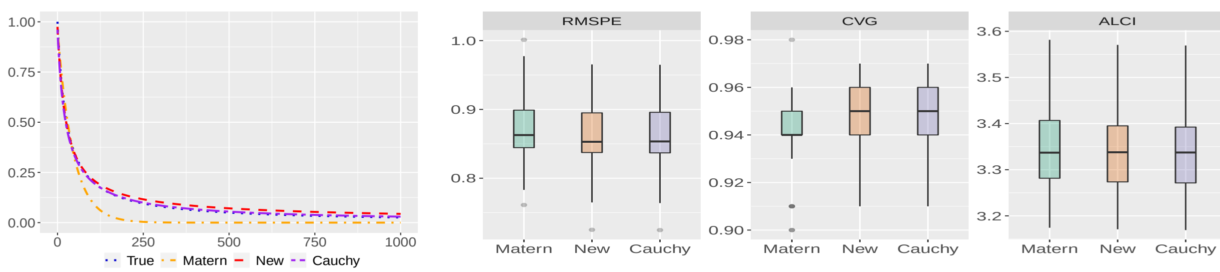


$\nu = 2.5, ER = 500$

Fig. S.4. Case 2: Comparison of predictive performance and estimated covariance structures when the true covariance is the CH class with 500 observations. The predictive performance is evaluated at 10-by-10 regular grids in the square domain. These figures summarize the predictive measures based on RMSPE, CVG and ALCI under 30 simulated realizations.

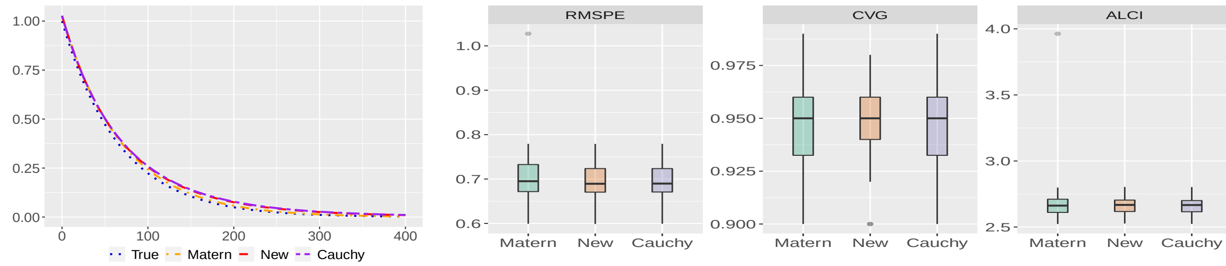


$\delta = 1, ER = 200$

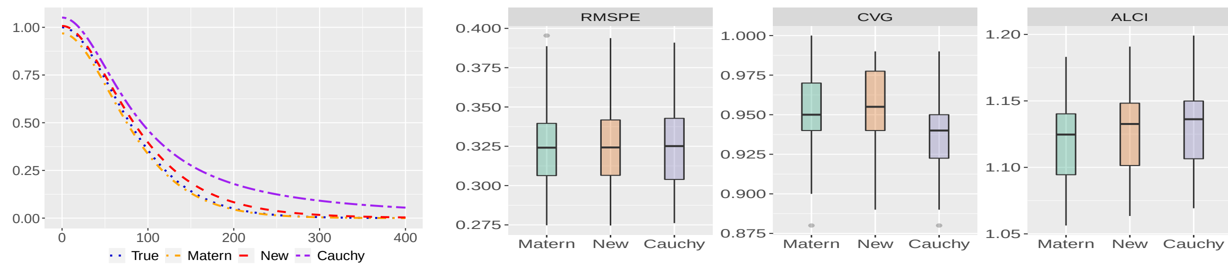


$\delta = 1, ER = 500$

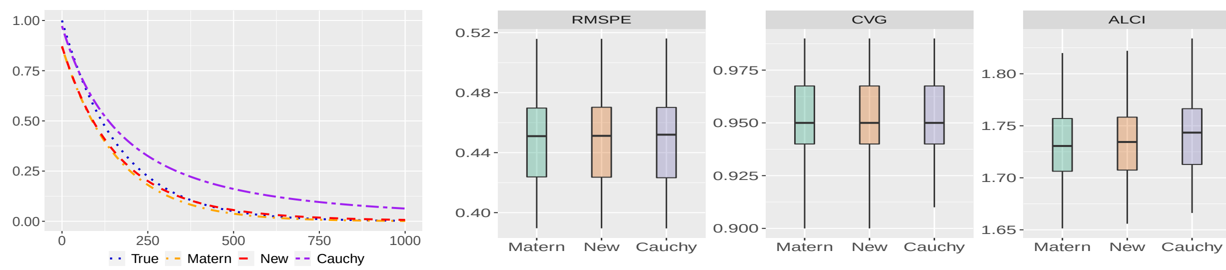
Fig. S.5. Case 3: Comparison of predictive performance and estimated covariance structures when the true covariance is the GC class with 500 observations. The predictive performance is evaluated at 10-by-10 regular grids in the square domain. These figures summarize the predictive measures based on RMSPE, CVG and ALCI under 30 simulated realizations.



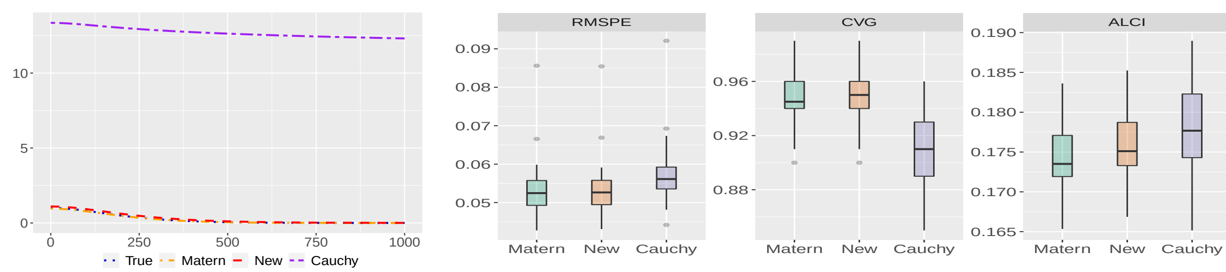
$\nu = 0.5, ER = 200$



$\nu = 2.5, ER = 200$

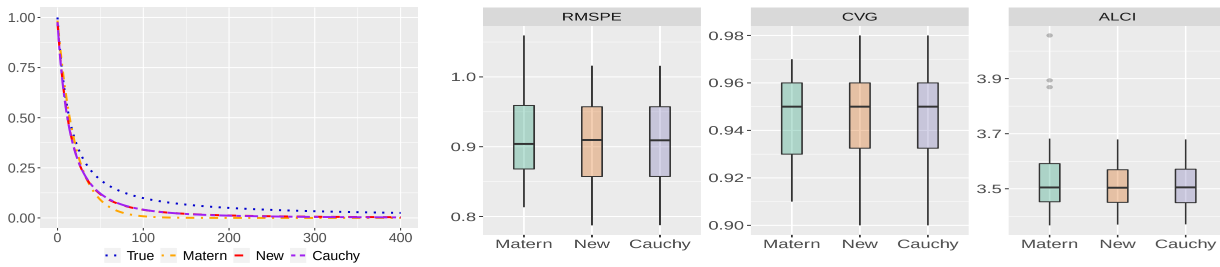


$\nu = 0.5, ER = 500$

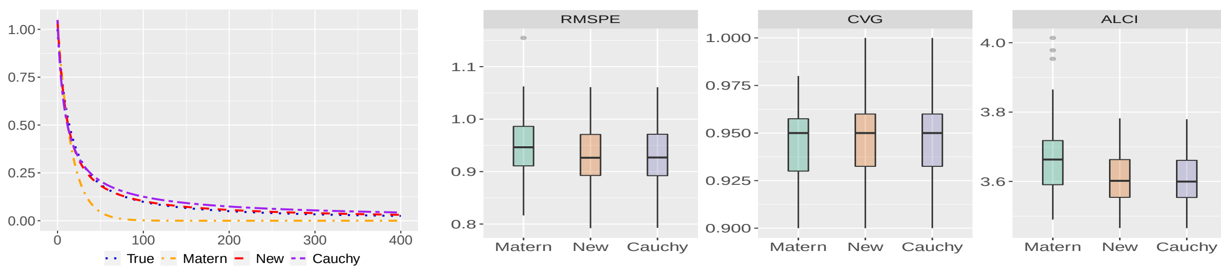


$\nu = 2.5, ER = 500$

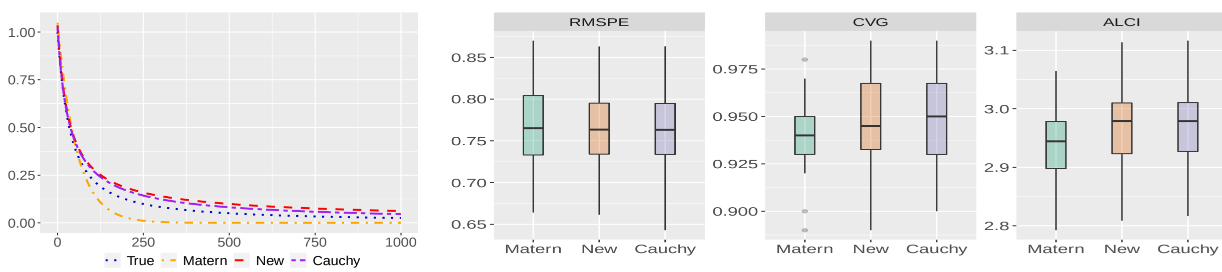
Fig. S.6. Case 1: Comparison of predictive performance and estimated covariance structures when the true covariance is the Matérn class with 1000 observations. The predictive performance is evaluated at 10-by-10 regular grids in the square domain. These figures summarize the predictive measures based on RMSPE, CVG and ALCI under 30 simulated realizations.



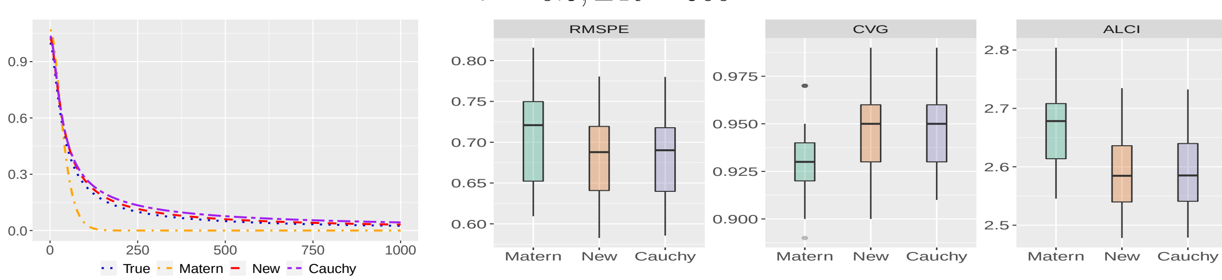
$\nu = 0.5, ER = 200$



$\nu = 2.5, ER = 200$

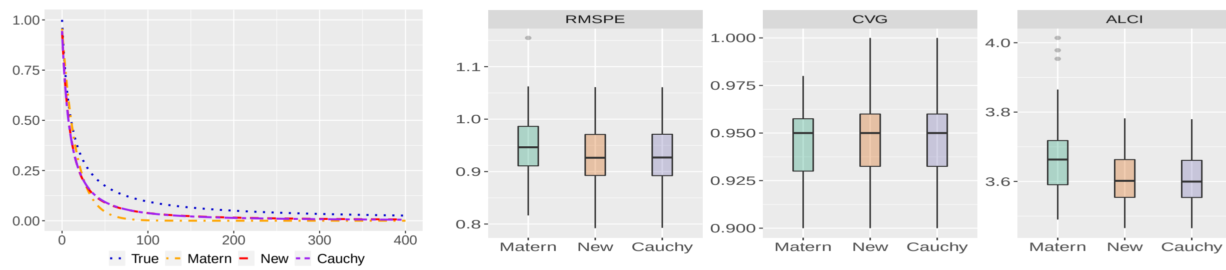


$\nu = 0.5, ER = 500$

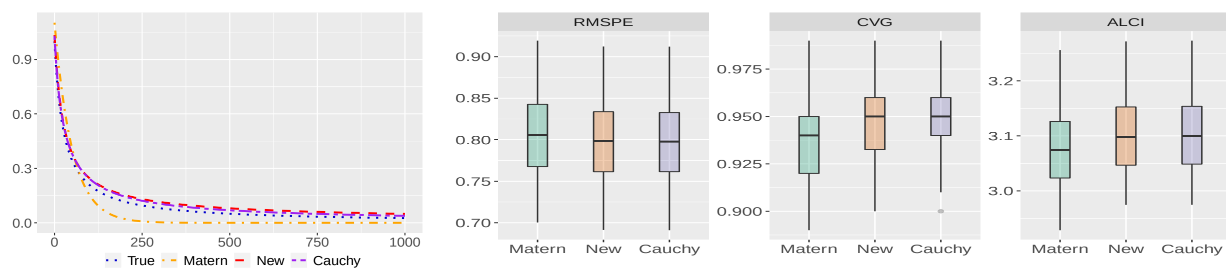


$\nu = 2.5, ER = 500$

Fig. S.7. Case 2: Comparison of predictive performance and estimated covariance structures when the true covariance is the CH class with 1000 observations. The predictive performance is evaluated at 10-by-10 regular grids in the square domain. These figures summarize the predictive measures based on RMSPE, CVG and ALCI under 30 simulated realizations.



$\delta = 1, ER = 200$



$\delta = 1, ER = 500$

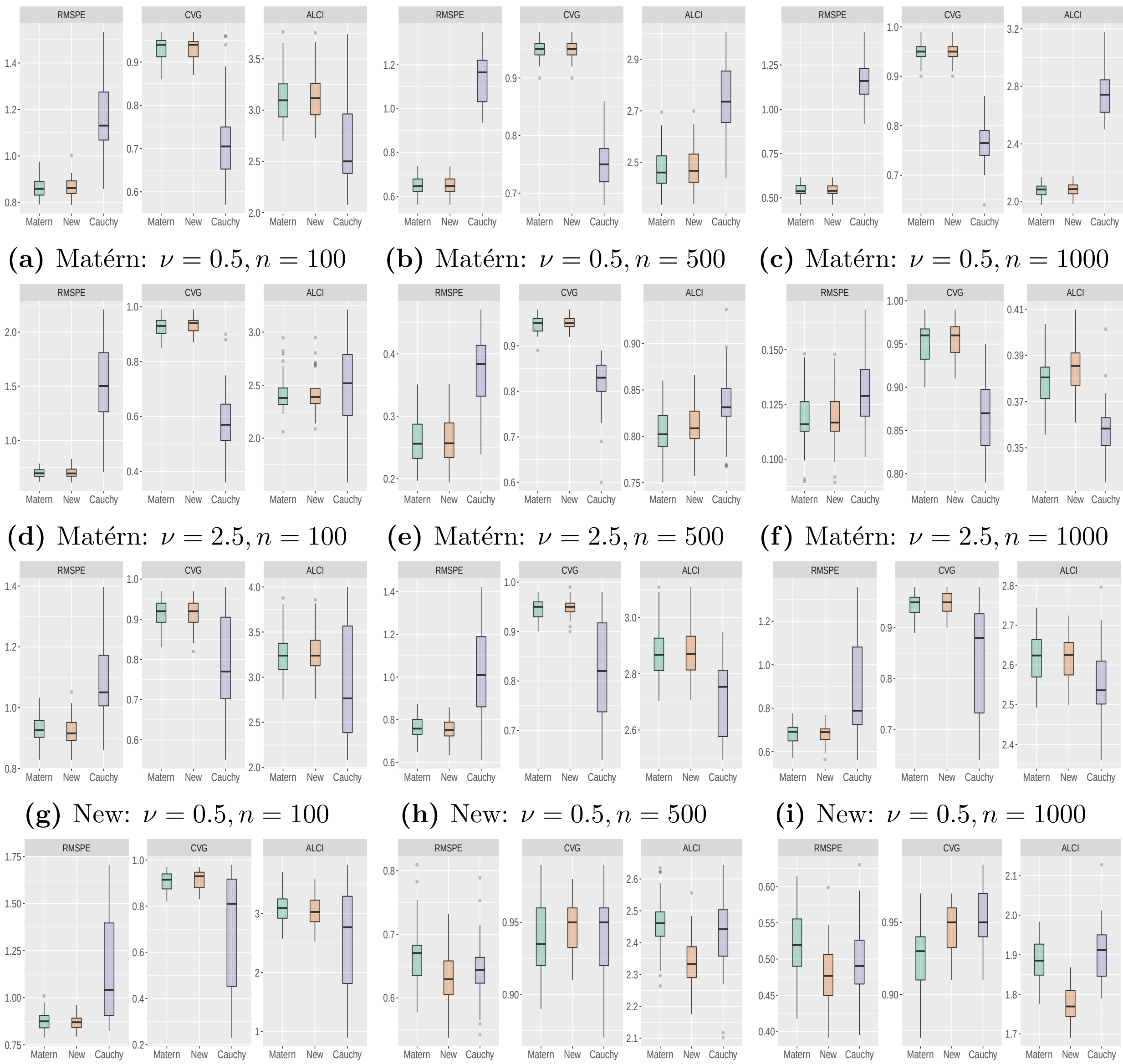
Fig. S.8. Case 3: Comparison of predictive performance and estimated covariance structures when the true covariance is the GC class with 1000 observations. The predictive performance is evaluated at 10-by-10 regular grids in the square domain. These figures summarize the predictive measures based on RMSPE, CVG and ALCI under 30 simulated realizations.

S.6.2 Simulation with a Tensor Product of Covariance Functions

In this section, we study the predictive performance of the CH class with a product form, i.e., $r(\|\mathbf{s}-\mathbf{u}\|) = \prod_{i=1}^{d} R(|s_i - u_i|; \boldsymbol{\theta}_i)$, where $R(\cdot; \boldsymbol{\theta}_i)$ is an isotropic covariance function with parameter $\boldsymbol{\theta}_i$. This product form of covariance functions allows different properties along different coordinate directions (or input space) and has been widely used in uncertainty quantification and machine learning.

We simulate the true processes under the Matérn class and the CH class with effective range fixed at 200 and 500. For the smoothness parameter, we consider $\nu = 0.5, 2.5$. The tail decay parameter in the CH class is chosen to be 0.5. As each dimension has a different range parameter or scale parameter, we choose these parameters in each dimension such that their correlation will be $0.5^{1/2}$ at distance 200 and 500. This will guarantee the overall effective range will be 200 and 500, respectively. For each simulation setting, the true process is simulated at $n = 100, 500, 1000$ locations. The GC class has a smoothness parameter that is specified as in Section 4. The prediction locations are the same as those in Section 4.

In the first case where the true process has a product of Matérn covariance functions, the prediction results under the Matérn class, the CH class and the GC class are shown in panels from (a) to (f) of Figure S.9. As expected, the Matérn class and the CH class yield indistinguishable predictive performance in terms of RMSPE, CVG, and ALCI. However, the GC class has much worse performance than the other two covariance classes. In the second case where the true process has a product of CH functions, the prediction results under these three covariance classes are shown in panels from (g) to (l) of Figure S.9. As expected, the CH class yields much better prediction results than the Matérn class, since the Matérn class has an exponentially decaying tail that is not able to capture the tail behavior in the CH class. It is worth noting that the GC class yields much worse predictive performance than the other two covariance classes. This is quite different from the situation when the true process does not have a product covariance form.



(a) Matérn: $\nu = 0.5, n = 100$ **(b)** Matérn: $\nu = 0.5, n = 500$ **(c)** Matérn: $\nu = 0.5, n = 1000$

(d) Matérn: $\nu = 2.5, n = 100$ **(e)** Matérn: $\nu = 2.5, n = 500$ **(f)** Matérn: $\nu = 2.5, n = 1000$

(g) New: $\nu = 0.5, n = 100$ **(h)** New: $\nu = 0.5, n = 500$ **(i)** New: $\nu = 0.5, n = 1000$

(j) New: $\nu = 2.5, n = 100$ **(k)** New: $\nu = 2.5, n = 500$ **(l)** New: $\nu = 2.5, n = 1000$

Fig. S.9. Predictive performance over 10-by-10 regular grids under three covariance classes when the true process has a product form of covariance structures. The predictive performance is studied under different smoothness parameters, effective ranges and number of observation locations.

S.7 Additional Numerical Results

This section contains parameter estimation results and figures referenced in Section 5. Figure S.10 shows the directional semivariograms for the OCO2 data. Table S.4 shows the estimated parameters under the Matérn covariance model and the CH covariance in the cross-validation study of Section 5. Figure S.11 compares kriging predictors and associated standard errors under the CH class and Matérn class in the cross-validation study of Section 5.

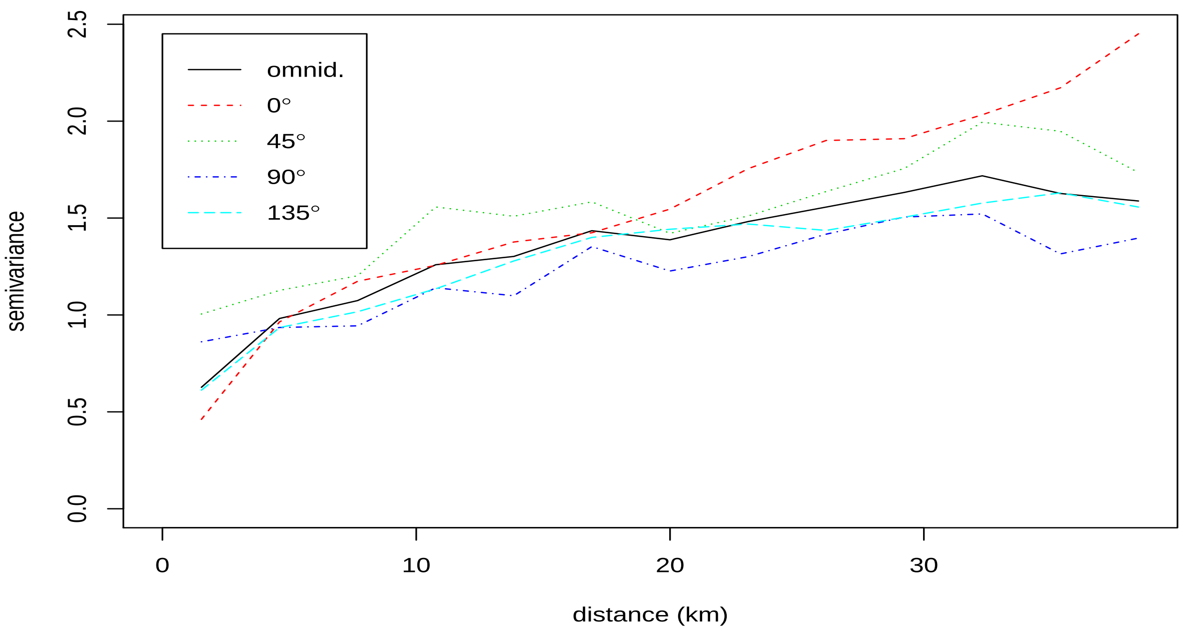
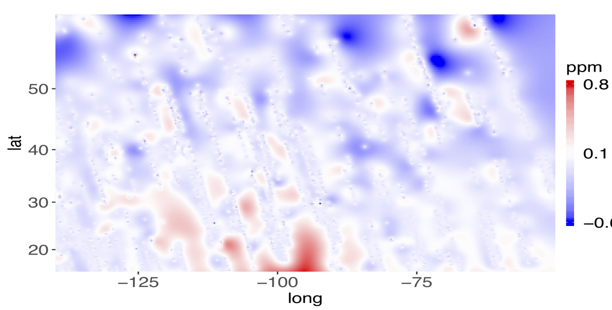


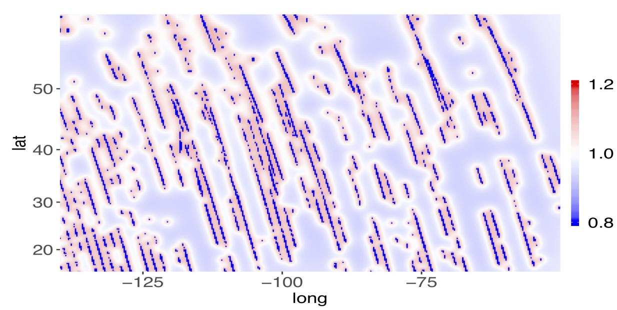
Fig. S.10. Graphical assessments of isotropy in the OCO2 data. The directional semivariograms do not appear to exhibit differences, indicating that the assumption of an isotropic covariance function is likely to be true.

Table S.4. Cross-validation results based on the Matérn covariance model and the CH covariance model.

	Matérn class		CH class	
	$\nu = 0.5$	$\nu = 1.5$	$\nu = 0.5$	$\nu = 1.5$
b	411.1	411.1	411.0	411.0
σ^2	1.679	1.439	1.750	1.585
ϕ	160.5	104.1	—	—
α	—	—	0.381	0.353
β	—	—	80.17	58.65



(a) Difference of kriging predictors



(b) Ratio of kriging standard errors.

Fig. S.11. Comparison of kriging predictions under the Matérn class and the CH class. The left panel shows the difference between kriging predictors under the CH class and those under the Matérn class. The right panel shows the ratio of kriging standard errors under the CH class to those under the Matérn class.

References

Abramowitz, M. and Stegun, I. A. (1965). *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, volume 55. Courier Corporation, North Chelmsford, MA.

Barndorff-Nielsen, O. E., Kent, J. T., and Sørensen, M. (1982). Normal variance-mean mixtures and z distributions. *International Statistical Review* **50,** 145–159.

Bevilacqua, M., Faouzi, T., Furrer, R., and Porcu, E. (2019). Estimation and prediction using generalized Wendland covariance functions under fixed domain asymptotics. *The Annals of Statistics* **47,** 828–856.

Ibragimov, I. A. and Rozanov, Y. A. (1978). *Conditions for Regularity of Stationary Random Processes*, pages 108–143. Springer New York, New York, NY.

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

Magnus, J. R. and Neudecker, H. (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics, Thrid Edition.* Probability and Statistics. John Wiley & SONS, New York.

Matérn, B. (1960). *Spatial variation*, Meddelanden fran Statens Skogsforskningsinstitut, 49, 5. Second ed. (1986), Lecture Notes in Statistics 36, New York: Springer.

Schoenberg, I. J. (1938). Metric Spaces and Completely Monotone Functions. *Annals of Mathematics* **39,** 811–841.

Stein, E. M. and Weiss, G. (1971). *Introduction to Fourier Analysis on Euclidean Spaces (PMS-32).* Princeton university press.

Stein, M. L. (1988). Asymptotically efficient prediction of a random field with a misspecified covariance function. *The Annals of Statistics* **16,** 55–63.

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

Stein, M. L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging.* Springer Science & Business Media, New York, NY.

Stein, M. L. and Handcock, M. S. (1989). Some asymptotic properties of kriging when the covariance function is misspecified. *Mathematical Geology* **21,** 171–190.

Wang, D. (2010). Fixed domain asymptotics and consistent estimation for Gaussian random field models in spatial statistics and computer experiments. PhD thesis, Department of Statistics and Applied Probability, National University of Singapore.

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

Yadrenko, M. I. (1983). *Spectral theory of random fields.* Translation series in mathematics and engineering. Optimization Software, New York, NY. Transl. from the Russian.

Yaglom, A. M. (1987). Correlation Theory of Stationary and Related Random Functions. *Springer Series in Statistics* .

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