The Annals of Statistics
2019, Vol. 47, No. 2, 828–856
https://doi.org/10.1214/17-AOS1652
© Institute of Mathematical Statistics, 2019

ESTIMATION AND PREDICTION USING GENERALIZED WENDLAND COVARIANCE FUNCTIONS UNDER FIXED DOMAIN ASYMPTOTICS

BY MORENO BEVILACQUA[*,1,4], TARIK FAOUZI[†], REINHARD FURRER[‡,2] AND EMILIO PORCU[§,¶,3,4]

Universidad de Valparaiso, Universidad del BioBio†, University of Zurich‡, Newcastle University§ and Universidad de Atacama¶*

We study estimation and prediction of Gaussian random fields with covariance models belonging to the generalized Wendland (GW) class, under fixed domain asymptotics. As for the Matérn case, this class allows for a continuous parameterization of smoothness of the underlying Gaussian random field, being additionally compactly supported. The paper is divided into three parts: first, we characterize the equivalence of two Gaussian measures with GW covariance function, and we provide sufficient conditions for the equivalence of two Gaussian measures with Matérn and GW covariance functions. In the second part, we establish strong consistency and asymptotic distribution of the maximum likelihood estimator of the microergodic parameter associated to GW covariance model, under fixed domain asymptotics. The third part elucidates the consequences of our results in terms of (misspecified) best linear unbiased predictor, under fixed domain asymptotics. Our findings are illustrated through a simulation study: the former compares the finite sample behavior of the maximum likelihood estimation of the microergodic parameter with the given asymptotic distribution. The latter compares the finite-sample behavior of the prediction and its associated mean square error when using two equivalent Gaussian measures with Matérn and GW covariance models, using covariance tapering as benchmark.

1. Introduction. Covariance functions cover a central aspect of inference and prediction of Gaussian fields defined over some (compact) set of $\mathbb{R}^d$. For instance, the best linear unbiased prediction at an unobserved site depends on the knowledge of the covariance function. Since a covariance function must be positive definite, practical estimation generally requires the selection of some parametric classes of covariances and the corresponding estimation of these parameters.

Received December 2016; revised August 2017.
[1]Supported by grant FONDECYT 1160280 from the Chilean government.
[2]Supported by URPP-GCB and SNSF-175529.
[3]Supported by grant FONDECYT 1130647 from the Chilean government.
[4]Supported by Iniciativa Cientfica Milenio–Minecon Nucleo Milenio MESCD.
MSC2010 subject classifications. Primary 62M30; secondary 62F12, 60G25.
Key words and phrases. Compactly supported covariance, spectral density, large dataset, microergodic parameter.

The maximum likelihood (ML) estimation method is generally considered best for estimating the parameters of covariance models. The study of asymptotic properties of ML estimators is complicated by the fact that more than one asymptotic framework can be considered when observing a single realization from a Gaussian field. In particular, under fixed domain asymptotics, one supposes that the sampling domain is bounded and that the sampling set becomes increasingly dense. Under increasing domain asymptotics, the sampling domain increases with the number of observed data, and the distance between any two sampling locations is bounded away from zero.

The asymptotic behavior of ML estimators of the covariance parameters can be quite different under these two frameworks [Zhang and Zimmerman (2005)]. Under increasing domain asymptotic framework and some mild regularity conditions, Mardia and Marshall (1984) give a general result. Specifically, they show that ML estimators are consistent and asymptotically Gaussian, with asymptotic covariance matrix equal to the inverse of the Fisher information matrix.

Equivalence of Gaussian measures [Ibragimov and Rozanov (1978), Skorokhod and Yadrenko (1973)] represents an essential tool to establish the asymptotic properties of both prediction and estimation of Gaussian fields under fixed domain asymptotics. In his *tour de force*, Stein (1988, 1990, 1993, 1999a, 2004) provides conditions under which predictions under a misspecified covariance function are asymptotically efficient, and mean square errors converge almost surely to their targets. Since Gaussian measures depend exclusively on their mean and covariance functions, practical evaluation of Stein's conditions translates into the fact that the true and the misspecified covariances must be compatible, that is, the induced Gaussian measures are equivalent.

Under fixed domain asymptotics, no general results are available for the asymptotic properties of ML estimators. Yet, some results have been obtained when assuming that the covariance belongs to the parametric family of Matérn covariance functions [Matérn (1960)] that has been very popular in spatial statistics for its flexibility with respect to continuous parameterization of smoothness, in the mean square sense, of the underlying Gaussian field. For a Gaussian field defined over a bounded and infinite set of $\mathbb{R}^d$, Zhang (2004) shows that when the smoothness parameter is known and fixed, not all parameters can be estimated consistently when $d = 1, 2, 3$. Instead, the ratio of variance and scale (to the power of the smoothing parameter), sometimes called microergodic parameter [Stein (1999b), Zhang and Zimmerman (2005)], is consistently estimable. In contrast for $d \geq 5$, Anderes (2010) proved the orthogonality of two Gaussian measures with different Matérn covariance functions, and hence, in this case, all the parameters can be consistently estimated under fixed-domain asymptotics. The case $d = 4$ is still open.

Asymptotic results for the ML estimator of the microergodic parameter of the Matérn model can be found in Zhang (2004), Du, Zhang and Mandrekar (2009), Wang and Loh (2011) and Kaufman and Shaby (2013). In particular, Kaufman

and Shaby (2013) give strong consistency and asymptotic distribution of the microergodic parameter when estimating jointly the scale and variance parameters, generalizing previous results in Zhang (2004) and Wang and Loh (2011) where the scale parameter is assumed known and fixed. Kaufman and Shaby (2013) show by means of simulation study that asymptotic approximation using a fixed scale parameter can be problematic when applied to finite samples, even for large sample sizes. In contrast, they show that performance is improved and asymptotic approximations are applicable for smaller sample sizes when the parameters are jointly estimated.

Under the Matérn family, similar results have been obtained for the covariance tapering (CT) method of estimation, as originally proposed in Kaufman, Schervish and Nychka (2008) and consisting of setting to zero the dependence after a given distance. This is in turn achieved by multiplying the Matérn covariance with a taper function, that is, a correlation function being additionally compactly supported over a ball with given radius. Thus, the resulting covariance tapered matrix is sparse, with the level of sparseness depending on the radius of compact support. Sparse matrix algorithms can then be used to evaluate efficiently an approximate likelihood where the original covariance matrix is replaced by the tapered matrix. The results proposed in Kaufman, Schervish and Nychka (2008) have then inspired the works in Du, Zhang and Mandrekar (2009), Wang and Loh (2011) and Kaufman and Shaby (2013), where asymptotic properties of the CT estimator of the Matérn microergodic parameter are given.

Using the Matérn family, Furrer, Genton and Nychka (2006) study CT when applied to the best unbiased linear predictor and show that under fixed domain asymptotics and some specific conditions of the taper function, asymptotically efficient prediction and asymptotically correct estimation of mean square error can be achieved using a tapered Matérn covariance function. Extensions have been discussed by, for example, Stein (2013) and Hirano and Yajima (2013). The basic message of CT method is that large data sets can be handled both for estimation and prediction exploiting sparse matrix algorithms when using the Matérn model.

Inspired by this idea, we focus on a covariance model that offers the strength of the Matérn family and allows the use of sparse matrices. Specifically, we study estimation and prediction of Gaussian fields under fixed domain asymptotics, using the generalized Wendland (GW) class of covariance functions [Gneiting (2002a), Zastavnyi (2006)], the members of which are compactly supported over balls of $\mathbb{R}^d$ with arbitrary radii, and additionally allows for a continuous parameterization of differentiability at the origin, in a similar way to the Matérn family.

In particular, we provide the following results. First, we characterize the equivalence of two Gaussian measures with covariance functions belonging to the GW class and sharing the same smoothness parameter. A consequence of this result is that, as in the Matérn case [Zhang (2004)], when the smoothness parameter is known and fixed, not all parameters can be estimated consistently under fixed domain asymptotics. Then we give sufficient conditions for the equivalence of two

Gaussian measures where the state of truth is represented by a member of the Matérn family and the other measure has a GW covariance model and vice versa.

We assess the asymptotic properties of the ML estimator of the microergodic parameter associated to the GW class. Specifically, for a fixed smoothness parameter, we establish strong consistency and asymptotic distribution of the microergodic parameter estimate assuming the compact support parameter fixed and known. Then we generalize these results when jointly estimating with ML the variance and the compact support parameter.

Finally, using results in Stein (1988, 1993), we study the implications of our results on prediction under fixed domain asymptotics. One remarkable implication is that when the true covariance belongs to the Matérn family, asymptotic efficiency prediction and asymptotically correct estimation of mean square error can be achieved using a compatible GW covariance model.

The remainder of the paper is organized as follows. In Section 2, we review some results of Matérn and GW covariance models. In Section 3, we first characterize the equivalence of Gaussian measure under the GW covariance model. Then we find a sufficient condition for the equivalence of two Gaussian measures with Matérn and GW covariance models. In Section 4, we establish strong consistency and asymptotic distribution of the ML estimator of the microergodic parameter of the GW models, under fixed domain asymptotics. Section 5 discusses the consequences of the results in Section 3 on prediction under fixed domain asymptotics. Section 6 provides two simulation studies: The first shows how well the given asymptotic distribution of the microergodic parameter applies to finite sample cases when estimating with ML a GW covariance model under fixed domain asymptotics. The second compares the finite-sample behavior of the prediction when using two compatible Matérn and GW models, using CT as a benchmark. The final section provides discussion on the consequence of our results and identifies problems for future research.

2. Matérn and generalized Wendland covariance models. We denote $\{Z(\boldsymbol{s}), \boldsymbol{s} \in D\}$ a zero mean Gaussian field on a bounded set D of $\mathbb{R}^d$, with stationary covariance function $C : \mathbb{R}^d \to \mathbb{R}$. We consider the class Φ_d of continuous mappings $\phi : [0, \infty) \to \mathbb{R}$ with $\phi(0) > 0$, such that

$$\operatorname{cov}\big(Z(\boldsymbol{s}), Z(\boldsymbol{s}')\big) = C(\boldsymbol{s}' - \boldsymbol{s}) = \phi(\|\boldsymbol{s}' - \boldsymbol{s}\|),$$

with $\boldsymbol{s}, \boldsymbol{s}' \in D$, and $\|\cdot\|$ denoting the Euclidean norm. Gaussian fields with such covariance functions are called weakly stationary and isotropic.

Schoenberg (1938) characterized the class Φ_d as scale mixtures of the characteristic functions of random vectors uniformly distributed on the spherical shell of $\mathbb{R}^d$, with any positive measure, F:

$$\phi(r) = \int_0^\infty \Omega_d(r\xi) F(\mathrm{d}\xi), \qquad r \geq 0,$$

with $\Omega_d(r) = r^{1-d/2} J_{d/2-1}(r)$ and J_ν a Bessel function of order ν. The class Φ_d is nested, with the inclusion relation $\Phi_1 \supset \Phi_2 \supset \cdots \supset \Phi_\infty$ being strict, and where $\Phi_\infty := \bigcap_{d\geq 1} \Phi_d$ is the class of continuous mappings ϕ, the radial version of which is positive definite on any d-dimensional Euclidean space.

The Matérn function, defined as

$$\mathcal{M}_{\nu,\alpha,\sigma^2}(r) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{r}{\alpha}\right)^\nu \mathcal{K}_\nu\left(\frac{r}{\alpha}\right), \qquad r \geq 0,$$

is a member of the class Φ_∞ for any positive values of α and ν. Here, $\mathcal{K}_\nu$ is a modified Bessel function of the second kind of order ν, σ^2 is the variance and α a positive scaling parameter. The parameter ν characterizes the differentiability at the origin and, as a consequence, the differentiability of the associated sample paths. In particular for a positive integer k, the sample paths are k times differentiable, in any direction, if and only if $\nu > k$.

When $\nu = 1/2 + m$ and m is a nonnegative integer, the Matérn function simplifies to the product of a negative exponential with a polynomial of degree m, and for ν tending to infinity, a rescaled version of the Matérn converges to a squared exponential model being infinitely differentiable at the origin. Thus, the Matérn function allows for a continuous parameterization of its associated Gaussian field in terms of smoothness.

We also define Φ_d^b as the class that consists of members of Φ_d being additionally compactly supported on a given interval, $[0, b]$, $b > 0$. Clearly, their radial versions are compactly supported over balls of $\mathbb{R}^d$ with radius b.

We now define GW correlation functions $\varphi_{\mu,\kappa}$ as introduced by Gneiting (2002b), Zastavnyi (2006) and Chernih and Hubbert (2014). For $\kappa > 0$, we define

$$\varphi_{\mu,\kappa}(r) := \begin{cases} \dfrac{1}{B(2\kappa, \mu+1)} \displaystyle\int_r^1 u\big(u^2 - r^2\big)^{\kappa-1}(1-u)^\mu \,\mathrm{d}u, & 0 \leq r < 1, \\ 0, & r \geq 1, \end{cases} \tag{1}$$

with B denoting the beta function. Arguments in Gneiting (2002b) and Zastavnyi (2006) show that, for a given $\kappa > 0$, $\varphi_{\mu,\kappa} \in \Phi_d^1$ if and only if

$$\mu \geq \lambda(d, \kappa) := (d+1)/2 + \kappa. \tag{2}$$

Throughout, we use λ instead of $\lambda(d, \kappa)$ whenever no confusion arises. Integration by parts shows that the first part of (1) can also be written as

$$\frac{1}{B(1+2\kappa, \mu)} \int_r^1 \big(u^2 - r^2\big)^\kappa (1-u)^{\mu-1} \,\mathrm{d}u, \qquad 0 \leq r < 1.$$

Note that $\varphi_{\mu,0}$ is not defined because κ must be strictly positive. In this special case, we consider the Askey function [Askey (1973)]

$$\mathcal{A}_\mu(r) = (1-r)_+^\mu = \begin{cases} (1-r)^\mu, & 0 \leq r < 1, \\ 0, & r \geq 1, \end{cases}$$

where $(\cdot)_+$ denotes the positive part. Arguments in Golubov (1981) show that $\mathcal{A}_\mu \in \Phi_d^1$ if and only if $\mu \geq (d+1)/2$ and we define $\varphi_{\mu,0} := \mathcal{A}_\mu$.

Finally, we define the GW covariance function, with compact support $\beta > 0$, variance σ^2 and smoothness parameter $\kappa > 0$ as

$$\varphi_{\mu,\kappa,\beta,\sigma^2}(r) := \sigma^2 \varphi_{\mu,\kappa}(r/\beta), \qquad r \geq 0, \tag{3}$$

and $\varphi_{\mu,\kappa,\beta,\sigma^2} \in \Phi_d^\beta$ if and only if $\mu \geq \lambda$. Accordingly, for $\kappa = 0$, we define

$$\varphi_{\mu,0,\beta,\sigma^2}(r) := \sigma^2 \varphi_{\mu,0}(r/\beta), \qquad r \geq 0. \tag{4}$$

When computing (3), numerical integration is obviously feasible, but could be cumbersome to (spatial) statisticians used to handle closed-form parametric covariance model. Nevertheless, closed-form solution of the integral in equation (1) can be obtained when $\kappa = k$, a positive integer. In this case, $\varphi_{\mu,k,1,1}(r) = \mathcal{A}_{\mu+k}(r) P_k(r)$, with P_k a polynomial of order k. These functions, termed (original) Wendland functions, were originally proposed by Wendland (1995). Other closed-form solutions of integral (1) can be obtained when $\kappa = k + 0.5$, using some results in Schaback (2011). Such solutions are called *missing Wendland* functions.

Recently, Porcu, Zastavnyi and Bevilacqua (2017) have shown that the GW class includes almost all classes of covariance functions with compact supports known to the geostatistical and numerical analysis communities. Not only original and Wendland functions, but also Wu's functions [Wu (1995)], which in turn include the spherical model [Wackernagel (2003)], as well as the Trigub splines [Zastavnyi (2006)]. Finally, Chernih, Sloan and Womersley (2014) show that, for κ tending to infinity, a rescaled version of the GW model converges to a squared exponential covariance model.

As noted by Gneiting (2002a), GW and Matérn functions exhibit the same behavior at the origin, with the smoothness parameters of the two covariance models related by the equation $\nu = \kappa + 1/2$.

Fourier transforms of radial versions of members of Φ_d, for a given d, have a simple expression, as reported in Yaglom (1987) or Stein (1999b). For a member ϕ of the class Φ_d, we define its isotropic spectral density as

$$\widehat{\phi}(z) = \frac{z^{1-d/2}}{(2\pi)^d} \int_0^\infty u^{d/2} J_{d/2-1}(uz)\phi(u)\,\mathrm{d}u, \qquad z \geq 0, \tag{5}$$

and throughout the paper, we use the notation: $\widehat{\mathcal{M}}_{\nu,\alpha,\sigma^2}$, and $\widehat{\varphi}_{\mu,\kappa,\beta,\sigma^2}$ for the radial parts of Fourier transforms of $\mathcal{M}_{\nu,\alpha,\sigma^2}$ and $\varphi_{\mu,\kappa,\beta,\sigma^2}$, respectively.

A well-known result about the spectral density of the Matérn model is the following:

$$\widehat{\mathcal{M}}_{\nu,\alpha,\sigma^2}(z) = \frac{\Gamma(\nu + d/2)}{\pi^{d/2}\Gamma(\nu)} \frac{\sigma^2 \alpha^d}{(1+\alpha^2 z^2)^{\nu+d/2}}, \qquad z \geq 0. \tag{6}$$

For two given nonnegative functions $g_1(x)$ and $g_2(x)$, with $g_1(x) \asymp g_2(x)$ we mean that there exist two constants c and C such that $0 < c < C < \infty$ and $cg_2(x) \le g_1(x) \le Cg_2(x)$ for each x. The next result follows from Zastavnyi (2006), Chernih and Hubbert (2014) and from standard properties of Fourier transforms. Their proofs are thus omitted. Let us first define the function ${}_1F_2$ as

$$ {}_1F_2(a; b, c; z) = \sum_{k=0}^{\infty} \frac{(a)_k z^k}{(b)_k (c)_k k!}, \qquad z \in \mathbb{R}, $$

which is a special case of the generalized hypergeometric functions ${}_qF_p$ [Abramowitz and Stegun (1965)], with $(q)_k = \Gamma(q+k)/\Gamma(q)$ for $k \in \mathbb{N} \cup \{0\}$, being the Pochhammer symbol.

THEOREM 1. *Let $\varphi_{\mu,\kappa,\beta,\sigma^2}$ be the function defined at equation (3) and let λ as defined through equation (2). Then, for $\kappa, \sigma^2, \beta > 0$ and $\mu \ge \lambda$:*

1. $\widehat{\varphi}_{\mu,\kappa,\beta,\sigma^2}(z) = \sigma^2 L^{\varsigma} \beta^d {}_1F_2(\lambda; \lambda + \frac{\mu}{2}, \lambda + \frac{\mu}{2} + \frac{1}{2}; -\frac{(z\beta)^2}{4})$, $z > 0$;
2. $\widehat{\varphi}_{\mu,\kappa,\beta,\sigma^2}(z) = \sigma^2 L^{\varsigma} \beta^d [c_3^{\varsigma}(z\beta)^{-2\lambda}\{1 + \mathcal{O}(z^{-2})\} + c_4^{\varsigma}(z\beta)^{-(\mu+\lambda)}\{\cos(z\beta - c_5^{\varsigma}) + \mathcal{O}(z^{-1})\}]$, *for* $z \to \infty$;
3. $\widehat{\varphi}_{\mu,\kappa,\beta,\sigma^2}(z) \asymp z^{-2\lambda}$, *for* $z \to \infty$,

where $c_3^{\varsigma} = \frac{\Gamma(\mu+2\lambda)}{\Gamma(\mu)}$, $c_4^{\varsigma} = \frac{\Gamma(\mu+2\lambda)}{\Gamma(\lambda)2^{\lambda-1}}$, $c_5^{\varsigma} = \frac{\pi}{2}(\mu+\lambda)$, $L^{\varsigma} = \frac{K^{\varsigma}\Gamma(\kappa)}{2^{1-\kappa}B(2\kappa,\mu+1)}$ *and*

$$ K^{\varsigma} = \frac{2^{-\kappa-d+1}\pi^{-\frac{d}{2}}\Gamma(\mu+1)\Gamma(2\kappa+d)}{\Gamma(\kappa+\frac{d}{2})\Gamma(\mu+2\lambda)}, $$

where $\varsigma := (\mu, \kappa, d)'$.

Point 1 has been shown by Zastavnyi (2006). Points 2 and 3 can be found in Chernih and Hubbert (2014). Note that the case $\kappa = 0$ is not included in Theorem 1. We consider it in the following result, whose proof follows the lines of Zastavnyi (2006) and Chernih and Hubbert (2014) for the case $\kappa > 0$.

THEOREM 2. *Let $\varphi_{\mu,0,\beta,\sigma^2}$ as being defined at equation (4). Then, for $\sigma^2, \beta > 0$, $\mu \ge (d+1)/2$:*

1. $\widehat{\varphi}_{\mu,0,\beta,\sigma^2}(z) = \sigma^2 K^{\varsigma} \beta^d {}_1F_2(\frac{d+1}{2}; \frac{d+1}{2} + \frac{\mu}{2}, \frac{d+1}{2} + \frac{\mu}{2} + \frac{1}{2}; -\frac{(z\beta)^2}{4})$, $z > 0$;
2. $\widehat{\varphi}_{\mu,0,\beta,\sigma^2}(z) = \sigma^2 K^{\varsigma} \beta^d [c_3^{\varsigma}(z\beta)^{-(d+1)}\{1 + \mathcal{O}(z^{-2})\} + c_4^{\varsigma}(z\beta)^{-(\mu+(d+1)/2)} \times \{\cos(z\beta - c_5^{\varsigma}) + \mathcal{O}(z^{-1})\}]$, *for* $z \to \infty$;
3. $\widehat{\varphi}_{\mu,0,\beta,\sigma^2}(z) \asymp z^{-(d+1)}$, *for* $z \to \infty$,

with c_3^{ς}, c_4^{ς}, c_5^{ς} *and* K^{ς} *defined as in Theorem 1 but with* $\varsigma := (\mu, 0, d)'$.

TABLE 1

GW correlation $\varphi_{\mu,\kappa,1,1}(r)$ and Matérn correlation $\mathcal{M}_{\nu,1,1}(r)$ with increasing smoothness parameters κ and ν. SP(k) means that the sample paths of the associated Gaussian field are k times differentiable

κ	$\varphi_{\mu,\kappa,1,1}(r)$	ν	$\mathcal{M}_{\nu,1,1}(r)$	$SP(k)$
0	$(1-r)_+^{\mu}$	0.5	e^{-r}	0
1	$(1-r)_+^{\mu+1}(1+r(\mu+1))$	1.5	$e^{-r}(1+r)$	1
2	$(1-r)_+^{\mu+2}(1+r(\mu+2)+r^2(\mu^2+4\mu+3)\frac{1}{3})$	2.5	$e^{-r}(1+r+\frac{r^2}{3})$	2
3	$(1-r)_+^{\mu+3}(1+r(\mu+3)+r^2(2\mu^2+12\mu+15)\frac{1}{5}$ $+r^3(\mu^3+9\mu^2+23\mu+15)\frac{1}{15})$	3.5	$e^{-r}(1+\frac{r}{2}+r^2\frac{6}{15}+\frac{r^3}{15})$	3

The spectral density and its decay for $z \to \infty$ in Theorems 1 and 2 are useful when studying some geometrical properties of a Gaussian field or its associated sample paths [Adler (1981)]. For instance, using Theorem 1 Point 1 or Theorem 2 Point 1, it is easy to prove that for a positive integer k, the sample paths of a Gaussian field with GW function are k times differentiable, in any direction, if and only if $\kappa > k - 1/2$. Table 1 compares the GW $\varphi_{\mu,\kappa,1,1}(r)$ for $\kappa = 0, 1, 2, 3$ with $\mathcal{M}_{\nu,1,1}(r)$ for $\nu = 0.5, 1.5, 2.5, 3.5$ with the associated degree of sample paths differentiability.

3. Equivalence of Gaussian measures with GW models. Equivalence and orthogonality of probability measures are useful tools when assessing the asymptotic properties of both prediction and estimation for Gaussian fields. Denote with P_i, $i = 0, 1$, two probability measures defined on the same measurable space $\{\Omega, \mathcal{F}\}$. P_0 and P_1 are called equivalent (denoted $P_0 \equiv P_1$) if $P_1(A) = 1$ for any $A \in \mathcal{F}$ implies $P_0(A) = 1$ and vice versa. On the other hand, P_0 and P_1 are orthogonal (denoted $P_0 \perp P_1$) if there exists an event A such that $P_1(A) = 1$ but $P_0(A) = 0$. For a stochastic process $\{Z(\boldsymbol{s}), \boldsymbol{s} \in \mathbb{R}^d\}$, to define previous concepts, we restrict the event A to the σ-algebra generated by $\{Z(\boldsymbol{s}), \boldsymbol{s} \in D\}$, where $D \subset \mathbb{R}^d$. We emphasize this restriction by saying that the two measures are equivalent on the paths of $\{Z(\boldsymbol{s}), \boldsymbol{s} \in D\}$.

Gaussian measures are completely characterized by their mean and covariance function. We write $P(\rho)$ for a Gaussian measure with zero mean and covariance function ρ. It is well known that two Gaussian measures are either equivalent or orthogonal on the paths of $\{Z(\boldsymbol{s}), \boldsymbol{s} \in D\}$ [Ibragimov and Rozanov (1978)].

Let $P(\rho_i)$, $i = 0, 1$ be two zero mean Gaussian measures with isotropic covariance function ρ_i and associated spectral density $\widehat{\rho}_i$, $i = 0, 1$, as defined through (5). Using results in Skorokhod and Yadrenko (1973) and Ibragimov and Rozanov (1978), Stein (2004) has shown that, if for some $a > 0$, $\widehat{\rho}_0(z)z^a$ is bounded away

from 0 and ∞ as $z \to \infty$, and for some finite and positive c,

$$\int_c^\infty z^{d-1} \left\{ \frac{\widehat{\rho}_1(z) - \widehat{\rho}_0(z)}{\widehat{\rho}_0(z)} \right\}^2 \mathrm{d}z < \infty, \tag{7}$$

then for any bounded subset $D \subset \mathbb{R}^d$, $P(\rho_0) \equiv P(\rho_1)$ on the paths of $Z(\boldsymbol{s})$, $\boldsymbol{s} \in D$.

For the reminder of the paper, we denote with $P(\mathcal{M}_{\nu,\alpha,\sigma^2})$ a zero mean Gaussian measure induced by a Matérn covariance function with associated spectral density $\widehat{\mathcal{M}}_{\nu,\alpha,\sigma^2}$, and with $P(\varphi_{\mu,\kappa,\beta,\sigma^2})$ a zero mean Gaussian measure induced by a GW covariance function with associated spectral density $\widehat{\varphi}_{\mu,\kappa,\beta,\sigma^2}$.

Using (7) and (6), Zhang (2004) established the following characterization concerning the equivalent conditions of two Gaussian measures with Matérn covariance models.

THEOREM 3 [Zhang (2004)]. *For a given $\nu > 0$, let $P(\mathcal{M}_{\nu,\alpha_i,\sigma_i^2})$, $i = 0, 1$, be two zero mean Gaussian measures. For any bounded infinite set $D \subset \mathbb{R}^d$, $d = 1, 2, 3$, $P(\mathcal{M}_{\nu,\alpha_0,\sigma_0^2}) \equiv P(\mathcal{M}_{\nu,\alpha_1,\sigma_1^2})$ on the paths of $Z(\boldsymbol{s})$, $\boldsymbol{s} \in D$, if and only if*

$$\sigma_0^2/\alpha_0^{2\nu} = \sigma_1^2/\alpha_1^{2\nu}. \tag{8}$$

The first relevant result of this paper concerns the characterization of the equivalence of two zero mean Gaussian measures under GW functions. The crux of the proof is the arguments in equation (7), coupled with the asymptotic expansion of the spectral density as in Theorems 1 and 2.

THEOREM 4. *For a given $\kappa \geq 0$, let $P(\varphi_{\mu,\kappa,\beta_i,\sigma_i^2})$, $i = 0, 1$, be two zero mean Gaussian measures and let $\mu > \lambda + d/2$, with λ as defined through equation (2). For any bounded infinite set $D \subset \mathbb{R}^d$, $d = 1, 2, 3$, $P(\varphi_{\mu,\kappa,\beta_0,\sigma_0^2}) \equiv P(\varphi_{\mu,\kappa,\beta_1,\sigma_1^2})$ on the paths of $Z(\boldsymbol{s})$, $\boldsymbol{s} \in D$ if and only if*

$$\sigma_0^2/\beta_0^{2\kappa+1} = \sigma_1^2/\beta_1^{2\kappa+1}. \tag{9}$$

PROOF. We first consider the case $\kappa > 0$. Let us start with the sufficient part of the assertion. From Theorem 1 (Point 3), there exist two positive constants c_i and C_i such that

$$c_i \leq z^{2\lambda} \widehat{\varphi}_{\mu,\kappa,\beta_i,\sigma_i^2}(z) \leq C_i, \qquad i = 0, 1.$$

In order to prove the sufficient part, we need to find conditions such that, for some positive and finite c,

$$\int_c^\infty z^{d-1} \left(\frac{\widehat{\varphi}_{\mu,\kappa,\beta_1,\sigma_1^2}(z) - \widehat{\varphi}_{\mu,\kappa,\beta_0,\sigma_0^2}(z)}{\widehat{\varphi}_{\mu,\kappa,\beta_0,\sigma_0^2}(z)} \right)^2 \mathrm{d}z < \infty. \tag{10}$$

We proceed by direct construction and, using Theorem 1 (Points 1 and 2), we find that, as $z \to \infty$,

$$\begin{aligned}
&\left|\frac{\widehat{\varphi}_{\mu,\kappa,\beta_1,\sigma_1^2}(z) - \widehat{\varphi}_{\mu,\kappa,\beta_0,\sigma_0^2}(z)}{\widehat{\varphi}_{\mu,\kappa,\beta_0,\sigma_0^2}(z)}\right| \\
&\quad \le L^{\varsigma} c_0^{-1} z^{2\lambda} \big|\sigma_1^2 \beta_1^d \big[c_3^{\varsigma}(\beta_1 z)^{-2\lambda}\{1 + \mathcal{O}(z^{-2})\} \\
&\qquad + c_4^{\varsigma}(z\beta_1)^{-(\mu+\lambda)}\{\cos(\beta_1 z - c_5^{\varsigma}) + \mathcal{O}(z^{-1})\}\big] \\
&\qquad - \sigma_0^2 \beta_0^d \big[c_3^{\varsigma}(\beta_0 z)^{-2\lambda}\{1 + \mathcal{O}(z^{-2})\} \\
&\qquad + c_4^{\varsigma}(z\beta_0)^{-(\mu+\lambda)}\{\cos(\beta_0 z - c_5^{\varsigma}) + \mathcal{O}(z^{-1})\}\big]\big| \\
&\quad \le L^{\varsigma} c_0^{-1} \big|c_3^{\varsigma}\big[\sigma_1^2 \beta_1^{-(1+2\kappa)} - \sigma_0^2 \beta_0^{-(1+2\kappa)}\big] + \mathcal{O}(z^{-2}) \\
&\qquad + c_4^{\varsigma} z^{\lambda-\mu}\big[\sigma_1^2 \beta_1^{\tilde{\lambda}} \cos(\beta_1 z - c_5^{\varsigma}) - \sigma_0^2 \beta_0^{\tilde{\lambda}} \cos(\beta_0 z - c_5^{\varsigma})\big] \\
&\qquad + c_4^{\varsigma} z^{\lambda-\mu} \mathcal{O}(z^{-1})\{\sigma_1^2 \beta_1^{\tilde{\lambda}} - \sigma_0^2 \beta_0^{\tilde{\lambda}}\}\big|,
\end{aligned}$$

where $\tilde{\lambda} = d - (\mu + \lambda)$. Let us now write

$$\begin{aligned}
A(z) &= c_3^{\varsigma}\big[\sigma_1^2 \beta_1^{-(1+2\kappa)} - \sigma_0^2 \beta_0^{-(1+2\kappa)}\big] + \mathcal{O}(z^{-2}), \\
B(z) &= c_4^{\varsigma} z^{\lambda-\mu}\big[\sigma_1^2 \beta_1^{\tilde{\lambda}} \cos(\beta_1 z - c_5^{\varsigma}) - \sigma_0^2 \beta_0^{\tilde{\lambda}} \cos(\beta_0 z - c_5^{\varsigma})\big], \quad \text{and} \\
D(z) &= c_4^{\varsigma} z^{\lambda-\mu} \mathcal{O}(z^{-1})\{\sigma_1^2 \beta_1^{\tilde{\lambda}} - \sigma_0^2 \beta_0^{\tilde{\lambda}}\}.
\end{aligned}$$

Then a sufficient condition for (10) is the following:

$$(L^{\varsigma}/c_0)^2 \int_c^{\infty} z^{d-1}\big(A(z) + B(z) + D(z)\big)^2 \,\mathrm{d}z < \infty. \tag{11}$$

Note that $A(z)$ is of order $\mathcal{O}(z^{-2})$ under Condition (9). We claim that (11) is satisfied if $\sigma_1^2 \beta_1^{-(1+2\kappa)} = \sigma_0^2 \beta_0^{-(1+2\kappa)}$ for $\mu > \lambda + d/2$, $d = 1, 2, 3$.

In fact, we have, for $z \to \infty$,

$$|B(z)| \le c_4^{\varsigma} z^{\lambda-\mu}\big[\sigma_1^2 \beta_1^{\tilde{\lambda}} + \sigma_0^2 \beta_0^{\tilde{\lambda}}\big] \le c_6 z^{\lambda-\mu}$$

and

$$\begin{aligned}
|D(z)| &\le c_4^{\varsigma} z^{\lambda-\mu} \mathcal{O}(z^{-1})\{\sigma_1^2 \beta_1^{\tilde{\lambda}} + \sigma_0^2 \beta_0^{\tilde{\lambda}}\} \\
&\le c_7 c_4^{\varsigma} z^{\lambda-\mu-1}\{\sigma_1^2 \beta_1^{\tilde{\lambda}} + \sigma_0^2 \beta_0^{\tilde{\lambda}}\} \\
&\le c_8 z^{\lambda-\mu-1}
\end{aligned}$$

with c_6, c_7 and c_8 being positive and finite constants. Expanding (11), we notice that the dominant terms are A^2 and B^2, independently on the cross products. These

are respectively of the order $\mathcal{O}(z^{-4})$ and $\mathcal{O}(z^{2(\lambda-\mu)})$. This in turn implies that the integral (11) is finite if $\sigma_1^2\beta_1^{-(1+2\kappa)} = \sigma_0^2\beta_0^{-(1+2\kappa)}$, for $\mu > \lambda + d/2$ and $d = 1, 2, 3$ and this implies that (10) is satisfied under the same conditions. The sufficient part of our claim is thus proved.

The necessary part follows the proof of Zhang (2004). For $\mu > \lambda + d/2$ and $d = 1, 2, 3$, we suppose $\sigma_1^2\beta_1^{-(1+2\kappa)} \neq \sigma_0^2\beta_0^{-(1+2\kappa)}$ and let $\sigma_2^2 = \sigma_0^2(\beta_0/\beta_1)^{-(1+2\kappa)}$. Then $\varphi_{\mu,\kappa,\beta_0,\sigma_0^2}$ and $\varphi_{\mu,\kappa,\beta_1,\sigma_2^2}$ define two equivalent Gaussian measures. We need to show that $\varphi_{\mu,\kappa,\beta_1,\sigma_2^2}$ and $\varphi_{\mu,\kappa,\beta_1,\sigma_1^2}$ define two orthogonal Gaussian measures. The rest of the proof follows the same arguments in Zhang (2004).

We omit the proof of the special case $\kappa = 0$, since is similar to the case $\kappa > 0$, but using the arguments in Theorem 2. $\square$

An immediate consequence of Theorem 4 is that for fixed κ and μ, the β and σ^2 parameters cannot be estimated consistently [Zhang (2004)]. Instead, the microergodic parameter $\sigma^2\beta^{-(1+2\kappa)}$ is consistently estimable. In Section 4, we establish the asymptotic properties of ML estimation associated to the microergodic parameter.

The next result depicts an interesting scenario in which a GW and Matérn model are considered and gives sufficient conditions for the compatibility of these two covariance models. We offer a constructive proof, the crux of the argument being again equation (7). We treat the cases $\kappa > 0$ and $\kappa = 0$ separately.

THEOREM 5. *For given $\nu > 1/2$ and $\kappa > 0$, let $P(\mathcal{M}_{\nu,\alpha,\sigma_0^2})$ and $P(\varphi_{\mu,\kappa,\beta,\sigma_1^2})$ be two zero mean Gaussian measures. If $\nu = \kappa + 1/2$, $\mu > \lambda + d/2$, with λ as defined through equation* (2), *and*

$$\sigma_0^2\alpha^{-2\nu} = C_{\nu,\kappa,\mu}\sigma_1^2\beta^{-(1+2\kappa)}, \tag{12}$$

where $C_{\nu,\kappa,\mu} = \frac{\mu 2^{-d}\Gamma(\nu)\Gamma(\kappa)\Gamma(2\kappa+d)}{\Gamma(\nu+d/2)\Gamma(\kappa+d/2)B(2\kappa,\mu+1)}$, then for any bounded infinite set $D \subset \mathbb{R}^d$, $d = 1, 2, 3$, $P(\mathcal{M}_{\nu,\alpha,\sigma_0^2}) \equiv P(\varphi_{\mu,\kappa,\beta,\sigma_1^2})$ on the paths of $Z(\boldsymbol{s})$, $\boldsymbol{s} \in D$.

PROOF. In order to prove Theorem 5, we need to find conditions such that for some positive and finite c,

$$\int_c^\infty z^{d-1}\left(\frac{\widehat{\varphi}_{\nu,\kappa,\beta,\sigma_1^2}(z) - \widehat{\mathcal{M}}_{\nu,\alpha,\sigma_0^2}(z)}{\widehat{\mathcal{M}}_{\nu,\alpha,\sigma_0^2}(z)}\right)^2 \mathrm{d}z < \infty. \tag{13}$$

It is known that $\widehat{\mathcal{M}}_{\nu,\alpha,\sigma_0^2}(z)z^a$ is bounded away from 0 and ∞ as $z \to \infty$ for some $a > 0$ [Zhang (2004)]. Using (6) and Theorem 1 (Points 1 and 2), we have,

as $z \to \infty$,

$$\begin{aligned}
&\left|\frac{\widehat{\varphi}_{\mu,\kappa,\beta,\sigma_1^2}(z) - \widehat{\mathcal{M}}_{\nu,\alpha,\sigma_0^2}(z)}{\widehat{\mathcal{M}}_{\nu,\alpha,\sigma_0^2}(z)}\right| \\
&\quad = \Bigg|\frac{\sigma_1^2\beta^d\Gamma(\nu)L^{\varsigma}}{\Gamma(\nu+d/2)\sigma_0^2\alpha^{-2\nu}\pi^{-\frac{d}{2}}}\big[c_3^{\varsigma}(\beta z)^{-2\lambda}\{1+\mathcal{O}(z^{-2})\} \\
&\qquad + c_4^{\varsigma}(z\beta)^{-(\mu+\lambda))}\{\cos(\beta z - c_5^{\varsigma}) + \mathcal{O}(z^{-1})\}\big](\alpha^{-2}+z^2)^{\nu+\frac{d}{2}} - 1\Bigg| \\
&\quad = \Bigg|\frac{\sigma_1^2\beta^d\Gamma(\nu)L^{\varsigma}}{\Gamma(\nu+d/2)\sigma_0^2\alpha^{-2\nu}\pi^{-\frac{d}{2}}}\big[c_3^{\varsigma}(\beta z)^{-2\lambda}\{1+\mathcal{O}(z^{-2})\} \\
&\qquad + c_4^{\varsigma}(z\beta)^{-(\mu+\lambda)}\{\cos(\beta z - c_5^{\varsigma}) + \mathcal{O}(z^{-1})\}\big]z^{2\nu+d}((\alpha z)^{-2}+1)^{\nu+\frac{d}{2}} - 1\Bigg| \\
&\quad = \big|w_1 z^{-2\lambda}\{1+\mathcal{O}(z^{-2})\}z^{2\nu+d}\big[1+(\nu+d/2)(\alpha z)^{-2}+\mathcal{O}(z^{-2})\big] \\
&\qquad + w_2 z^{-(\mu+\lambda)}z^{2\nu+d}\big[1+(\nu+d/2)(\alpha z)^{-2}+\mathcal{O}(z^{-2})\big] \\
&\qquad \times \{\cos(\beta z - c_5^{\varsigma}) + \mathcal{O}(z^{-1})\} - 1\big|,
\end{aligned}$$

where $w_1 = \frac{L^{\varsigma}\sigma_1^2\beta^{-(1+2\kappa)}\Gamma(\nu)c_3^{\varsigma}}{\Gamma(\nu+d/2)\sigma_0^2\alpha^{-2\nu}\pi^{-d/2}}$, $w_2 = w_1 c_4^{\varsigma}\beta^{\lambda-\mu}/c_3^{\varsigma}$. Since $z^{2\nu+d}[(\nu+d/2)\times(\alpha z)^{-2}+\mathcal{O}(z^{-2})] = \mathcal{O}(z^{2\nu+d-2})$, we have

$$\begin{aligned}
&\int_c^{\infty} z^{d-1}\left|\frac{\widehat{\varphi}_{\mu,\kappa,\beta,\sigma_1^2}(z) - \widehat{\mathcal{M}}_{\nu,\alpha,\sigma_0^2}(z)}{\widehat{\mathcal{M}}_{\nu,\alpha,\sigma_0^2}(z)}\right|^2 \mathrm{d}z \\
&\quad = \int_c^{\infty} z^{d-1}\big|w_1 z^{-2\lambda}\mathcal{O}(z^{2\nu+d-2}) + \{w_1 z^{2\nu-(1+2\kappa)} - 1\} \\
&\qquad + w_1 z^{-2\lambda}\{\mathcal{O}(z^{2\nu+d-2}) + \mathcal{O}(z^{2\nu+d-4})\} \\
&\qquad + w_2 z^{-(\mu+\lambda)}\{\mathcal{O}(z^{2\nu+d-2}) + z^{2\nu+d}\}\{\cos(\beta z - c_5^{\varsigma}) + \mathcal{O}(z^{-1})\}\big|^2 \mathrm{d}z.
\end{aligned}$$

For assessing the last integral, the following is relevant:

(i) $w_1 z^{2\nu-(1+2\kappa)} - 1 = 0$ if $\nu = \kappa + 1/2$ and $w_1 = 1$.
(ii) $\int_c^{\infty} z^{d-1}(w_1 z^{-2\lambda}\mathcal{O}(z^{2\nu+d-2}))^2\,\mathrm{d}z < \infty$ if $d = 1, 2, 3$ and $\nu = \kappa + 1/2$.
(iii) $\int_c^{\infty} z^{d-1}(w_1 z^{-2\lambda}\{\mathcal{O}(z^{2\nu+d-2}) + \mathcal{O}(z^{2\nu+d-4})\})^2\,\mathrm{d}z < \infty$ if $d = 1, 2, 3$ and $\nu = \kappa + 1/2$.
(iv)

$$\int_c^{\infty} z^{d-1}(w_2 z^{-(\mu+\lambda)}\{\mathcal{O}(z^{2\nu+d-2}) + z^{2\nu+d}\}\{\cos(\beta z - c_5^{\varsigma}) + \mathcal{O}(z^{-1})\})^2\,\mathrm{d}z < \infty$$

if $\mu > \lambda + d/2$ and $\nu = \kappa + 1/2$.

(v)

$$\int_c^\infty z^{d-1}\big(w_1 z^{-2\lambda}\mathcal{O}(z^{2\nu+d-2})\big)\big(w_1 z^{-2\lambda}\mathcal{O}(z^{-2})\big(\mathcal{O}(z^{2\nu+d-2})+z^{2\nu+d}\big)\big)\,\mathrm{d}z<\infty$$

if $d=1,2,3$ and $\nu=\kappa+1/2$.

(vi) $\int_c^\infty z^{d-1}(w_1 z^{-2\lambda}\mathcal{O}(z^{2\nu+d-2}))(w_2 z^{-(\mu+\lambda)}\{\mathcal{O}(z^{2\nu+d-2})+z^{2\nu+d}\}\{\cos(\beta z-c_5^{\varsigma})+\mathcal{O}(z^{-1})\})\,\mathrm{d}z<\infty$ if $\mu>\lambda+d-2$ and $\nu=\kappa+1/2$.

(vii)

$$\int_c^\infty z^{d-1}\big(w_1 z^{-2\lambda}\mathcal{O}(z^{-2})\{\mathcal{O}(z^{2\nu+d-2})+z^{2\nu+d}\}\big)$$
$$\times\big(w_2 z^{-(\mu+\lambda)}\{\mathcal{O}(z^{2\nu+d-2})+z^{2\nu+d}\}\{\cos(\beta z-c_5^{\varsigma})+\mathcal{O}(z^{-1})\}\big)\,\mathrm{d}z<\infty$$

if $\mu>\lambda+d-2$ and $\nu=\kappa+1/2$.

This allows us to conclude that, for a given $\kappa>0$, if $w_1=1$, $\nu=\kappa+1/2$, $\mu>\lambda+d/2$ and $d=1,2,3$, then (13) holds and thus $P(\mathcal{M}_{\nu,\alpha,\sigma_0^2})\equiv P(\varphi_{\mu,\kappa,\beta,\sigma_1^2})$.

Condition $w_1=1$ is equivalent to

$$L^{\varsigma}c_3^{\varsigma}\sigma_1^2\beta^{-(1+2\kappa)}=\pi^{-d/2}\Gamma(\nu+d/2)\Gamma(\nu)^{-1}\sigma_0^2\alpha^{-2\nu},$$

and from the definition of c_3^{ς} and L^{ς}, the previous condition can be rewritten as $\sigma_0^2\alpha^{-2\nu}=C_{\nu,\kappa,\mu}\sigma_1^2\beta^{-(1+2\kappa)}$. □

THEOREM 6. *Let $P(\mathcal{M}_{1/2,\alpha,\sigma_0^2})$ and $P(\varphi_{\mu,0,\beta,\sigma_1^2})$ be two zero mean Gaussian measures. If $\mu>d+1/2$ and*

$$\sigma_0^2\alpha^{-2\nu}=R_\mu\sigma_1^2\beta^{-1}, \tag{14}$$

where $R_\mu=\frac{\mu 2^{1-d}\Gamma(1/2)\Gamma(d)}{\Gamma(1/2+d/2)\Gamma(d/2)}$, then for any bounded infinite set $D\subset\mathbb{R}^d$, $d=1,2,3$, $P(\mathcal{M}_{1/2,\alpha,\sigma_0^2})\equiv P(\varphi_{\mu,0,\beta,\sigma_1^2})$ on the paths of $Z(\boldsymbol{s})$, $\boldsymbol{s}\in D$.

PROOF. The proof follows the same arguments exposed for the case $\kappa>0$ in Theorem 5, but using (6) and Theorem 2 (Points 1 and 2). In this case, it can be shown that if $\mu>d+1/2$, $d=1,2,3$ and $(\frac{\mu 2^{1-d}\Gamma(1/2)\Gamma(d)}{\Gamma(1/2+d/2)\Gamma(d/2)})\sigma_1^2\beta^{-1}=\sigma_0^2\alpha^{-2\nu}$ then (13) holds. □

REMARK. In Theorems 5 and 6 since $\nu=\kappa+1/2$ for $\kappa\geq 0$, using the duplication formula of the gamma function, we easily obtain $C_{\kappa+1/2,\kappa,\mu}=\mu\Gamma(2\kappa+\mu+1)/\Gamma(\mu+1)$, and $R_\mu=\mu$ in (12) and (14), respectively. Then the sufficient condition for $P(\mathcal{M}_{\nu,\alpha,\sigma_0^2})\equiv P(\varphi_{\mu,\kappa,\beta,\sigma_1^2})$ can be simplified as

$$\sigma_0^2\alpha^{-2\nu}=\left(\frac{\mu\Gamma(2\kappa+\mu+1)}{\Gamma(\mu+1)}\right)\sigma_1^2\beta^{-(1+2\kappa)}, \tag{15}$$

$\nu=\kappa+1/2$, $\mu>\lambda+d/2$ and $d=1,2,3$ for $\kappa\geq 0$.

4. Asymptotic properties of the ML estimation for the GW model. We now focus on the microergodic parameter $\sigma^2\beta^{-(1+2\kappa)}$ associated to the GW family. The following results fix the asymptotic properties of its ML estimator. In particular, we will show that the microergodic parameter can be estimated consistently, and then we will assess the asymptotic distribution of the ML estimator.

Let $D \subset \mathbb{R}^d$ be a bounded subset of $\mathbb{R}^d$ and $S_n = \{\boldsymbol{s}_1, \dots, \boldsymbol{s}_n \in D \subset \mathbb{R}^d\}$ denote any set of distinct locations. Let $\boldsymbol{Z}_n = (Z(\boldsymbol{s}_1), \dots, Z(\boldsymbol{s}_n))'$ be a finite realization of a zero mean stationary Gaussian field with a given parametric covariance function $\sigma^2\phi(\cdot;\boldsymbol{\tau})$, with $\sigma^2 > 0$, $\boldsymbol{\tau}$ a parameter vector and ϕ a member of the class Φ_d, with $\phi(0;\boldsymbol{\tau}) = 1$.

We then write $R_n(\boldsymbol{\tau}) = [\phi(\|\boldsymbol{s}_i - \boldsymbol{s}_j\|;\boldsymbol{\tau})]_{i,j=1}^n$ for the associated correlation matrix. The Gaussian log-likelihood function is defined as

$$\mathcal{L}_n(\sigma^2,\boldsymbol{\tau}) = -\frac{1}{2}\left(n\log(2\pi\sigma^2) + \log(|R_n(\boldsymbol{\tau})|) + \frac{1}{\sigma^2}\boldsymbol{Z}_n' R_n(\boldsymbol{\tau})^{-1}\boldsymbol{Z}_n\right). \tag{16}$$

Under the Matérn model, the Gaussian log-likelihood is obtained with $\phi(\cdot;\boldsymbol{\tau}) \equiv \mathcal{M}_{\nu,\alpha,1}$ and $\boldsymbol{\tau} = (\nu,\alpha)'$. Since in what follows ν is assumed known and fixed, for notation convenience, we write $\tau = \alpha$. Let $\hat{\sigma}_n^2$ and $\hat{\alpha}_n$ be the maximum likelihood estimator obtained maximizing $\mathcal{L}_n(\sigma^2,\alpha)$ for a fixed ν.

Below, we report a result that establishes strong consistency and asymptotic distribution of the ML estimation of the microergodic parameter of the Matérn model.

THEOREM 7 [Kaufman and Shaby (2013)]. *Let $Z(\boldsymbol{s}), \boldsymbol{s} \in D \subset \mathbb{R}^d$, $d = 1, 2, 3$, be a zero mean Gaussian field with a Matérn covariance model $\mathcal{M}_{\nu,\alpha_0,\sigma_0^2}$. Suppose $(\sigma_0^2,\alpha_0)' \in (0,\infty) \times [\alpha_L,\alpha_U]$, for any $0 < \alpha_L < \alpha_U < \infty$. Let $(\hat{\sigma}_n^2,\hat{\alpha}_n)'$ maximize (16) over $(0,\infty) \times [\alpha_L,\alpha_U]$. Then as $n \to \infty$:*

1. $\hat{\sigma}_n^2/\hat{\alpha}_n^{2\nu} \xrightarrow{a.s.} \sigma_0^2/\alpha_0^{2\nu}$, *and*
2. $\sqrt{n}(\hat{\sigma}_n^2/\hat{\alpha}_n^{2\nu} - \sigma_0^2/\alpha_0^{2\nu}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 2(\sigma_0^2/\alpha_0^{2\nu})^2)$.

Analogous results can be found in [Wang and Loh (2011), Zhang (2004)], when $\hat{\alpha}_n$ is replaced by α, an arbitrary positive fixed constant. Kaufman and Shaby (2013) show, through simulation study, that asymptotic approximation using a fixed scale parameter can be problematic when applied to finite samples, even for large sample sizes. In contrast, they show that performance is improved and asymptotic approximations are applicable for smaller sample sizes, when the parameters are jointly estimated.

Now, let us consider the Gaussian log-likelihood under the GW model, so that $\boldsymbol{\tau} = (\mu,\kappa,\beta)'$ and $\phi(\cdot;\boldsymbol{\tau}) = \varphi_{\mu,\kappa,\beta,1}(\cdot)$ according to the previous notation. Since in what follows κ and μ are assumed known and fixed, for notation convenience we write $\tau = \beta$. To prove the analogue of Theorem 7 for the GW case, we consider

two types of estimators. The first maximizes (16) with respect to σ^2 for a fixed arbitrary compact support $\beta > 0$, obtaining the following estimator:

$$\hat{\sigma}_n^2(\beta) = \arg\max_{\sigma^2} \mathcal{L}_n(\sigma^2, \beta) = \boldsymbol{Z}_n' R_n(\beta)^{-1} \boldsymbol{Z}_n / n. \tag{17}$$

Here, $R_n(\beta)$ is the correlation matrix coming from the GW family $\varphi_{\mu,\kappa,\beta,1}$. The following result offers some asymptotic properties of the sequence of random variables $\hat{\sigma}_n^2(\beta)/\beta^{(1+2\kappa)}$.

THEOREM 8. *Let $Z(\boldsymbol{s})$, $\boldsymbol{s} \in D \subset \mathbb{R}^d$, $d = 1, 2, 3$, be a zero mean Gaussian field with GW covariance model $\varphi_{\mu,\kappa,\beta_0,\sigma_0^2}$, with $\mu > \lambda + d/2$. Suppose $(\sigma_0^2, \beta_0) \in (0, \infty) \times (0, \infty)$. For a fixed $\beta > 0$, let $\hat{\sigma}_n^2(\beta)$ as defined through equation (17). Then, as $n \to \infty$:*

1. $\hat{\sigma}_n^2(\beta)/\beta^{1+2\kappa} \xrightarrow{a.s.} \sigma_0^2(\beta_0)/\beta_0^{1+2\kappa}$ *and*
2. $\sqrt{n}(\hat{\sigma}_n^2(\beta)/\beta^{1+2\kappa} - \sigma_0^2(\beta_0)/\beta_0^{1+2\kappa}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 2(\sigma_0^2(\beta_0)/\beta_0^{1+2\kappa})^2)$.

PROOF. The proof of the first assertion follows the same arguments of the proof of Theorem 3 in Zhang (2004), and we omit it. The proof of the second assertion is quite technical and long and it has been deferred to the Supplementary Materials [Bevilacqua et al. (2019)]. □

The second type of estimation technique considers the joint maximization of (16) with respect to $(\sigma^2, \beta) \in (0, \infty) \times I$, where $I = [\beta_L, \beta_U]$ and $0 < \beta_L < \beta_U < \infty$. The solution of this optimization problem is given by $(\hat{\sigma}_n^2(\hat{\beta}_n), \hat{\beta}_n)$ where

$$\hat{\sigma}_n^2(\hat{\beta}_n) = \boldsymbol{Z}_n' R_n(\hat{\beta}_n)^{-1} \boldsymbol{Z}_n / n$$

and $\hat{\beta}_n = \arg\max_{\beta \in I} \mathcal{PL}_n(\beta)$. Here, $\mathcal{PL}_n(\beta)$ is the profile log-likelihood:

$$\mathcal{PL}_n(\beta) = -\frac{1}{2}\big(\log(2\pi) + n \log\big(\hat{\sigma}_n^2(\beta)\big) + \log\big|R_n(\beta)\big| + n\big). \tag{18}$$

In order to establish strong consistency and asymptotic distribution of the sequence of random variables $\hat{\sigma}_n^2(\hat{\beta}_n)/\hat{\beta}_n^{1+2\kappa}$, we use the following lemma that establishes the monotone behavior of $\hat{\sigma}_n^2(\beta)/\beta^{1+2\kappa}$ when viewed as a function of $\beta \in I$ under specific condition on the μ parameter.

LEMMA 1. *For any $\beta_1 < \beta_2 \in I$ and for each n, $\hat{\sigma}_n^2(\beta_1)/\beta_1^{1+2\kappa} \le \hat{\sigma}_n^2(\beta_2)/\beta_2^{1+2\kappa}$ if and only if $\mu \ge \lambda + 3$.*

PROOF. The proof follows Kaufman and Shaby (2013). Let $0 < \beta_1 < \beta_2$, with $\beta_1, \beta_2 \in I$. Then, for any $\boldsymbol{Z}_n$,

$$\begin{aligned}
&\hat{\sigma}_n^2(\beta_1)/\beta_1^{1+2\kappa} - \hat{\sigma}_n^2(\beta_2)/\beta_2^{1+2\kappa} \\
&\quad = \frac{1}{n} \boldsymbol{Z}_n' \big(R_n(\beta_1)^{-1} \beta_1^{-(1+2\kappa)} - R_n(\beta_2)^{-1} \beta_2^{-(1+2\kappa)}\big) \boldsymbol{Z}_n
\end{aligned}$$

is nonnegative if the matrix $R_n(\beta_1)^{-1}\beta_1^{-(1+2\kappa)} - R_n(\beta_2)^{-1}\beta_2^{-(1+2\kappa)}$ is positive semidefinite and this happens if and only if the matrix $B = R_n(\beta_2)\beta_2^{1+2\kappa} - R_n(\beta_1)\beta_1^{1+2\kappa}$ with generic element

$$B_{ij} = \beta_2^{1+2\kappa}\varphi_{\mu,\kappa,\beta_2,1}(\|\boldsymbol{s}_i - \boldsymbol{s}_j\|) - \beta_1^{1+2\kappa}\varphi_{\mu,\kappa,\beta_1,1}(\|\boldsymbol{s}_i - \boldsymbol{s}_j\|)$$

is positive semidefinite. From Theorem 2 in Porcu, Zastavnyi and Bevilacqua (2017), this happens if and only if $\mu \geq \lambda + 3$. $\square$

THEOREM 9. *Let $Z(\boldsymbol{s})$, $\boldsymbol{s} \in D \subset \mathbb{R}^d$, $d = 1, 2, 3$, be a zero mean Gaussian field with a GW covariance model $\varphi_{\mu,\kappa,\beta_0,\sigma_0^2}$ with $\mu \geq \lambda + 3$. Suppose $(\sigma_0^2, \beta_0) \in (0, \infty) \times I$ where $I = [\beta_L, \beta_U]$ with $0 < \beta_L < \beta_U < \infty$. Let $(\hat{\sigma}_n^2, \hat{\beta}_n)'$ maximize (16) over $(0, \infty) \times I$. Then as $n \to \infty$:*

1. $\hat{\sigma}_n^2(\hat{\beta}_n)/\hat{\beta}_n^{1+2\kappa} \xrightarrow{a.s.} \sigma_0^2(\beta_0)/\beta_0^{1+2\kappa}$ *and*
2. $\sqrt{n}(\hat{\sigma}_n^2(\hat{\beta}_n)/\hat{\beta}_n^{1+2\kappa} - \sigma_0^2(\beta_0)/\beta_0^{1+2\kappa}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 2(\sigma_0^2(\beta_0)/\beta_0^{1+2\kappa})^2)$.

PROOF. The proof follows Kaufman and Shaby (2013) which uses the same arguments in the Matérn case. Let $\mathcal{G}_n(x) = \hat{\sigma}_n^2(x)/x^{1+2\kappa}$ and define the sequences $\mathcal{G}_n(\beta_L)$ and $\mathcal{G}_n(\beta_U)$. Since $\beta_L \leq \hat{\beta}_n \leq \beta_U$ for every n, then using Lemma 1, $\mathcal{G}_n(\beta_U) \leq \mathcal{G}_n(\hat{\beta}_n) \leq \mathcal{G}_n(\beta_L)$ for all n with probability one. Combining this with Theorem 8 implies the result. $\square$

5. Prediction using GW model. We now consider prediction of a Gaussian field at a new location $\boldsymbol{s}_0$, using the GW model, under fixed domain asymptotics. Specifically, we focus on two properties: asymptotic efficiency prediction and asymptotically correct estimation of prediction variance. Stein (1988) shows that both asymptotic properties hold when the Gaussian measures are equivalent. Let $P(\varphi_{\mu,\kappa,\beta_i,\sigma_i^2})$, $i = 1, 2$, be two probability zero mean Gaussian measures. Under $P(\varphi_{\mu,\kappa,\beta_0,\sigma_0^2})$, and using Theorem 4, both properties hold when $\sigma_0^2\beta_0^{-(1+2\kappa)} = \sigma_1^2\beta_1^{-(1+2\kappa)}$, $\mu > \lambda + d/2$ and $d = 1, 2, 3$.

Similarly, let $P(\mathcal{M}_{\nu,\alpha,\sigma_2^2})$ and $P(\varphi_{\mu,\kappa,\beta_1,\sigma_1^2})$ be two Gaussian measures with the Matérn and GW model. Under $P(\mathcal{M}_{\nu,\alpha,\sigma_2^2})$, both properties hold when (15) is true, $\mu > \lambda + d/2$, $d = 1, 2, 3$. Actually, Stein (1993) gives a substantially weaker condition for asymptotic efficiency prediction based on the asymptotic behavior of the ratio of the isotropic spectral densities. Now, let

$$\widehat{Z}_n(\mu, \kappa, \beta) = \boldsymbol{c}_n(\mu, \kappa, \beta)' R_n(\mu, \kappa, \beta)^{-1} \boldsymbol{Z}_n \tag{19}$$

be the best linear unbiased predictor at an unknown location $\boldsymbol{s}_0 \in D \subset \mathbb{R}^d$, under the misspecified model $P(\varphi_{\mu,\kappa,\beta,\sigma^2})$, where $\boldsymbol{c}_n(\mu, \kappa, \beta) = [\varphi_{\mu,\kappa,\beta,1}(\|\boldsymbol{s}_0 - \boldsymbol{s}_i\|)]_{i=1}^n$ and $R_n(\mu, \kappa, \beta) = [\varphi_{\mu,\kappa,\beta,1}(\|\boldsymbol{s}_i - \boldsymbol{s}_j\|)]_{i,j=1}^n$ is the correlation matrix.

If the correct model is $P(\varphi_{\mu,\kappa,\beta_0,\sigma_0^2})$, then the mean squared error of the predictor is given by

$$\begin{aligned}
&\mathrm{Var}_{\mu,\kappa,\beta_0,\sigma_0^2}\big[\widehat{Z}_n(\mu,\kappa,\beta) - Z(\boldsymbol{s}_0)\big] \\
&\quad = \sigma_0^2\big(1 - 2\boldsymbol{c}_n(\mu,\kappa,\beta)' R_n(\mu,\kappa,\beta)^{-1}\boldsymbol{c}_n(\mu,\kappa,\beta_0) \\
&\qquad + \boldsymbol{c}_n(\mu,\kappa,\beta)' R_n(\mu,\kappa,\beta)^{-1} R_n(\mu,\kappa,\beta_0) R_n(\mu,\kappa,\beta)^{-1}\boldsymbol{c}_n(\mu,\kappa,\beta)\big).
\end{aligned} \tag{20}$$

In the case that $\beta_0 = \beta$, that is, true and wrong models coincide, this expression simplifies to

$$\begin{aligned}
&\mathrm{Var}_{\mu,\kappa,\beta_0,\sigma_0^2}\big[\widehat{Z}_n(\mu,\kappa,\beta_0) - Z(\boldsymbol{s}_0)\big] \\
&\quad = \sigma_0^2\big(1 - \boldsymbol{c}_n(\mu,\kappa,\beta_0)' R_n(\mu,\kappa,\beta_0)^{-1}\boldsymbol{c}_n(\mu,\kappa,\beta_0)\big).
\end{aligned} \tag{21}$$

Similarly, $\mathrm{Var}_{\nu,\alpha,\sigma_2^2}[\widehat{Z}_n(\mu,\kappa,\beta) - Z(\boldsymbol{s}_0)]$ and $\mathrm{Var}_{\nu,\alpha,\sigma_2^2}[\widehat{Z}_n(\nu,\alpha) - Z(\boldsymbol{s}_0)]$ can be defined under $P(\mathcal{M}_{\nu,\alpha,\sigma_2^2})$, where $\widehat{Z}_n(\nu,\alpha)$ is the best linear unbiased predictor using the Matérn model. The following results are an application of Theorems 1 and 2 of Stein (1993).

THEOREM 10. *Let $P(\varphi_{\mu,\kappa,\beta_0,\sigma_0^2})$, $P(\varphi_{\mu,\kappa,\beta_1,\sigma_1^2})$, $P(\mathcal{M}_{\nu,\alpha,\sigma_2^2})$ be three Gaussian probability measures on $D \subset \mathbb{R}^d$ and let $\mu > \lambda$. Then, for all $\boldsymbol{s}_0 \in D$:*

1. *Under $P(\varphi_{\mu,\kappa,\beta_0,\sigma_0^2})$, as $n \to \infty$,*

$$\frac{\mathrm{Var}_{\mu,\kappa,\beta_0,\sigma_0^2}[\widehat{Z}_n(\mu,\kappa,\beta_1) - Z(\boldsymbol{s}_0)]}{\mathrm{Var}_{\mu,\kappa,\beta_0,\sigma_0^2}[\widehat{Z}_n(\mu,\kappa,\beta_0) - Z(\boldsymbol{s}_0)]} \longrightarrow 1, \tag{22}$$

   for any fixed $\beta_1 > 0$.
2. *Under $P(\mathcal{M}_{\nu,\alpha,\sigma_2^2})$, if $\nu = \kappa + 1/2$ as $n \to \infty$,*

$$\frac{\mathrm{Var}_{\nu,\alpha,\sigma_2^2}[\widehat{Z}_n(\mu,\kappa,\beta_1) - Z(\boldsymbol{s}_0)]}{\mathrm{Var}_{\nu,\alpha,\sigma_2^2}[\widehat{Z}_n(\nu,\alpha) - Z(\boldsymbol{s}_0)]} \longrightarrow 1, \tag{23}$$

   for any fixed $\beta_1 > 0$.
3. *Under $P(\varphi_{\mu,\kappa,\beta_0,\sigma_0^2})$, if $\sigma_0^2\beta_0^{-(1+2\kappa)} = \sigma_1^2\beta_1^{-(1+2\kappa)}$, then as $n \to \infty$,*

$$\frac{\mathrm{Var}_{\mu,\kappa,\beta_1,\sigma_1^2}[\widehat{Z}_n(\mu,\kappa,\beta_1) - Z(\boldsymbol{s}_0)]}{\mathrm{Var}_{\mu,\kappa,\beta_0,\sigma_0^2}[\widehat{Z}_n(\mu,\kappa,\beta_1) - Z(\boldsymbol{s}_0)]} \longrightarrow 1. \tag{24}$$

4. *Under $P(\mathcal{M}_{\nu,\alpha,\sigma_2^2})$, if $\mu\Gamma(2\kappa+\mu+1)/\Gamma(\mu+1) \times \sigma_1^2\beta_1^{-(1+2\kappa)} = \sigma_2^2\alpha^{-2\nu}$, $\nu = \kappa + 1/2$, then as $n \to \infty$,*

$$\frac{\mathrm{Var}_{\mu,\kappa,\beta_1,\sigma_1^2}[\widehat{Z}_n(\mu,\kappa,\beta_1) - Z(\boldsymbol{s}_0)]}{\mathrm{Var}_{\nu,\alpha,\sigma_2^2}[\widehat{Z}_n(\mu,\kappa,\beta_1) - Z(\boldsymbol{s}_0)]} \longrightarrow 1. \tag{25}$$

PROOF. Since $\widehat{\varphi}_{\mu,\kappa,\beta_0,\sigma_0^2}(z)$ is bounded away from zero and infinity and as $z \to \infty$,

$$\frac{\widehat{\varphi}_{\mu,\kappa,\beta_1,\sigma_1^2}(z)}{\widehat{\varphi}_{\mu,\kappa,\beta_0,\sigma_0^2}(z)} = \frac{\sigma_1^2\beta_1^d[c_3^\varsigma\beta_1^{-2\lambda}\{1+\mathcal{O}(z^{-2})\}+c_4^\varsigma\beta_1^{-(\mu+\lambda)}z^{\lambda-\mu}\{\cos(z\beta_1-c_5^\varsigma)+\mathcal{O}(z^{-1})\}]}{\sigma_0^2\beta_0^d[c_3^\varsigma\beta_0^{-2\lambda}\{1+\mathcal{O}(z^{-2})\}+c_4^\varsigma\beta_0^{-(\mu+\lambda)}z^{\lambda-\mu}\{\cos(z\beta_0-c_5^\varsigma)+\mathcal{O}(z^{-1})\}]}$$

then, for $\mu > \lambda$, we have

$$\lim_{z\to\infty}\frac{\widehat{\varphi}_{\mu,\kappa,\beta_1,\sigma_1^2}(z)}{\widehat{\varphi}_{\mu,\kappa,\beta_0,\sigma_0^2}(z)} = \frac{\sigma_1^2\beta_1^{-(1+2\kappa)}}{\sigma_0^2\beta_0^{-(1+2\kappa)}} \tag{26}$$

and, using Theorem 1 of Stein (1993), we obtain (22). If $\sigma_1^2\beta_1^{-(1+2\kappa)} = \sigma_0^2\beta_0^{-(1+2\kappa)}$ and using Theorem 2 of Stein (1993), we obtain (24).

Similarly, since $\widehat{\mathcal{M}}_{\nu,\alpha,\sigma_2^2}(z)$ is bounded away from zero and infinity and as $z \to \infty$,

$$\begin{aligned}
&\frac{\widehat{\varphi}_{\mu,\kappa,\beta_1,\sigma_1^2}(z)}{\widehat{\mathcal{M}}_{\nu,\alpha,\sigma_2^2}(z)} \\
&\quad= \frac{\sigma_1^2\beta^d\Gamma(\nu)L^\varsigma}{\Gamma(\nu+d/2)\sigma_0^2\alpha^{-2\nu}\pi^{-\frac{d}{2}}}\big[c_3^\varsigma(\beta z)^{-2\lambda}\{1+\mathcal{O}(z^{-2})\}+c_4^\varsigma(z\beta)^{-(\mu+\lambda)} \\
&\qquad\times\{\cos(\beta z-c_5^\varsigma)+\mathcal{O}(z^{-1})\}\big](\alpha^{-2}+z^2)^{\nu+\frac{d}{2}} \\
&\quad= \frac{\sigma_1^2\beta^d\Gamma(\nu)L^\varsigma}{\Gamma(\nu+d/2)\sigma_0^2\alpha^{-2\nu}\pi^{-\frac{d}{2}}}\big[c_3^\varsigma(\beta z)^{-2\lambda}\{1+\mathcal{O}(z^{-2})\}+c_4^\varsigma(z\beta)^{-(\mu+\lambda)} \\
&\qquad\times\{\cos(\beta z-c_5^\varsigma)+\mathcal{O}(z^{-1})\}\big]z^{2\nu+d}\big[1+(\nu+d/2)(\alpha z)^{-2}+\mathcal{O}(z^{-2})\big] \\
&\quad= \frac{\sigma_1^2\beta^d\Gamma(\nu)L^\varsigma}{\Gamma(\nu+d/2)\sigma_0^2\alpha^{-2\nu}\pi^{-\frac{d}{2}}}\big[c_3^\varsigma\beta^{-2\lambda}z^{2\nu-2\lambda+d}\{1+\mathcal{O}(z^{-2})\} \\
&\qquad+c_4^\varsigma\beta^{-(\mu+\lambda)}z^{2\nu-(\mu+\lambda)+d}\{\cos(\beta z-c_5^\varsigma)+\mathcal{O}(z^{-1})\}\big] \\
&\qquad\times\big[1+(\nu+d/2)(\alpha z)^{-2}+\mathcal{O}(z^{-2})\big]
\end{aligned}$$

then, if $2\nu + d = 2\lambda$, that is, $\kappa + 1/2 = \nu$, $\mu > \lambda$ and considering the Remark then

$$\lim_{z\to\infty}\frac{\widehat{\varphi}_{\mu,\kappa,\beta_1,\sigma_1^2}(z)}{\widehat{\mathcal{M}}_{\nu,\alpha,\sigma_2^2}(z)} = \frac{\sigma_1^2\beta_1^{-(1+2\kappa)}}{\sigma_2^2\alpha^{-2\nu}}\left(\mu\frac{\Gamma(2\kappa+\mu+1)}{\Gamma(\mu+1)}\right). \tag{27}$$

Using Theorem 1 of Stein (1993), we obtain (23). If $\sigma_1^2\beta_1^{-(1+2\kappa)}(\mu\frac{\Gamma(2\kappa+\mu+1)}{\Gamma(\mu+1)}) = \sigma_2^2\alpha^{-2\nu}$ and using Theorem 2 of Stein (1993), we obtain (25). $\square$

The implication of Point 1 is that under $P(\varphi_{\mu,\kappa,\beta_0,\sigma_0^2})$, prediction with $\varphi_{\mu,\kappa,\beta_1,\sigma_0^2}$ with an arbitrary $\beta_1 > 0$ gives asymptotic prediction efficiency, if the correct value of κ and μ are used and $\mu > \lambda$. By virtue of Point 2, under $P(\mathcal{M}_{\nu,\alpha,\sigma_2^2})$, prediction with $\varphi_{\mu,\kappa,\beta_1,\sigma_0^2}$, with an arbitrary $\beta_1 > 0$, gives asymptotic prediction efficiency, if $\nu = \kappa + 1/2$, $\mu > \lambda$. For instance, if $\sigma_2^2 \mathrm{e}^{-r/\alpha}$ is the true covariance, asymptotic prediction efficiency can be achieved with $\sigma_0^2(1 - r/\beta_1)_+^\mu$, using an arbitrary β_1, and $\mu > 1.5$ when $d = 2$. In view of Point 3, under $P(\varphi_{\mu,\kappa,\beta_0,\sigma_0^2})$, prediction with $\varphi_{\mu,\kappa,\beta_1,\sigma_1^2}$, when $\sigma_0^2\beta_0^{-(1+2\kappa)} = \sigma_1^2\beta_1^{-(1+2\kappa)}$ provides asymptotic prediction efficiency and asymptotically correct estimates of error variance, if $\mu > \lambda$. Finally, Point 4 implies that under $P(\mathcal{M}_{\nu,\alpha,\sigma_2^2})$, prediction using $\varphi_{\mu,\kappa,\beta_1,\sigma_1^2}$, under the conditions $\mu\Gamma(2\kappa + \mu + 1)/\Gamma(\mu + 1)\sigma_1^2\beta_1^{-(1+2\kappa)} = \sigma_2^2\alpha^{-2\nu}$, $\nu = \kappa + 1/2$ and $\mu > \lambda$, provides asymptotic prediction efficiency and asymptotically correct estimates of error variance.

For instance, if $\sigma_2^2 \mathrm{e}^{-r/\alpha}$ is the true covariance and $d = 2$, asymptotic prediction efficiency and asymptotically correct estimates of variance error can be achieved with $\sigma_1^2(1 - r/\beta_1)_+^\mu$ setting $\beta_1 = \mu\alpha\sigma_1^2\sigma_2^{-2}$, and $\mu > 1.5$. Setting $\sigma_2^2 = \sigma_1^2 = 1$, $\mu = 3$, $\alpha = x/3$ (x in this case is the so-called practical range, that is, the correlation is lower than 0.05 when $r > x$), the *equivalent* compact support is $\beta_1 = x$. Note that in this special case, the practical range of the exponential model and the compact support of the Askey function coincide. Figure 1 shows the Matérn correlation model with $\nu = 0.5, 1, 1.5$ and practical range equal to 0.6, and two compatible GW correlation models when $d = 2$ with $\kappa = \nu - 0.5$, $\mu = \lambda + 1 + x$, with $x = 0.5, 2$ and the associated compact supports are obtained using the equivalence condition. They are 0.601, 0.595, 0.624 for $\kappa = 0, 0.5, 1$, respectively, when $x = 0.5$ and 0.901, 0.821, 0.815 for $\kappa = 0, 0.5, 1$, respectively, when $x = 2$.

In practice, covariance parameters are unknown, so it is common to estimate them and then plug into (19) and (21). Nevertheless, the asymptotic properties of

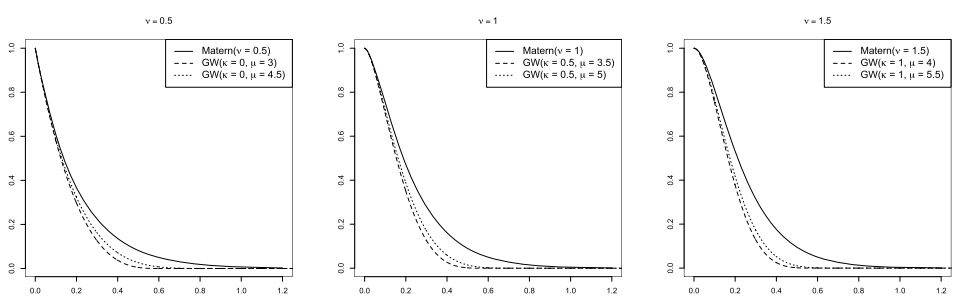

FIG. 1. *Compatible correlation models for the case $d = 2$: The Matérn model when $\nu = 0.5, 1, 1.5$ (from left to right) and the practical range is 0.6 and two compatibles GW models. For the GW models $\kappa = \nu - 0.5$, $\mu = \lambda + 1 + x$, with $x = 0.5, 2$ and the compact support is fixed using the equivalence condition.*

this procedure are quite difficult to obtain [Putter and Young (2001)]. Instead, most theoretical results have been given under a framework in which plug-in parameters are fixed, rather than being estimated from observations.

As in Theorem 4 of Kaufman and Shaby (2013), our Points 3 and 4 may be extended to include estimation of the variance parameter. Specifically, let $\widehat{\sigma}_n^2 = \boldsymbol{Z}_n' R_n(\mu, \kappa, \beta_1)^{-1} \boldsymbol{Z}_n / n$. Then as $n \to \infty$,

$$\frac{\mathrm{Var}_{\mu,\kappa,\beta_1,\widehat{\sigma}_n^2}[\widehat{Z}_n(\mu,\kappa,\beta_1) - Z(\boldsymbol{s}_0)]}{\mathrm{Var}_{\mu,\kappa,\beta_0,\sigma_0^2}[\widehat{Z}_n(\mu,\kappa,\beta_1) - Z(\boldsymbol{s}_0)]} \to 1, \tag{28}$$

$$\frac{\mathrm{Var}_{\mu,\kappa,\beta_1,\widehat{\sigma}_n^2}[\widehat{Z}_n(\mu,\kappa,\beta_1) - Z(\boldsymbol{s}_0)]}{\mathrm{Var}_{\nu,\alpha,\sigma_2^2}[\widehat{Z}_n(\mu,\kappa,\beta_1) - Z(\boldsymbol{s}_0)]} \to 1. \tag{29}$$

The proof follows the lines of Kaufman and Shaby (2013), and we omit it. As outlined in Kaufman and Shaby (2013), we also conjecture that (28) and (29) hold if β_1 is replaced by its maximum likelihood estimator.

6. Simulations and illustrations. The main goals of this section are twofold: on the one hand, we compare the finite sample behavior of the ML estimation of the microergodic parameter of the GW model with the asymptotic distributions given in Theorems 8 and 9. On the other hand, we compare the finite sample behavior of MSE prediction of a zero mean Gaussian field with Matérn covariance model, using both a Matérn and a compatible GW covariance model, using CT applied to a Matérn model as a benchmark.

Regarding the first goal, we simulate, using Cholesky decomposition, and then we estimate with ML, 1000 realizations from a zero mean Gaussian field with GW model. Sampling locations are constructed as in Kaufman, Schervish and Nychka (2008), using a perturbed regular grid. A perturbed grid helps to get more stable estimates because different sets of small distances are available to estimate the parameters. Specifically, we have considered a regular grid with increments 0.03 over $[0, 1]^d$, $d = 2$. Then the grid points have been perturbed, adding a uniform random value on $[-0.01, 0.01]$ to each coordinate. Figure 2 shows the perturbed grid considered, from which we randomly choose $n = 50, 100, 250, 500, 1000$ locations without replacement.

For the GW covariance model $\varphi_{\mu,\kappa,\beta_0,\sigma_0^2}$, we use different values of the compact support and smoothness parameters, that is, $\beta_0 = 0.2, 0.4, 0.6$, $\kappa = 0, 0.5, 1$, and fix $\sigma_0^2 = 1$ and, in view of Theorem 9, $\mu = \lambda(2, \kappa) + 3$. For each simulation, we consider κ and μ as known and fixed, and we estimate with ML the variance and compact support parameters, obtaining $\hat{\sigma}_i^2$ and $\hat{\beta}_i$, $i = 1, \dots, 1000$. To estimate, we first maximize the profile log-likelihood (18) to get $\hat{\beta}_i$. Then we obtain $\hat{\sigma}_i^2(\hat{\beta}_i) = z_i' R(\hat{\beta}_i)^{-1} z_i / n$, where z_i is the data vector of simulation i.

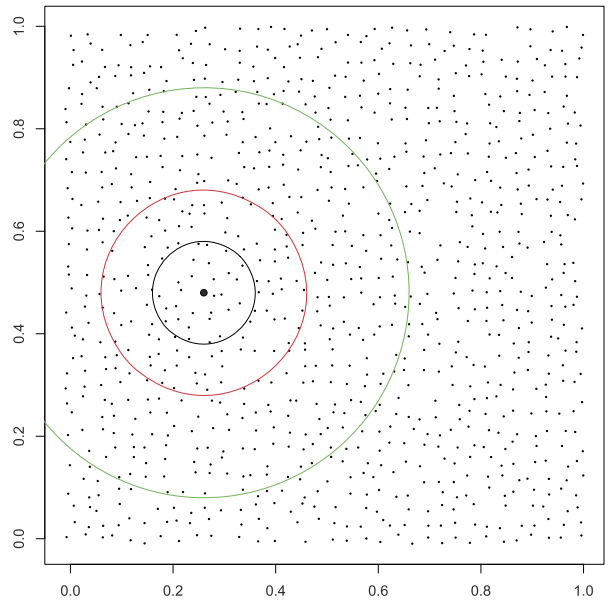

FIG. 2. *Perturbated grid consisting of* $n = 1156$ *considered in the simulation study. The black dot has coordinates* $(0.26, 0.48)$. *In the circles* (*from smaller to larger*), *the location sites involved in prediction with GW with compact support equal to* $0.1, 0.2, 0.4$.

Optimization was carried out using the R [R Development Core Team (2016)] function *optimize* where, following Kaufman and Shaby (2013), the compact support parameter was restricted to the interval $[\varepsilon, 15\beta]$ and ε is slightly larger than machine precision, about 10^{-15} here.

Using the asymptotic distributions stated in Theorems 8 and 9, Table 2 compares the sample quantiles of order $0.05, 0.25, 0.5, 0.75, 0.95$, mean and variance of $\sqrt{n/2}(\widehat{\sigma}_i^2(x)\beta_0^{1+2\kappa}/(\sigma_0^2 x^{1+2\kappa}) - 1)$ for $x = \widehat{\beta}_i, \beta_0, 0.5\beta_0, 2\beta_0$ with the associated theoretical values of the standard Gaussian distribution, for $\beta_0 = 0.4$, $\kappa = 0, 0.5, 1$ and $n = 250, 500, 1000$.

As expected, the best approximation is achieved overall when using the true compact support, that is, $x = \beta_0$, with little difference between the different values of β and κ. In the case of $x = \widehat{\beta}_i$, the asymptotic distribution given in Theorem 9 is a satisfactory approximation of the sample distribution, visually improving when increasing n. The value of κ has less impact compared to β_0. In general, smaller values lead to better results.

When using compact supports that are too small or too large with respect to the true compact support ($x = 0.5\beta_0, 2\beta_0$), the convergence of the asymptotic distribution given in Theorem 8 is very slow. In particular, when $x = 0.5\beta_0$, the asymptotic approximation is not satisfactory even for $n = 1000$. In other words, confidence intervals for the microergodic parameter, based on Theorem 8, that is, fixing

TABLE 2

Sample quantiles, mean and variance of $\sqrt{n/2}(\widehat{\sigma}_i^2(x)\beta_0^{1+2\kappa}/(\sigma_0^2 x^{1+2\kappa}) - 1)$, $i = 1, \ldots, 1000$, *for* $x = \widehat{\beta}_i, \beta_0, 0.5\beta_0, 2\beta_0$ *for different values of* κ, *when* $\beta_0 = 0.4$ *and* $n = 250, 500, 1000$, *compared with the associated theoretical values of the standard Gaussian distribution*

κ	x	n	5%	25%	50%	75%	95%	Mean	Var
0		250	−1.699	−0.721	−0.020	0.798	2.084	0.072	1.375
	$\widehat{\beta}$	500	−1.680	−0.677	0.027	0.758	1.966	0.071	1.212
		1000	−1.614	−0.666	0.062	0.767	1.788	0.057	1.104
		250	−1.548	−0.670	−0.039	0.675	1.833	0.025	1.058
	β_0	500	−1.632	−0.665	0.001	0.661	1.754	0.027	1.047
		1000	−1.629	−0.690	0.020	0.698	1.627	0.011	1.009
		250	3.224	4.953	6.163	7.471	9.370	6.234	3.493
	$0.5\beta_0$	500	3.399	4.762	5.948	7.018	8.879	5.979	2.840
		1000	2.792	4.063	5.059	5.984	7.516	5.088	2.088
		250	−2.443	−1.698	−1.128	−0.490	0.610	−1.065	0.898
	$2\beta_0$	500	−2.485	−1.576	−0.941	−0.313	0.718	−0.904	0.947
		1000	−2.324	−1.438	−0.759	−0.107	0.819	−0.757	0.949
0.5		250	−1.761	−0.786	0.019	0.807	2.271	0.072	1.506
	$\widehat{\beta}$	500	−1.774	−0.714	0.027	0.822	1.978	0.063	1.309
		1000	−1.609	−0.700	0.047	0.761	1.840	0.051	1.152
		250	−1.548	−0.670	−0.039	0.675	1.833	0.025	1.058
	β_0	500	−1.632	−0.665	0.001	0.661	1.754	0.027	1.047
		1000	−1.629	−0.690	0.020	0.698	1.627	0.011	1.009
		250	11.462	14.603	16.995	19.573	23.414	17.155	12.818
	$0.5\beta_0$	500	11.133	13.624	15.459	17.592	21.090	15.697	9.060
		1000	9.192	11.051	12.578	14.187	16.904	12.733	5.560
		250	−3.166	−2.469	−1.914	−1.315	−0.260	−1.860	0.784
	$2\beta_0$	500	−3.136	−2.258	−1.628	−1.037	−0.029	−1.604	0.883
		1000	−2.851	−1.999	−1.353	−0.707	0.207	−1.342	0.907
1		250	−1.825	−0.868	0.042	0.836	2.389	0.078	1.661
	$\widehat{\beta}$	500	−1.869	−0.770	0.027	0.820	2.092	0.059	1.412
		1000	−1.679	−0.719	0.058	0.762	1.836	0.045	1.199
		250	−1.548	−0.670	−0.039	0.675	1.833	0.025	1.058
	β_0	500	−1.632	−0.665	0.001	0.661	1.754	0.027	1.047
		1000	−1.629	−0.690	0.020	0.698	1.627	0.011	1.009
		250	28.654	34.704	39.574	44.651	52.477	39.856	51.483
	$0.5\beta_0$	500	27.166	31.848	35.553	39.808	46.519	35.992	34.995
		1000	22.055	25.398	28.218	31.256	36.451	28.565	19.929
		250	−3.949	−3.312	−2.806	−2.262	−1.288	−2.750	0.666
	$2\beta_0$	500	−3.876	−3.050	−2.445	−1.862	−0.925	−2.427	0.809
		1000	−3.524	−2.675	−2.065	−1.419	−0.532	−2.047	0.856
$N(0, 1)$			−1.645	−0.674	0	0.674	1.645	0	1

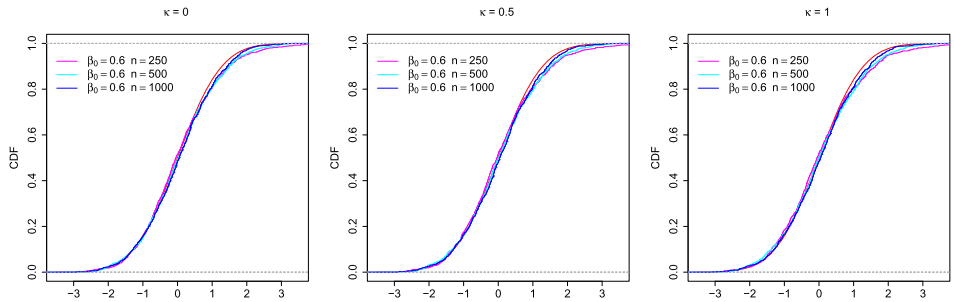

FIG. 3. *Empirical CDF of the simulated ML estimation of the standardized microergodic parameter vs. CDF of a standard Gaussian distribution* (*red line*) *when* $\sigma_0^2 = 1$, $\kappa = 0, 0.5, 1$ (*from left to right*), $\beta_0 = 0.6$ *and* $n = 250, 500, 1000$

an arbitrary compact support, can be problematic when applied to finite samples, even for large sample sizes. We strongly recommend jointly estimating variance and compact support and using the asymptotic distribution give in Theorem 9 or, alternatively, choosing β conservatively.

As a graphical example, Figure 3 compares the empirical CDF of the ML estimates of the standardized microergodic parameter with the CDF of the standard Gaussian distribution when $\sigma_0^2 = 1$, $\kappa = 0, 0.5, 1$, $\beta_0 = 0.6$ and $n = 250, 500, 1000$. Finally, our numerical results are consistent with the results in Kaufman and Shaby (2013), in the Matérn case.

As for the second goal, using the results given in Theorem 10 Points 2 and 4, we now specifically compare asymptotic prediction efficiency and asymptotically correct estimation of prediction variance using ratios (23) and (25), respectively. As a benchmark, we also consider the same ratios using a tapered Matérn model.

More precisely, we consider a Matérn model $\mathcal{M}_{\nu,\alpha,\sigma_2^2}$ setting $\sigma_2^2 = 1$, $\nu = 0.5, 1, 1.5$ and $\alpha = y/c_\nu$ with $y = 0.1, 0.2, 0.4$ if $\nu = 0.5$, $y = 0.101, 0.202, 0.404$ if $\nu = 1$ and $y = 0.097, 0.193, 0.385$ if $\nu = 1.5$. Here, c_ν is a scalar depending on ν such that $\mathcal{M}_{\nu,1,1}(r)$ is lower than 0.05 when $r > c_\nu$ that is, y is the practical range.

Let us define the ratios (23) and (25) as $U_1(\beta_1)$ and U_2, respectively. For each ν and α, we randomly select $n_j = 50, 100, 250, 500, 1000$, $j = 1, \dots, 500$ location sites without replacement from the perturbed grid in Figure 2. For each j, we compute the ratio $U_{1j}(\beta_1)$ and the ratio U_{2j}, $j = 1, \dots, 500$, using closed-form expressions in equations (20) and (21) when predicting the location site $(0.26, 0.48)'$ (black dot in Figure 2). Specifically for each U_{2j}, following the conditions in Theorem 10 Point 4, we set $\sigma_1^2 = 1$, $\kappa = \nu - 1/2$, $\mu = \lambda + 1.5$. The "equivalent" compact support is obtained as

$$\beta_1^* = \left[\left(\mu \frac{\Gamma(2\kappa + \mu + 1)}{\Gamma(\mu + 1)}\right) \frac{\sigma_1^2 \alpha^{-2\nu}}{\sigma_2^2}\right]^{1/(1+2\kappa)}.$$

Under this specific setting, the "equivalent" compact support associated to the (varying with ν) practical range is approximately $\beta_1^* = 0.1, 0.2, 0.4$, irrespectively of ν. Figure 2 shows the location sites involved in the prediction using GW functions with $\beta_1^* = 0.1, 0.2, 0.4$.

For each $U_{1j}(\beta)$, following Theorem 10 Point 2, we fix $\kappa = \nu - 1/2$, $\mu = \lambda + 1.5$ and $\beta = \beta_1^*$. Then, to investigate the effect of considering an arbitrary compact support on the convergence of ratio (23), we also consider, $U_{1j}(0.2\beta_1^*)$ and $U_{1j}(5\beta_1^*)$. For each combination of ν, α, Table 3 shows the empirical means $\bar{U}_1(x\beta_1^*) = \sum_{j=1}^{500} U_{1j}(x\beta_1^*)/500$ for $x = 1, 0.5, 2$, and $\bar{U}_2 = \sum_{j=1}^{500} U_{2j}/500$ when increasing n.

As a benchmark, we also compute the empirical means replacing the GW model with a tapered Matérn covariance model, that is, considering the model $\mathcal{M}_{\nu,\alpha,\sigma_2^2} K_{x\beta_1^*}$, and we denote these means by $\bar{U}_1^T(x\beta_1^*)$, $x = 1, 0.5, 2$ and $\bar{U}_2^T$. Here, $K_{x\beta_1^*}$ is a known compactly supported correlation function called the taper function. Following Furrer, Genton and Nychka (2006), as the taper function, we use $K_{x\beta_1^*} = \varphi_{2,0,x\beta_1^*,1}$ if $\nu = 0.5$, $K_{x\beta_1^*} = \varphi_{3,1,x\beta_1^*,1}$ if $\nu = 1$ and $K_{x\beta_1^*} = \varphi_{4,2,x\beta_1^*,1}$ if $\nu = 1.5$. for $x = 1, 0.5, 2$.

These specific choices of taper functions guarantee the convergence of ratios (23) and (25), using a tapered Matérn model instead of the GW model [see Theorem 2 in Furrer, Genton and Nychka (2006)]. In Table 3, the percentages of nonzero elements in the covariance matrices are also reported in all scenarios and for each n when using the compact support β_1^*.

Table 3 shows that $\bar{U}_2$ clearly overall outperforms $\bar{U}_2^T$ in terms of speed of convergence in particular when increasing β_1^*. This implies that in terms of finite sample, if the Matérn model is the state of nature, prediction efficiency and correct estimation of prediction variance are better achieved when predicting with the (compatible) GW model with respect to the so-called naive CT predictor [Furrer, Genton and Nychka (2006)], sharing the same compact support.

Comparing $\bar{U}_1(x\beta_1^*)$ with $\bar{U}_1^T(x\beta_1^*)$ for $x = 1, 0.5, 2$ note that when $x = 1$, $\bar{U}_1(\beta_1^*)$ overall slightly outperforms $\bar{U}_1^T(\beta_1^*)$ and when $x = 0.5$, the convergence of both ratios seems to be very slow, in particular for larger ν. This suggests that taking an arbitrary compact support too small with respect to the "equivalent" compact support β_1^* can seriously affect the prediction efficiency both for tapered Matérn and GW models. This kind of problem disappears when $x = 2$, as expected. By the tapering effect, that is, inducing a covariance with an apparent shorter range, $\bar{U}_1^T(2\beta_1^*)$ slightly outperforms $\bar{U}_1(2\beta_1^*)$.

7. Concluding remarks. Parameter estimation for interpolation of spatially or spatio-temporally correlated random processes is used in many areas and often requires particular models or careful implementation. In recent years, the dataset

TABLE 3

$\bar{U}_1(x)$, $\bar{U}_1^T(x)$, $x = 0.5\beta_1^*, 2\beta_1^*, \beta_1^*$ and $\bar{U}_2$, $\bar{U}_2^T$, as defined in Section 6, when considering a Matérn model with increasing practical range y, smoothness parameter ν and n. Here β_1^* is the compact support parameter of the GW model computed using the equivalence condition. The column % reports the mean of percentages of non-zero elements in the covariance matrices involved when considering β_1^*

		$\nu = 0.5$								$\nu = 1$								$\nu = 1.5$								
	n	$\bar{U}_1(0.5\beta_1^*)$	$\bar{U}_1(2\beta_1^*)$	$\bar{U}_1(\beta_1^*)$	$\bar{U}_1^T(0.5\beta_1^*)$	$\bar{U}_1^T(2\beta_1^*)$	$\bar{U}_1^T(\beta_1^*)$	$\bar{U}_2$	$\bar{U}_2^T$	$\bar{U}_1(0.5\beta_1^*)$	$\bar{U}_1(2\beta_1^*)$	$\bar{U}_1(\beta_1^*)$	$\bar{U}_1^T(0.5\beta_1^*)$	$\bar{U}_1^T(2\beta_1^*)$	$\bar{U}_1^T(\beta_1^*)$	$\bar{U}_2$	$\bar{U}_2^T$	$\bar{U}_1(0.5\beta_1^*)$	$\bar{U}_1(2\beta_1^*)$	$\bar{U}_1(\beta_1^*)$	$\bar{U}_1^T(0.5\beta_1^*)$	$\bar{U}_1^T(2\beta_1^*)$	$\bar{U}_1^T(\beta_1^*)$	$\bar{U}_2$	$\bar{U}_2^T$	%
	$\alpha = \frac{y}{c_\nu}$	$y = 0.1$								$y = 0.101$								$y = 0.097$								
$\beta_1^* = 0.1$	50	1.051	1.029	1.009	1.051	1.008	1.025	1.019	1.029	1.098	1.047	1.018	1.101	1.009	1.041	1.038	1.048	1.124	1.054	1.024	1.134	1.012	1.057	1.050	1.056	4.67
	100	1.096	1.043	1.014	1.096	1.013	1.043	1.035	1.056	1.189	1.076	1.028	1.195	1.013	1.072	1.073	1.097	1.246	1.095	1.039	1.266	1.019	1.105	1.098	1.112	3.70
	250	1.182	1.046	1.019	1.183	1.018	1.069	1.064	1.118	1.379	1.097	1.038	1.393	1.016	1.121	1.138	1.204	1.521	1.156	1.059	1.567	1.025	1.197	1.197	1.241	3.12
	500	1.267	1.030	1.015	1.268	1.016	1.077	1.081	1.211	1.608	1.065	1.030	1.639	1.011	1.132	1.187	1.372	1.928	1.116	1.051	2.039	1.020	1.253	1.300	1.481	2.92
	1000	1.325	1.015	1.009	1.330	1.010	1.061	1.073	1.332	1.858	1.032	1.016	1.923	1.005	1.088	1.168	1.586	2.549	1.054	1.025	2.820	1.008	1.209	1.300	1.877	2.82
	$\alpha = \frac{y}{c_\nu}$	$y = 0.2$								$y = 0.202$								$y = 0.193$								
$\beta_1^* = 0.2$	50	1.151	1.044	1.016	1.157	1.016	1.058	1.053	1.134	1.316	1.090	1.032	1.326	1.014	1.094	1.119	1.217	1.448	1.139	1.048	1.505	1.021	1.149	1.177	1.288	11.95
	100	1.209	1.032	1.013	1.221	1.015	1.066	1.068	1.235	1.471	1.068	1.027	1.491	1.011	1.103	1.162	1.377	1.730	1.120	1.045	1.848	1.018	1.186	1.266	1.534	11.04
	250	1.227	1.012	1.007	1.247	1.008	1.046	1.060	1.397	1.578	1.026	1.013	1.614	1.004	1.061	1.146	1.590	2.085	1.049	1.022	2.363	1.007	1.137	1.271	1.954	10.48
	500	1.152	1.005	1.003	1.178	1.003	1.021	1.040	1.513	1.415	1.009	1.005	1.447	1.002	1.022	1.092	1.625	1.945	1.017	1.009	2.321	1.004	1.051	1.174	2.069	10.27
	1000	1.061	1.002	1.001	1.083	1.001	1.007	1.024	1.586	1.145	1.003	1.002	1.152	1.002	1.014	1.052	1.497	1.358	1.005	1.003	1.554	1.003	1.029	1.093	1.728	10.18
	$\alpha = \frac{y}{c_\nu}$	$y = 0.4$								$y = 0.404$								$y = 0.385$								
$\beta_1^* = 0.4$	50	1.208	1.016	1.008	1.226	1.010	1.050	1.060	1.372	1.507	1.035	1.016	1.530	1.005	1.072	1.148	1.519	1.900	1.066	1.027	2.088	1.010	1.152	1.271	1.823	34.97
	100	1.151	1.006	1.004	1.174	1.004	1.026	1.041	1.499	1.399	1.013	1.007	1.421	1.002	1.030	1.100	1.583	1.846	1.024	1.011	2.106	1.004	1.071	1.196	2.006	34.18
	250	1.050	1.001	1.001	1.066	1.001	1.006	1.020	1.598	1.128	1.003	1.002	1.141	1.001	1.011	1.048	1.454	1.328	1.006	1.003	1.491	1.003	1.028	1.091	1.691	33.90
	500	1.013	1.000	1.000	1.019	1.000	1.002	1.011	1.633	1.024	1.001	1.001	1.044	1.001	1.009	1.025	1.314	1.053	1.002	1.001	1.121	1.001	1.019	1.047	1.373	33.66
	1000	1.003	1.000	1.000	1.005	1.000	1.000	1.006	1.649	1.003	1.000	1.000	1.034	1.000	1.006	1.014	1.208	1.005	1.000	1.000	1.077	1.000	1.009	1.024	1.184	33.59

sizes have steadily increased such that straightforward statistical tools are computationally too expensive to use. The use of covariance functions with an (inherent or induced) compact support, leading to sparse matrices, is a very accessible and scalable approach. In this paper, we studied estimation and prediction of Gaussian fields with covariance models belonging to the GW class, under fixed domain asymptotics.

Specifically, we first characterize the equivalence of two Gaussian measures with GW models, and then we establish strong consistency and asymptotic Gaussianity of the ML estimator of the associated microergodic parameter when considering both an arbitrary and an estimated compact support. Simulation results show that for a finite sample, the choice of an arbitrary compact support can result in a very poor approximation of the asymptotic distribution. These results are consistent with those in Kaufman and Shaby (2013) in the Matérn case.

In a second aspect, we give a sufficient condition for the equivalence of two Gaussian measures with Matérn and GW model and we study the effect on prediction when using these two covariance models under fixed domain asymptotics. A first consequence of our results is that GW model is more than a valid competitor of the Matérn model. It allows, as in the Matérn case, a continuous parameterization of smoothness of the underlying Gaussian field and, under fixed domain asymptotics, prediction and mean square error prediction obtained with a Matérn model can be achieved using a GW model inducing an equivalent Gaussian measure, using our condition. For this reason, we advocate the GW class when working with (not necessarily) large or huge spatial datasets since well established and implemented algorithms for sparse matrices can be used when estimating the covariance parameters and/or predicting at unknown locations [e.g., Furrer and Sain (2010)]. Alternatively, for covariances which are analytic away from the origin as the Matérn model, in some circumstances a hierarchical factorization scheme as proposed for instance in Ambikasaran et al. (2016), is a possible solution in order to handle sample sizes that cannot be handled by straightforward Cholesky factorization.

As the theoretical and numerical results illustrate, CT for prediction is essentially an obsolete approach. When comparing both approaches with the same sensible compact support, the tapered CT is less efficient. For estimation, one has to distinguish between a so-called one-taper or two-taper approach, that is, a proper likelihood or an estimating function approach, Kaufman, Schervish and Nychka (2008). Fixing again the support, a GW model can approximate a Matérn covariance function much better than a tapered one. Thus, the GW is in an estimation setting superior to a one-taper CT. In both approaches, one needs to be aware of the resulting biases, which can be substantial. In the case of (kriging) predictions based on plug-in estimates, the biases are largely canceled [Furrer, Bachoc and Du (2016)]. Finally, the two-taper approach is conceptually a different approach and,

as it is computationally very expensive, it would not be fair to compare it with the GW model.

Similar to the Matérn model with smoothness parameter different to $p + 1/2$, $p \in \mathbb{N}$, the GW does not have a closed-form expression when its smoothness parameter is different to p, and low level software implementations are needed for a computationally efficient use.

SUPPLEMENTARY MATERIAL

Supplement to "Estimation and prediction using generalized Wendland covariance functions under fixed domain asymptotics" (DOI: 10.1214/17-AOS1652SUPP; .pdf). The Supplement contains the proof of Assertion 2 in Theorem 8.

REFERENCES

ABRAMOWITZ, M. and STEGUN, I. A., eds. (1965). *Handbook of Mathematical Functions, with Formulas, Graphs, and Mathematical Tables. National Bureau of Standards Applied Mathematics Series* **55**. Superintendent of Documents, U.S. Government Printing Office, Washington, DC. MR0177136

ADLER, R. J. (1981). *The Geometry of Random Fields*. Wiley, Chichester. MR0611857

AMBIKASARAN, S., FOREMAN-MACKEY, D., GREENGARD, L., HOGG, D. W. and O'NEIL, M. (2016). Fast Direct Methods for Gaussian Processes. *IEEE Trans. Pattern Anal. Mach. Intell.* **38** 252–265.

ANDERES, E. (2010). On the consistent separation of scale and variance for Gaussian random fields. *Ann. Statist.* **38** 870–893. MR2604700

ASKEY, R. (1973). Radial characteristic functions. Technical Report, Research Center, Univ. Wisconsin-Madison, Madison, WI.

BEVILACQUA, M., FAOUZI, T., FURRER, R. and PORCU, E. (2019). Supplement to "Estimation and prediction using generalized Wendland covariance functions under fixed domain asymptotics." DOI:10.1214/17-AOS1652SUPP.

CHERNIH, A. and HUBBERT, S. (2014). Closed form representations and properties of the generalised Wendland functions. *J. Approx. Theory* **177** 17–33. MR3128401

CHERNIH, A., SLOAN, I. H. and WOMERSLEY, R. S. (2014). Wendland functions with increasing smoothness converge to a Gaussian. *Adv. Comput. Math.* **40** 17–33.

DU, J., ZHANG, H. and MANDREKAR, V. S. (2009). Fixed-domain asymptotic properties of tapered maximum likelihood estimators. *Ann. Statist.* **37** 3330–3361. MR2549562

FURRER, R., BACHOC, F. and DU, J. (2016). Asymptotic properties of multivariate tapering for estimation and prediction. *J. Multivariate Anal.* **149** 177–191.

FURRER, R., GENTON, M. G. and NYCHKA, D. (2006). Covariance tapering for interpolation of large spatial datasets. *J. Comput. Graph. Statist.* **15** 502–523.

FURRER, R. and SAIN, S. R. (2010). spam: A sparse matrix R package with emphasis on MCMC methods for Gaussian Markov random fields. *J. Stat. Softw.* **36** 1–25.

GNEITING, T. (2002a). Compactly supported correlation functions. *J. Multivariate Anal.* **83** 493–508.

GNEITING, T. (2002b). Nonseparable, stationary covariance functions for space-time data. *J. Amer. Statist. Assoc.* **97** 590–600. MR1941475

GOLUBOV, B. I. (1981). On Abel–Poisson type and Riesz means. *Anal. Math.* **7** 161–184.

HIRANO, T. and YAJIMA, Y. (2013). Covariance tapering for prediction of large spatial data sets in transformed random fields. *Ann. Inst. Statist. Math.* **65** 913–939. MR3105802

IBRAGIMOV, I. D. A. and ROZANOV, Y. A. (1978). *Gaussian Random Processes. Applications of Mathematics* **9**. Springer, New York. MR0543837

KAUFMAN, C. G., SCHERVISH, M. J. and NYCHKA, D. W. (2008). Covariance tapering for likelihood-based estimation in large spatial data sets. *J. Amer. Statist. Assoc.* **103** 1545–1555.

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

MARDIA, K. V. and MARSHALL, J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika* **71** 135–146.

MATÉRN, B. (1960). *Spatial Variation*. Meddelanden Från Statens Skogsforskningsinstitut, Band 49, Nr 5., Stockholm.

PORCU, E., ZASTAVNYI, V. P. and BEVILACQUA, M. (2017). Buhmann covariance functions, their compact supports, and their smoothness. *Dolomites Res. Notes Approx.* **10** 33–42. MR3708624

PUTTER, H. and YOUNG, G. A. (2001). On the effect of covariance function estimation on the accuracy of kriging predictors. *Bernoulli* **7** 421–438. MR1836738

R DEVELOPMENT CORE TEAM (2016). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

SCHABACK, R. (2011). The missing Wendland functions. *Adv. Comput. Math.* **34** 67–81.

SCHOENBERG, I. J. (1938). Metric spaces and completely monotone functions. *Ann. of Math.* (2) **39** 811–841.

SKOROKHOD, A. V. and YADRENKO, M. I. (1973). On absolute continuity of measures corresponding to homogeneous Gaussian fields. *Theory Probab. Appl.* **18** 27–40.

STEIN, M. (1988). Asymptotically efficient prediction of a random field with a misspecified covariance function. *Ann. Statist.* **16** 55–63. MR0924856

STEIN, M. (1990). Uniform asymptotic optimality of linear predictions of a random field using an incorrect second-order structure. *Ann. Statist.* **18** 850–872. MR1056340

STEIN, M. L. (1993). A simple condition for asymptotic optimality of linear predictions of random fields. *Statist. Probab. Lett.* **17** 399–404. MR1237787

STEIN, M. L. (1999a). Predicting random fields with increasing dense observations. *Ann. Appl. Probab.* **9** 242–273. MR1682572

STEIN, M. L. (1999b). *Interpolation of Spatial Data*: *Some Theory for Kriging*. Springer, New York. MR1697409

STEIN, M. L. (2004). Equivalence of Gaussian measures for some nonstationary random fields. *J. Statist. Plann. Inference* **123** 1–11. MR2058118

STEIN, M. L. (2013). Statistical properties of covariance tapers. *J. Comput. Graph. Statist.* **22** 866–885. MR3173747

WACKERNAGEL, H. (2003). *Multivariate Geostatistics*: *An Introduction with Applications*, 3rd ed. Springer, New York.

WANG, D. and LOH, W.-L. (2011). On fixed-domain asymptotics and covariance tapering in Gaussian random field models. *Electron. J. Stat.* **5** 238–269. MR2792553

WENDLAND, H. (1995). Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. *Adv. Comput. Math.* **4** 389–396. MR1366510

WU, Z. (1995). Compactly supported positive definite radial functions. *Adv. Comput. Math.* **18** 283–292.

YAGLOM, A. M. (1987). *Correlation Theory of Stationary and Related Random Functions. Volume I*: *Basic Results*. Springer, New York.

ZASTAVNYI, V. P. (2006). On some properties of Buhmann functions. *Ukrainian Math. J.* **58** 1184–1208.

ZHANG, H. (2004). Inconsistent estimation and asymptotically equivalent interpolations in model-based geostatistics. *J. Amer. Statist. Assoc.* **99** 250–261.

ZHANG, H. and ZIMMERMAN, D. L. (2005). Towards reconciling two asymptotic frameworks in spatial statistics. *Biometrika* **92** 921–936. MR2234195

M. BEVILACQUA
DEPARTMENT OF STATISTICS
UNIVERSIDAD DE VALPARAISO
2360102 VALPARAISO
CHILE
E-MAIL: moreno.bevilacqua@uv.cl

R. FURRER
DEPARTMENT OF MATHEMATICS AND
DEPARTMENT OF COMPUTATIONAL SCIENCE
UNIVERSITY OF ZURICH
8057 ZURICH
SWITZERLAND
E-MAIL: reinhard.furrer@math.uzh.ch

T. FAOUZI
DEPARTMENT OF STATISTICS
UNIVERSIDAD DEL BIOBIO
4081112 CONCEPCION
CHILE
E-MAIL: tfaouzi@ubiobio.cl

E. PORCU
SCHOOL OF MATHEMATICS,
STATISTICS AND PHYSICS
NEWCASTLE UNIVERSITY
NE1 7RU NEWCASTLE UPON TYNE
UNITED KINGDOM
AND
DEPARTMENT DE MATEMATICÁS
UNIVERSIDAD DE ATACAMA
COPIAPÓ
CHILE
E-MAIL: emilio.porcu@newcastle.ac.uk