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

Abstract

The Matérn covariance function is a popular choice for prediction in spatial statistics and uncertainty quantification literature. A key benefit of the Matérn class is that it is possible to get precise control over the degree of mean-square differentiability of the random process. However, the Matérn class possesses exponentially decaying tails, and thus may not be suitable for modeling polynomially decaying dependence. This problem can be remedied using polynomial covariances; however one loses control over the degree of mean-square differentiability of corresponding processes, in that random processes with existing polynomial covariances are either infinitely mean-square differentiable or nowhere mean-square differentiable at all. We construct a new family of covariance functions called the *Confluent Hypergeometric* (CH) class using a scale mixture representation of the Matérn class where one obtains the benefits of both Matérn and polynomial covariances. The resultant covariance contains two parameters: one controls the degree of mean-square differentiability near the origin and the other controls the tail heaviness, independently of each other. Using a spectral representation, we derive theoretical properties of this new covariance including equivalent measures and asymptotic behavior of the maximum likelihood estimators under infill asymptotics. The improved theoretical properties of the CH class are verified via extensive simulations. Application using NASA's Orbiting Carbon Observatory-2 satellite data confirms the advantage of the CH class over the Matérn class, especially in extrapolative settings.

Keywords: Equivalent measures; Gaussian process; Gaussian scale mixture; Polynomial covariance; XCO2.

1 Introduction

Kriging, also known as spatial best linear unbiased prediction, is a term coined by Matheron (1963) in honor of the South African mining engineer D. G. Krige (Cressie, 1990). With origins in geostatistics, applications of kriging have permeated fields as diverse as spatial statistics (e.g., Banerjee et al., 2014; Berger et al., 2001; Cressie, 1993; Journel and Huijbregts, 1978; Matérn, 1960; Stein, 1999), uncertainty quantification or UQ (e.g., Berger and Smith, 2019; Sacks et al., 1989; Santner et al., 2018) and machine learning (Williams and Rasmussen, 2006). Suppose that $\{Z(\mathbf{s}) \in \mathbb{R} : \mathbf{s} \in \mathcal{D} \subset \mathbb{R}^d\}$ is a stochastic process with a covariance function $\text{cov}(Z(\mathbf{s}), Z(\mathbf{s}+\mathbf{h})) = C(\mathbf{h})$ that is solely a function of the increment $\mathbf{h}$. Then $C(\cdot)$ is said to be second-order stationary (or weakly stationary). Further, if $C(\cdot)$ is a function of $|\mathbf{h}|$ with $|\cdot|$ denoting the Euclidean norm, then $C(\cdot)$ is called isotropic. If the process $Z(\cdot)$ possesses a constant mean function and a weakly stationary (resp. isotropic) covariance function, the process $Z(\cdot)$ is called weakly stationary (resp. isotropic). Further, $Z(\cdot)$ is a Gaussian process (GP) if every finite-dimensional realization $Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n)$ jointly follows a multivariate normal distribution for $\mathbf{s}_i \in \mathcal{D}$ and every n.

The Matérn covariance function (Matérn, 1960) has been widely used in spatial statistics due to its flexible local behavior and nice theoretical properties (Stein, 1999) with increasing popularity in the UQ and machine learning literature (Gu et al., 2018; Guttorp and Gneiting, 2006). The Matérn covariance function is of the form:

$$\mathcal{M}(h; \nu, \phi, \sigma^2) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}}{\phi} h \right)^{\nu} \mathcal{K}_\nu \left(\frac{\sqrt{2\nu}}{\phi} h \right), \tag{1}$$

where $\sigma^2 > 0$ is the variance parameter, $\phi > 0$ is the range parameter, and $\nu > 0$ is the smoothness parameter that controls the mean-square differentiability of associated random processes. We denote by $\mathcal{K}_\nu(\cdot)$ the modified Bessel function of the second kind with the asymptotic expansion $\mathcal{K}_\nu(h) \asymp (\pi/(2h))^{1/2} \exp(-h)$ as $h \to \infty$, where $f(x) \asymp g(x)$ denotes $\lim_{x\to\infty} f(x)/g(x) = c \in (0, \infty)$. Further, we use the notation $f(x) \sim g(x)$ if $c = 1$. Thus, using this asymptotic expression

of $\mathcal{K}_\nu(h)$ for large h from Section 6 of Barndorff-Nielsen et al. (1982), the tail behavior of the Matérn covariance function is given by:

$$\mathcal{M}(h;\nu,\phi,\sigma^2) \asymp h^{\nu-1/2}\exp\left(-\frac{\sqrt{2\nu}}{\phi}h\right), \quad h\to\infty.$$

Eventually, the $\exp(-\sqrt{2\nu}h/\phi)$ term dominates, and the covariance decays exponentially for large h. This exponential decay may make it unsuitable for capturing polynomially decaying dependence. This problem with the Matérn covariance can be remedied by using covariance functions that decay polynomially, such as the generalized Wendland (Gneiting, 2002) and generalized Cauchy covariance functions (Gneiting, 2000), but in using these polynomial covariance functions one loses a key benefit of the Matérn class: that of the degree of mean-square differentiability of the process. Random processes with a Matérn covariance function are exactly $\lfloor\nu\rfloor$ times mean-square differentiable, whereas the random processes with a generalized Cauchy covariance function are either non-differentiable (very rough) or infinitely differentiable (very smooth) in the mean-square sense, without any middle ground (Stein, 2005). The generalized Wendland covariance family also has limited flexibility near the origin compared to the Matérn class and has compact support (Gneiting, 2002).

Stochastic processes with polynomial-tailed dependences are ubiquitous in many scientific disciplines including geophysics, meteorology, hydrology, astronomy, agriculture and engineering; see Beran (1992) for a survey. In UQ, Gaussian stochastic processes have been often used for computer model emulation and calibration (Santner et al., 2018). In some applications, certain inputs may have little impact on output from a computer model, and these inputs are called *inert inputs*; see Chapter 7 of Santner et al. (2018) for detailed discussion. Power-law covariance functions can allow for large correlations among distant observations and hence are more suitable for modeling these inert inputs. Most often, computer model outputs can have different smoothness properties due to the behavior of the physical process to be modeled. Thus, power-law covariances with the

possibility of controlling the mean-square differentiability of stochastic processes are very desirable for modeling such output.

In spatial statistics, polynomial covariances have been studied in a limited number of works (e.g., Gay and Heyde, 1990; Gneiting, 2000; Haslett and Raftery, 1989). In the rest of the paper, we focus on investigation of polynomial covariances in spatial settings. For spatial modeling, polynomial covariances can improve prediction accuracy over large missing regions. A covariance function with polynomially decaying tail can be useful to model highly correlated observations. As a motivating example, Figure 1 shows a 16-day repeat cycle of NASA's Level 3 data product of the column-averaged carbon dioxide dry air mole fraction (XCO2) at 0.25° and 0.25° collected from the Orbiting Carbon Observatory-2 (OCO-2) satellite. The XCO2 data are collected over longitude bands and have large missing gaps between them. Predicting the true process over these large missing gaps based on a spatial process model is challenging. If the covariance function only allows exponentially decaying dependence, the predicted true process will be dominated by the mean function in the spatial process model with the covariance function having negligible impact over these large missing gaps. However, if the covariance function can model polynomially decaying dependence, the predicted true process over these missing gaps will carry more information from distant locations where observations are available, presumably resulting in better prediction. Thus, it is of fundamental and practical interest to develop a covariance function with polynomially decaying tails, without sacrificing the control over the smoothness behavior of the process realizations.

In this paper we propose a new family of *interpretable* covariance functions called the *Confluent Hypergeometric* (CH) class that bridges this gap between the Matérn covariance and polynomial covariances. The proposed covariance class is obtained by mixing the Matérn covariance over its range parameter ϕ. This is done by recognizing the Bessel function in the Matérn covariance function as proportional to the normalizing constant of the generalized inverse Gaussian distribution (e.g., Barndorff-Nielsen, 1977), which then allows analytically tractable calculations with respect to

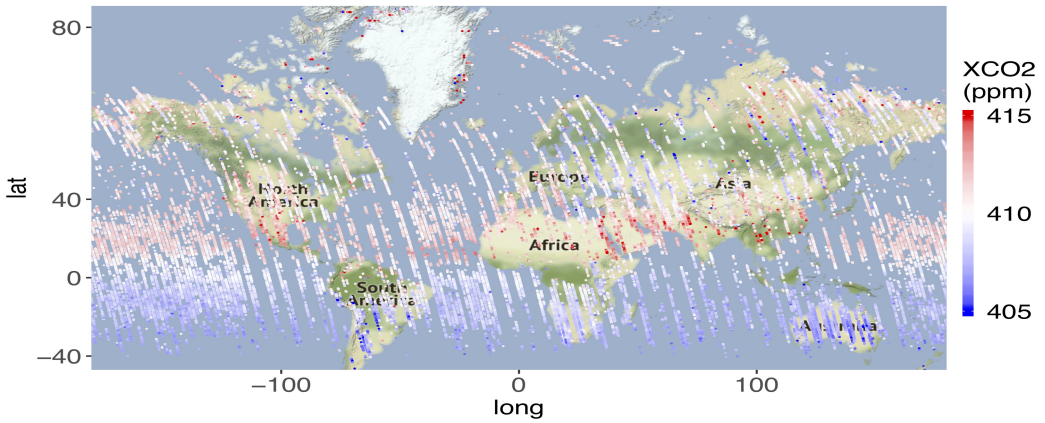


Fig. 1. XCO2 data from June 1 to June 16, 2019. The units are parts per millions (ppm).

a range of choices for mixing densities, resulting in valid covariance functions with varied features. Apart from this technical innovation, the key benefit is that this mixing does not affect the origin behavior and thus allows one to retain the precise control over the smoothness of process realizations as in Matérn. However, the tail is inflated due to mixing, and, in fact, the mixing distribution can be chosen in a way so that the tail of the resultant covariance function displays regular variation, with precise control over the tail decay parameter α. A function $f(\cdot)$ is said to have a regularly decaying right tail with index α if it satisfies $f(x) \asymp x^{-\alpha}L(x)$ as $x \to \infty$ for some $\alpha > 0$ where $L(\cdot)$ is a slowly varying function at infinity with the property $\lim_{x\to\infty} L(tx)/L(x) = 1$ for all $t \in (0, \infty)$ (Bingham et al., 1989). Unlike a generalized Cauchy covariance function, this CH class is obtained without sacrificing the control over the degree of mean-square differentiability of the process, which is still controlled solely by ν, and the resulting process is still exactly $\lfloor \nu \rfloor$ times mean-square differentiable, independent of α. Moreover, regular variation is preserved under several commonly used transformations, such as sums or products. Thus, it is possible to exploit these properties of regular variation to derive CH covariances with similar features from the original covariance function that is obtained via a mixture of the Matérn class.

The rest of the paper is organized as follows. Section 2 begins with the construction of the proposed covariance function as a mixture of the Matérn covariance function over its range parameter. We verify that such construction indeed results in a valid covariance function. Moreover, we demonstrate that the behaviors of this covariance function near the origin and in the tails are

characterized by two distinct parameters, which in turn control the smoothness and the degree of polynomially decaying dependence, respectively. Section 3 presents the main theoretical results for the CH class. We first derive the spectral representation of this CH class and characterize its high-frequency behavior, and then show theoretical properties concerning equivalence classes under Gaussian measures and asymptotic properties related to parameter estimation and prediction. The resultant theory is extensively verified via simulations in Section 4. In Section 5, we use this CH covariance to analyze NASA's OCO-2 data, and demonstrate better prediction results over the Matérn covariance. Section 6 concludes with some discussions for future investigations. All the technical proofs can be found in the Supplementary Material.

2 The CH Class as a Mixture of the Matérn Class

Our starting point in mixing over the range parameter ϕ in the Matérn covariance function is the correspondence between the form of the Matérn covariance function and the normalizing constant of the generalized inverse Gaussian distribution (e.g., Barndorff-Nielsen, 1977). The generalized inverse Gaussian distribution has density on $(0, \infty)$ given by:

$$\pi_{GIG}(x) = \frac{(a/b)^{p/2}}{2\mathcal{K}_p(\sqrt{ab})} x^{(p-1)} \exp\{-(ax + b/x)/2\}; \quad a, b > 0, p \in \mathbb{R}.$$

Thus,

$$\mathcal{K}_p(\sqrt{ab}) = \frac{1}{2}(a/b)^{p/2} \int_0^\infty x^{(p-1)} \exp\{-(ax + b/x)/2\} dx.$$

Take $a = \phi^{-2}, b = 2\nu h^2$ and $p = \nu$. Then we have the following representation of the Matérn covariance function with range parameter ϕ and smoothness parameter ν:

$$\mathcal{M}(h; \nu, \phi, \sigma^2) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}}{\phi} h \right)^\nu \mathcal{K}_\nu \left(\frac{\sqrt{2\nu}}{\phi} h \right)$$

$$= \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}h}{\phi} \right)^{\nu} \frac{1}{2} \left(\frac{1}{\sqrt{2\nu}h\phi} \right)^{\nu} \int_0^\infty x^{(\nu-1)} \exp\{-(x/\phi^2 + 2\nu h^2/x)/2\} dx$$
$$= \frac{\sigma^2}{2^\nu \phi^{2\nu} \Gamma(\nu)} \int_0^\infty x^{(\nu-1)} \exp\{-(x/\phi^2 + 2\nu h^2/x)/2\} dx.$$

Thus, the mixture over ϕ^2 with respect to a mixing measure $G(\phi^2)$ on $(0, \infty)$ can be written as

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

The resultant covariance via this mixture is quite general with different choices for the mixing measure $G(\phi^2)$. When the mixing measure $G(\phi^2)$ admits a probability density function, say $\pi(\phi^2)$, the inner integral may be recognized as a mixture of gamma integrals (by change of variable $u = \phi^{-2}$), which is analytically tractable for many choices of $\pi(\phi^2)$; see for example the chapter on gamma integrals in Abramowitz and Stegun (1965). More importantly, as we show below, the mixing density $\pi(\phi^2)$ can be chosen to achieve precise control over certain features of the resulting covariance function.

Theorem 1. *Let $X \sim \mathcal{IG}(a, b)$ denote an inverse gamma random variable using the shape–scale parameterization with density $\pi_{IG}(x) = \{b^a/\Gamma(a)\} x^{-a-1} \exp(-b/x)$; $a, b > 0$. Assume that $\phi^2 \sim \mathcal{IG}(\alpha, \beta^2/2)$ and that $\mathcal{M}(h; \nu, \phi, \sigma^2)$ is the Matérn covariance function in Equation (1). Then $C(h; \nu, \alpha, \beta, \sigma^2) := \int_0^\infty \mathcal{M}(h; \nu, \phi, \sigma^2) \pi(\phi^2; \alpha, \beta) d\phi^2$ is a positive-definite covariance function on $\mathbb{R}^d$ with the following form:*

$$C(h; \nu, \alpha, \beta, \sigma^2) = \frac{\sigma^2 \beta^{2\alpha} \Gamma(\nu + \alpha)}{\Gamma(\nu)\Gamma(\alpha)} \int_0^\infty x^{(\nu-1)} (x + \beta^2)^{-(\nu+\alpha)} \exp(-\nu h^2/x) dx, \tag{3}$$

where $\sigma^2 > 0$ is the variance parameter, $\alpha > 0$ is called the tail decay parameter, $\beta > 0$ is called the scale parameter, and $\nu > 0$ is called the smoothness parameter.

REMARK 1. The Matérn covariance is sometimes parameterized differently. The mixing density can be chosen accordingly to arrive at results identical to ours. For instance, with parameterization of the Matérn class given in Stein (1999), a gamma mixing density with shape parameter α and rate parameter $\beta^2/2$ would lead to an alternative route to the same representation of the CH class. The limiting case of the Matérn class is the squared exponential (or Gaussian) covariance when its smoothness parameter ν goes to ∞. In this case, mixing over the inverse gamma distribution in Theorem 1 yields the Cauchy covariance.

REMARK 2. The Matérn covariance arises as a limiting case of the proposed covariance in Theorem 1 when the mixing distribution on ϕ^2 is a point mass. Indeed, standard calculations show that the mode of the inverse gamma distribution $\mathcal{IG}(\phi^2 \mid \alpha, \beta^2/2)$ is $\beta^2/\{2(\alpha+1)\}$ and its variance is $\beta^4/(\{4(\alpha-1)^2(\alpha-2)\}$. Thus, if one takes $\beta^2 = 2(\alpha+1)\gamma^2$ for fixed $\gamma > 0$ and allows α to be large, the entire mass of the distribution $\mathcal{IG}(\alpha, (\alpha+1)\gamma^2)$ is concentrated at the fixed quantity γ^2 as $\alpha \to \infty$, which gives the Matérn covariance $\mathcal{M}(h; \nu, \gamma, \sigma^2)$ as the limiting case of the covariance function $C(h; \nu, \alpha, \sqrt{2(\alpha+1)}\gamma, \sigma^2)$ as $\alpha \to \infty$.

Having established in Theorem 1 the resultant mixture as a valid covariance function, one may take a closer look at its properties. To begin, although the final form of the CH class involves an integral, and thus may not appear to be in closed form at a first glance, the situation is indeed not too different from that of Matérn, where the associated Bessel function is available in an algebraically closed form only for certain special cases; otherwise it is available as an integral. In addition, this representation of CH class is sufficient for numerically evaluating the covariance function as a function of h via either quadrature or Monte Carlo methods. Additionally, with a certain change of variable, the above integral can be identified as belonging to a certain class of special functions that can be computed efficiently. More precisely, we have the following elegant representation of the CH class, justifying its name.

Corollary 1. *The proposed covariance function in Equation* (3) *can also be represented in terms of the confluent hypergeometric function of the second kind:*

$$C(h;\nu,\alpha,\beta,\sigma^2)=\frac{\sigma^2\Gamma(\nu+\alpha)}{\Gamma(\nu)}\mathcal{U}\left(\alpha,1-\nu,\nu\left(\frac{h}{\beta}\right)^2\right), \tag{4}$$

where $\sigma^2>0,\alpha>0,\beta>0$*, and* $\nu>0$*. We name the proposed covariance class as the Confluent Hypergeometric (CH) class after the confluent hypergeometric function.*

Proof. By making the change of variable $x=\beta^2/t$, standard calculation yields that

$$C(h;\nu,\alpha,\beta,\sigma^2)=\frac{\sigma^2\Gamma(\nu+\alpha)}{\Gamma(\nu)\Gamma(\alpha)}\int_0^\infty t^{\alpha-1}(t+1)^{-(\nu+\alpha)}\exp(-\nu h^2 t/\beta^2)dt.$$

Thus, the conclusion follows by recognizing the form of the confluent hypergeometric function of the second kind $\mathcal{U}(a,b,c)$ from Chapter 13.2 of Abramowitz and Stegun (1965). □

Equation (4) provides a convenient way to evaluate the CH covariance function, since efficient numerical calculation of the confluent hypergeometric function is implemented in various libraries such as the GNU scientific library (Galassi et al., 2002) and softwares including R and MATLAB, facilitating its practical deployment; see Section S.1 of the Supplementary Material for an illustration of computing times for Bessel function and confluent hypergeometric function. For certain special parameter values, the evaluation of the confluent hypergeometric covariance function can be as easy as the Matérn covariance function; see Chapter 13.6 of Abramowitz and Stegun (1965). Besides the computational convenience, the CH covariance function in Equation (3) also allows us to make precise statements concerning the origin and tail behaviors of the resultant mixture. The next theorem makes the origin and tail behaviors explicit.

Theorem 2. *The CH class has the following two properties:*

(a) ***Origin behavior****: The differentiability of the CH class is solely controlled by* ν *in the same way as the Matérn class given in Equation* (1)*.*

(b) ***Tail behavior****:* $C(h;\nu,\alpha,\beta,\sigma^2)\sim\frac{\sigma^2\beta^{2\alpha}\Gamma(\nu+\alpha)}{\nu^\alpha\Gamma(\nu)}|h|^{-2\alpha}L(h^2)$ *as* $h\to\infty$*, where* $L(x)$ *is a slowly varying function at* ∞ *of the form* $L(x)=\{x/(x+\beta^2/(2\nu))\}^{\nu+\alpha}$*.*

Theorem 2 indicates that random processes with the CH covariance function are $\lfloor \nu \rfloor$ times mean-square differentiable. The local behavior of this CH covariance function is very flexible in the sense that the parameter ν can allow for any degree of mean-square differentiability of a weakly stationary process in the same way as the Matérn class. However, its tail behavior is quite different from that of the Matérn class, since the CH covariance has a polynomial tail that decays slower than the exponential tail in the Matérn covariance. This is natural since mixture inflates the tails in general, and in our particular case, changes the exponential decay to a polynomial one. The rate of tail decay is controlled by the parameter α. Thus, the CH class is more suitable for modeling polynomially decaying dependence which any exponentially decaying covariance function fails to capture. Moreover, the control over the degree of smoothness of process realizations is not lost. Theorem 2 also establishes a very desirable property that the degrees of differentiability near origin and the rate of decay of the tail for the CH covariance are controlled by two different parameters, ν and α, independently of each other. Each of these parameters can allow any degrees of flexibility.

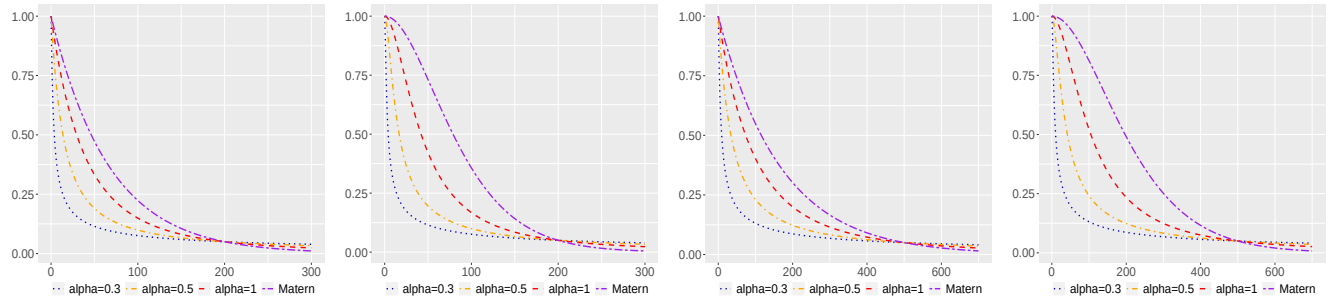
Remark 3. Porcu and Stein (2012) point out that it is possible to obtain a covariance function with flexible origin behavior and polynomial tails by simply taking a sum of a Matérn and a Cauchy covariance, which is again a valid covariance function. There are three major difficulties in this approach compared to ours: (a) the individual covariances in such a finite sum are not identifiable and hence practical interpretation becomes difficult, although prediction may still be feasible, (b) this summed covariance has five parameters, hindering its practical use in both frequentist and Bayesian settings, since numerical optimization of the likelihood function is costly and judicious prior elicitation is likely to be difficult. In contrast, our covariance has four parameters, each of which has a well-defined role. Finally, (c) the microergodic parameter under such a summed covariance is not likely available in closed form, in contrast to ours, as derived later in Section 3.

Remark 4. Our approach to constructing the CH covariance by mixing over ϕ^2 leads to a well-defined covariance class that can be used for Gaussian process modeling. The resulting covariance has four parameters and inference can be performed either via maximum likelihood or Bayesian

approaches, although we solely focus on the former in the current work. This construction should not be confused with Bayesian spatial modeling where a standard practice is to put a prior on the spatial range parameter in the Matérn covariance. More importantly, the likelihood under the CH covariance is fundamentally different from the posterior that is proportional to the product of the likelihood under the Matérn covariance and the prior on the spatial range parameter, where the prior could either be discrete or inverse gamma.

REMARK 5. It is also worthing noting that our construction yields a covariance that is fundamentally different from a finite sum of Matérn covariances where the range parameter is assigned a discrete prior, since the latter does not possess polynomially decaying dependence and is undesirable for modeling spatial data in practice due to costly computation and lack of practical motivation. Moreover, individual covariances in the finite sum are not identifiable.

Example 1. *This example visualizes the difference between the CH class and the Matérn class. We fix the* effective range *(ER) at* 200 *and* 500*, where ER is defined as the distance at which a correlation function has value approximately* 0.05*. For the CH class, we find the corresponding scale parameter β such that the ER corresponds to* 200 *and* 500 *under different smoothness parameters $\nu \in \{0.5, 2.5\}$ and different tail decay parameters $\alpha \in \{0.3, 0.5, 1\}$. For the Matérn class, we find the corresponding range parameter ϕ such that the ER corresponds to* 200 *and* 500 *under smoothness parameters $\nu \in \{0.5, 2.5\}$. These correlation functions are visualized in Figure 2. As the CH correlation has a polynomial tail, it drops much faster than the Matérn correlation in order to reach the same correlation 0.05 at the same ER. As α becomes smaller, the new correlation has a heavier tail, and hence it drops more quickly to reach the correlation 0.05 at the same effective range. After the ER, the CH correlation with a smaller α decays slower than those with larger α. The faster decay of the tail in the Matérn class is indicated by the behavior after the ER. Corresponding 1-dimensional process realizations can be found in Section S.2 of the Supplementary Material.*



(a) $\nu = 0.5, ER = 200$ **(b)** $\nu = 2.5, ER = 200$ **(c)** $\nu = 0.5, ER = 500$ **(d)** $\nu = 2.5, ER = 500$

Fig. 2. Correlation functions for the CH class and the Matérn class. The panels (a) and (b) show the correlation functions with the effective range (ER) at 200. The panels (c) and (d) show the correlation functions with the effective range (ER) at 500. ER is defined as the distance at which correlation is approximately 0.05.

3 Theoretical Properties of the CH Class

For an isotropic random field, the properties of a covariance function can be characterized by its spectral density. The tail behavior of the spectral density can be used to derive properties of the theoretical results shown in later sections. The following proposition characterizes the tail behavior of the spectral density for the CH covariance function in Equation (3).

Proposition 1 (**Tail behavior of the spectral density**). *The spectral density of the CH covariance function in Equation* (3) *admits the following tail behavior:*

$$f(\omega) \sim \frac{\sigma^2 2^{2\nu} \nu^\nu \Gamma(\nu+\alpha)}{\pi^{d/2} \beta^{2\nu} \Gamma(\alpha)} \omega^{-(2\nu+d)} L(\omega^2), \quad \omega \to \infty,$$

where $L(x) = \{x/(x + \beta^2/(2\nu))\}^{\nu+d/2}$ *is a slowly varying function at* ∞.

Recall that the spectral density of the Matérn class is proportional to $\omega^{-(2\nu+d)}$ for large ω. By mixing over the range parameter with an inverse gamma mixing density, the high-frequency behavior of the CH class differs from that of the Matérn class by a slowly varying function $L(\omega^2)$ up to a constant that does not depend on any frequency.

3.1 Equivalence Results

Let $(\Omega, \mathcal{F})$ be a measurable space with sample space Ω and σ-algebra $\mathcal{F}$. Two probability measures $\mathcal{P}_1, \mathcal{P}_2$ defined on the same measurable space $(\Omega, \mathcal{F})$ are said to be *equivalent* if $\mathcal{P}_1$ is absolutely continuous with respect to $\mathcal{P}_2$ and $\mathcal{P}_2$ is absolutely continuous with respect to $\mathcal{P}_1$. Suppose that the σ-algebra $\mathcal{F}$ is generated by a random process $\{Z(\mathbf{s}) : \mathbf{s} \in \mathcal{D}\}$. If $\mathcal{P}_1$ is equivalent to $\mathcal{P}_2$ on the σ-algebra $\mathcal{F}$, the probability measures $\mathcal{P}_1$ and $\mathcal{P}_2$ are then said to be equivalent on the realizations of the random process $\{Z(\mathbf{s}) : \mathbf{s} \in \mathcal{D}\}$. It follows immediately that the equivalence of two probability measures on the σ-algebra $\mathcal{F}$ implies that their equivalence on any σ-algebra $\mathcal{F}' \subset \mathcal{F}$. The equivalence of probability measures has important applications to statistical inferences on parameter estimation and prediction according to Zhang (2004). The equivalence between $\mathcal{P}_1$ and $\mathcal{P}_2$ implies that $\mathcal{P}_1$ cannot be correctly distinguished from $\mathcal{P}_2$ with probability 1 under measure $\mathcal{P}_1$ for any realizations. Let $\{\mathcal{P}_\theta : \theta \in \Theta\}$ be a collection of equivalent measures indexed by θ in the parameter space Θ. Let $\hat{\theta}_n$ be an estimator for θ based on n observations. Then $\hat{\theta}_n$ cannot converge to θ in probability regardless of what is observed (Zhang, 2004). Otherwise, for any fixed $\theta \in \Theta$, there exists a subsequence $\{\hat{\theta}_{n_k}\}_{k\geq 1}$ such that $\hat{\theta}_{n_k}$ converges to θ with probability 1 under measure $\mathcal{P}_\theta$ (see, e.g., Dudley, 2002, p. 288). For any $\theta' \in \Theta$ with $\theta' \neq \theta$, it follows from the property of equivalent measures that $\hat{\theta}_{n_k}$ also converges to θ with probability 1 under measure $\mathcal{P}_{\theta'}$. This further implies that there exists a subsubsequence $\{\hat{\theta}_{n_{k_r}}\}_{r\geq 1}$ that converges to θ' with probability 1 under measure $\mathcal{P}_{\theta'}$. Hence, the subsequence $\{\hat{\theta}_{n_k}\}_{k\geq 1}$ and its subsubsequence $\{\hat{\theta}_{n_{k_r}}\}_{r\geq 1}$ converge to two different values under the same measure $\mathcal{P}_{\theta'}$. By contradiction, $\hat{\theta}_n$ cannot converge to θ in probability. This implies that individual parameters cannot be estimated consistently under equivalent measures. The second application of equivalent measures concerns the asymptotic efficiency of predictors that is discussed in Section 3.3.

The tail behavior of the spectral densities in Proposition 1 can be used to check the equivalence of probability measures generated by stationary Gaussian random fields. The details of equivalence of Gaussian measures and the condition for equivalence are given in Section S.3 of the

Supplementary Material. Any zero-mean Gaussian process with a covariance function defines a corresponding Gaussian probability measure. In what follows, we say that the Gaussian probability measure defined under a covariance function implies that such Gaussian measure is defined through a Gaussian process on a bounded domain with mean zero and a covariance function. Our first result on equivalence of two Gaussian measures under the CH class is given in Theorem 3.

Theorem 3. *Let $\mathcal{P}_i$ be the Gaussian probability measure corresponding to the covariance $C(h; \nu, \alpha_i, \beta_i, \sigma_i^2)$ with $\alpha_i > d/2$ for $i = 1, 2$. Then $\mathcal{P}_1$ and $\mathcal{P}_2$ are equivalent on the realizations of $\{Z(\mathbf{s}) : \mathbf{s} \in \mathcal{D}\}$ for any fixed and bounded set $\mathcal{D} \subset \mathbb{R}^d$ with infinite locations in $\mathcal{D}$ and $d = 1, 2, 3$ if and only if*

$$\frac{\sigma_1^2 \Gamma(\nu + \alpha_1)}{\beta_1^{2\nu} \Gamma(\alpha_1)} = \frac{\sigma_2^2 \Gamma(\nu + \alpha_2)}{\beta_2^{2\nu} \Gamma(\alpha_2)}. \tag{5}$$

An immediate consequence of Theorem 3 is that for fixed ν; the tail decay parameter α, the scale parameter β and the variance parameter σ^2 cannot be estimated consistently under the infill asymptotics. Instead, the quantity $\sigma^2 \beta^{-2\nu} \Gamma(\nu + \alpha)/\Gamma(\alpha)$ is consistently estimable and has been referred to as the *microergodic parameter*. We refer the readers to page 163 of Stein (1999) for the definition of microergodicity.

Theorem 3 gives the result on equivalent measures within the CH class. The CH class can allow the same smoothness behavior as the Matérn class, but it has a polynomially decaying tail that is quite different from the Matérn class. One may ask whether there is an analogous result on the Gaussian measures under the CH class and the Matérn class. Theorem 4 provides an answer to this question.

Theorem 4. *Let $\mathcal{P}_1$ be the Gaussian probability measure under the CH covariance $C(h; \nu, \alpha, \beta, \sigma_1^2)$ with $\alpha > d/2$ and $\mathcal{P}_2$ be the Gaussian probability measure under the Matérn covariance function $\mathcal{M}(h; \nu, \phi, \sigma_2^2)$. If*

$$\sigma_1^2 (\beta^2/2)^{-\nu} \Gamma(\nu + \alpha)/\Gamma(\alpha) = \sigma_2^2 \phi^{-2\nu}, \tag{6}$$

then $\mathcal{P}_1$ and $\mathcal{P}_2$ are equivalent on the realizations of $\{Z(\mathbf{s}) : \mathbf{s} \in \mathcal{D}\}$ for any fixed and bounded set

$\mathcal{D} \subset \mathbb{R}^d$ *with* $d = 1, 2, 3$.

Theorem 4 gives the conditions under which the Gaussian measures under the CH class and the Matérn class are equivalent. If the condition in Equation (6) is satisfied, the Gaussian measure under the CH class cannot be distinguished from the Gaussian measure under the Matérn class, regardless of what is observed. This shows the robustness property for statistical inference under the CH class when the underlying true covariance model is the Matérn class.

In Section 3.2, the microergodic parameter of the CH class can be shown to be consistently estimated under infill asymptotics for a Gaussian process under the CH model with fixed and known ν. Moreover, one can show that the maximum likelihood estimator of this microergodic parameter converges to a normal distribution.

3.2 Asymptotic Normality

Let $\{Z(\mathbf{s}) : \mathbf{s} \in \mathcal{D}\}$ be a zero mean Gaussian process with the covariance function $C(h; \nu, \alpha, \beta, \sigma^2)$, where $\mathcal{D} \subset \mathbb{R}^d$ is a bounded subset of $\mathbb{R}^d$ with $d = 1, 2, 3$. Let $\mathbf{Z}_n := (Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n))^\top$ be a partially observed realization of the process $Z(\cdot)$ at n distinct locations in $\mathcal{D}$, denoted by $\mathcal{D}_n := \{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$. Then the log-likelihood function is

$$\ell_n(\sigma^2, \boldsymbol{\theta}) = -\frac{1}{2}\left\{ n \log(2\pi\sigma^2) + \log|\mathbf{R}_n(\boldsymbol{\theta})| + \frac{1}{\sigma^2}\mathbf{Z}_n^\top \mathbf{R}_n^{-1}(\boldsymbol{\theta})\mathbf{Z}_n \right\}, \tag{7}$$

where $\boldsymbol{\theta} := \{\alpha, \beta\}$ and $\mathbf{R}_n(\boldsymbol{\theta}) = [R(|\mathbf{s}_i - \mathbf{s}_j|; \boldsymbol{\theta})]_{i,j=1,\ldots,n}$ is an $n \times n$ correlation matrix with the correlation function $R(h) := C(h)/\sigma^2$.

In what follows, ν is assumed to be known and fixed. Let $\hat{\sigma}_n^2$ and $\hat{\boldsymbol{\theta}}_n$ be the maximum likelihood estimators (MLE) for σ^2 and $\boldsymbol{\theta}$ by maximizing the log-likelihood function in Equation (7). To show the consistency and asymptotic normality results for the microergodic parameter, we first obtain an estimator for σ^2 when $\boldsymbol{\theta}$ is fixed: $\hat{\sigma}_n^2 = \mathbf{Z}_n^\top \mathbf{R}_n^{-1}(\boldsymbol{\theta})\mathbf{Z}_n/n$. Then, let $\hat{c}_n(\boldsymbol{\theta})$ be the maximum likelihood

estimator of $c(\boldsymbol{\theta}) := \sigma^2\beta^{-2\nu}\Gamma(\nu+\alpha)/\Gamma(\alpha)$, as a function of $\boldsymbol{\theta}$, given by

$$\hat{c}_n(\boldsymbol{\theta}) = \hat{c}_n(\alpha,\beta) = \frac{\hat{\sigma}_n^2\Gamma(\nu+\alpha)}{\beta^{2\nu}\Gamma(\alpha)} = \frac{\mathbf{Z}_n^\top \mathbf{R}_n^{-1}(\boldsymbol{\theta})\mathbf{Z}_n\Gamma(\nu+\alpha)}{n\beta^{2\nu}\Gamma(\alpha)}.$$

For notational convenience, we use $c(\alpha,\beta)$ instead of $c(\boldsymbol{\theta})$ to denote the microergodic parameter in what follows. We discuss three situations. In the first situation, we consider joint estimation of β and σ^2 for fixed α. The MLE of β will be denoted by $\hat{\beta}_n$, and the MLE of the microergodic parameter is $\hat{c}_n(\alpha,\hat{\beta}_n)$. In the second situation, we consider joint estimation of α and σ^2 for fixed β. The MLE of α will be denoted by $\hat{\alpha}_n$ and the MLE of the microergodic parameter is $\hat{c}_n(\hat{\alpha}_n,\beta)$. In the third situation, we consider joint estimation of all parameters α,β,σ^2. The corresponding MLE for $c(\boldsymbol{\theta})$ is denoted by $\hat{c}_n(\hat{\boldsymbol{\theta}}_n)$, where $\hat{\boldsymbol{\theta}}_n := \{\hat{\alpha}_n,\hat{\beta}_n\}$. Note that the MLEs of either α or β (or both) are typically computed numerically, since there is no closed-form expression. We have the following results on the asymptotic properties of $\hat{c}_n(\boldsymbol{\theta})$ for various scenarios of α and β under the infill asymptotics.

Theorem 5 (**Asymptotics of the MLE**). *Let $\mathcal{P}_0$ be the Gaussian measure defined under the covariance function $C(h;\nu,\alpha_0,\beta_0,\sigma_0^2)$ with $\sigma_0^2 > 0$, and let $\boldsymbol{\theta}_0 := \{\alpha_0,\beta_0\}$. Let Z_n be the set of observations generated under $\mathcal{P}_0$. Then the following results can be established:*

(a) Suppose that $\alpha_0 > d/2$ and $\beta_0 \in [\beta_L,\beta_U]$, where β_L,β_U are fixed constants such that $0 < \beta_L < \beta_U$. For any fixed $\alpha > d/2$, if $(\hat{\sigma}_n^2,\hat{\beta}_n)$ maximizes the log-likelihood function (7) over $(0,\infty)\times[\beta_L,\beta_U]$, then as $n\to\infty$, $\hat{c}_n(\alpha,\hat{\beta}_n) \xrightarrow{a.s.} c(\boldsymbol{\theta}_0)$ under $\mathcal{P}_0$ and $\sqrt{n}\left\{\hat{c}_n(\alpha,\hat{\beta}_n) - c(\boldsymbol{\theta}_0)\right\} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 2[c(\boldsymbol{\theta}_0)]^2)$.

(b) Suppose that $\alpha_0 \in [\alpha_L,\alpha_U]$ and $\beta_0 > 0$, where α_L,α_U are fixed constants such that $d/2 < \alpha_L < \alpha_U$. For any fixed $\beta > 0$, if $(\hat{\sigma}_n^2,\hat{\alpha}_n)$ maximizes the log-likelihood function (7) over $(0,\infty)\times[\alpha_L,\alpha_U]$, then as $n\to\infty$, $\hat{c}_n(\hat{\alpha}_n,\beta) \xrightarrow{a.s.} c(\boldsymbol{\theta}_0)$ under $\mathcal{P}_0$ and $\sqrt{n}\left\{\hat{c}_n(\hat{\alpha}_n,\beta) - c(\boldsymbol{\theta}_0)\right\} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 2[c(\boldsymbol{\theta}_0)]^2)$.

(c) Suppose that $\alpha_0 \in [\alpha_L,\alpha_U]$ and $\beta_0 \in [\beta_L,\beta_U]$ where $\alpha_L,\alpha_U,\beta_L,\beta_U$ are fixed constants such that $d/2 < \alpha_L < \alpha_U$ and $0 < \beta_L < \beta_U$. If $(\hat{\sigma}_n^2,\hat{\alpha}_n,\hat{\beta}_n)$ maximizes the log-likelihood func-

tion (7) *over* $(0,\infty)\times[\alpha_L,\alpha_U]\times[\beta_L,\beta_U]$, *then as* $n\to\infty$, $\hat{c}_n(\hat{\boldsymbol{\theta}}_n)\xrightarrow{a.s.}c(\boldsymbol{\theta}_0)$ *under* $\mathcal{P}_0$ *and* $\sqrt{n}\left\{\hat{c}_n(\hat{\boldsymbol{\theta}}_n)-c(\boldsymbol{\theta}_0)\right\}\xrightarrow{\mathcal{L}}\mathcal{N}\left(0,2[c(\boldsymbol{\theta}_0)]^2\right)$.

The first two results of Theorem 5 imply that the microergodic parameter can be estimated consistently by fixing $\alpha > d/2$ or $\beta > 0$ in compact sets. In practice, fixing either α or β may be too restrictive for modeling spatial processes. We would expect that the finite sample prediction performance can be improved by jointly estimating all covariance parameters for the CH class, which should be the preferred approach for practical purposes.

The third result of Theorem 5 establishes that the microergodic parameter can be consistently estimated by jointly maximizing the log-likelihood (7) over α and β. However, the current result requires that $\alpha > d/2$. This means that the CH covariance cannot decay too slowly in its tail in order to establish the consistency result. Nevertheless, this result shows a significant improvement over existing asymptotic normality results for other types of polynomially decaying covariance functions. For instance, it was shown by Bevilacqua and Faouzi (2019) that the microergodic parameter in the generalized Cauchy class can be estimated consistently under infill asymptotics. However, their results assume that the parameter that controls the tail behavior is fixed. This is similar to the first result of Theorem 5. Unlike their results, a theoretical improvement in Theorem 5 is that the asymptotic results for the MLE of the microergodic parameter $c(\boldsymbol{\theta})$ can be obtained for joint estimation of all three parameters, including the parameter that controls the decay of the tail. We provide extensive numerical evidence in support of Theorem 5 in Section S.5 of the Supplementary Material.

3.3 Asymptotic Prediction Efficiency

This section is focused on studying the prediction problem of Gaussian process at a new location $\mathbf{s}_0 \in \mathcal{D}\cap\mathcal{D}_n^c$. This problem has been studied extensively when an incorrect covariance model is used. Our focus here is to show the asymptotic efficiency and asymptotically correct estimation of prediction variance in the context of the CH class. Stein (1988) shows that both of these two

properties hold when the Gaussian measure under a misspecified covariance model is equivalent to the Gaussian measure under the true covariance model. In the case of the CH class, Theorem 3 gives the conditions for equivalence of two Gaussian measures in the light of the microergodic parameter $c(\boldsymbol{\theta}) = \sigma^2\beta^{-2\nu}\Gamma(\nu+\alpha)/\Gamma(\alpha)$. As in Section 3.2, ν will be assumed to be fixed.

With observations generated under the CH model $C(h;\nu,\alpha,\beta,\sigma^2)$, we define the best linear unbiased predictor for $Z(\mathbf{s}_0)$ to be

$$\hat{Z}_n(\boldsymbol{\theta}) = \mathbf{r}_n^\top(\boldsymbol{\theta})\mathbf{R}_n^{-1}(\boldsymbol{\theta})\mathbf{Z}_n, \tag{8}$$

where $\mathbf{r}_n(\boldsymbol{\theta}) := [R(|\mathbf{s}_0 - \mathbf{s}_i|;\boldsymbol{\theta})]_{i=1,\ldots,n}$ is an n-dimensional vector. This predictor depends only on correlation parameters $\{\alpha,\beta\}$. If the true covariance is $C(h;\nu,\alpha_0,\beta_0,\sigma_0^2)$, the mean squared error of the predictor in Equation (8) is given by $\mathrm{Var}_{\nu,\boldsymbol{\theta}_0,\sigma_0^2}\{\hat{Z}_n(\boldsymbol{\theta}) - Z(\mathbf{s}_0)\} = \sigma_0^2\left\{1 - 2\mathbf{r}_n^\top(\boldsymbol{\theta})\mathbf{R}_n^{-1}(\boldsymbol{\theta})\mathbf{r}_n(\boldsymbol{\theta}_0)\right.$ $\left.+\mathbf{r}_n^\top(\boldsymbol{\theta})\mathbf{R}_n^{-1}(\boldsymbol{\theta})\mathbf{R}_n(\boldsymbol{\theta}_0)\mathbf{R}_n^{-1}(\boldsymbol{\theta})\mathbf{r}_n(\boldsymbol{\theta})\right\}$. If $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, i.e., $\alpha = \alpha_0$ and $\beta = \beta_0$, the above expression simplifies to

$$\mathrm{Var}_{\nu,\boldsymbol{\theta}_0,\sigma_0^2}\{\hat{Z}_n(\boldsymbol{\theta}_0) - Z(\mathbf{s}_0)\} = \sigma_0^2\left\{1 - \mathbf{r}_n^\top(\boldsymbol{\theta}_0)\mathbf{R}_n^{-1}(\boldsymbol{\theta}_0)\mathbf{r}_n(\boldsymbol{\theta}_0)\right\}. \tag{9}$$

If the true model is $\mathcal{M}(h;\nu,\phi,\sigma^2)$, analogous expressions can be derived for $\mathrm{Var}_{\nu,\phi_0,\sigma_0^2}\{\hat{Z}_n(\boldsymbol{\theta}) - Z(\mathbf{s}_0)\}$.

Let $\mathcal{P}_0$ be the Gaussian measure defined under the true covariance model and $\mathcal{P}_1$ be the Gaussian measure defined under the misspecified covariance model. The following results concern the asymptotic equivalence between the best linear predictor (BLP) under a misspecified probability measure $\mathcal{P}_1$ and the BLP under the the true measure $\mathcal{P}_0$.

Theorem 6. *Suppose that $\mathcal{P}_0, \mathcal{P}_1$ are two Gaussian probability measures defined by a zero mean Gaussian process with the CH class $C(h;\nu,\alpha_i,\beta_i,\sigma_i^2)$ for $i = 1,2$ on $\mathcal{D}$. The following results hold true:*

(a) For any fixed $\boldsymbol{\theta}_1$, under $\mathcal{P}_0$, as $n \to \infty$,

$$\frac{Var_{\nu,\boldsymbol{\theta}_0,\sigma_0^2}\{\hat{Z}_n(\boldsymbol{\theta}_1) - Z(\mathbf{s}_0)\}}{Var_{\nu,\boldsymbol{\theta}_0,\sigma_0^2}\{\hat{Z}_n(\boldsymbol{\theta}_0) - Z(\mathbf{s}_0)\}} \to 1.$$

(b) Moreover, if $\sigma_0^2\beta_0^{-2\nu}\Gamma(\nu+\alpha_0)/\Gamma(\alpha_0) = \sigma_1^2\beta_1^{-2\nu}\Gamma(\nu+\alpha_1)/\Gamma(\alpha_1)$*, then under* $\mathcal{P}_0$*, as* $n \to \infty$*,*

$$\frac{Var_{\nu,\boldsymbol{\theta}_1,\sigma_1^2}\{\hat{Z}_n(\boldsymbol{\theta}_1) - Z(\mathbf{s}_0)\}}{Var_{\nu,\boldsymbol{\theta}_0,\sigma_0^2}\{\hat{Z}_n(\boldsymbol{\theta}_1) - Z(\mathbf{s}_0)\}} \to 1.$$

(c) Let $\hat{\sigma}_n^2 = \mathbf{Z}_n^\top \mathbf{R}_n^{-1}(\boldsymbol{\theta}_1)\mathbf{Z}_n/n$*. It follows that almost surely under* $\mathcal{P}_0$*, as* $n \to \infty$*,*

$$\frac{Var_{\nu,\boldsymbol{\theta}_1,\hat{\sigma}_n^2}\{\hat{Z}_n(\boldsymbol{\theta}_1) - Z(\mathbf{s}_0)\}}{Var_{\nu,\boldsymbol{\theta}_0,\sigma_0^2}\{\hat{Z}_n(\boldsymbol{\theta}_1) - Z(\mathbf{s}_0)\}} \to 1.$$

Part (a) of Theorem 6 implies that if the smoothness parameter ν is correctly specified, any values for α and β will result in asymptotically efficient predictors. The condition $\sigma_0^2\beta_0^{-2\nu}\Gamma(\nu+\alpha_0)/\Gamma(\alpha_0) = \sigma_1^2\beta_1^{-2\nu}\Gamma(\nu+\alpha_1)/\Gamma(\alpha_1)$ is not necessary for asymptotic efficiency, but it provides asymptotically correct estimate of the mean squared prediction error (MSPE). The quantity $\mathrm{Var}_{\nu,\boldsymbol{\theta}_1,\sigma_1^2}\{\hat{Z}_n(\boldsymbol{\theta}_1) - Z(\mathbf{s}_0)\}$ is the MSPE for $\hat{Z}_n(\boldsymbol{\theta}_1)$ under the model $C(h;\nu,\alpha_1,\beta_1,\sigma_1^2)$, while the quantity $\mathrm{Var}_{\nu,\boldsymbol{\theta}_0,\sigma_0^2}\{\hat{Z}_n(\boldsymbol{\theta}_1) - Z(\mathbf{s}_0)\}$ is the true MSPE for $\hat{Z}_n(\boldsymbol{\theta}_1)$ under the true model $C(h;\nu,\alpha_0,\beta_0,\sigma_0^2)$. In practice, it is common to estimate model parameters and then prediction is made by plugging these estimates into Equations (8) and (9). Part (c) shows the same convergence results when $\boldsymbol{\theta}$ is fixed at $\boldsymbol{\theta}_1$, but σ^2 is estimated via the maximum likelihood method.

One can conjecture that the result in Part (c) of Theorem 6 still holds if $\boldsymbol{\theta}_1$ is replaced by its maximum likelihood estimator, but its proof seems elusive. Theorem 6 demonstrates the asymptotic prediction efficiency for the CH class. The following results are established to show the asymptotic efficiency of the best linear predictor under the CH class when the true Gaussian measure is defined by a zero-mean Gaussian process under the Matérn class.

Theorem 7. *Let* $\mathcal{P}_0$ *be the Gaussian probability measure under the Matérn covariance* $\mathcal{M}(h;\nu,\phi,\sigma_0^2)$ *and* $\mathcal{P}_1$ *be the Gaussian probability measure under the CH covariance* $C(h;\nu,\alpha,\beta,\sigma_1^2)$ *on* $\mathcal{D}$*.* $\hat{Z}_n(\alpha,\beta)$ *be the kriging predictor under* $C(h;\nu,\alpha,\beta,\sigma_1^2)$ *and* $\hat{Z}_n(\phi)$ *be the kriging predictor under* $\mathcal{M}(h;\nu,\phi,\sigma_0^2)$*. If the condition in Equation* (6) *is satisfied, then it follows that under the Gaussian measure* $\mathcal{P}_0$*, as* $n \to \infty$*,*

$$\frac{Var_{\nu,\alpha,\beta,\sigma_1^2}\{\hat{Z}_n(\alpha,\beta) - Z(\mathbf{s}_0)\}}{Var_{\nu,\phi,\sigma_0^2}\{\hat{Z}_n(\phi) - Z(\mathbf{s}_0)\}} \to 1,$$

for any fixed $\alpha > 0$ *and* $\beta > 0$.

A key consequence of Theorem 7 is that when a true Gaussian process is generated by the Matérn covariance model, the CH covariance model (3) can yield an asymptotically equivalent BLP. The practical implication is when the true model is generated from the Matérn class, the predictive performance under the CH class is indistinguishable from that under the Matérn class as the number of observations gets larger in a fixed domain. Both Theorem 6 and Theorem 7 imply that the kriging predictor under the CH class can allow robust prediction property even if the underlying true covariance model is misspecified.

4 Numerical Illustrations

In this section, we use simulated examples to study the properties of the CH class and compare with alternative covariance models. In what follows, we compare the CH model with the other two covariance models: the Matérn class and the generalized Cauchy class. The predictive performance is evaluated based on root mean-squared prediction errors (RMSPE), coverage probability of the 95% percentile confidence intervals (CVG), and the average length of the predictive confidence intervals (ALCI) at held-out locations.

The goal of this section is to study the finite sample predictive performance under the CH model in interpolative settings. Specifically, we consider three different cases, where the true covariance model is specified as the Matérn covariance (Case 1), the CH covariance (Case 2) and the generalized Cauchy (GC) covariance (Case 3), respectively. The Matérn class is very flexible near origin and has an exponentially decaying tail, the CH class is also very flexible near origin but has a polynomially decaying tail, and the GC class is either non-differentiable or infinitely differentiable and has a polynomially decaying tail. The GC covariance has the form $C(h) = \sigma^2 \left\{1 + (h/\phi)^\delta\right\}^{-\lambda/\delta}$, where

$\sigma^2 > 0$ is the variance parameter, $\phi > 0$ is the range parameter, $\lambda \in (0, d]$ is the parameter controlling the degree of polynomial decay, and $\delta \in (0, 2]$ is the smoothness parameter. When $\delta \in (0, 2)$, the corresponding process is nowhere mean-square differentiable. When $\delta = 2$, it corresponds to the Cauchy covariance, whose process is infinitely mean-square differentiable. For each case, predictive performance is compared at held-out locations with estimated covariance structures.

We simulate data in the square domain $\mathcal{D} = [0, 2000] \times [0, 2000]$ from mean zero Gaussian processes with three different covariance models: the Matérn covariance (Case 1), the CH covariance (Case 2), and the GC covariance (Case 3) for a variety of settings. We simulate $n = 2000$ data points via maximin Latin hypercube design (Stein, 1987) for parameter estimation and evaluate predictive performance at 10-by-10 regular grid points in $\mathcal{D}$. We fix the variance parameter at 1 and consider moderate spatial dependence with effective range (ER) at 200 and 500 for the underlying true covariances. For each of these simulation settings, we use 30 different random number seeds to generate the realizations. We always choose the same smoothness parameter for the Matérn class and the CH class. For the GC covariance, we fix its smoothness parameter to be $\delta = \min\{2\nu, d\}$, since the Gaussian measure with the Matérn class could be equivalent to that with the GC class as pointed by Bevilacqua and Faouzi (2019). However, the smoothness parameter δ in the GC class cannot be greater than 2, otherwise the GC class is no longer a valid covariance function.

4.1 Case 1: Examples with the Matérn Class as Truth

In Case 1, we simulate Gaussian process realizations from the Matérn model with smoothness parameter ν fixed at 0.5 and 2.5 and effective range at 200 and 500. The parameters in each covariance model are estimated based on profile likelihood as described in Section 3.2. Figure 3 shows the estimated covariance structures and summary of prediction results. Regardless of the smoothness behavior and strength of dependence in the underlying true process, there is no clear difference between the CH class and the Matérn class in terms of estimated covariance structures

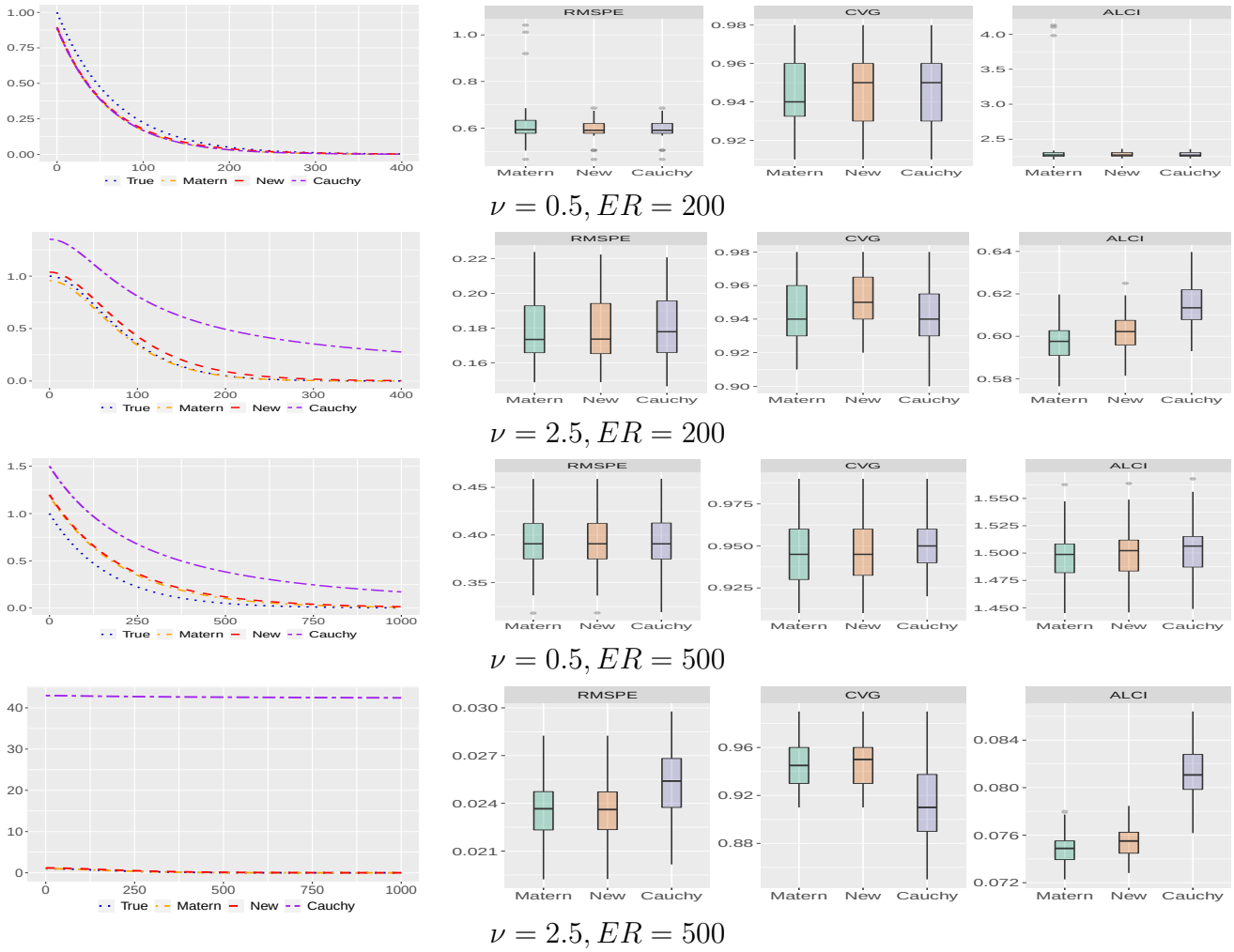


Fig. 3. Case 1: Comparison of predictive performance and estimated covariance structures when the true covariance is the Matérn class with 2000 observations. The predictive performance is evaluated at 10-by-10 regular grids in the square domain. These figures summarize the predictive measures based on RMSPE, CVG and ALCI under 30 simulated realizations.

and prediction performance. In contrast, the estimated GC covariance structure only performs as accurately as the Matérn class when $\nu = 0.5$. When the process is twice mean-square differentiable ($\nu = 2.5$), as expected, the GC class cannot mimic such behavior, and hence, yields worse estimates of the covariance structures and prediction results compared to both the Matérn class and the CH class. The CH covariance is able to capture the true covariance structure as implied by Theorem 7. In terms of RMSPE, there is no clear difference between the estimated CH covariance and the estimated Matérn covariance. However, the CVG and ALCI based on the CH class are slightly larger than those based on the estimated Matérn covariance.

4.2 Case 2: Examples with the CH Class as Truth

In Case 2, we simulate Gaussian process realizations from the CH covariance model with smoothness parameter ν fixed at 0.5 and 2.5, tail decay parameter fixed at 0.5, and effective range fixed at 200 and 500. Figure 4 shows the estimated covariance structures and summary of prediction results. As expected, when the underlying true process is simulated from a process with polynomially decaying dependence, the Matérn class cannot be expected to capture such behavior. The prediction results also indicate that the Matérn class performs much worse than the other two covariance models. When the underlying true process is not differentiable ($\nu = 0.5$), there is no clear difference between the estimates under the GC covariance structure and the estimates under the CH covariance structure. However, when the underlying true process is twice differentiable ($\nu = 2.5$), it is obvious that the estimated GC covariance structure is not as accurate as the estimated CH covariance structure. This makes sense because the GC class is either non-differentiable or infinitely differentiable. In terms of prediction performance, the CH covariance class performs better than the GC class in terms of coverage probability.

4.3 Case 3: Examples with the GC Class as Truth

In Case 3, we simulate Gaussian process realizations from the GC class with the smoothness parameter $\delta = 1$ and $\lambda = 1$ under ER=200 and 500. The corresponding process is non-differentiable and corresponds to the smoothness parameter $\nu = 0.5$ in both the Matérn class and the CH class. The parameter λ in the GC class is fixed at 1 so that it corresponds to the tail parameter $\alpha = 0.5$ in the CH class. We did not consider Gaussian processes that are infinitely differentiable, since such processes are unrealistic for environmental processes. Figure 5 shows the estimated covariance structures and prediction results. As expected, the Matérn class performs much worse than the CH class and the GC class for the same reason as in Case 2. Between the CH class and the GC class, no difference is seen in terms of estimated covariance structures and predictive performances. This is not surprising, since the CH class has a tail decay parameter α that is able to capture the

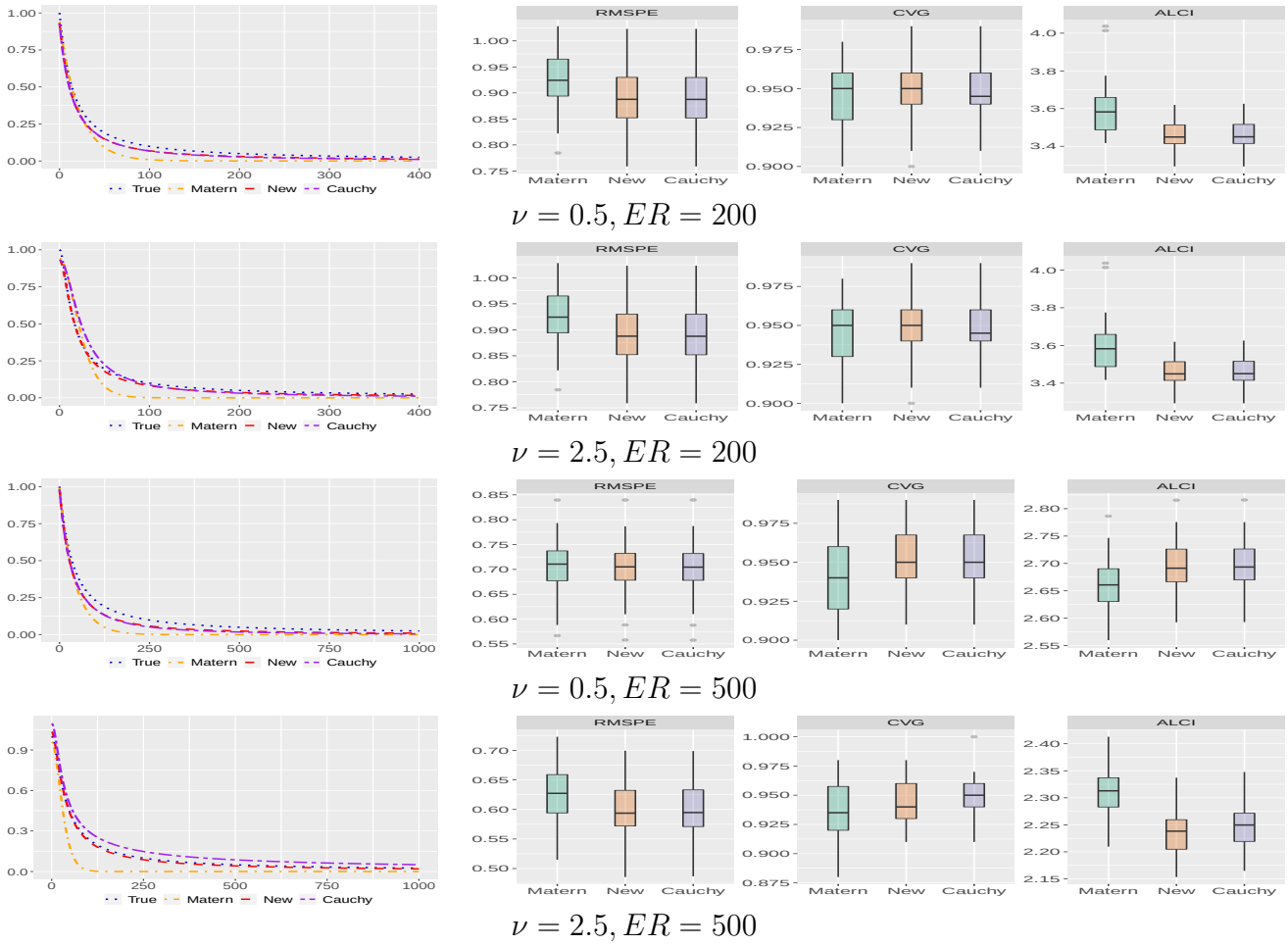


Fig. 4. Case 2: Comparison of predictive performance and estimated covariance structures when the true covariance is the CH class with 2000 observations. The predictive performance is evaluated at 10-by-10 regular grids in the square domain. These figures summarize the predictive measures based on RMSPE, CVG and ALCI under 30 simulated realizations.

tail behavior in the GC class.

5 Application to the OCO-2 Data

In this section, the proposed CH class is used to model spatial data collected from NASA's Orbiting Carbon Observatory-2 (OCO-2) satellite and comparisons are made in kriging performances with alternative covariances. The OCO-2 satellite is NASA's first dedicated remote sensing earth satellite to study atmospheric carbon dioxide from space with the primary objective to estimate the global geographic distribution of CO_2 sources and sinks at Earth's surface; see Cressie (2017);

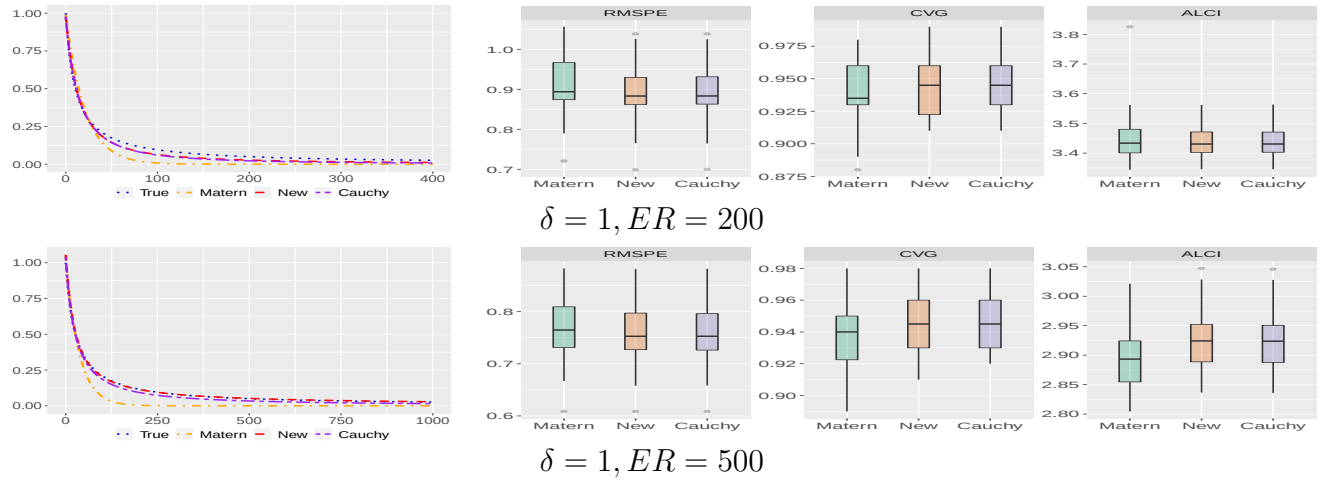
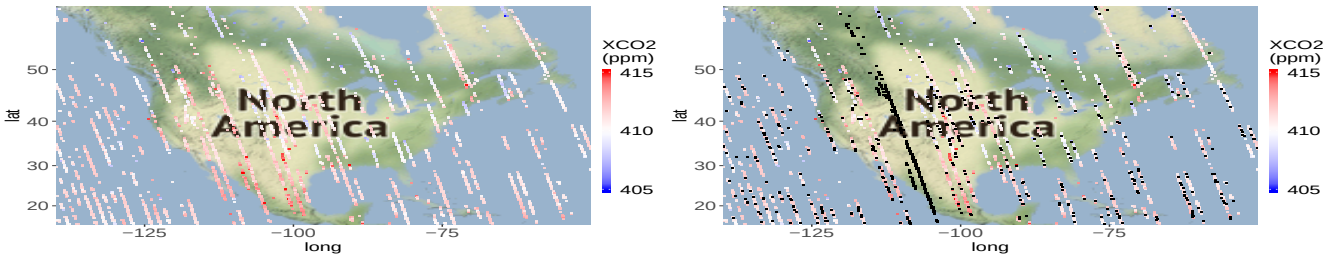


Fig. 5. Case 3: Comparison of predictive performance and estimated covariance structures when the true covariance is the GC class with 2000 observations. The predictive performance is evaluated at 10-by-10 regular grids in the square domain. These figures summarize the predictive measures based on RMSPE, CVG and ALCI under 30 simulated realizations.

Wunch et al. (2011) for detailed discussions. The OCO-2 satellite carries three high-resolution grating spectrometers designed to measure the near-infrared absorption of reflected sunlight by carbon dioxide and molecular oxygen and orbits over a 16-day repeat cycle. In this application, we consider NASA's Level 3 data product of the XCO2 at $0.25° \times 0.25°$ spatial resolution over one repeat cycle from June 1 to June 16, 2019. These gridded data were processed based on Level 2 data product by the OCO-2 project at the Jet Propulsion Laboratory, California Technology, and obtained from the OCO-2 data archive maintained at the NASA Goddard Earth Science Data and Information Services Center. They can be downloaded at `https://co2.jpl.nasa.gov/#mission=OCO-2`.

This Level 3 data product consists of 43,698 measurements. We focus on the study region that covers the entire United States with longitudes between 140W and 50W and latitudes between 15N and 60N. This region includes 3,682 measurements; see panel (a) of Figure 6. These data points are very sparse in space. As the OCO-2 satellite has swath width 10.6 kilometers, large missing gaps can be observed between swaths. Predicting the underlying geophysical process based on data with such patterns requires the statistical model not only to interpolate in space (prediction near observed locations) but also to extrapolate in space (prediction away from observed locations).

Given the data $\mathbf{Z} := (Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n))^\top$, we assume a typical spatial process model:



(a) XCO2 data in the study region. **(b)** XCO2 testing data in black.

Fig. 6. XCO2 measurements from June 1 to June 16, 2019 in the study region.

$$Z(\mathbf{s}) = Y(\mathbf{s}) + \epsilon(\mathbf{s}), \quad \mathbf{s} \in \mathcal{D},$$

where $Y(\cdot)$ is assumed to be a Gaussian process with mean function $\mu(\cdot)$ and covariance function $C(\cdot, \cdot)$. The term $\epsilon(\cdot)$ is assumed to be a spatial white-noise process accounting for the nugget effect with $\text{var}(\epsilon(\mathbf{s})) = \tau^2 > 0$. The goal of this analysis is to predict the process $Y(\mathbf{s}_0)$ for any $\mathbf{s}_0 \in \mathcal{D}$ based on the data $\mathbf{Z}$. Exploratory analysis indicates no clear trend, so we assume a constant trend for the mean function $\mu(\mathbf{s}) = b$. For this particular dataset, the assumption of an isotropic covariance function seems to be reasonable based on directional semivariograms in Figure S.10 of the Supplementary Material. For the covariance function $C(\cdot, \cdot)$, we assume the CH model with parameters $\{\sigma^2, \alpha, \beta, \nu\}$, where the smoothness parameter ν is fixed at 0.5 and 1.5, indicating the resulting process is non-differentiable or once differentiable, respectively. Here we fix ν in the Matérn and CH classes over a grid of values, since (a) estimating ν requires intensive computations and it has often been fixed in practice (e.g., Banerjee et al., 2014; Berger et al., 2001) and (b) the likelihood can be nearly flat (Berger et al., 2001; Gu et al., 2018; Stein, 1999, p. 173; Zhang, 2004) and hence it is notoriously difficult to estimate covariance parameters including ν with either profile or integrated likelihood functions (Gu et al., 2018). However, estimating ν may improve prediction accuracy in practice, which is left for future investigation.

To evaluate the performance of the CH class, we perform cross-validation and make comparisons with the Matérn class. The testing dataset consists of (1) a complete longitude band across the United States, which will be referred to as missing by design (MBD) and (2) randomly selected 15% of remaining XCO2 measurements, which will be referred to as missing at random (MAR). Panel (b) of Figure 6 highlights these testing data with black grid points. This dataset is used

for evaluating out-of-sample predictive performance in interpolative and extrapolative settings. The remaining data points are used for parameter estimation under the Matérn covariance and the CH covariance. The parameters are estimated based on the restricted maximum likelihood (Harville, 1974). Table 1 shows the predictive measures and estimated nugget parameters. The CH model with the smoothness parameter $\nu = 0.5$ yields the smallest estimated nugget parameter among all the models. This suggests that the CH model with $\nu = 0.5$ best captures the spatial dependence structure among all the models. The kriging predictions under the CH model show lots of fine-scale or micro-scale variations, which are more desired for accurate spatial prediction. In an interpolative setting, the Matérn covariance yields slightly smaller (but indistinguishable) RMSPE and ALCI over randomly selected locations than the CH covariance, which indicates that the Matérn covariance has slightly better interpolative prediction skill than the CH model in this application. The empirical coverage probability is closer to the nominal value of 0.95 under the Matérn covariance model. In contrast, in an extrapolative setting, the CH model yields much smaller RMSPE and ALCI than the Matérn covariance model with indistinguishable empirical coverage probabilities, which indicates that the CH model has a better extrapolative prediction skill than the Matérn covariance model. These prediction results are not surprising, since the Matérn class can only model exponentially decaying dependence while the CH class can offer considerable benefits for extrapolative predictions while maintains the same interpolative prediction skill as the Matérn class. The difference in interpolative prediction performance between the CH class and the Matérn class is negligible, in part because the CH class can yield asymptotically equivalent best linear predictors as the Matérn class under conditions established in Theorem 7. Notice that the empirical coverage probabilities under all the models are less than the nominal coverage probability 0.95, this is partly because uncertainties due to parameter estimation are not accounted for in the predictive distribution. A fully Bayesian analysis may remedy this issue.

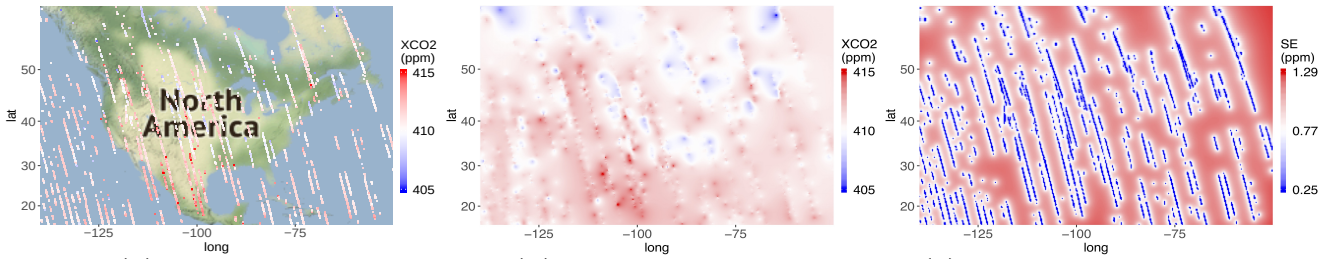
For other model parameters shown in Table S.4 of the Supplementary Material, we notice that the estimates of the regression parameters under the two different covariance models are very

similar. As expected, the estimated variance parameter (partial sill) is larger under the CH class than the one estimated under the Matérn class. Perhaps the most interesting parameter is the tail decay parameter in the CH class, which is estimated to be around 0.38. This clearly indicates that the underlying true process has a polynomially decaying dependence structure. As Gneiting (2013) points out, the Matérn class is positive definite on sphere only if $\nu \leq 0.5$ with great circle distance. To avoid this technical difficulty, we use chordal distance for modeling spatial data on sphere when $\nu > 0.5$, since it was pointed out on pages 71-77 of Yadrenko (1983) that chordal distance can guarantee the positive definiteness of a covariance function on $\mathbb{S}^d \times \mathbb{S}^d$ when the original covariance function is positive definite on $\mathbb{R}^{d+1} \times \mathbb{R}^{d+1}$.

Table 1. Cross-validation results on the XCO2 data based on the Matérn covariance model and the CH covariance model. The measures in the first coordinate correspond to those based on MAR locations for interpolative prediction, and the measures in the second coordinate correspond to those based on MBD locations for extrapolative prediction.

	Matérn class		CH class	
	$\nu = 0.5$	$\nu = 1.5$	$\nu = 0.5$	$\nu = 1.5$
τ^2 (nugget)	0.0642	0.2215	**0.0038**	0.1478
RMSPE	0.672, 1.478	0.675, 1.599	0.676, 1.263	0.735, 1.227
CVG(95%)	0.952, 0.929	0.952, 0.951	0.944, 0.921	0.878, 0.937
ALCI(95%)	2.533, 5.095	2.536, 5.044	2.543, 4.722	2.098, 4.855

Next, we predict the process $Y(\cdot)$ at $0.25^\circ \times 0.25^\circ$ grid in the study region. The parameters are estimated based on all the data points under the CH class and the Matérn class with the smoothness parameter fixed at 0.5. In Figure S.11 of the Supplementary Material, we observe that the optimal kriging predictors over these grid points under the CH covariance model generally yield smaller values than those under the Matérn covariance function model in large missing gaps except for certain regions such as the Gulf of Mexico. More importantly, we also observe that the CH covariance model yields 10% to 20% smaller kriging standard errors than the Matérn covariance model in the observed spatial locations and contiguous missing regions. This indicates that the CH covariance model has an advantage over the Matérn covariance in terms of in-sample prediction skills and in an extrapolative setting (such as large missing gaps). Prediction in an



(a) XCO2 data. **(b)** Kriging predictors. **(c)** Kriging standard errors.

Fig. 7. XCO2 data and kriging predictions based on the CH model.

interpolative setting (such as locations near the observed locations) shows that the CH class yields indistinguishable (no more than 2%) kriging standard errors compared to the Matérn class. It is clear to see that the CH class is able to show lots of fine-scale variations in the kriging map, which is a desirable property for prediction accuracy. This is partly because the nugget parameter under the CH covariance is estimated to be much smaller than that under the Matérn covariance and partly because the polynomially decaying dependence exhibited in the CH class can better utilize information at both nearby locations and distant locations to infer such fine-scale variations. Finally, Figure 7 shows the optimal kriging predictors and associated kriging standard errors at $0.25^\circ \times 0.25^\circ$ grid in the study region. These kriging maps help create a complete NASA Level 3 data product with associated uncertainties, which can be further used for downstream applications such as CO2 flux inversion.

6 Concluding Remarks

This paper introduces a new class of interpretable covariance functions called the *Confluent Hypergeometric* class that can allow precise and simultaneous control of the origin and tail behaviors with well-defined roles for each covariance parameter. Our approach in constructing the CH class is to mix over the range parameter of the Matérn class. As expected, the origin behavior of this CH class is as flexible as the Matérn class. The high-frequency behavior of the CH class is also similar to that of the Matérn class, since they differ by a slowly varying function up to a multiplicative constant. Unlike the Matérn class, however, this CH class has a polynomially decaying tail, which

allows for modeling power-law stochastic processes.

The advantage of the CH class is examined in theory and numerical examples. Conditions for equivalence of two Gaussian measures based on the CH class are established. We derive the conditions on the asymptotic efficiency of kriging predictors based on an increasing number of observations in a bounded region when the CH covariance is misspecified. We also show that the CH class can yield an asymptotically efficient kriging predictor under the infill asymptotics framework when the true covariance belongs to the Matérn class. It is worth noting that the CH class itself is valid and can allow any degrees of decaying tail, while the asymptotic results of the MLE for the microergodic parameters are proven for $\alpha > d/2$. Investigation of the similar theoretical result on the MLE is elusive for the case $\alpha \in (0, d/2]$. Extensive simulation results show that when the underlying true process is generated from either the Matérn covariance or the GC covariance, the CH covariance can allow robust prediction property. We also noticed in simulation study that the Matérn class gives worse performance than the CH class when the underlying true covariance has a polynomially decaying tail. In the real data analysis, we found significant advantages of the CH class when prediction is made in an extrapolative setting while the difference in terms of interpolative prediction is indistinguishable, which is implied by our theoretical results. This feature is practically important for spatial modeling especially with large missing patterns. Future work along the theoretical side is to establish theoretical results of the CH class under the increasing domain asymptotics.

This paper mainly focuses on theoretical contributions and practical advantage of the CH class. Common challenges in spatial statistics include modeling large spatial data and spatial nonstationarity, which are often tackled based on the Matérn class in recent developments (e.g., Lindgren et al., 2011; Ma and Kang, 2020). The proposed CH class can be used as a substantially improved starting point over the Matérn class to develop more complicated covariance models to tackle these challenges. Several extensions can be pursued. It is interesting to extend the proposed CH class for modeling dependence on sphere, space-time dependence, and/or multivariate dependence (e.g.,

Apanasovich et al., 2012; Ma and Kang, 2019; Ma et al., 2019). Prior elicitation for the CH class could be challenging. It is also interesting to develop objective priors such as reference prior to facilitate default Bayesian analysis for analyzing spatial data or computer experiments (Berger et al., 2001; Ma, 2020).

The CH class not only plays an important role in spatial statistics, but also is of particular interest in UQ. In the UQ community, a covariance function that is of a product form (e.g., Sacks et al., 1989; Santner et al., 2018) has been widely used to model dependence structures for computer model output to allow for different physical interpretations in each input dimension. The product form of this CH covariance can not only control the smoothness of the process realizations in each direction but also allow polynomially decaying dependence in each direction. The simulation example in Section S.6.2 of the Supplementary Material shows significant improvement of the CH class over the Matérn class and the GC class. Predicting real-world processes often relies on computer models whose output can have different smoothness properties and can be insensitive to certain inputs. This CH class can not only allow flexible control over the smoothness of the physical process of interest, but also allow near constant behavior along these inert inputs. Most often, predicting the real-world process involves extrapolation away from the original input space. The CH covariance should be useful in dealing with such challenging applications.

Supplementary Materials

The Supplementary Material contains seven parts: (1) illustration of timing for Bessel function and confluent hypergeometric function, (2) 1-dimensional process realizations for the Matérn class and CH class, (3) ancillary results that are used to prove the main theorems, (4) technical proofs omitted in the main text, (5) simulation results that verify asymptotic normality, (6) additional simulation examples referenced in Section 4, and (7) parameter estimation results and figures referenced in Section 5. Computer code for the real data analysis is also available as a **.zip** archive.

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.

Apanasovich, T. V., Genton, M. G., and Sun, Y. (2012). A valid Matérn class of cross-covariance functions for multivariate random fields with any number of components. *Journal of the American Statistical Association*, 107(497):180–193.

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2014). *Hierarchical Modeling and Analysis for Spatial Data, Second Edition*. CRC Press, Boca Raton, FL.

Barndorff-Nielsen, O. E. (1977). Exponentially decreasing distributions for the logarithm of particle size. *Royal Society of London Proceedings Series A*, 353:401–419.

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

Beran, J. (1992). Statistical methods for data with long-range dependence. *Statistical Science*, 7(4):404–416.

Berger, J. O., De Oliveira, V., and Sanso, B. (2001). Objective Bayesian analysis of spatially correlated data. *Journal of the American Statistical Association*, 96(456):1361–1374.

Berger, J. O. and Smith, L. A. (2019). On the statistical formalism of uncertainty quantification. *Annual Review of Statistics and Its Application*, 6(1):433–460.

Bevilacqua, M. and Faouzi, T. (2019). Estimation and prediction of Gaussian processes using generalized Cauchy covariance model under fixed domain asymptotics. *Electronic Journal of Statistics*, 13(2):3025–3048.

Bingham, N. H., Goldie, C. M., and Teugels, J. L. (1989). *Regular Variation*, volume 27 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge.

Cressie, N. (1990). The origins of kriging. *Mathematical Geology*, 22(3):239–252.

Cressie, N. (1993). *Statistics for Spatial Data*. John Wiley & Sons, New York, revised edition.

Cressie, N. (2017). Mission CO_2ntrol: A statistical scientist's role in remote sensing of atmospheric carbon dioxide. *Journal of the American Statistical Association*, 113(521):152–168.

Dudley, R. M. (2002). *Real Analysis and Probability*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2 edition.

Galassi, M., Davies, J., Theiler, J., Gough, B., Jungman, G., Alken, P., Booth, M., Rossi, F., and Ulerich, R. (2002). *GNU scientific library*. Network Theory Limited.

Gay, R. and Heyde, C. C. (1990). On a class of random field models which allows long range dependence. *Biometrika*, 77(2):401–403.

Gneiting, T. (2000). Power-law correlations, related models for long-range dependence and their simulation. *Journal of Applied Probability*, 37(4):1104–1109.

Gneiting, T. (2002). Compactly supported correlation functions. *Journal of Multivariate Analysis*, 83(2):493–508.

Gneiting, T. (2013). Strictly and non-strictly positive definite functions on spheres. *Bernoulli*, 19(4):1327–1349.

Gu, M., Wang, X., and Berger, J. O. (2018). Robust Gaussian stochastic process emulation. *The Annals of Statistics*, 46(6A):3038–3066.

Guttorp, P. and Gneiting, T. (2006). Studies in the history of probability and statistics XLIX on the Matérn correlation family. *Biometrika*, 93(4):989–995.

Harville, D. A. (1974). Bayesian inference for variance components using only error contrasts. *Biometrika*, 61(2):383–385.

Haslett, J. and Raftery, A. E. (1989). Space-time modelling with long-memory dependence: Assessing Ireland's wind power resource. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 38(1):1–50.

Journel, A. G. and Huijbregts, C. J. (1978). *Mining Geostatistics*. Academic Press, Cambridge, MA.

Lindgren, F., Rue, H., and Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498.

Ma, P. (2020). Objective Bayesian analysis of a cokriging model for hierarchical multifidelity codes. *SIAM/ASA Journal on Uncertainty Quantification*, 8(4):1358–1382.

Ma, P. and Kang, E. L. (2019). Spatio-temporal data fusion for massive sea surface temperature data from MODIS and AMSR-E instruments. *Environmetrics*, 31(2):73.

Ma, P. and Kang, E. L. (2020). A fused Gaussian process model for very large spatial data. *Journal of Computational and Graphical Statistics*, 29(3):479–489.

Ma, P., Konomi, B. A., and Kang, E. L. (2019). An additive approximate Gaussian process model for large spatio-temporal data. *Environmetrics*, 30(8):1838.

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

Matheron, G. (1963). Principles of geostatistics. *Economic Geology*, 58(8):1246–1266.

Porcu, E. and Stein, M. L. (2012). On some local, global and regularity behaviour of some classes of covariance functions. In Porcu, E., Montero, J.-M., and Schlather, M., editors, *Advances and Challenges in Space-time Modelling of Natural Events*, pages 221–238, Berlin, Heidelberg. Springer Berlin Heidelberg.

Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989). Design and analysis of computer experiments. *Statistical Science*, 4(4):409–423.

Santner, T. J., Williams, B. J., and Notz, W. I. (2018). *The design and analysis of computer experiments; 2nd ed.* Springer series in statistics. Springer, New York.

Stein, M. L. (1987). Large sample properties of simulations using Latin hypercube sampling. *Technometrics*, 29(2):143–151.

Stein, M. L. (1988). Asymptotically efficient prediction of a random field with a misspecified covariance function. *The Annals of Statistics*, 16(1):55–63.

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

Stein, M. L. (2005). Nonstationary spatial covariance functions. Unpublished Report. URL: `https://cfpub.epa.gov/ncer_abstracts/index.cfm/fuseaction/display.files/fileID/14471`.

Williams, C. K. and Rasmussen, C. E. (2006). *Gaussian processes for machine learning*. MIT press Cambridge, MA.

Wunch, D., Toon, G. C., Blavier, J.-F. L., Washenfelder, R. A., Notholt, J., Connor, B. J., Griffith, D. W. T., Sherlock, V., and Wennberg, P. O. (2011). The total carbon column observing network. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 369(1943):2087–2112.

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

Zhang, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association*, 99(465):250–261.