Compatibility of space-time kernels with full, dynamical, or compact support

Faouzi, Tarik ; Furrer, Reinhard ; Porcu, Emilio

DOI: https://doi.org/10.1002/mma.10379

Originally published at:
Faouzi, Tarik; Furrer, Reinhard; Porcu, Emilio (2024). Compatibility of space-time kernels with full, dynamical, or compact support. Mathematical Methods in The Applied Sciences:Epub ahead of print.
DOI: https://doi.org/10.1002/mma.10379

Received: 10 June 2023 | Revised: 4 June 2024 | Accepted: 21 July 2024

DOI: 10.1002/mma.10379

RESEARCH ARTICLE

WILEY

Compatibility of space-time kernels with full, dynamical, or compact support

Tarik Faouzi[1] | **Reinhard Furrer**[2] | **Emilio Porcu**[3,4]

[1]Department of Mathematics and Computer Science, USACH, Santiago, Chile

[2]Department of Mathematical Modeling and Machine Learning, University of Zurich, Zurich, Switzerland

[3]Department of Mathematics and Center for Biotechnology (BTC), Khalifa University, Abu Dhabi, United Arab Emirates

[4]ADIA Lab, Abu Dhabi, United Arab Emirates

Correspondence
Emilio Porcu, Department of Mathematics and Center for Biotechnology (BTC), Khalifa University, Abu Dhabi, United Arab Emirates.
Email: emilio.porcu@ku.ac.ae

Communicated by: P. Agarwal

Present address
Emilio Porcu, Khalifa University of Science and Technology, Campus, Abu Dhabi, United Arab Emirates

Funding information
Fondo Nacional de Desarrollo Científico y Tecnológico, Grant/Award Number: 11200749; Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung

This paper deals with compatibility of space-time kernels with (either) full, spatially dynamical, or space-time compact support. We deal with the dilemma of statistical accuracy versus computational scalability, which are in a notorious trade-off. Apparently, models with full support ensure maximal information but are computationally expensive, while compactly supported models achieve computational scalability at the expense of loss of information. Hence, an inspection of whether these models might be compatible is necessary. The criterion we use for such an inspection is based on equivalence of Gaussian measures. We provide sufficient conditions for space-time compatibility. As a corollary, we deduce implications in terms of maximum likelihood estimation and misspecified kriging prediction under fixed domain asymptotics. Some results of independent interest relate about the space-time spectrum associated with the classes of kernels proposed in the paper.

KEYWORDS

fixed-domain asymptotics, Matérn covariance, maximum likelihood, microergodic parameter, prediction, space-time generalized Wendland family

MSC CLASSIFICATION

60G60, 62M40, 42A82

1 | INTRODUCTION

This paper provides a comparison of space-time covariance functions on the basis of their support. We target three covariance models having a full, a dynamically varying, or a compact support. The basis for comparison is their compatibility, which in turn translates into conditions for equivalence or orthogonality of Gaussian measures under given classes of kernels.

The reason for such a comparison is motivated by the ubiquitous interest from the statistical community in comparing models with full support versus those with compact support. While such a discussion has been largely conducted for spatial domains only, we are unaware of any analog for the space-time case. Probably the main reason is the lack of space-time models with compact support. Before digging into this aspect, we contextualize the main reasons for a discussion about supports in spatial statistics.

The main dispute is connected with the role of compact support in mitigating the computational burden induced by huge spatial datasets while keeping a reasonable level of statistical efficiency. A wealth of statistical approaches has been proposed to circumvent this problem, and it is not within the scope of this paper to report all of them. We shall instead remind the reader of the review by [1], with the references therein.

While being aware that there are many competitive frameworks to deal with the dilemma of statistical accuracy versus computational scalability, we believe that compact support-based approaches are quite unique in that they embrace many important aspects of space-time modeling that go beyond the mentioned trade-off. To mention a few, we have the following:

1. Fixed domain asymptotics has an important role in quantifying the impact of given parametric classes of covariance functions into both maximum likelihood estimation and best linear unbiased prediction, termed kriging in the spatial statistics literature [2]. The literature in this direction has been centered on the Matérn class of covariance functions [3, 4], which allows for indexing mean square differentiability of the associated Gaussian random field. Recent work [5] has shown that, under some mild regularity conditions, a parametric class of compactly supported covariance functions allows asymptotically achieving the same level of estimation and prediction accuracy while guaranteeing computational savings.
2. Screening effect historically refers to the situation when the observations located far from the predictand (the value to be predicted at some target point) receive a small (ideally, zero) kriging weight. The screening effect phenomenon is certainly multifactorial: spatial design, the dimension of the spatial domain, and the geometric properties of the associated random field are all having an impact on the screening effect.

   Stein [6] deviates from earlier literature and adopts an asymptotic approach to assess the screening effect problem. The general suggestion is to use the Matérn model to assess the screening effect, and an example in [3] argues against models with compact support. A recent contribution by [7] shows that some classes of compactly supported covariance functions allow for screening effect when working in either regular or irregular settings of the spatial design. Further, their numerical studies suggest that the screening effect under compact support might be even stronger than the screening effect under a Matérn model.
3. Recent findings [8] prove that the Matérn class is a limiting case of a reparameterized version of a class of compactly supported covariance functions, called Generalized Wendland class [5]. This implies that the (reparametrized) generalized Wendland model is more flexible than the Matérn model, having an additional parameter that allows for switching from compactly to globally supported covariance functions. A thorough analysis of the state of the art for the Matérn covariance function can be found in [9].
4. Tapering is a very popular technique in spatial statistics. It consists of multiplying a parametric class of covariance functions with a (normalized) class of covariance functions having additionally compact support. Here, the compact support is not estimated from data but fixed in such a way as to guarantee a desired level of sparseness for the covariance matrix (obtained from the tapered model applied to the observations). In turn, sparsity prompts a much faster inversion of the covariance matrix. This is crucial to implement both estimation and prediction. The impact of tapering on prediction has been celebrated in [10]. Tapering for maximum likelihood estimation under fixed domain asymptotics has been studied by [11]. A review of tapering for estimation and prediction is provided by [12].

1.1 | Context and challenges

A discussion about different supports in concert with the implications on the related covariance matrix is provided in Section 2. We anticipate that fully supported models have no zeros in the related covariance matrix, while dynamically supported models have zeros only in the diagonal blocks of the space-time covariance matrix. The models with compact support allow for sparsity, which means that the related covariance matrix can have many zero entries.

Our interest in this paper is in understanding whether space-time models having either full, dynamical, or compact support might be compatible. Whenever this happens, there are precise consequences in terms of kriging efficiency under misspecified covariance models as well as maximum likelihood estimation under fixed domain asymptotics.

Faouzi et al. [13] have started such a comparison between a class having full support [14] against a class with dynamical supports [15]. Their results prove that compatibility is possible under suitable parametric restrictions.

Unfortunately, there are no space-time models with compact support. An apparently simple solution is to taper a dynamically supported model with a temporal covariance model having compact support. While the validity of such construction is guaranteed by classical arguments based on properties of positive definite functions, the related spectral density—required to show compatibility conditions—is challenging. The solution to this problem requires Fourier arguments in concert with involved computations, and we defer this part to a technical Supplementary Material (see Section A therein) to avoid mathematical obfuscation. The paper centers on the conceptual exposition of the compatibility results, which are, in turn, based on the theory for equivalence of Gaussian measures.

1.2 | Contribution

This paper provides the following contributions. We engage in the analytic closed form associated with the tapered spectral density. We then provide the asymptotic properties of the tapered spectrum. This opens the study of the parametric conditions ensuring compatibility of the involved classes of covariance functions.

The plan of the paper is the following. Section 2 contains a succinct and simplified mathematical background. Section 3 provides the proposal of this paper. Section 4 inspects conditions for space-time covariance compatibility. The supporting information (SI throughout) is rich and contains an extended background, technical lemmas, technical results, and proofs.

2 | BACKGROUND MATERIAL

2.1 | Space-time covariance functions

For the remainder of the paper, we let d be a positive integer. Throughout, $\mathcal{D}$ is a bounded set in $\mathbb{R}^d$ and mimics the role of the spatial domain. Here, $\mathcal{T}$ is a subset of the real line and plays the role of time. We denote by $Z = \{Z(ß, t), (ß, t) \in D \times \mathcal{T}\}$ a zero mean Gaussian random field with index set on $D \times \mathcal{T}$, with stationary covariance function $C : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$. Covariance functions are positive definite: for every arbitrary collection $\{(\pi_i, t_l)\}$, $i = 1, \dots, N$, $l = 1, \dots, M$ and for every arbitrary finite system $\{c_{il}\}$ of real constants, we have

$$\sum_{i,j=1}^{N} \sum_{l,m=1}^{M} c_{il} C\left(\pi_i - \pi_j, t_l - t_m\right) c_{jm} \geq 0.$$

The paper works under the following assumptions.

Condition 1.

1. The covariance functions are spatially isotropic and temporally symmetric.
2. The covariance functions are continuous and absolutely integrable.

The implication of Condition 1.1 is that

$$C(\boldsymbol{h}, u) = K(\|\boldsymbol{h}\|, |u|), \qquad \boldsymbol{h} \in \mathbb{R}^d,\ u \in \mathbb{R},$$

for some suitable function K that guarantees positive definiteness. Observe that, by Bochner's theorem [16], C (or equivalently, K) is the uniquely determined Fourier transform of a positive and bounded measure F. In view of Condition 1.2, F is absolutely continuous, and we call its derivative a space-time spectral density and use the notation $\hat{C}$, or $\hat{K}$, whenever there is no confusion. Section A (SI) contains mathematical details about $\hat{K}$, which is radial in the first argument and symmetric in the second.

For a space-time covariance function K, the margins $K_{\mathrm{S}}(\cdot) = K(\cdot, 0)$ and $K_{\mathrm{T}}(\cdot) = K(0, \cdot)$ are called spatial and temporal covariance functions, respectively. Spatial and temporal spectral densities are described in Section A (SI).

A space-time covariance function is called separable if $K(h, u) = K_{\mathrm{S}}(h) K_{\mathrm{T}}(u)$, where K_{S} and K_{T} are spatial and temporal covariance functions, respectively. In all the other cases, K is called nonseparable. The function K is termed compactly supported if a pair (h_o, t_o) of positive real numbers exists such that $K(h, t) = 0$ whenever $h \geq h_o$ and $t \geq t_o$. Consequently, the pair (h_o, t_o) is called space-time compact support, and h_o and t_o are spatial and temporal compact supports.

The mapping K is called dynamically supported if there exists a function $\psi : [0, \infty) \to \mathbb{R}_+$ such that for every fixed temporal lag t_o, the spatial margin $K(h, t_o)$ is compactly supported with radius $\psi(t_o)$. In all the other cases, the mapping K will be called globally supported. The substantial differences between the three cases are the following:

1. Let $\{(\pi_i, t_l)\}$, $i = 1, \dots, N$, $l = 1, \dots, M$ be an $N \times M$-dimensional collection of space-time points. Let $\boldsymbol{\Sigma}$ be the square $N \times M$ dimensional matrix with elements $\Sigma_{il,jm} = \text{cov}\left(Z(\pi_i, t_l), Z(\pi_j, t_m)\right)$. If K is globally supported, then $\boldsymbol{\Sigma}$ is full, in the sense that $\boldsymbol{\Sigma}$ has no zeros.
2. If K is dynamically supported, then $\boldsymbol{\Sigma}$ is blockwise sparse, in the sense that the diagonal blocks in $\boldsymbol{\Sigma}$ will have as many zeros as soon as the distance between any pair of spatial points is greater than $\psi(t_o)$.
3. If K is compactly supported, then $\boldsymbol{\Sigma}$ will be sparse, and it can be chosen to be as sparse as desired depending on the space-time compact support (h_o, t_o).

2.2 | Spatial and temporal margins

The paper centers on two spatial isotropic covariance models, C_S. The first model is termed Matérn class and is globally supported. The second is termed Generalized Wendland and is compactly supported. The two models have similar behavior in terms of differentiability at the origin, which makes them compatible in terms of spatial kriging prediction under fixed domain asymptotics [5].

The Matérn class of functions, $\mathcal{M}(\cdot; \alpha, \nu)$, is defined through

$$\mathcal{M}(r; \alpha, \nu) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{r}{\alpha}\right)^{\nu} \mathcal{K}_{\nu}\left(\frac{r}{\alpha}\right), \qquad r \geq 0, \tag{2.1}$$

with $\alpha > 0$ a scaling parameter and ν measuring smoothness. Here, $\mathcal{K}_\nu$ is the MacDonald function [17]. The Matérn class is a parametric class of isotropic parts of covariance functions that allow for continuously indexing the mean squared differentiability and the fractal dimension of the associated Gaussian random field. The isotropic spectral density, $\widehat{\mathcal{M}}$, associated with the Matérn model is reported in Section A (SI).

The generalized Wendland class $\mathcal{GW}(\cdot; \beta, \mu, \kappa) : [0, \infty) \to \mathbb{R}$ is defined as [18, 19]

$$\mathcal{GW}(r; \beta, \mu, \kappa) := \begin{cases} \frac{1}{B(2\kappa, \mu+1)} \int_{r/\beta}^{1} u(u^2 - (r/\beta)^2)^{\kappa-1}(1-u)^{\mu} u, & 0 \leq r < \beta, \\ 0, & r \geq \beta, \end{cases} \tag{2.2}$$

where $\kappa \geq 0$, $\mu \geq (d+1)/2 + \kappa$ (such a condition is required to ensure positive definiteness in $\mathbb{R}^d$, see [20]) and where $\beta > 0$ is the compact support parameter, and B denotes the Beta function.

The $\mathcal{GW}$ model is compactly supported on a ball embedded in $\mathbb{R}^d$ with radius β. The parameter κ determines the smoothness at the origin, similar to the Matérn model. The parameter μ is a convergence parameter, where the following fact justifies this name. Let

$$\widetilde{\mathcal{GW}}(r; \beta, \mu, \kappa) = \mathcal{GW}\left(r; \beta\left(\frac{\Gamma(\mu + 2\kappa + 1)}{\Gamma(\mu)}\right)^{\frac{1}{1+2\kappa}}, \mu, \kappa\right), \qquad r \geq 0. \tag{2.3}$$

Arguments in [8] prove that

$$\lim_{\mu \to \infty} \widetilde{\mathcal{GW}}(r; \beta, \mu, \kappa) = \mathcal{M}(r; \beta, \kappa + 1/2), \ \kappa \geq 0,$$

with uniform convergence over the set $r \in (0, \infty)$. This result proves two facts: On the one hand, the Matérn model is a limit case of a rescaled version of the $\mathcal{GW}$ model. On the other hand, the $\mathcal{GW}$ is a very flexible model because it allows indexing differentiability in the same fashion as the Matérn model and additionally has a parameter (μ) that allows switching between compact and global supports.

As for the temporal margins, C_T, we shall center on the following three cases:

a. C_T belongs to the Matérn class, $\mathcal{M}$, as defined through (2.1);
b. C_T belongs to the Cauchy class, $\mathcal{C}$, defined as

$$\mathcal{C}(t; \xi, \delta, \gamma) = \left(1 + \left(\frac{t}{\xi}\right)^{\delta}\right)^{-\gamma}, \qquad t \geq 0, \tag{2.4}$$

where ξ is a the temporal scale, and where $\gamma > 0$ decides on the long memory of the process. The parameter $\delta \in (0, 2]$ decides on the local properties of the temporal trajectories in terms of fractal dimension.

c. C_{T} is the product of a Cauchy function, $\mathcal{C}$, with a special case of the $\mathcal{GW}$ class in (2.2) obtained for $\kappa = 0$.

3 | THE $\mathcal{DM}$, THE $\mathcal{DGW}$, AND THE $\mathcal{DGW}_{\mathbf{Tap}}$ FAMILIES OF SPACE-TIME COVARIANCE FUNCTIONS

This paper deals with three parametric classes of nonseparable covariance functions. The amount of parameters and the complexities of the algebraic form need some explanation, and Table 1 helps to do the job.

Some comments are in order. There is no free lunch: These three families have features and drawbacks. The $\mathcal{DM}$ family has a parameter, ϵ, that allows to continuously index nonseparability. The separable case is attained for $\epsilon = 0$. The spatial and temporal margins are both of Matérn type. Unfortunately, no algebraically closed forms are available except for the separable case. Another drawback is scalability: The associated covariance matrix $\boldsymbol{\Sigma}$ is full. On the other hand, both $\mathcal{DGW}$ and $\mathcal{DGW}_{\mathrm{Tap}}$ are available in algebraically closed forms. Their associated covariance matrices $\boldsymbol{\Sigma}$ are block diagonal sparse, and sparse, respectively. For the latter, any level of sparsity (the percentage of zeros in the matrix $\boldsymbol{\Sigma}$) can be achieved by fixing ξ ad hoc. Another relevant comment is that the temporal margin C_{T} is either nondifferentiable or infinitely differentiable under the $\mathcal{DGW}$ model, while is it always nondifferentiable at the origin for the $\mathcal{DGW}_{\mathrm{Tap}}$ model. A final comment is that both $\mathcal{DGW}$ and $\mathcal{DGW}_{\mathrm{Tap}}$ models have no parameter that allows switching from separability to nonseparability.

Our effort here goes in the direction of whether we can increase the sparsity of the covariance matrix $\boldsymbol{\Sigma}$ from $\mathcal{DM}$ to $\mathcal{DGW}_{\mathrm{Tap}}$ at little loss in terms of statistical accuracy. This is measured through the compatibility of these models under fixed domain asymptotics. We are now ready to formally define the three models:

1. *The space-time Matérn model*, denoted $\mathcal{DM}$.

   Define FT_{T} to be the Fourier transform with respect to the temporal component. Mathematical details are contained in the Section A (SI). Let $g_{\mathcal{M}}(\cdot, \cdot; \boldsymbol{\theta}) : [0, \infty)^2 \to \mathbb{R}$ be the function defined at (A.3) in Section A (SI). Specifically, $g_{\mathcal{M}}$ depends on the spatial distance in the first coordinate and on the time frequency in the second. The parameter vector contains the parameters described in the first row of Table 1. Hence, $\boldsymbol{\theta} = (\nu, \xi, \upsilon, \epsilon, \sigma)^{\top}$ with $\top$ denoting the transpose of a vector. Section A (SI) explains how the partial Fourier transform with respect to time, $F_{\mathrm{T|S}}$ as defined through (A.2) in SI, provides a covariance function

$$\mathcal{DM}(r, t; \boldsymbol{\theta}) = \mathrm{FT}_{\mathrm{T|S}}\left[g_{\mathcal{M}}(\cdot, \tau; \boldsymbol{\theta})\right](r, t), \qquad (r, t) \in [0, \infty), \tag{3.1}$$

   that depends on the spatial and temporal distances (r, t). The fact that it is a valid space-time covariance function is proved by [21]. The spatial and temporal margins of this covariance function are both of the Matérn types.

   It is worth remarking that the function $g_{\boldsymbol{\theta}}$ has a factor $\ell(\boldsymbol{\theta})$, having a mathematically involved expression. To avoid mathematical obfuscation, we report its expression in Equation (A.4) in Section A (SI).

2. *Dynamical generalized Wendland model*, denoted $\mathcal{DGW}$. Using the same notation as much as in [15], we now introduce the $\mathcal{DGW}$ class of space-time covariance functions, defined as

TABLE 1 The three space-time models used in this paper.

Family	C_{S}	C_{T}	$\boldsymbol{\Sigma}$	$\mathbf{Scale}_{\mathrm{S}}$	$\mathbf{Scale}_{\mathrm{T}}$	$\mathbf{Smooth}_{\mathrm{S}}$	$\mathbf{Smooth}_{\mathrm{T}}$	Nonsep
$\mathcal{DM}$ (3.1)	$\mathcal{M}$ (2.1)	$\mathcal{M}$ (2.1)	Full	$\zeta > 0$	$\upsilon > 0$	$\nu > 0$	$\nu > 0$	$\epsilon \in [0, 1]$
$\mathcal{DGW}$ (3.2)	$\mathcal{GW}$ (2.2)	$\mathcal{C}$ (2.4)	Bloc Diag sparse	$\beta > 0$	$\xi > 0$	$\kappa \geq 0$	$\delta \in (0, 2)$	None
$\mathcal{DGW}_{\mathrm{Tap}}$ (3.2)	$\mathcal{GW}$ (2.2)	$\mathcal{C} \times \mathcal{GW}$	Sparse	$\beta > 0$	$\xi > 0$	$\kappa \geq 0$	0	None

Note: For all of them, the parameter $\sigma^2 > 0$ denotes the variance of the associated process. The fourth column describes the sparsity of the associated covariance matrix $\boldsymbol{\Sigma}$. Here, $\mathbf{Scale}_{\mathrm{S}}$ (resp. $\mathbf{Scale}_{\mathrm{T}}$) stands for spatial or temporal scales (or compact supports), respectively. Also, $\mathbf{Smooth}_{\mathrm{S}}$ ($\mathbf{Smooth}_{\mathrm{T}}$) stands for the smoothness of the spatial (temporal) margin, respectively. Finally, **Nonsep** describes a nonseparability parameter. We note that the second and third rows have two additional parameters: We call the parameter μ a convergence parameter. The range of μ is defined in Condition 2. The parameter $\gamma > 0$ induces long or short memory (Hurst effect), and its role is not major in this paper. The parameter γ is normally fixed and not estimated.

$$\mathcal{DGW}(r,t;\chi)=\sigma^2\mathcal{C}\left(t;\xi,\delta,\gamma\right)\mathcal{GW}\left(\frac{r}{\mathcal{C}\left(t;\xi,\delta,1\right)};\beta,\mu,\kappa\right),\qquad r,t\geq 0, \tag{3.2}$$

The meaning of each parameter is explained in the second row of Table 1. Hence, the resulting parameter χ amounts to $\chi=(\sigma^2,\beta,\mu,\kappa,\xi,\delta,\gamma)^\top$.

We note that the parametric conditions ensuring the validity of the $\mathcal{DGW}$ model are the ones in Table 1. Yet, some technical additional conditions are required, and we formalize them below.

Condition 2. Let $\eta=(d+1)/2+\kappa$. For the remainder of the paper, we always work under the parametric condition

$$\mu>\max\left((d+5)/2+\kappa+\varsigma^*,\eta+(d+1)/2\right)\qquad\text{and}\qquad\gamma\geq\max\left((d+3)/2+2\kappa,2\kappa+3\right). \tag{3.3}$$

The strictly positive constant ς^* plays no role in this paper, and we omit its specification while referring to Table 2 and Section 3.4 in [15] for details.

3. Tapered dynamical generalized Wendland model, denoted $\mathcal{DGW}_{\text{Tap}}$. This model is obtained by temporal tapering of the $\mathcal{DGW}$ model. The resulting equation is

$$\mathcal{DGW}_{\text{Tap}}(r,t;\chi)=\mathcal{DGW}(r,t;\chi)\times\mathcal{GW}(t;\xi,4,0),\qquad r,t\geq 0. \tag{3.4}$$

We note that

$$\mathcal{GW}(t;\xi,4,0)=\left(1-\frac{t}{\xi}\right)_+^4,\qquad t\geq 0.$$

The exponent 4 is the minimal exponent that guarantees our theoretical results hold. The results can be shown for any exponent that is greater or equal to 4, at the expense of additional notation. We keep things simple here.

Clearly, the construction (3.4) is positive definite as it is the product of two positive definite functions. The resulting covariance matrix, $\boldsymbol{\Sigma}_{\text{Tap}}$, is the Kronecker product of the matrix $\boldsymbol{\Sigma}$, associated with $\mathcal{DGW}$, with the temporal matrix $\boldsymbol{\Sigma}_{\text{Tap}}$, associated with $\mathcal{GW}(\cdot;\xi,4,0)$. On the other hand, the spectral density associated with $\mathcal{DGW}_{\text{Tap}}$ has a very complicated expression, being the convolution (over time) of the space-time spectral density associated with $\mathcal{DGW}$ with temporal spectral density associated with $\mathcal{GW}(\cdot;\xi,4,0)$.

Another relevant comment is the following: let

$$\widetilde{\mathcal{DGW}}(r,t;\chi)=\sigma^2\mathcal{C}\left(t;\xi,\delta,\gamma\right)\widetilde{\mathcal{GW}}\left(\frac{r}{\mathcal{C}\left(t;\xi,\delta,\gamma\right)};\beta,\mu,\kappa\right),\qquad r,t\geq 0.$$

Then, convergence arguments as much as in [8] prove that

$$\lim_{\mu\to\infty}\widetilde{\mathcal{DGW}}(r,t;\chi)=\sigma^2\mathcal{C}\left(t;\xi,\delta,\gamma\right)\mathcal{M}\left(\frac{r}{\mathcal{C}\left(t;\xi,\delta,\gamma\right)};\beta,\kappa+1/2\right),\qquad r,t\geq 0,$$

uniformly for all r,t. The right-hand side of the above equation responds to the Gneiting class of covariance functions [22], justifying calling μ a convergence parameter (Table 1).

We note that the parametric ranges ensuring positive definiteness of $\mathcal{DGW}$ and $\mathcal{DGW}_{\text{Tap}}$ are reported in Section A (SI).

4 | COMPATIBILITY THEOREMS FOR SPACE-TIME COVARIANCE FUNCTIONS

We illustrate compatibility here, and for a formal introduction, the reader is referred to Section A in SI.

Equivalence and orthogonality of probability measures are useful tools when assessing the asymptotic properties of both prediction and estimation for stochastic processes. 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 $Z = \{Z(ß, t), (ß, t) \in \mathbb{R}^d \times \mathbb{R}\}$, to define previous concepts, we restrict the event A to the σ-algebra generated by $\{Z(ß, t), (ß, t) \in D \times \mathcal{T}\}$ where $D \times \mathcal{T} \subset \mathbb{R}^d \times \mathbb{R}$. We emphasize this restriction by saying that the two measures are equivalent on the paths of Z.

Gaussian measures are completely characterized by their mean and covariance functions. We write $P(C)$ for a Gaussian measure with zero mean and covariance function C. It is well-known that two Gaussian measures are either equivalent or orthogonal on the paths of Z [23]. Conditions for equivalence of Gaussian measures are explained in Section A (SI).

The information above allows for providing the following formal statement:

Remark 1. Two space-time covariance functions are called compatible if their induced Gaussian measures are equivalent on the paths of $Z(ß, t), (ß, t) \in D \times \mathcal{T}$.

Some notation is now necessary. We define $\chi_i = (\sigma_i^2, \beta_i, \mu, \kappa_i, \xi_i, \delta_i, \gamma_i)^\top$, $i = 0, 1$.

Throughout, we always suppose that Condition 3.3 holds. We avoid stating this explicitly in every formal statement for the sake of simplicity.

Theorem 1. *Let $\eta_i = (d+1)/2 + \kappa_i$, $i = 0, 1$ and $\gamma_0 = \gamma_1 = \gamma$. Let*

$$a_{1,i} = 5\sqrt{\frac{2}{\pi}}\frac{\Gamma(\mu + 2\eta_i)}{\Gamma(\mu)} \qquad \text{and} \qquad a_{2,i} = a_{1,i}\frac{\sqrt{\pi}\,(\gamma + d - 2\eta_i)\Gamma(\delta_i/2 + 1)}{5\Gamma(-\delta_i/2)\Gamma(6)^2},$$

with

$$L_i = \frac{\Gamma(\kappa_i + (d+1)/2)\Gamma(2\kappa_i + \mu + 1)}{\pi^{d/2}\Gamma(\kappa_i + 1/2)\Gamma(\mu + 2\eta_i)}.$$

Consider the covariance models: $\mathcal{DGW}_{Tap}(\cdot, \cdot; \chi_0)$ and $\mathcal{DGW}_{Tap}(\cdot, \cdot; \chi_1)$. For a given $\kappa_1 = \kappa_0$ and $\delta_i \in \{1, 2\}$, and for any bounded infinite set $D \subset \mathbb{R}^d$, $d = 1, 2$, the following compatibility assertions are true:

1. *For $\delta_0 = \delta_1 = 1$ and $\kappa_0 = \kappa_1$, $\mathcal{DGW}_{Tap}(\cdot, \cdot; \chi_0)$ and $\mathcal{DGW}_{Tap}(\cdot, \cdot; \chi_1)$ are compatible if and only if*

$$L_1\left(\frac{4a_{2,1}}{\xi_1^2} + \frac{a_{1,1}}{\xi_1}\right)\frac{\sigma_1^2}{\beta_1^{2\kappa_1+1}} = L_0\left(\frac{4a_{2,0}}{\xi_0^2} + \frac{a_{1,0}}{\xi_0}\right)\frac{\sigma_0^2}{\beta_0^{2\kappa_0+1}}.$$

2. *For $\delta_0 = \delta_1 = 2$ and $\kappa_0 = \kappa_1$, $\mathcal{DGW}_{Tap}(\cdot, \cdot; \chi_0)$ and $\mathcal{DGW}_{Tap}(\cdot, \cdot; \chi_1)$ are compatible if and only if*

$$\sigma_1^2 L_1\frac{\beta_1^{d-2\eta_1}}{\xi_1} = \sigma_0^2 L_0\frac{\beta_0^{d-2\eta_0}}{\xi_0}.$$

The following result relates to the compatibility of the $\mathcal{DGW}$ model with parameter vector χ_0 with the corresponding tapered version $\mathcal{DGW}_{\text{Tap}}$ and a misspecified parameter vector χ_1. All the constants appearing in the result below are available in theorems given in Theorem B.1 in SI.

Theorem 4.2. *For given $\delta_1 = 1$, $\gamma_0 > 2\kappa_0 + 3$ and $\gamma_1 > 2\kappa_1 + 3$. Consider the models $\mathcal{DGW}_{Tap}(\cdot, \cdot; \chi_1)$ and $\mathcal{DGW}(\cdot, \cdot; \chi_0)$. Let*

$$\varrho_{\gamma_0,\eta} = \frac{(d + \gamma_0 - 2\eta)\Gamma(\delta_0 + 1)\sin(\frac{\pi\delta_0}{2})}{\xi_0^{\delta_0}\pi}.$$

Then, the compatibility of the two models for any bounded infinite set $D \times \mathcal{T} \subset \mathbb{R}^d \times \mathbb{R}$, $d = 1, 2$, holds if $\delta_0 = 1 + 2(\kappa_1 - \kappa_0)$ and

$$\varrho_{\gamma_0,\eta}\sqrt{\frac{\pi}{2}}\frac{a_{1,0}\ L_0}{5}\ \frac{\sigma_0^2}{\beta_0^{-(1+2\kappa_0)}} = \frac{a_{1,1}L_1}{\sqrt{2\pi}}\ \frac{\sigma_1^2}{\beta_1^{-(1+2\kappa_1)}},$$

with all the relevant constants as being defined through Theorem 1.

The final result relates the compatibility of the tapered version of the $\mathcal{DGW}$ model to the space-time Matérn model $\mathcal{DM}$. Some additional notation is necessary. We call $\boldsymbol{\theta}_0 = (\nu, \zeta, \upsilon, \epsilon, \sigma_0)^\top$ from the parameterization related to the $\mathcal{DM}$ model.

Theorem 4.3. *For given $\nu > (d+1)/2$ and $\epsilon \in (0, 1]$, consider the models $\mathcal{DM}(\cdot, \cdot; \boldsymbol{\theta}_0)$ and $\mathcal{DGW}_{Tap}(\cdot, \cdot; \chi)$). Let $\ell(\boldsymbol{\theta}_0)$ be as defined at (A.4) in Section A (SI).*

If

$$\left[\frac{\sigma^2 L}{\sqrt{2\pi}\beta^{1+2\kappa}}\left(\frac{4a_2}{\xi^2}+\frac{a_1}{\xi_1}\right)\right]\mathbf{1}_{\delta=1}=\ell(\boldsymbol{\theta}_0)\epsilon^{-2\nu},$$

with $2\nu = \eta + 1$, *and* $\kappa > d - 1/2$ *then, for any bounded infinite set* $D \times \mathcal{T} \subset \mathbb{R}^d \times \mathbb{R}$, $d = 1, 2$, *the two models are compatible.*

These compatibility theorems have severe implications for both ML estimation and kriging prediction under fixed domain asymptotics.

Following the arguments in [4], an immediate consequence of Theorem 1 is that for fixed δ, μ, and γ, the parameters σ^2, β, κ, and ξ cannot be estimated consistently. Instead, for $\boldsymbol{\tau} = (\beta, \xi)^\top$, the microergodic parameter

$$\left(\frac{4a_2}{\xi^2}+\frac{a_1}{\xi_1}\right)\frac{\sigma^2(\boldsymbol{\tau})L}{\beta^{2\kappa+1}}$$

is consistently estimable. The proof comes straight by using the same arguments as in [13], and is hence omitted. Hence, we proved that the maximum likelihood estimator for the single parameters cannot be consistent under conditions for equivalence. On the other hand, we proved that the microergodic parameter can be consistently estimated under maximum likelihood, and hence this should be the quantity to focus on to evaluate the performance of maximum likelihood estimators for these classes of covariance functions.

The second implication is in terms of kriging prediction. When working under fixed domain asymptotics, the kriging variance under the misspecified model will tend to be equal to the kriging variance under the correct model. Proofs are obtained by mimicking [5] and hence excluded from this paper. This is a very relevant property that allows the practitioner to be ensured that, even if selecting a wrong kernel, best linear unbiased prediction is asymptotically unaffected under the conditions provided in this paper.

Several parameters in the $\mathcal{DGW}_{\text{Tap}}$ model framework affect the sparsity of the resulting covariance matrix, as shown in Table 1. Together with the compatibility theorems, this enables handling very large datasets, despite the nontrivial form of the generalized Wendland class. (As demonstrated in [24], fast implementations of the covariance functions exist.) Similar to the spatial setting, the best results are achieved for both estimation and prediction settings when balancing the smoothness and decay of the misspecified covariance function. Still, there is a nontrivial interplay of the parameters and the behavior at the origin (see https://shiny.math.uzh.ch/user/furrer/shinyas/GenWendTap/). The optimal parameter choices strongly depend on the specific dataset, making it unwise to provide recommendations for the choices thereof.

Most of the stochastic expansions that are typically used in the framework of equivalence of Gaussian measures are based on univariate orthogonal polynomials. It would be of great interest to find other classes of more general expansions. The concepts and the related formalism of the theory of bi-orthogonal polynomials have been discussed to a very limited extent. By starting from the orthogonality properties satisfied from the ordinary and generalized Hermite polynomials, [25] has derived a very flexible family in concert with related generalizations.

This paper has addressed a crucial gap in the infill-asymptotics framework by providing compatibility theorems for arbitrary compact support in both space and time.

5 | CONCLUSION

Our effort has allowed us to quantify statistical accuracy in terms of estimation and prediction when tapering the temporal part of a covariance function that is spatially dynamically supported. Arguments in [26] suggest that better performance in terms of prediction accuracy might be achieved by using temporal tapers that are smoother at the origin. Yet, the convoluted space-time spectral density becomes analytically intractable even in the case of a temporal taper that is once differentiable at the origin. Since this paper is more oriented to analytical solutions rather than heuristics, we prefer to leave the issue of improving temporal differentiability as an open problem.

Future developments include multivariate random fields, for which the covariance is a matrix-valued function, and functional processes, for which the theory of tapering for covariance operators is still elusive.

Recent applications in statistics and machine learning include processes that are defined over manifolds (for instance, the sphere) cross time. We are not aware of any theoretical study regarding tapering for this case, and a big effort is needed, starting with the theory on the equivalence of Gaussian measures for processes defined over spheres cross time.

Another direction of research that is unexplored up to now is related to stochastic processes that are continuously indexed over networks. While the case of a linear network is more straightforward (because of classical embedding theorems), that of generalized networks requires special attention. In fact, no spectral representation is available for this case. Hence, new techniques should be used to provide conditions for equivalence of Gaussian measures. Certainly this will be a very important line of research for the future.

AUTHOR CONTRIBUTIONS

Tarik Faouzi: Conceptualization; methodology; formal analysis; writing—original draft. **Reinhard Furrer:** con-Ceptualization; methodology; formal analysis; writing—original draft; validation. **Emilio Porcu:** Conceptualization; investigation; methodology; validation; formal analysis; writing—original draft.

ACKNOWLEDGEMENTS

Tarik Faouzi is supported by FONDECYT grant 11200749, Chile. Reinhard Furrer was supported by the Swiss National Science Foundation SNSF-175529. Emilio Porcu is supported by FSU grant number FSU-2021-016.

ORCID

Tarik Faouzi https://orcid.org/0000-0003-3497-050X
Emilio Porcu https://orcid.org/0000-0001-9934-4409

REFERENCES

1. M. J. Heaton, A. Datta, A. O. Finley, R. Furrer, J. Guinness, R. Guhaniyogi, F. Gerber, R. B. Gramacy, D. Hammerling, M. Katzfuss, F. Lindgren, D. W. Nychka, F. Sun, and A. Zammit-Mangion, *A case study competition among methods for analyzing large spatial data*, J. Agric. Biol. Environ. Stat. **24** (2019), no. 3, 398–425.
2. M. L. Stein, *Uniform asymptotic optimality of linear predictions of a random field using an incorrect second order structure*, The Ann. Stat. **19** (1990), 850–872.
3. M. L. Stein, *Statistical interpolation of spatial data: some theory for kriging*, Springer, New York, 1999.
4. H. Zhang, *Inconsistent estimation and asymptotically equivalent interpolations in model-based geostatistics*, J. Am. Stat. Assoc. **99** (2004), 250–261.
5. M. Bevilacqua, T. Faouzi, R. Furrer, and E. Porcu, *Estimation and prediction using generalized Wendland functions under fixed domain asymptotics*, Ann. Stat. **47** (2019), 828–856.
6. M. L. Stein, *When does the screening effect not hold?* Spatial Stat. **11** (2015), 65–80.
7. E. Porcu, V. Zastavnyi, M. Bevilacqua, and X. Emery, *Stein hypothesis and screening effect for covariances with compact support*, Electron. J. Stat. **14** (2020), no. 2, 2510–2528.
8. M. Bevilacqua, C. Caamaño-Carrillo, and E. Porcu, *Unifying compactly supported and Matérn covariance functions in spatial statistics*, J. Multivar. Anal. **189** (2022), 104949.
9. E. Porcu, M. Bevilacqua, R. Schaback, and C. J. Oates, *The mat\'ern model: a journey through statistics, numerical analysis and machine learning*, 2023. arXiv preprint arXiv:2303.02759.
10. R. Furrer, M. G. Genton, and D. Nychka, *Covariance tapering for interpolation of large spatial datasets*, J. Comput. Graph. Stat. **15** (2006), 502–523.
11. J. Du, H. Zhang, and V. S. Mandrekar, *Fixed-domain asymptotic properties of tapered maximum likelihood estimators*, the Ann. Stat. **37** (2009), no. 6A, 3330–3361.
12. H. Zhang and J. Du, *Covariance tapering in spatial statistics*, Posit. Definite Funct.: From Schoenberg to Space-Time Chall. (2008), 181–196.
13. T. Faouzi, E. Porcu, and M. Bevilacqua, *Space-time estimation and prediction under fixed-domain asymptotics with compactly supported covariance functions*, Stat. Sin. **32** (2022), 1–17.
14. R. H. L. Ip and W. K. Li, *On some matérn covariance functions for spatio-temporal random fields*, Stat. Sin. **27** (2017), 805–822.
15. E. Porcu, M. Bevilacqua, and M. G. Genton, *Nonseparable, space-time covariance functions with dynamical compact supports*, Stat. Sin. **30** (2020), no. 2, 719–739.
16. S. Bochner, *Harmonic analysis and the theory of probability*, California Monographs in mathematical sciences, University of California Press, 1955.
17. I. S. Gradshteyn and I. M. Ryzhik, *Tables of integrals, series, and products*, seventh, Academic Press, Amsterdam, 2007.
18. T. Gneiting, *Compactly supported correlation functions*, J. Multivar. Anal. **83** (2002), 493–508.
19. V. P. Zastavnyi, *Positive-definite radial functions and splines*, Doklady Math. **66** (2002), 213–216.
20. H. Wendland, *Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree*, Adv. Comput. Math. **4** (1995), 389–396.

21. M. Fuentes, L. Chen, and J. M. Davis, *A class of nonseparable and nonstationary spatial temporal covariance functions*, Environ.: The Off. J. Int. Environ. Soc. **19** (2008), no. 5, 487–507.
22. T. Gneiting, *Nonseparable, stationary covariance functions for space–time data*, J. Am. Stat. Assoc. **97** (2002), no. 458, 590–600.
23. I. A. Ibragimov and Y. A. Rozanov, *Gaussian random processes*, Springer, New York, 1978.
24. T. C. Fischer, R. Furrer, and J. Stocker, *Generalizedwendland: fully parameterized generalized Wendland covariance function*, 2022. R package version 0.5-2. https://CRAN.R-project.org/package=GeneralizedWendland
25. C. Cesarano, *A note on bi-orthogonal polynomials and functions*, Fluids **5** (2020), no. 3, 105.
26. R. Furrer, M. G. Genton, and D. Nychka, *Covariance tapering for interpolation of large spatial datasets*, J. Comput. Graph. Stat. **15** (2006), no. 3, 502–523.
27. E. Porcu, P. Gregori, and J. Mateu, *Nonseparable stationary anisotropic space–time covariance functions*, Stoch. Environ. Res. Risk Assess. **21** (2006), no. 2, 113–122.
28. N. Cressie and H. C. Huang, *Classes of nonseparable, spatio-temporal stationary covariance functions*, J. Am. Statist. Ass. **94** (1999), no. 448, 1330–1340.
29. T. Gneiting, *Nonseparable, stationary covariance functions for space-time data*, J. Am. Statist. Ass. **97** (2002), 590–600.
30. V. P. Zastavnyi and E. Porcu, *Characterization theorems for the Gneiting class of space–time covariance functions*, Bernoulli **17** (2011), no. 1, 456–465.
31. A. V. Skorokhod and M. I. Yadrenko, *On absolute continuity of measures corresponding to homogeneous Gaussian fields*, Theory Probab. Appl. **18** (1973), 27–40.
32. M. L. Stein, *Equivalence of Gaussian measures for some nonstationary random fields*, J. Stat. Plan. Infer. **123** (2004), 1–11.
33. M. Abramowitz and I. A. Stegun Eds., *Handbook of mathematical functions Edited by* M. Abramowitz and I. A. Stegun, Dover, New York, 1970.
34. J. L. González-Santander, *Closed-form expressions for derivatives of bessel functions with respect to the order*, J. Math. Anal. Appl. **466** (2018), no. 1, 1060–1081.
35. V. Nijimbere, *Evaluation of some non-elementary integrals involving sine, cosine, exponential and logarithmic integrals: part I*, 2017. arXiv preprint arXiv:1703.01907.
36. V. P. Zastavnyi, *On some properties of Buhmann functions*, Ukrainian Math. J. **58** (2006), no. 8, 1184–1208.
37. L. A. Borisov and R. P. Horja, *Mellin–Barnes integrals as Fourier–Mukai transforms*, Adv. Math. **207** (2006), no. 2, 876–927.

How to cite this article: T. Faouzi, R. Furrer, and E. Porcu, *Compatibility of space-time kernels with full, dynamical, or compact support*, Math. Meth. Appl. Sci. (2024),1–10, DOI 10.1002/mma.10379.