Inference for Gaussian Processes with Matérn Covariogram on Compact Riemannian Manifolds

Didong Li   DIDONGLI@UNC.EDU
Department of Biostatistics
University of North Carolina at Chapel Hill
Chapel Hill, NC 27599, USA

Wenpin Tang   WT2319@COLUMBIA.EDU
Department of Industrial Engineering and Operations Research,
Columbia University
New York, NY 10027, USA

Sudipto Banerjee   SUDIPTO@UCLA.EDU
Department of Biostatistics
University of California, Los Angeles
Los Angeles, CA 90095 USA

Editor: Marc Peter Deisenroth

Abstract

Gaussian processes are widely employed as versatile modelling and predictive tools in spatial statistics, functional data analysis, computer modelling and diverse applications of machine learning. They have been widely studied over Euclidean spaces, where they are specified using covariance functions or covariograms for modelling complex dependencies. There is a growing literature on Gaussian processes over Riemannian manifolds in order to develop richer and more flexible inferential frameworks for non-Euclidean data. While numerical approximations through graph representations have been well studied for the Matérn covariogram and heat kernel, the behaviour of asymptotic inference on the parameters of the covariogram has received relatively scant attention. We focus on asymptotic behaviour for Gaussian processes constructed over compact Riemannian manifolds. Building upon a recently introduced Matérn covariogram on a compact Riemannian manifold, we employ formal notions and conditions for the equivalence of two Matérn Gaussian random measures on compact manifolds to derive the parameter that is identifiable, also known as the microergodic parameter, and formally establish the consistency of the maximum likelihood estimate and the asymptotic optimality of the best linear unbiased predictor. The circle is studied as a specific example of compact Riemannian manifolds with numerical experiments to illustrate and corroborate the theory.

Keywords: Equivalence of Gaussian measures, Identifiability and consistency, Laplace–Beltrami operator, Microergodic parameters.

1. Introduction

Gaussian processes are pervasive in spatial statistics, functional data analysis, computer modelling and machine learning applications because of the flexibility and richness they allow in modelling complex dependencies (Rasmussen and Williams, 2006; Stein, 1999;

Gelfand et al., 2010; Cressie and Wikle, 2011; Banerjee et al., 2015). For example, in spatial statistics Gaussian processes are widely used to model spatial dependencies in geostatistical models and perform spatial prediction or interpolation ("kriging") (Matheron, 1963). In non-parametric regression models Gaussian processes are used to model unknown functions and, specifically in Bayesian contexts, act as priors over functions (Ghosal and van der Vaart, 2017). A typical modelling framework assumes $y(x) = \mu(x) + Z(x) + \epsilon(x)$ for inputs x (e.g., spatial coordinates; functional inputs) over a domain $\mathcal{D}$, where $y(x)$ is a dependent variable of interest, $\mu(x)$ is a mean function, $Z(x)$ is a zero-mean Gaussian process and $\epsilon(x)$ is a noise process[1]. These frameworks can also be adapted to deal with discrete outcomes and applied to classification problems (Neal, 1999). Gaussian processes are also being increasingly employed in deep learning and reinforcement learning (Damianou and Lawrence, 2013; Deisenroth et al., 2013). The current manuscript focuses upon inferential properties of $Z(x)$ when $\mathcal{D}$ is not necessarily Euclidean but a compact Riemannian manifold.

A Gaussian process is determined by its covariogram, also known as the covariance function. In Euclidean space, the Matérn covariogram (Matérn, 1986) is especially popular in spatial statistics and machine learning (see, e.g., Stein, 1999, for an extensive discussion on the theoretical properties of the Matérn covariogram). A key attraction of the Matérn covariogram is the availability of a smoothness parameter for the process. Several simpler covariograms, such as the exponential, arise as special cases of the Matérn.

This article is motivated by the emergence of non-Euclidean data, especially manifold data, in a variety of scientific fields over the last decade. As a consequence, inference for Gaussian processes on manifolds have been attracting attention in spatial statistics and machine learning in settings where the data generating process is more appropriately modelled over non-Euclidean spaces. Taking climate science as an example, geographic data involving geopotential height, temperature and humidity are measured at global scales and are more appropriately treated as (partial) realisations of a spatial process over a sphere (see, e.g., Banerjee, 2005; Jun and Stein, 2008; Jeong and Jun, 2015a). Data arising over domains with irregular shapes or examples in biomedical imaging where the domain is a three-dimensional shape of an organ comprise other examples where inference for Gaussian processes over manifolds will be relevant (see, e.g., Gao et al., 2019, and references therein). Motivated by isotropic covariograms in Euclidean space, it is natural to replace Euclidean distance by an appropriate geodesic distance to define a "Matérn" covariogram on Riemannian manifolds. However, this formal generalisation is not valid for the squared exponential covariogram, or Matérn with $\nu = \infty$ (Feragen et al., 2015), unless the manifold is flat. For Matérn with $\nu \in (1/2, \infty)$, this naive generalisation is not even valid on the sphere (Gneiting, 2013). Recently, valid covariograms for smooth Gaussian processes on general Riemannian manifolds have been constructed based upon heat equations, Brownian motion and diffusion models on manifolds (Castillo et al., 2014; Niu et al., 2019; Dunson et al., 2020). However, these covariograms lack flexibility, especially in terms of modelling smoothness.

Whittle (1963) proposed a new representation of GP by stochastic partial differential equations. Following this path, Lindgren et al. (2011) introduced a "Matérn" family on generic compact Riemannian manifolds with three parameters involved in the covariogram.

---

1. This article does not consider the noise process, which introduces additional difficulties that are beyond the scope of the current manusript; see Tang et al. (2021) for related developments in Euclidean space.

Since such Matérn covariograms involve the spectrum of the Laplace-Beltrami operator, a numerical approximation to the covariogram is needed for most nontrivial manifolds. There is a rich literature focusing on approximations to the covariogram using tools from harmonic analysis, graph Laplacians, and stochastic partial differential equations (Sanz-Alonso and Yang, 2022a,b). However, the study of statistical inference for the parameters in the Matérn covariogram remains relatively sparse.

In Euclidean domains $\mathbb{R}^d$ with $d \leq 3$, while not all parameters in the Matérn covariogram are consistently estimable within the paradigm of "fixed-domain" or "in-fill" asymptotic inference (see, e.g. Stein, 1999; Zhang, 2004), certain parameters, customarily referred to as microergodic parameters, which can identify Gaussian processes specified by Matérn covariograms are consistently estimable (see Section 2). Furthermore, the maximum likelihood estimator of the spatial variance under any misspecified decay parameter is consistently and asymptotically normally distributed (Du et al., 2009; Kaufman et al., 2008; Wang and Loh, 2011), while predictive inference is also asymptotically optimal using maximum likelihood estimators (Kaufman and Shaby, 2013). Recently, Bevilacqua et al. (2019) and Ma and Bhadra (2022) considered more general classes of covariance functions outside of the Matérn family and studied the consistency and asymptotic normality of the maximum likelihood estimator for the corresponding microergodic parameters.

Our current contribution develops asymptotic inference for a flexible and rich Matérn-type covariogram on compact Riemannian manifolds. We review the Matérn covariogram (Section 3.1) on general compact Riemannian manifolds from the perspective of stochastic partial differential equations with reasonably tractable covariograms and spectral densities (Borovitskiy et al., 2020). Our specific results emanate from a sufficient and necessary condition for the equivalence of two Gaussian random measures on compact Riemannian manifolds with Matérn or squared exponential covariograms (Section 3.2). We subsequently establish (Section 3.3) that for Gaussian measures with Matérn covariograms the smoothness parameter is identifiable, while the spatial variance and decay parameters are not identifiable when $d \leq 3$, where d is the dimension of the manifold. For $d \geq 4$, all three parameters are identifiable. For squared exponential covariograms on manifolds with arbitrary dimension, we show that both parameters are identifiable. Again, this problem is still open in Euclidean spaces. For Matérn covariograms on manifolds with $d \leq 3$, we formally establish that the maximum likelihood estimate of the spatial variance with a misspecified decay parameter is still consistent. Next, we turn to predictive inference (Section 3.4) and show that for any misspecified decay parameter in the Matérn covariogram, the best linear unbiased predictor derived from the maximum likelihood estimate is asymptotically optimal. Finally, for spheres with dimension less than 4, we explicitly study the Matérn covariogram, the microergodic parameter, the consistency of the maximum likelihood estimate and the optimality of the best linear unbiased predictor (Section 4). Proofs and mathematical details surrounding our main results are provided in the Appendix.

2. Gaussian Processes in Euclidean spaces

Let $Z = \{Z(x) : x \in \mathcal{M} \subset \mathbb{R}^d\}$ be a zero-mean Gaussian process on a bounded domain $\mathcal{M}$. The process $Z(\cdot)$ is characterised by its covariogram $k(x,y) = \mathbb{E}(Z(x)Z(y))$, $x, y \in \mathcal{M}$ so that for any finite collection of points, say $x_1, \cdots, x_n \in \mathcal{M}$, we have $(Z(x_1), \cdots, Z(x_n))^{\mathrm{T}} \sim$

$\mathcal{N}(0, \Sigma)$, where Σ is the $n \times n$ covariance matrix with (i,j)-th entry $\Sigma_{ij} = k(x_i, x_j)$. The Matérn process is a zero-mean stationary Gaussian process specified by the covariogram[2],

$$k(x,y) = \frac{\sigma^2 (\alpha\|x-y\|)^\nu}{\Gamma(\nu)2^{\nu-1}} K_\nu(\alpha\|x-y\|), \quad x, y \in \mathcal{M} \subset \mathbb{R}^d, \tag{1}$$

where $\sigma^2 > 0$ is called the partial sill or spatial variance, $\alpha > 0$ is the scale or decay parameter, $\nu > 0$ is a smoothness parameter, $\Gamma(\cdot)$ is the Gamma function, and $K_\nu(\cdot)$ is the modified Bessel function of the second kind of order ν (Abramowitz and Stegun, 1965, Section 10). The Matérn covariogram in (1) is isotropic and its spectral density (also known as the Hankel-Fourier transform, Genton (2002)) is given by

$$f(u) = \frac{\sigma^2 \alpha^{2\nu}}{\pi^{d/2}(\alpha^2 + u^2)^{\nu+d/2}}, \quad u \geq 0.$$

2.1 Identifiability

Let P_0 and P_1 be Gaussian measures corresponding to Matérn parameters $\{\sigma_0^2, \alpha_0, \nu\}$ and $\{\sigma_1^2, \alpha_1, \nu\}$, respectively. Two measures are said to be equivalent, denoted by $P_0 \equiv P_1$, if they are absolutely continuous with respect to each other. Two equivalent measures cannot be distinguished no matter how dense the observations are. Zhang (2004) showed that when $d < 4$, P_0 is equivalent to P_1 if and only if $\sigma_0^2 \alpha_0^{2\nu} = \sigma_1^2 \alpha_1^{2\nu}$. Hence, σ^2 and α do not admit asymptotically consistent estimators, while $\sigma^2 \alpha^{2\nu}$, also known as a *microergodic parameter*, is consistently estimable. For $d > 4$, Anderes (2010) proved that both σ^2 and α are consistently estimable. The case for $d = 4$ remains unresolved. The integral test offers a sufficient (but not necessary) condition on the spectral densities to determine whether two measures are equivalent. While unidentifiable parameters are never consistently estimable, identifiable parameters may be consistently estimable. However, deriving an explicit construction for such a consistent estimator is often challenging and is beyond the scope of the current manuscript; we identify this as an area of future research.

2.2 Parameter estimation

In practice, the maximum likelihood estimate is customarily used to estimate unknown parameters in the covariogram. Let $L_n(\sigma^2, \alpha)$ be the likelihood function:

$$L_n(\sigma^2, \alpha) = (2\pi\sigma^2)^{-n/2} \det(\Gamma_n(\alpha))^{-1/2} \exp\left\{-\frac{1}{2\sigma^2} Z_n^{\mathrm{T}} \Gamma_n(\alpha)^{-1} Z_n\right\}, \tag{2}$$

where $Z_n = (Z(x_1), \cdots, Z(x_n))^{\mathrm{T}}$ and $(\Gamma_n(\alpha))_{i,j} = \frac{(\alpha\|x_i - x_j\|)^\nu}{\Gamma(\nu)2^{\nu-1}} K_\nu(\alpha\|x_i - x_j\|)$ is independent of σ^2. Given α, the maximum likelihood estimation of σ^2 is given by (Stein, 1999)

$$\widehat{\sigma}^2 = \frac{Z_n^{\mathrm{T}} \Gamma_n(\alpha)^{-1} Z_n}{n}.$$

Let $\{\sigma_0^2, \alpha_0\}$ be the data generating parameters with observations $Z(x_1), \cdots, Z(x_n)$. For any misspecified α_1, if $\widehat{\sigma}_{1,n}^2$ is the maximum likelihood estimation of $L_n(\sigma^2, \alpha_1)$, then $\widehat{\sigma}_{1,n}^2 \alpha_1^{2\nu} \to$

---

2. Solin and Kok (2019) provides an alternative definition based on PDEs with boundary conditions.

$\sigma_0^2 \alpha_0^{2\nu}$ as $n \to \infty$ with probability 1 under P_0 when $\cup_{n=1}^{\infty}\{x_n\}$ is bounded and infinite (Zhang, 2004; Kaufman et al., 2008). Moreover, $\sqrt{n}\left(\frac{\widehat{\sigma}_{1,n}^2 \alpha_1^{2\nu}}{\sigma_0^2 \alpha_0^{2\nu}} - 1\right) \to \mathcal{N}(0,2)$ as $n \to \infty$ (Du et al., 2009; Wang and Loh, 2011; Kaufman and Shaby, 2013). As a result, even if we do not know the true parameters $\{\alpha_0, \sigma_0^2\}$, we can choose an arbitrary, possibly misspecified, decay parameter α_1 and find the maximum likelihood estimate of the spatial variance $\widehat{\sigma}_{1,n}^2$. The resulting Gaussian measure is asymptotically equivalent to the Gaussian measure corresponding to the true parameter.

2.3 Prediction and kriging

Gaussian processes are widely deployed in spatial or nonparametric regression models to carry out model-based predictive inference. Given a new location x_0, the best linear unbiased predictor (BLUP) for $Z_0 = Z(x_0)$ is given by

$$\widehat{Z}_n(\alpha) = \gamma_n(\alpha)^{\mathrm{T}} \Gamma_n(\alpha)^{-1} Z_n,$$

where $(\gamma_n(\alpha))_i = \frac{(\alpha\|x_0 - x_i\|)^{\nu}}{\Gamma(\nu)2^{\nu-1}} K_{\nu}(\alpha\|x_0 - x_i\|)$. Then

$$\frac{\mathbb{E}_{\sigma_0^2,\alpha_0}(\widehat{Z}_n(\alpha_1) - Z_0)^2}{\mathbb{E}_{\sigma_0^2,\alpha_0}(\widehat{Z}_n(\alpha_0) - Z_0)^2} \xrightarrow{n\to\infty} 1, \quad \frac{\mathbb{E}_{\widehat{\sigma}_{1,n}^2,\alpha_1}(\widehat{Z}_n(\alpha_1) - Z_0)^2}{\mathbb{E}_{\sigma_0^2,\alpha_0}(\widehat{Z}_n(\alpha_1) - Z_0)^2} \xrightarrow{n\to\infty} 1,$$

where $\mathbb{E}$ is the expectation with respect to the measure characterised by the parameter or spectral density (see Section 3) in the subscript. As a result, any misspecified α still yields an asymptotically optimal BLUP as long as σ^2 is replaced by its maximum likelihood estimate (Stein, 1993; Kaufman and Shaby, 2013). In the current manuscript, we develop parallel results for the d dimensional compact Riemannian manifold $\mathcal{M}$.

3. Gaussian processes on compact Riemannian manifold

Henceforth, we assume that our domain of interest is a d-dimensional compact Riemannian manifold $\mathcal{M}$ equipped with a Riemannian metric g. We denote the Laplace–Beltrami operator on $\mathcal{M}$ by $-\Delta_g$ with eigenvalues λ_n and eigenfunctions f_n, the volume form by $\mathrm{d}V_g$ and the volume of $\mathcal{M}$ by $V_{\mathcal{M}}$ (see, e.g., Kobayashi and Nomizu, 1963; Lee, 2018; do Carmo, 1992, for further details on operators and spectral theory on Riemannian manifolds).

3.1 Matérn covariogram on compact Riemannian manifolds

On a Riemannian manifold, where the linear structure of $\mathbb{R}^d$ is missing, the standard definition of the Matérn covariogram is no longer valid. A natural extension of the Matérn covariogram to manifolds will consider replacing the Euclidean norm $\|x - y\|$ in (1) by the geodesic distance $d(x, y)$. Unfortunately, this naive generalisation is not valid for $\nu = \infty$ (Feragen et al., 2015), unless the manifold is flat. If we restrict ourselves to spheres, Matérn with $\nu \in (1/2, \infty)$ is still invalid (Gneiting, 2013). Instead, some Matérn-like covariograms including chordal, circular and Legendre Matérn covariograms and other families of covariograms have been studied (Jeong and Jun, 2015b; Porcu et al., 2016; Guinness and Fuentes, 2016; Guella et al., 2018; Clarke De la Cerda et al., 2018; Alegría et al., 2021). However,

these covariograms are constructed specifically with respect to the geometry of the sphere and do not generalise to generic compact Riemannian manifolds.

Whittle (1963) showed that the Matérn covariogram in Euclidean space admits a representation through a stochastic partial differential equation involving white noise and the Laplace operator Δ. Lindgren et al. (2011) built on this stochastic partial differential equation approach to define the Matérn covariogram on manifolds involving the Laplace–Beltrami operator Δ_g. This idea was further developed, both theoretically and practically, by several scholars (see, e.g., Bolin and Lindgren, 2011; Lang and Schwab, 2015; Herrmann et al., 2020; Borovitskiy et al., 2020, 2021, among others). We state the definition of the Matérn covariogram in the stochastic partial differential equation sense, which is a valid positive definite function for any ν on any compact Riemannian manifold $\mathcal{M}$.

Definition 1 *Let f_l be the orthonormal eigenfunctions of $-\Delta_g$ and $\lambda_l \geq 0$ be the corresponding eigenvalues in ascending order. The Matérn covariogram is defined by*

$$k(x,y) = \frac{\sigma^2}{C_{\nu,\alpha}} \sum_{l=0}^{\infty} \left(\alpha^2 + \lambda_l\right)^{-\nu-\frac{d}{2}} f_l(x) f_l(y),$$

where $C_{\nu,\alpha} = \displaystyle\sum_{l=0}^{\infty}(\alpha^2 + \lambda_l)^{-\nu-d/2}$ is a constant such that the average variance is $\sigma^2 = \frac{1}{V_{\mathcal{M}}} \int_{\mathcal{M}} k(x,x) \mathrm{dV_g}(x)$. The corresponding spectral density is

$$\rho(l) = \frac{\sigma^2}{C_{\nu,\alpha}} (\alpha^2 + \lambda_l)^{-\nu-\frac{d}{2}}.$$

Similarly, the squared exponential covariogram is

$$k(x,y) = \frac{\sigma^2}{C_{\infty,\alpha}} \sum_{l=0}^{\infty} e^{-\frac{\lambda_l}{2\alpha^2}} f_l(x) f_l(y),$$

where $C_{\infty,\alpha} = \displaystyle\sum_{l=0}^{\infty} e^{-\frac{1}{2\alpha^2}\lambda_l}$ is a constant such that the average variance is $\sigma^2 = \frac{1}{V_{\mathcal{M}}} \int_{\mathcal{M}} k(x,x) \mathrm{dV_g}(x)$. The corresponding spectral density is

$$\rho(n) = \frac{\sigma^2}{C_{\infty,\alpha}} e^{-\frac{\lambda_l}{2\alpha^2}}.$$

Remark 2 *There are several commonly used parametric representations of the Matérn covariogram. In particular, this article adopts the same parametric representation as the one in Zhang (2004), but different from Borovitskiy et al. (2021).*

If $\mathcal{M}$ is a sphere, the covariograms defined above coincide with the Matérn-like covariograms on spheres provided by Guinness and Fuentes (2016) and Kirchner and Bolin (2022). As a result, we focus on a non-trivial generalisation to generic compact Riemannian manifolds. The relation between the three parameters (α, σ^2, ν) in the above definition and the coefficients in the stochastic partial differential equation representation is not straightforward

(see Lindgren et al., 2011, for details). Note that for any (α, σ^2, ν), the covariogram shares the same eigenbasis with the Laplace–Beltrami operator Δ_g. This property is not deemed restrictive for our ensuing development since we primarily focus on the Matérn and squared exponential covariograms. Furthermore, this property offers crucial analytic tractability for several results developed subsequently. Hence, we refer to the Matérn and squared exponential covariograms as in Definition 1 in the following sections.

3.2 Identifiability

In Euclidean domains, the integral test (Yadrenko, 1983; Stein, 1999) is a powerful tool to determine the equivalence of two Gaussian measures. However, such tests do not carry through to non-Euclidean domains as the spectrum on such manifolds is discrete. Alegría et al. (2021) studied the so called $\mathcal{F}-$family of covariograms on spheres and numerically deduced, without proof, the consistency of the maximum likelihood estimate of some parameters for this family. Arafat et al. (2018) derived the equivalence of Gaussian measures on spheres and derived microergodic parameters of some covariograms excluding the Matérn. All of the above results are built upon the Feldman–Hájek Theorem (Da Prato and Zabczyk, 2014), which is valid for any metric space and, hence, applicable to compact Riemannian manifolds. Here, we generalise the above results to a Gaussian process with Matérn and squared exponential covariograms on arbitrary compact Riemannian manifolds, also motivated by the Feldman–Hájek theorem. Therefore, we can still study the identifiability of these parameters by finding the microergodic parameters.

Lemma 3 *Let P_i ($i = 1, 2$) be mean zero Matérn/squared exponential Gaussian random measures with spectral densities ρ_i. Then, $P_1 \equiv P_2$ if and only if*

$$\sum_l \left| \frac{\rho_2(l) - \rho_1(l)}{\rho_1(l)} \right|^2 < \infty.$$

Proof See Appendix A. ∎

From Definition 1, ρ_i is strictly positive so the denominator is always non-zero. The series test is a sufficient and necessary condition. This is a significant enhancement over the integral test in Euclidean spaces, which offers only a sufficient condition. Its importance to us will become clear after Theorem 4. Subsequently, we consider microergodic parameters of Gaussian processes on a manifold with the Matérn covariogram. This is analogous to Theorem 2 in Zhang (2004) for compact Riemannian manifolds.

Theorem 4 *Let P_i, $i = 1, 2$, denote two Gaussian measures with the Matérn covariogram parametrized by $\theta_i = \{\sigma_i^2, \alpha_i, \nu_i\}$. Then the following results hold.*

(A) If $d \leq 3$, then $P_1 \equiv P_2$ if and only if $\sigma_1^2/C_{\nu_1,\alpha_1} = \sigma_2^2/C_{\nu_2,\alpha_2}$, $\nu_1 = \nu_2$.

(B) If $d \geq 4$, then $P_1 \equiv P_2$ if and only if $\sigma_1^2 = \sigma_2^2$ and $\alpha_1 = \alpha_2$, $\nu_1 = \nu_2$.

Proof See Appendix B. ∎

Part (A) of Theorem 4 implies that if $d \leq 3$, then neither σ^2 nor α are identifiable or

consistently estimable, while ν is identifiable. Part (B) implies that when $d \geq 4$, all three parameters—σ^2, α and ν—are identifiable. In Euclidean space, the smoothness parameter ν is typically assumed to be known and fixed when discussing fixed-domain asymptotic inference. In this specific Euclidean setting, assuming $\nu_1 = \nu_2 = \nu$, (A) still holds while (B) holds for $d > 4$; $d = 4$ is still an unresolved problem in Euclidean space unless the domain is assumed to be bounded (Bolin and Kirchner, 2021). This difference in behaviour between (A) and (B) can be attributed to the integral test being a sufficient condition in Euclidean spaces, which ensures only the equivalence of measures when $d \leq 3$; (see Zhang, 2004, for details). In $d > 4$, Anderes (2010) estimated the principal irregular term without the integral test and constructed consistent estimators for α and σ^2 directly. However, this construction does not hold for $d = 4$.

In contrast, the series test in Lemma 3 is a *sufficient and necessary* condition so that we can provide a condition for the equivalence of two measures with Matérn covariograms over any dimension. The dimension also plays an important role in the manifold setting due to Weyl's Law (Li, 1987; Canzani, 2013). That is, the growth of the eigenvalues and their multiplicities are intertwined with the dimension d; further details are provided within the proof in Appendix B. Another benefit of the sufficient and necessary condition is that the series test can be applied to the squared exponential covariogram, also known as the radial basis function, which can be viewed as a limiting case of the Matérn covariogram when $\nu \to \infty$, as introduced in Definition 1. Since the spectral density is not a polynomial, the integral test over Euclidean domains is invalid and the conditions for the equivalence of two squared exponential covariograms are intractable. In contrast, the following theorem resolves the equivalence of squared exponential covariograms on a compact manifold $\mathcal{M}$.

Theorem 5 *Let P_i, for $i = 1, 2$, be Gaussian measures with squared exponential covariograms parametrised by $\theta_i = \{\sigma_i^2, \alpha_i\}$. Then $P_1 \equiv P_2$ if and only if $\sigma_1^2 = \sigma_2^2$ and $\alpha_1 = \alpha_2$.*

Proof See Appendix C. ■

Theorem 5 shows that it is possible to have consistent estimators for both σ^2 and α. So far we have developed formal results on the identifiability of parameters in the covariogram on a compact Riemannian manifold. Inference for identifiable parameters will proceed in customary fashion so we turn our attention to non-identifiable settings, i.e., the Matérn covariogram with known ν on manifolds with dimension $d \leq 3$.

3.3 Consistency of maximum likelihood estimation

Since $\mathcal{M}$ is compact, there is no increasing-domain asymptotic framework and $\cup_{n=1}^{\infty}\{x_n\}$ is always bounded. In the remaining sections, we assume that $\cup_{n=1}^{\infty}\{x_n\}$ is infinite, which is the standard assumption also known as the increasing sequence assumption (also see Stein, 1999; Zhang, 2004; Kaufman and Shaby, 2013). Let $\{\sigma_0, \alpha_0\}$ be the data generating parameter (oracle) and let $\hat{\sigma}_{1,n}^2$ be the maximum likelihood estimate of σ^2 obtained by maximising $L_n(\sigma^2, \alpha_1)$ with a misspecified α_1. The following theorem is analogous to Theorem 3 in Zhang (2004) for compact Riemannian manifolds.

Theorem 6 *Under the setting of Theorem 4, assuming $\cup_{n=1}^\infty \{x_n\}$ is infinite, we obtain*

$$\frac{\widehat{\sigma}_{1,n}^2}{C_{\nu,\alpha_1}} \xrightarrow{n\to\infty} \frac{\sigma_0^2}{C_{\nu,\alpha_0}}, \ P_0 \ a.s.$$

Proof See Appendix D. ∎

In Euclidean space, $\widehat{\sigma}_{1,n}^2/C_{\nu,\alpha_1}$ is asymptotically Gaussian. We conjecture that this asymptotic normality still holds on Riemannian manifolds. However, this result relies on specific constructions in Euclidean space (Wang, 2010), which become invalid for manifolds. A formal proof is beyond the scope of the current manuscript and we intend to pursue this development in future investigations. In Section 4 we present a numerical simulation experiment to demonstrate the asymptotic (normal) behaviour of this parameter on spheres.

3.4 Prediction

Given a new location $x_0 \in \mathcal{M}\backslash\{x_i\}_{i=1}^n$, the best linear unbiased predictor for $Z_0 = Z(x_0)$ under a covariance function k_ρ characterised by its spectral density ρ is given by

$$\widehat{Z}_n(\rho) = \gamma_n(\rho)^{\mathrm{T}} \Gamma_n(\rho)^{-1} Z_n,$$

where $\gamma_n(\rho) = \frac{1}{\sigma^2} k_\rho(x_0, x_i)$ and $\{\Gamma_n(\rho)\}_{ij} = \frac{1}{\sigma^2} k_\rho(x_i, x_j)$.

Kirchner and Bolin (2022) and Bolin and Kirchner (2021) generalise the results of asymptotic optimality of the BLUP based on a misspecified scale parameter in Euclidean spaces (Stein, 1993) to metric spaces. That is, the prediction error of the BLUP under a misspecified scale parameter is asymptotically the same as the error of the BLUP under the true parameter. If the domain is a compact Riemannian manifold and the covariograms are Matérn, then two covariance operators share the same eigenbasis; this is the setting described in Section 5.1 of Kirchner and Bolin (2022) as a special case of Theorem 3.1 therein. We rephrase it in the following lemma with some modifications to fit the Matérn covariograms on a compact Riemannian manifold with a different and simpler proof.

Lemma 7 *Let ρ_0, ρ_1 be the spectral densities of two Gaussian measures on $\mathcal{M}$ with Matérn covariograms. Given $x_0 \in \mathcal{M}\backslash\{x_i\}_{i=1}^n$, let $\widehat{Z}_n(\rho_i)$ be the best linear unbiased predictor of $Z_0 \coloneqq Z(x_0)$ based on observations $\{Z(x_1), \cdots, Z(x_n)\}$ with $\{x_i\}_{i=1}^\infty$ being infinite and having x_0 as an accumulation point, where ρ_i is the spectral density of $Z(\cdot)$. If there exists a real number c such that $\lim\limits_{m\to\infty} \frac{\rho_1(m)}{\rho_0(m)} = c$, then:*

(i) $\dfrac{\mathbb{E}_{\rho_0}(\widehat{Z}_n(\rho_1) - Z_0)^2}{\mathbb{E}_{\rho_0}(\widehat{Z}_n(\rho_0) - Z_0)^2} \xrightarrow{n\to\infty} 1,$

(ii) $\dfrac{\mathbb{E}_{\rho_1}(\widehat{Z}_n(\rho_1) - Z_0)^2}{\mathbb{E}_{\rho_0}(\widehat{Z}_n(\rho_1) - Z_0)^2} \xrightarrow{n\to\infty} c.$

Proof See Appendix E. ∎

Focusing on the parameters in a Matérn covariogram, let $\widehat{\sigma}_{1,n}^2$ be the maximum likelihood estimate of $L_n(\sigma^2, \alpha_1)$ and ρ_i be the spectral density of the Matérn covariogram with decay parameter α_i.

Theorem 8 *Under the same conditions as in Theorem 4 and Lemma 7, let $\sigma_1^2 = \sigma_0^2 C_{\nu,\alpha_1}/C_{\nu,\alpha_0}$, then*

$$\frac{\mathbb{E}_{\sigma_0^2,\alpha_0}(\widehat{Z}_n(\alpha_1) - Z_0)^2}{\mathbb{E}_{\sigma_0^2,\alpha_0}(\widehat{Z}_n(\alpha_0) - Z_0)^2} \xrightarrow{n\to\infty} 1, \quad \frac{\mathbb{E}_{\widehat{\sigma}_{1,n}^2,\alpha_1}(\widehat{Z}_n(\alpha_1) - Z_0)^2}{\mathbb{E}_{\sigma_0^2,\alpha_0}(\widehat{Z}_n(\alpha_1) - Z_0)^2} \xrightarrow[P_0 \ a.s.]{n\to\infty} 1.$$

Proof See Appendix F. ∎

Note that Lemma 7 and Theorem 8 offer the manifold versions of Theorems 3 and 4 in Kaufman and Shaby (2013).

4. Matérn on spheres

We now consider Gaussian processes with the Matérn covariogram on the d-dimensional sphere S^d, including two popular manifolds in spatial statistics: the circle S^1 and sphere S^2. We show that all theorems in the previous sections hold for S^d with $d = 1, 2, 3$. As earlier, we assume that P_i, $i = 1, 2$, are two Gaussian measures on S^d with Matérn covariogram parameters $\{\sigma_i^2, \alpha_i, \nu\}$.

Theorem 9 *For spheres with dimension $d = 1, 2, 3$, the following results are true:*

1. *$P_1 \equiv P_2$ if and only if $\sigma_1^2/C_{\nu,\alpha_1} = \sigma_2^2/C_{\nu,\alpha_2}$, so neither σ^2 nor α can be consistently estimated.*

2. *Let the data generating parameters be $\{\sigma_0, \alpha_0\}$ and $\widehat{\sigma}_{1,n}^2$ be the maximum likelihood estimation of $L_n(\sigma^2, \alpha_1)$ with misspecified α_1 based on increasing sequence $\{x_i\}_{i=1}^n$. Then,*

$$\frac{\widehat{\sigma}_{1,n}^2}{C_{\nu,\alpha_1}} \xrightarrow{n\to\infty} \frac{\sigma_0^2}{C_{\nu,\alpha_0}}, \ P_0 \ a.s.$$

3. *Given $x_0 \in \mathcal{M}\backslash\{x_i\}_{i=1}^n$, let $\widehat{Z}_n$ be the best linear unbiased predictor of $Z_0 := Z(x_0)$ based on observations $\{Z(x_1), \cdots, Z(x_n)\}$ with $\{x_i\}_{i=1}^\infty$ being infinite, then*

$$\frac{\mathbb{E}_{\widehat{\sigma}_{1,n}^2,\alpha_1}(\widehat{Z}_n(\alpha_1) - Z_0)^2}{\mathbb{E}_{\sigma_0^2,\alpha_0}(\widehat{Z}_n(\alpha_1) - Z_0)^2} \xrightarrow{n\to\infty} 1, \ P_0 \ a.s.$$

Proof See Appendix G. ∎

Next, we consider two concrete examples: the circle S^1 and the sphere S^2.

4.1 Matérn covariogram on circle

First, we recall the simplified form of the Matérn covariogram on S^1 (Borovitskiy et al., 2020):

Lemma 10 *When $\mathcal{M} = S^1 \subset \mathbb{R}^2$ and $\nu = 1/2 + s$, $s \in \mathbb{N}$, the Matérm covariogram is given by*

$$k(x,y) = \frac{\sigma^2}{C'_{\nu,\alpha}} \sum_{k=0}^s a_{s,k}(\alpha(|x-y| - 1/2))^k \mathrm{hyp}^k(\alpha(|x-y| - 1/2)), \quad x, y \in S^1, \qquad (3)$$
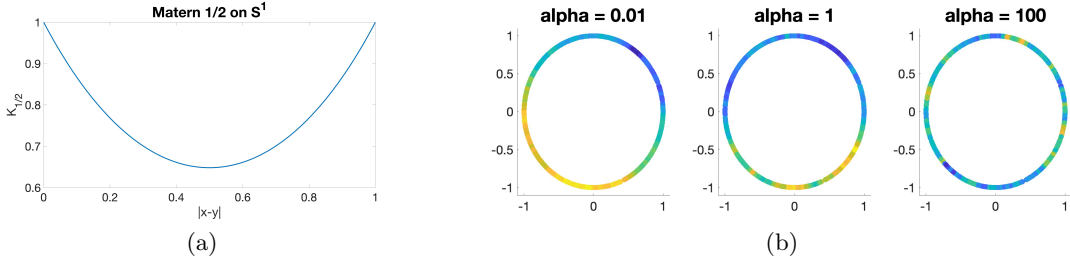
Figure 1: (a) Covariogram of Matérn 1/2 on S^1; (b): Sample fields with $\sigma^2 = 0.1$, $\nu = 1/2$, $\alpha \in \{0.01, 1, 100\}$.

where $C'_{\nu,\alpha}$ is chosen so that $k(x,x) = \sigma^2$, hyp^k is cosh when k is even and sinh when k is odd, $a_{s,k}$ are constants depending on ν and α; see Borovitskiy et al. (2020) for details.

Note that $x - y := \theta_x - \theta_y \mod 1$ for $x = e^{2\pi i \theta_x}$ and $y = e^{2\pi i \theta_y}$. Therefore, the Matérn covariogram is "stationary" with respect to this group addition instead of the standard addition in Euclidean space. The corresponding spectral density is given by

$$\rho(n) = \frac{2\sigma^2 \alpha \sinh(\alpha/2)}{C'_{\nu,\alpha}(2\pi)^{1-2\nu}} \left(\alpha^2 + 4\pi^2 n^2\right)^{-\nu-1/2}, \quad n \in \mathbb{Z}. \tag{4}$$

In particular, when $\nu = 1/2$, the covariogram and spectral densities admit simple forms:

$$k(x,y) = \frac{\sigma^2}{\cosh(\alpha/2)} \cosh\left(\alpha(|x-y| - 1/2)\right),$$

$$\rho(n) = 2\sigma^2 \alpha \tanh(\alpha/2)(\alpha^2 + 4\pi^2 n^2)^{-1}.$$

Figure 1(a) depicts a covariogram with $\nu = 1/2$, $\alpha = 2$, and $\sigma^2 = 1$. Note that $|x - y| = 1/2$ means that x and y are antipodal points so the correlation attains a minimum. Figure 1(b) shows a set of simulated Z's with different values of α. It is clear that the smaller values of α generate smoother random fields as the correlation grows larger.

Corollary 11 Let $\nu = 1/2$, then $P_1 \equiv P_2$ if and only if $\sigma_1^2 \alpha_1 \tanh(\alpha_1/2) = \sigma_2^2 \alpha_2 \tanh(\alpha_2/2)$, so neither σ^2 nor α can be consistently estimated.

For a general $\nu = 1/2 + s$, $s \in \mathbb{N}$, the normalising constant is

$$C'_{\nu,\alpha} = \sum_{k=0}^{s} a_{s,k}(-\alpha/2)^k \text{hyp}^k(-\alpha/2).$$

We point out that this $C'_{\nu,\alpha}$ is different from the $C'_{\nu,\alpha}$ in Definition 1 when $\mathcal{M} = S^1$. Although we cannot express $C'_{\nu,\alpha}$ as an elementary function, we can still find the microergodic parameter for any $\nu = s + 1/2$, $s \in \mathbb{Z}$:

Corollary 12 Let $\nu = 1/2 + s$, $s \in \mathbb{Z}$, then $P_1 \equiv P_2$ if and only if $\sigma_1^2 \alpha_1 \sinh(\alpha_1/2)/C'_{\nu,\alpha_1} = \sigma_2^2 \alpha_2 \sinh(\alpha_2/2)/C'_{\nu,\alpha_2}$, so neither σ^2 nor α can be consistently estimated.

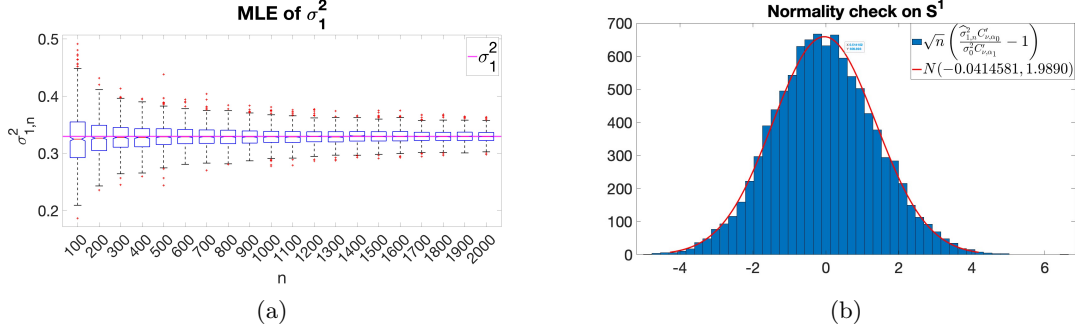Figure 2: (a) $\widehat{\sigma}_{1,n}^2$ v.s. σ_1^2; (b): Distribution of $\sqrt{n}\left(\dfrac{\widehat{\sigma}_{1,n}^2}{\sigma_1^2}-1\right)$.

Figure 2 shows that $\widehat{\sigma}_{1,n}^2 \to \sigma_1^2 := \dfrac{\sigma_0^2 \alpha_0 \sinh(\alpha_0/2)}{C'_{\nu,\alpha_0}} \dfrac{C'_{\nu,\alpha_1}}{\alpha_1 \sinh(\alpha_1/2)}$ as shown by the horizontal line and the empirical distribution of $\sqrt{n}\left(\dfrac{\widehat{\sigma}_{1,n}^2}{\sigma_1^2}-1\right)$ is $N(0,2)$, for $\nu = 1/2$, $\sigma_0 = 0.1$, $\alpha_0 = 2 \neq \alpha_1 = 1$. Panel (a) supports Theorem 9 empirically. That is, although (σ^2, α, ν) are not consistently estimable, the microergodic parameter $\dfrac{\sigma^2 \alpha \sinh(\alpha/2)}{C'_{\nu,\alpha}}$ is consistently estimable. Panel (b) supports our conjecture after Theorem 6 empirically.

4.2 Matérn covariogram on the sphere

On a sphere S^2, the Matérn covariogram is more complicated (Borovitskiy et al., 2020):

Lemma 13 *The Matérn covariogram on $\mathcal{M} = S^2$ with $\nu > 0$ is*

$$k(x,y) = \frac{\sigma^2}{C_{\nu,\alpha}} \sum_{l=0}^{\infty} \left(\alpha^2 + l(l+1)\right)^{-\nu-1} c_l \mathcal{L}_l(\cos(d_M(x,y)))$$

and its spectral density is given by

$$\rho(l) = \frac{\sigma^2}{C_{\nu,\alpha}} \left(\alpha^2 + l(l+1)\right)^{-\nu-1} ,$$

where $d_M(\cdot,\cdot)$ is the geodesic distance on S^2, $\mathcal{L}_l$ is the Legendre polynomial of degree l:

$$\mathcal{L}_l(z) = \sum_{k=0}^{\lfloor l/2 \rfloor} (-1)^k \frac{l!(l-k-\frac{1}{2})!}{k!(l-2k)!}(2z)^{n-2k} , \quad and$$

$$c_l = \frac{(2l+1)\Gamma(3/2)}{2\pi^{3/2}}, \quad C_{\nu,\alpha} = \frac{\Gamma(3/2)}{8\pi^{5/2}} \sum_{l=0}^{\infty}(2l+1)\left(2\nu\alpha^2 + l(l+1)\right)^{-\nu-1}$$

Remark 14 *The index l in the above covariance function is different from the index l in Definition 1. In fact, each Legendre polynomial corresponds to multiple spherical harmonics, so the spectral density does not contain the c_l constants anymore.*

Unlike Lemma 10, where ν is required to be a half-integer, here ν can be any positive number. However, the covariogram now involves an infinite series, which needs to be approximated when $x \neq y$. Approximating a function on S^2 is known as the "scatter data interpolation problem" (Narcowich et al., 1998) and preserving the positive definiteness is known as the stability problem (Kunis, 2009). For the Matérn covargioram considered in this manuscript, we adopt a natural and simple approximation using the partial sum of an infinite series. The following theorem controls the approximation error and ensures the positive definiteness of the approximated covariogram.

Theorem 15 *For the partial sum*

$$k^L(x,y) = \frac{\sigma^2}{C_{\nu,\alpha}} \sum_{l=0}^{L} \left(\alpha^2 + l(l+1)\right)^{-\nu-1} c_l \mathcal{L}_l(\cos(d_M(x,y))),$$

the approximation error is controlled by

$$|k^L(x,y) - k(x,y)| \leq \epsilon := \frac{12\pi\sigma^2}{\sum_l (2l+1)(\alpha^2 + l(l+1))^{-\nu-1}} L^{-2\nu}.$$

Given observations $x_1, \cdots, x_n$ with minimal separation $q = \inf_{i \neq j} d(x_i, x_j)$, the approximated covariance matrix $\{k^L(x_i, x_j)\}_{ij}$ is positive definite for any

$$L > \left(\frac{12\pi n\sigma^2}{\xi_\rho(q)\sum_l (2l+1)(\alpha^2 + l(l+1))^{-\nu-1}}\right)^{\frac{1}{2\nu}},$$

where $\xi_\rho(q)$ is a constant depending on the spectral density ρ and minimal separation q; see the proof for more details.

Proof See Appendix H. ∎

The above result implies that the computational cost is of order $\epsilon^{-\frac{1}{2\nu}}$ as $\epsilon \to 0$. Larger values of ν imply smoother random fields that require smaller values of N to approximate the covariogram. In practice, we can first calculate $\xi_\rho(q)$, which is computationally practicable because of the closed-form representation (see Appendix H for details), and then choose N.

Figure 3(a) presents the covariogram with $\nu = 1/2$, $\alpha = 1$, and $\sigma^2 = 1$. Note that $d(x,y) = \pi$ means that x and y are antipodal points so the correlation reaches the minimum. Figure 3(b) shows some simulated Z's with different α's. Similar to $\mathcal{M} = S^1$, smaller values of α lead to smoother random fields.

However, due to the bias introduced by the partial sum, we do not have access to the ground truth covariogram, so the analogue of Figure 2 is not available anymore. Similar issues arise in approximations to the Matérn on a compact manifold (Sanz-Alonso and Yang, 2022a). Instead, we show the theoretical results on microergodic parameters analogous to Corollary 12:

Corollary 16 $P_{\theta_1} \equiv P_{\theta_2}$ *if and only if* $\sigma_1^2/C_{\nu,\alpha_1} = \sigma_2^2/C_{\nu,\alpha_2}$, *so neither* σ^2 *nor* α *can be consistently estimated.*
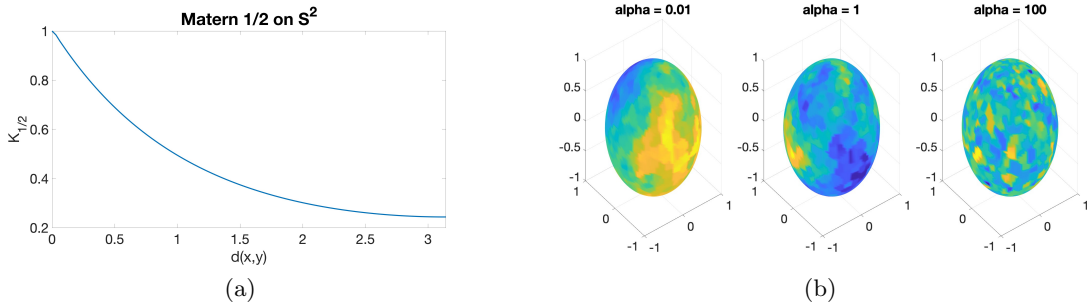
Figure 3: (a) Covariogram of Matérn $1/2$ on S^2; (b): Sample fields with $\sigma^2 = 0.1$, $\nu = 1/2$, $\alpha \in \{0.01, 1, 100\}$.

5. Discussion

This article has formally developed some theoretical results on statistical inference for Gaussian processes with Matérn covariograms on compact Riemannian manifolds. Our focus has primarily been on the identifiability and consistency (or lack thereof) of the covariogram parameters and of spatial predictions. For the Matérn and squared exponential covariograms, we provide a sufficient and necessary condition for the equivalence of two Gaussian random measures through a series test and derive identifiable and consistently estimable microergodic parameters for an arbitrary dimension d. Specifically for $d \leq 3$, we formally establish the consistency of maximum likelihood estimates of the parameters and the asymptotic normality of the best linear unbiased predictor under a misspecified decay parameter. The circle and sphere are analysed as two examples with corroborative numerical experiments.

We anticipate that the results developed here will generate substantial future work in this domain. For example, as we have alluded to earlier in the article, in Euclidean spaces we know that the maximum likelihood estimate of σ^2 is asymptotically normal: $\sqrt{n}\left(\frac{\widehat{\sigma}_{1,n}^2}{C_{\nu,\alpha_1}} - \frac{\sigma_0^2}{C_{\nu,\alpha_0}}\right) \to N(0,2)$. While our numerical experiments lead us to conjecture that an analogous result holds for compact Riemannian manifolds, a formal proof may well require substantial new machinery that we intend to explore further. Next, we conjecture that two measures with the Matérn covariogram are equivalent on $\mathbb{R}^4$ if and only if they have the same decay and spatial variance parameters. We know this result holds for manifolds with $d = 4$, but a formal proof for $\mathbb{R}^4$ has not yet been established. Based upon similar reasonings we conjecture that two measures with squared exponential covariograms are equivalent on Euclidean spaces if and only if they have the same decay and spatial variance parameters.

Another future generalisation is to consider covariograms on compact Riemannian manifolds that are not simultaneously diagonalisable, whose asymptotically optimal linear predictor has been studied in Kirchner and Bolin (2022). Nevertheless, issues pertaining to the equivalence of measures, derivation of microergodic parameters and consistency of maximum likelihood estimates remain unresolved. Furthermore, covariograms that offer scientific interpretation in practical inference need to be explored. In this regard, it is worth remarking that although our results are primarily concerned with maximum likelihood estimates, they

will provide useful insights into Bayesian learning on manifolds. For example, the failure to consistently estimate certain (non-microergodic) parameters will inform Bayesian modellers that inference for such parameters will always be sensitive to their prior specifications. This will open up new avenues of research in specifying prior distributions for microergodic parameters. Formal investigations into the consistency of the posterior distributions of Matérn covariogram parameters on manifolds are of inferential interest and may benefit from some of our developments in the current manuscript.

Other avenues for future developments will relate to computational efficiency of Gaussian processes on manifolds. Here, a natural candidate for explorations is the tapered covariogram on manifold to introduce sparsity in the covariance matrix (Furrer et al., 2006). Since our domain in the current manuscript is compact, unlike in Euclidean domains, further compact truncation is redundant. One can explore the development of new "tapered" covariograms that achieve positive-definiteness and sparsity. Other approaches that induce dimension reduction based on conditional expectations, such as Gaussian predictive processes (Banerjee et al., 2008), may be explored on compact Riemannian manifolds since these low-dimensional processes are induced by any valid probability measure, although the choice of inputs to define the lower dimensional subspace will need to be addressed. On the other hand, sparse processes resulting from approximations using directed acyclic graphs (Datta et al., 2016b) are less natural for modelling data on manifolds since they depend on well-defined neighbours of inputs, which are less obvious to define outside of Euclidean spaces. Nevertheless, Datta et al. (2016a) developed adaptive Nearest-Neighbour Gaussian processes for massive space-time data sets on Euclidean spaces that selected neighbours using the covariance kernel as a metric for proximity. Such an approach holds promise in modelling massive data sets on manifolds.

In addition, asymptotic properties of estimates under tapering are of interest and have, hitherto, been explored only in Euclidean domains (Kaufman et al., 2008; Du et al., 2009) and without the presence of measurement error processes ("nuggets"). Inference for Gaussian process models with measurement errors (nuggets) on compact manifolds also present novel challenges and can constitute future work. Identifiability and consistency of the nugget in Euclidean spaces have only recently started receiving attention (Tang et al., 2021). However, the developments for Euclidean spaces do not easily apply to compact Riemannian manifolds; hence new tools will need to be developed. On complex or unknown domains, the eigenvalues and eigenfunctions of the Laplacian operator need to be estimated (Belkin and Niyogi, 2007). Asymptotic analysis of estimation in the spectral domain should be closely related to the frequency domain. Finally, since compact manifolds are distinct from non-compact manifolds, both geometrically and topologically, generalisation to non-compact Riemannian manifolds is of interest, where the spectrum is not discrete. Analytic tools on non-compact manifolds will need to be developed.

Acknowledgments

Appendix A. Proof of Lemma 3

Before proving Lemma 3, we recall the following lemma (Proposition B, Chapter III Yadrenko, 1983), also known as the Feldman–Hájek theorem:

Lemma 17 $P_1 \equiv P_2$ *if and only if*

1. *Operator $D = B_1^{-1/2} B_2 B_1^{-1/2} - \mathrm{I}$ is Hilbert–Schmidt;*

2. *Eigenvalues of D are strictly greater than -1,*

where B_i is the correlation operator of P_i defined by:

$$(B_i h)(x) := \int_{\mathcal{M}} \sum_{l=0}^{\infty} \rho_i(l) f_l(x) f_l(y) h(y) \mathrm{d}V_{\mathrm{g}}(y), \ h \in L^2(\mathcal{M}).$$

Proof of Lemma 3. By Lemma 17, it suffices to check conditions 1 and 2. Let γ_n^i be the eigenvalue of B_i and d_n be the eigenvalue of D. Observe that f_n is an eigenfunction of B_i with eigenvalue $\rho_i(n)$:

$$
\begin{aligned}
(B_i f_l)(x) &= \int_{\mathcal{M}} \sum_m \rho_i(m) f_m(x) f_m(y) f_l(y) \mathrm{d}V_{\mathrm{g}}(y) \\
&= \sum_m \rho_i(m) f_m(x) \int_M f_m(y) f_l(y) \mathrm{d}V_{\mathrm{g}}(y) \\
&= \sum_m \rho_i(m) f_m(x) \langle f_m, f_l \rangle_{\mathcal{M}} \\
&= \sum_m \rho_i(m) f_m(x) \delta_{nm} \\
&= \rho_i(l) f_l(x),
\end{aligned}
$$

where δ is the Kronecker delta and $\langle \cdot, \cdot \rangle_{\mathcal{M}}$ is the L^2 inner product on $\mathcal{M}$ with f_n being orthonormal basis.

Since B_i's share the same eigenfunctions and hence commute, we have $d_l = \frac{\gamma_l^2}{\gamma_l^1} - 1 = \frac{\rho_2(l)}{\rho_1(l)} - 1 > -1$, so condition 2 holds by the definition of ρ_i. For condition 1, observe that $d_l = \frac{\rho_2(l) - \rho_1(l)}{\rho_1(l)}$, so

$$D \text{ is Hilbert–Schmidt} \iff \sum_l d_l^2 < \infty \iff \sum_l \left| \frac{\rho_2(l) - \rho_1(l)}{\rho_1(l)} \right|^2 < \infty.$$

∎

Appendix B. Proof of Theorem 4

Proof We start with (A). First assume that $\nu_1 = \nu_2 = \nu$ and $\sigma_1^2/C_{\nu,\alpha_1} = \sigma_2^2/C_{\nu,\alpha_2}$, then observe

$$
\begin{aligned}
\left| \frac{\rho_2(l) - \rho_1(l)}{\rho_1(l)} \right| &= \left| \frac{(\alpha_1^2 + \lambda_l)^{\nu+d/2}}{(\alpha_2^2 + \lambda_l)^{\nu+d/2}} - 1 \right| \\
&\leq \left| (\alpha_1^2 + \lambda_l)^{\nu+d/2} - (\alpha_2^2 + \lambda_l)^{\nu+d/2} \right| / \lambda_l^{\nu+d/2} \\
&\leq \left| ((\alpha_1^2/\lambda_l + 1)^{\nu+d/2} - ((\alpha_2^2/\lambda_l + 1)^{\nu+d/2} \right|.
\end{aligned}
$$

Note that $(1/x+1)^a = 1 + a/x + O(x^{-2})$ as $x \to \infty$, then when l is sufficiently large so that $\lambda_l > 0$,

$$
\left| (\alpha_1^2/\lambda_l + 1)^{\nu+d/2} - (\alpha_2^2/\lambda_l + 1)^{\nu+d/2} \right| \leq (\nu + d/2)(\alpha_1^2 - \alpha_2^2)\lambda_l^{-1} + O(\lambda_l^{-2}) = O(\lambda_l^{-1}).
$$

As a result,

$$
\sum_l \left| \frac{\rho_2(l) - \rho_1(l)}{\rho_1(l)} \right|^2 \lesssim \sum_l \lambda_l^{-2}.
$$

By Weyl's law (equation (4.1) in Grebenkov and Nguyen (2013)), $\lambda_l \sim l^{2/d}$, so we have $\lambda_l^{-2} \sim l^{-d/4}$ hence $-4/d < -1$ when $d \leq 3$. By the series test in Lemma 3, $P_1 \equiv P_2$.

For the other direction, observe that

$$
\left| \frac{\rho_2(l)}{\rho_1(l)} \right| = \left| \frac{\sigma_2^2 C_{\nu_1,\alpha_1}(\alpha_1^2 + \lambda_l)^{\nu_1+d/2}}{\sigma_1^2 C_{\nu_2,\alpha_2}(\alpha_2^2 + \lambda_l)^{\nu_2+d/2}} \right| \to \begin{cases} \infty & \nu_1 < \nu_2 \\ 1 & \nu_1 = \nu_2 \\ 0 & \nu_1 > \nu_2. \end{cases}
$$

As a result, if $\nu_1 \neq \nu_2$, $\sum_l \left| \frac{\rho_2(l)}{\rho_1(l)} - 1 \right| \to \infty$ so $P_1 \not\equiv P_2$ by the series test.

Then assume $\nu_1 = \nu_2 = \nu$ and $\sigma_1^2/C_{\nu,\alpha_1} \neq \sigma_2^2/C_{\nu,\alpha_2}$. Let $\sigma_0^2 = \sigma_2^2 \frac{C_{\nu,\alpha_1}}{C_{\nu,\alpha_2}} \neq \sigma_1^2$, then

$$
\sigma_0^2/C_{\nu,\alpha_1} = \sigma_2^2/C_{\nu,\alpha_2},
$$

so $k(\cdot; \sigma_0^2, \alpha_1)$ and $k(\cdot; \sigma_2^2, \alpha_2)$ define two equivalent measures, denoted by P_0 and P_2. Observe that

$$
k(x, y; \sigma_1^2, \alpha_1) = \frac{\sigma_1^2}{\sigma_0^2} k(x, y; \sigma_0^2, \alpha_1),
$$

then the corresponding spectral densities ρ_0 and ρ_1 only differ by a multiplicative scalar $\frac{\sigma_1^2}{\sigma_0^2}$ so $\sum_l \left| \frac{\rho_1(l)-\rho_0(l)}{\rho_1(l)} \right|^2 = \sum_l \left| \frac{\sigma_1^2-\sigma_0^2}{\sigma_1^2} \right|^2 = \infty$. So by Lemma 3, P_0 is orthogonal to P_1, so is P_2, which is equivalent to P_0. Now we conclude that $P_1 \equiv P_2$ if and only if $\sigma_1^2/C_{\nu,\alpha_1} = \sigma_2^2/C_{\nu,\alpha_2}$ and $\nu_1 = \nu_2$.

Then we show (B). As proved in (A), $P_1 \not\equiv P_2$ if $\nu_1 \neq \nu_2$ so we assume $\nu_1 = \nu_2 = \nu$. Recall that $\lambda_n \to \infty$, so when n is sufficiently large, $\lambda_n > \alpha^2$, then

$$
\begin{aligned}
\left| \frac{\rho_2(l) - \rho_1(l)}{\rho_1(l)} \right| &= \left| \frac{\sigma_2^2 C_{\nu,\alpha_1}(\alpha_1^2 + \lambda_l)^{\nu+d/2}}{\sigma_1^2 C_{\nu,\alpha_2}(\alpha_2^2 + \lambda_l)^{\nu+d/2}} - 1 \right| \\
&\geq \frac{\left| \sigma_2^2 C_{\nu,\alpha_1}(\alpha_1^2 + \lambda_l)^{\nu+d/2} - \sigma_1^2 C_{\nu,\alpha_2}(\alpha_2^2 + \lambda_l)^{\nu+d/2} \right|}{\sigma_1^2 C_{\nu,\alpha_2}(2\lambda_l)^{\nu+d/2}} \\
&= 2^{-\nu-d/2} \left| \frac{\sigma_2^2 C_{\nu,\alpha_1}}{\sigma_1^2 C_{\nu,\alpha_2}} \left(\alpha_1^2/\lambda_l + 1 \right)^{\nu+d/2} - \left(\alpha_2^2/\lambda_l + 1 \right)^{\nu+d/2} \right| \\
&= 2^{-\nu-d/2} \left| \frac{\sigma_2^2 C_{\nu,\alpha_1}}{\sigma_1^2 C_{\nu,\alpha_2}} - 1 + \left(\nu + \frac{d}{2} \right) \left(\frac{\sigma_2^2 C_{\nu,\alpha_1}}{\sigma_1^2 C_{\nu,\alpha_2}} \alpha_1^2 - \alpha_2^2 \right) \lambda_l^{-1} + O(\lambda_l^{-2}) \right|.
\end{aligned}
\tag{5}
$$

When $\sigma_1^2 \neq \sigma_2^2$ or $\alpha_1 \neq \alpha_2$, the constant term $\frac{\sigma_2^2 C_{\nu,\alpha_1}}{\sigma_1^2 C_{\nu,\alpha_2}} - 1$ and the linear coefficient $\frac{\sigma_2^2 C_{\nu,\alpha_1}}{\sigma_1^2 C_{\nu,\alpha_2}} \alpha_1^2 - \alpha_2^2$ in Equation (5) do not vanish at the same time hence $\left| \frac{\rho_2(l)-\rho_1(l)}{\rho_1(l)} \right| \gtrsim \lambda_l^{-1}$. Then

$$
\sum_l \left| \frac{\rho_2(l) - \rho_1(l)}{\rho_1(l)} \right|^2 \gtrsim \sum_l \lambda_l^{-2} = \sum_l l^{-4/d} = \infty
$$

since $d \geq 4$. By the series test, $P_1 \not\equiv P_2$. When $\sigma_1^2 = \sigma_2^2$ and $\alpha_1 = \alpha_2$, $P_1 = P_2$ so $P_1 \equiv P_2$, which finises the proof of (B). ∎

Appendix C. Proof of Theorem 5

Proof First assume $\alpha_1 \neq \alpha_2$, or $\alpha_1 < \alpha_2$ without loss of generality, then

$$
\left| \frac{\rho_2(l) - \rho_1(l)}{\rho_1(l)} \right| = \left| \frac{\sigma_2^2 C_{\alpha_1}}{\sigma_1^2 C_{\alpha_2}} e^{-\frac{\lambda_l}{2} \left(\frac{1}{\alpha_2^2} - \frac{1}{\alpha_1^2} \right)} - 1 \right| \to \infty
$$

since $\lambda_l \to \infty$. As a result,

$$
\sum_l \left| \frac{\rho_2(l) - \rho_1(l)}{\rho_1(l)} \right|^2 = \infty.
$$

Then assume $\alpha_1 = \alpha_2$ but $\sigma_1^2 \neq \sigma_2^2$, similarly,

$$
\sum_l \left| \frac{\rho_2(l) - \rho_1(l)}{\rho_1(l)} \right|^2 = \sum_l \left(\frac{\sigma_2^2}{\sigma_1^2} - 1 \right)^2 = \infty.
$$

Then the series test applies. ∎

Appendix D. Proof of Theorem 6

Proof Let $\sigma_1^2 = \frac{\sigma_0^2 C_{\nu,\alpha_1}}{C_{\nu,\alpha_0}}$ so $P_0 \equiv P_1$ by Theorem 4. It suffices to show $\widehat{\sigma}_{1,n}^2 \to \sigma_1^2$, P_1 a.s. Recall that $\widehat{\sigma}_{1,n}^2 = \frac{Z_n^{\mathrm{T}} \Gamma_n^{-1}(\alpha_1) Z_n}{n}$ and $Z_n \sim N(0, \sigma_1^2 \Gamma_n(\alpha_1))$ under P_1, where $(\Gamma_n(\alpha))_{i,j} = \frac{1}{C_{\nu,\alpha}} \sum_{l=0}^{\infty} \left(2\nu\alpha^2 + \lambda_l\right)^{-\nu-\frac{d}{2}} f_l(x_i) f_l(x_j)$. As a result, $\widehat{\sigma}_{1,n}^2 = \sigma_1^2 \frac{\chi_n^2}{n} \to \sigma_1^2$, P_1 a.s., as $n \to \infty$. $\blacksquare$

Appendix E. Proof of Lemma 7

Proof The logic of the proof is similar to the proof of Theorem 1 and 2 in Stein (1993). However, these two theorems are not directly applicable due to the discreteness of spectrum in our case. To be more specific, the key construction in the proof of Stein (1993) is the following. By the assumption, for any $\varepsilon > 0$, there exists $M_\varepsilon > 0$ such that $\sup_{m \geq M_\varepsilon} \left| \frac{\rho_1(m)}{c\rho_0(m)} - 1 \right| < \varepsilon$. We define

$$\eta_\varepsilon(m) := \begin{cases} \frac{1}{c}\rho_1(m) & m \leq M_\varepsilon \\ \rho_0(m) & m > M_\varepsilon \end{cases}.$$

That is, η_ε differs from ρ_0 only on a bounded subset of $\mathbb{N}$. Note that in Stein (1993), the key step is to show $P_{\eta_\varepsilon} \equiv P_{\rho_0}$, and the rest of the proof will not rely on any special structure of the Euclidean domain anymore. That is, it suffices to show $P_{\eta_\varepsilon} \equiv P_{\rho_0}$, which is a direct consequence of the series test in Lemma 3. The rest of the proof of (i) naturally follows the proof of Theorem 1 in Stein (1993) while the proof of (ii) follows the proof of Theorem 2 in Stein (1993), where $e(x_0, n, f_1)$ in Stein (1993) corresponds to $Z_0 - \widehat{Z}_n(\rho_1)$ in our paper. $\blacksquare$

Appendix F. Proof of Theorem 8

Proof For $\sigma_1^2 = \frac{\sigma_0^2 C_{\nu,\alpha_1}}{C_{\nu,\alpha_0}}$, let ρ_1 and ρ_0 be the spectral density of the Gaussian process parametrised by (α_1, σ_1^2) and (α_0, σ_0^2) hence $\rho_1/\rho_0 \to 1$. Then by (ii) in Lemma 7,

$$\frac{\mathbb{E}_{\sigma_1^2,\alpha_1}(\widehat{Z}_n(\alpha_1) - Z_0)^2}{\mathbb{E}_{\sigma_0^2,\alpha_0}(\widehat{Z}_n(\alpha_1) - Z_0)^2} \to 1.$$

Observe that

$$\frac{\mathbb{E}_{\widehat{\sigma}_{1,n}^2,\alpha_1}(\widehat{Z}_n(\alpha_1) - Z_0)^2}{\mathbb{E}_{\sigma_0^2,\alpha_0}(\widehat{Z}_n(\alpha_1) - Z_0)^2} = \frac{\mathbb{E}_{\widehat{\sigma}_{1,n}^2,\alpha_1}(\widehat{Z}_n(\alpha_1) - Z_0)^2}{\mathbb{E}_{\sigma_1^2,\alpha_1}(\widehat{Z}_n(\alpha_1) - Z_0)^2} \frac{\mathbb{E}_{\sigma_1^2,\alpha_1}(\widehat{Z}_n(\alpha_1) - Z_0)^2}{\mathbb{E}_{\sigma_0^2,\alpha_0}(\widehat{Z}_n(\alpha_1) - Z_0)^2}. \tag{6}$$

The second term in Equation (6) tends to 1. For the first term, by the definition of $\widehat{Z}_n$, we obtain

$$\mathbb{E}_{\widehat{\sigma}_{1,n}^2,\alpha_1}(\widehat{Z}_n(\alpha_1) - Z_0)^2 = \widehat{\sigma}_{1,n}^2 \left(1 - \gamma_n(\alpha_1)^{\mathrm{T}} \Gamma_n(\alpha_1)^{-1} \gamma_n(\alpha_1)\right).$$

Hence, the first term in Equation (6) is $\frac{\hat{\sigma}_{1,n}^2}{\sigma_1^2}$. Similar to the proof of Theorem 6, $\hat{\sigma}_{1,n}^2 = \sigma_1^2 \frac{\chi_n^2}{n} \to \sigma_1^2$, $P_1 := P_{\sigma_1^2, \alpha_1}$ a.s. By Theorem 4, $P_0 \equiv P_1$, so the left hand side of Equation (6) tends to 1, P_0 a.s. ∎

Appendix G. Proof of Theorem 9

Proof d-dimensional spheres are compact Riemannian manifolds. The eigenfunctions of the Laplace operator on S^d are known as spherical harmonics, denoted by S_m^l, $m = 0, 1, \cdots$, $l = 1, \cdots, t_d(m)$. The corresponding eigenvalues are $l(l + d - 1) = O(l^2)$ with multiplicity (Müller, 1966; Efthimiou and Frye, 2014)

$$\frac{2l + d - 1}{l}\binom{l + d - 2}{l - 1} = O(l^{d-1}).$$

So 1, 2, 3 follow directly from Theorem 4, 6 and 8 respectively. ∎

Appendix H. Proof of Theorem 15

Proof First we reformulate the covariogram as

$$k(x, y) = \frac{\sigma^2}{C_{\nu,\alpha}} \sum_{l=0}^{\infty} \left(\alpha^2 + l(l + 1)\right)^{-\nu-1} c_l \mathcal{L}_l(\cos(d_M(x, y))) = C \sum_{l=0}^{\infty} a_l(z),$$

where $C = \frac{\Gamma(3/2)\sigma^2}{2\pi^{3/2} C_{\nu,\alpha}} = \frac{4\pi\sigma^2}{\sum_l (2l+1)(\alpha^2+l(l+1))^{-\nu-1}}$, $z = \cos(d_M(x, y))$ and

$$a_l(z) = \left(\alpha^2 + l(l + 1)\right)^{-\nu-1}(2l + 1)\mathcal{L}_l(z).$$

Observe that $\mathcal{L}_l(z) \in [-1, 1]$. Therefore,

$$|a_l(z)| \leq \left(\alpha^2 + l(l + 1)\right)^{-\nu-1}(2l + 1).$$

As a result,

$$|k^L(x, y) - k(x, y)| \leq C \sum_{l=L+1}^{\infty} |a_l(z)| \leq C \sum_{l=L+1}^{\infty} \left(\alpha^2 + l(l + 1)\right)^{-\nu-1}(2l + 1)$$

$$\leq C \sum_{l=L+1}^{\infty} (l^2)^{-\nu-1}(3l) = 3C \sum_{l=L+1}^{\infty} l^{-2\nu-1} \leq 3C \int_{L+1}^{\infty} t^{-2\nu-1} dt$$

$$\leq \frac{3C}{2\nu} L^{-2\nu} = \frac{6\pi\sigma^2}{\nu \sum_l (2l + 1)(\alpha^2 + l(l + 1))^{-\nu-1}} L^{-2\nu}.$$

That is, if the target approximation error is ϵ, then we can truncate the infinite sum at

$$L = \left\lfloor \left(\frac{6\pi\sigma^2}{\epsilon\nu \sum_l (2l + 1)(\alpha^2 + l(l + 1))^{-\nu-1}}\right)^{\frac{1}{2\nu}} \right\rfloor + 1.$$

To prove positive definiteness, we first find the lower bound of the minimal eigenvalue of the covariance matrix $\Sigma := \{k(x_i, x_j)\}_{ij}$, denoted by $\lambda_{\min}$. By Theorem 2.8 (i) in Narcowich et al. (1998),

$$\lambda_{\min} \geq \xi_\rho(q) := \Gamma_\rho(K) \left(1 - \frac{\pi^3 k(q)}{4q} \left(\frac{\sin(q/2)}{q/2}\right)^K\right),$$

where

$$k(q) = \max\left\{\frac{8\pi^2}{q}, \frac{25\pi}{2}\right\}, \quad K = \operatorname*{arg\,min}_{m \in \mathbb{N}} \left\{m : 1 - \frac{\pi^3 k(q)}{4q} \left(\frac{\sin(q/2)}{q/2}\right)^m > 0\right\},$$

and $\Gamma_\rho(K)$ is determined by the spectral density ρ, K and the B-spline, see Equation (2.41) in Narcowich et al. (1998) for further details (where $m = 2$ in our setting). Let the truncated covariance function be $\Sigma^L := \{k^L(x_i, x_j)\}_{ij}$ with minimal eigenvalue $\lambda_{\min}^N$, then by the first half of the proof,

$$\lambda_{\min}^L \geq \lambda_{\min} - \|\Sigma - \Sigma^L\| \geq \xi_\rho(q) - n\|\Sigma - \Sigma^L\|_{\max} \geq \xi_\rho(q) - \frac{6\pi n\sigma^2 L^{-2\nu}}{\nu \sum_l (2l+1)(\alpha^2 + l(l+1))^{-\nu-1}}$$

The second inequality follows from a matrix norm equivalence: $\|A\|_{\max} \leq \|A\| \leq n\|A\|_{\max}$ for any $n \times n$ matrix A. The first inequality relies on the fact that $\operatorname{eig}_{\min}(A) \geq \operatorname{eig}_{\min}(B) - \|A - B\|$ for symmetric matrices A and B. Note that $\operatorname{eig}_{\min}(A) = \min_{\|x\|=1} x^\top A x$ and let x_0 be the eigenvector of A associated with the smallest eigenvalue, that is, $Ax_0 = \operatorname{eig}_{\min}(A)x_0$. By the same observation, $x_0^\top B x_0 \geq \operatorname{eig}_{\min}(B)$. Then,

$$\operatorname{eig}_{\min}(A) = x_0^\top A x_0 = x_0^\top (B + A - B)x_0 = x_0^\top B x_0 + x_0^\top (A-B)x_0 \geq \operatorname{eig}_{\min}(B) + x_0^\top (A-B)x_0.$$

For the last term, since $\|A - B\| = \max_{\|x\|=1} |x^\top (A-B)x|$, we have $|x_0^\top (A-B)x_0| \leq \|A-B\|$, hence $x_0^\top (A-B)x_0 \geq -\|A-B\|$ as desired. Let $\xi_\rho(q) - \frac{6\pi n\sigma^2 L^{-2\nu}}{\nu \sum_l (2l+1)(\alpha^2 + l(l+1))^{-\nu-1}} > 0$, we have

$$L > \left(\frac{6\pi n\sigma^2}{\nu \xi_\rho(q) \sum_l (2l+1)(\alpha^2 + l(l+1))^{-\nu-1}}\right)^{\frac{1}{2\nu}}$$

∎

References

M. Abramowitz and A. Stegun. *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables.* Dover, 1965.

A. Alegría, F. Cuevas-Pacheco, P. Diggle, and E. Porcu. The $\mathcal{F}$-family of covariance functions: a Matérn analogue for modeling random fields on spheres. *Spat. Stat.*, 43:Paper No. 100512, 25, 2021.

Ethan Anderes. On the consistent separation of scale and variance for Gaussian random fields. *Annals of Statistics*, 38(2):870–893, 2010.

Ahmed Arafat, Emilio Porcu, Moreno Bevilacqua, and Jorge Mateu. Equivalence and orthogonality of Gaussian measures on spheres. *Journal of Multivariate Analysis*, 167: 306–318, 2018.

Sudipto Banerjee. On geodetic distance computations in spatial modeling. *Biometrics*, 61 (2):617–625, 2005.

Sudipto Banerjee, Alan E. Gelfand, Andrew O. Finley, and Huiyan Sang. Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Methodology)*, 70(4):825–848, 2008.

Sudipto Banerjee, Bradley P. Carlin, and Alan E. Gelfand. *Hierarchical modeling and analysis for spatial data*, volume 135 of *Monographs on Statistics and Applied Probability*. CRC Press, Boca Raton, FL, second edition, 2015.

Mikhail Belkin and Partha Niyogi. Convergence of Laplacian eigenmaps. In *NIPS*, pages 129–136, 2007.

Moreno Bevilacqua, Tarik Faouzi, Reinhard Furrer, and Emilio Porcu. Estimation and prediction using generalized Wendland covariance functions under fixed domain asymptotics. *Annals of Statistics*, 47(2):828–856, 2019.

David Bolin and Kristin Kirchner. Equivalence of measures and asymptotically optimal linear prediction for Gaussian random fields with fractional-order covariance operators. *arXiv*, 2021. arXiv:2101.07860.

David Bolin and Finn Lindgren. Spatial models generated by nested stochastic partial differential equations, with an application to global ozone mapping. *Annals of Applied Statistics*, 5(1):523–550, 2011.

Viacheslav Borovitskiy, Alexander Terenin, Peter Mostowsky, and Marc Deisenroth. Matérn Gaussian processes on Riemannian manifolds. In *NIPS*, pages 12426–12437, 2020.

Viacheslav Borovitskiy, Iskander Azangulov, Alexander Terenin, Peter Mostowsky, Marc Deisenroth, and Nicolas Durrande. Matérn Gaussian processes on graphs. In *AISTATS*, pages 2593–2601, 2021.

Yaiza Canzani. Analysis on manifolds via the Laplacian, 2013. URL `https://www.math.mcgill.ca/toth/spectral%20geometry.pdf`.

Ismaël Castillo, Gérard Kerkyacharian, and Dominique Picard. Thomas Bayes' walk on manifolds. *Probab. Theory Related Fields*, 158(3-4):665–710, 2014.

Jorge Clarke De la Cerda, Alfredo Alegría, and Emilio Porcu. Regularity properties and simulations of Gaussian random fields on the sphere cross time. *Electronic Journal of Statistics*, 12(1):399–426, 2018.

Noel Cressie and Christopher K. Wikle. *Statistics for spatio-temporal data*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ, 2011.

Giuseppe Da Prato and Jerzy Zabczyk. *Stochastic equations in infinite dimensions*, volume 152 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, second edition, 2014.

Andreas Damianou and Neil D Lawrence. Deep Gaussian processes. In *AISTATS*, pages 207–215, 2013.

A. Datta, S. Banerjee, A. O. Finley, N. A. S. Hamm, and M. Schaap. Non-separable dynamic nearest-neighbor gaussian process models for large spatio-temporal data with an application to particulate matter analysis. *Annals of Applied Statistics*, 10:1286–1316, 2016a.

Abhirup Datta, Sudipto Banerjee, Andrew O. Finley, and Alan E. Gelfand. Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. *Journal of the American Statistical Association*, 111(514):800–812, 2016b.

Marc Peter Deisenroth, Dieter Fox, and Carl Edward Rasmussen. Gaussian processes for data-efficient learning in robotics and control. *IEEE Transactions in Pattern Analysis and Machine Intelligence*, 37(2):408–423, 2013.

Manfredo Perdigão do Carmo. *Riemannian geometry.* Mathematics: Theory & Applications. Birkhäuser Boston, Inc., Boston, MA, 1992.

Juan Du, Hao Zhang, and V. S. Mandrekar. Fixed-domain asymptotic properties of tapered maximum likelihood estimators. *Annals of Statistics*, 37(6A):3330–3361, 2009.

David B. Dunson, Hau-Tieng Wu, and Nan Wu. Diffusion based Gaussian processes on restricted domains. *arXiv*, 2020. arXiv:2010.07242.

Costas Efthimiou and Christopher Frye. *Spherical harmonics in p dimensions.* World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2014.

Aasa Feragen, François Lauze, and Søren Hauberg. Geodesic exponential kernels: When curvature and linearity conflict. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3032–3042, 2015. doi: 10.1109/CVPR.2015.7298922.

Reinhard Furrer, Marc G. Genton, and Douglas Nychka. Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics*, 15(3): 502–523, 2006.

Tingran Gao, Shahar Z Kovalsky, and Ingrid Daubechies. Gaussian process landmarking on manifolds. *SIAM Journal of Mathematics of Data Science*, 1(1):208–236, 2019.

Alan E. Gelfand, Peter J. Diggle, Montserrat Fuentes, and Peter Guttorp, editors. *Handbook of spatial statistics.* Chapman & Hall/CRC Handbooks of Modern Statistical Methods. CRC Press, Boca Raton, FL, 2010.

Marc G. Genton. Classes of kernels for machine learning: a statistics perspective. *Journal of Machine Learning Research*, 2(2):293–312, 2002.

Subhashis Ghosal and Aad van der Vaart. *Fundamentals of nonparametric Bayesian inference*, volume 44 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 2017.

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

Denis S Grebenkov and B-T Nguyen. Geometrical structure of laplacian eigenfunctions. *SIAM Review*, 55(4):601–667, 2013.

Jean Carlo Guella, Valdir Antonio Menegatto, and Emilio Porcu. Strictly positive definite multivariate covariance functions on spheres. *Journal of Multivariate Analysis*, 166:150–159, 2018.

Joseph Guinness and Montserrat Fuentes. Isotropic covariance functions on spheres: some properties and modeling considerations. *Journal of Multivariate Analysis*, 143:143–152, 2016.

Lukas Herrmann, Kristin Kirchner, and Christoph Schwab. Multilevel approximation of Gaussian random fields: Fast simulation. *Mathematical Models and Methods in Applied Sciences*, 30(01):181–223, 2020.

Jaehong Jeong and Mikyoung Jun. A class of Matérn-like covariance functions for smooth processes on a sphere. *Spatial Statistics*, 11:1–18, 2015a.

Jaehong Jeong and Mikyoung Jun. Covariance models on the surface of a sphere: when does it matter? *Stat*, 4(1):167–182, 2015b.

Mikyoung Jun and Michael L. Stein. Nonstationary covariance models for global data. *Annals of Applied Statistics*, 2(4):1271–1289, 2008.

C. G. Kaufman and B. A. Shaby. The role of the range parameter for estimation and prediction in geostatistics. *Biometrika*, 100(2):473–484, 2013.

Cari G. Kaufman, Mark J. Schervish, and Douglas W. Nychka. Covariance tapering for likelihood-based estimation in large spatial data sets. *Journal of the American Statistical Association*, 103(484):1545–1555, 2008.

Kristin Kirchner and David Bolin. Necessary and sufficient conditions for asymptotically optimal linear prediction of random fields on compact metric spaces. *Annals of Statistics*, 50(2):1038–1065, 2022.

Shoshichi Kobayashi and Katsumi Nomizu. *Foundations of differential geometry. Vol I*. Interscience Publishers, New York-London, 1963.

Stefan Kunis. A note on stability results for scattered data interpolation on euclidean spheres. *Advances in Computational Mathematics*, 30(4):303–314, 2009.

Annika Lang and Christoph Schwab. Isotropic Gaussian random fields on the sphere: regularity, fast simulation and stochastic partial differential equations. *Annals of Applied Probability*, 25(6):3047–3094, 2015.

John M. Lee. *Introduction to Riemannian manifolds*, volume 176 of *Graduate Texts in Mathematics*. Springer, Cham, 2018.

Peter Li. Book Review: Eigenvalues in Riemannian geometry. *Bulletin of the American Mathematical Society (N.S.)*, 16(2):324–325, 1987.

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

Pulong Ma and Anindya Bhadra. Beyond matérn: On a class of interpretable confluent hypergeometric covariance functions. *Journal of the American Statistical Association*, 0 (0):1–14, 2022. doi: 10.1080/01621459.2022.2027775.

Bertil Matérn. *Spatial variation*, volume 36 of *Lecture Notes in Statistics*. Springer-Verlag, Berlin, second edition, 1986.

Georges Matheron. Principles of geostatistics. *Econ. Geol.*, 58(8):1246–1266, 1963.

Claus Müller. *Spherical harmonics*, volume 17 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin-New York, 1966.

Francis J. Narcowich, N. Sivakumar, and Joseph D. Ward. Stability results for scattered-data interpolation on euclidean spheres. *Advances in Computational Mathematics*, 8(3): 137–163, 1998.

Radford M. Neal. Regression and classification using Gaussian process priors. In *Bayesian Statistics, 6 (Alcoceber, 1998)*, pages 475–501. Oxford Univ. Press, New York, 1999.

Mu Niu, Pokman Cheung, Lizhen Lin, Zhenwen Dai, Neil Lawrence, and David Dunson. Intrinsic Gaussian processes on complex constrained domains. *Journal of the Royal Statistical Society: Series B (Methodology)*, 81(3):603–627, 2019.

Emilio Porcu, Moreno Bevilacqua, and Marc G. Genton. Spatio-temporal covariance and cross-covariance functions of the great circle distance on a sphere. *Journal of the American Statistical Association*, 111(514):888–898, 2016.

Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian processes for machine learning*. Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA, 2006.

Daniel Sanz-Alonso and Ruiyi Yang. The SPDE Approach to Matérn Fields: Graph Representations. *Statist. Sci.*, 37(4):519–540, 2022a.

Daniel Sanz-Alonso and Ruiyi Yang. Finite Element Representations of Gaussian Processes: Balancing Numerical and Statistical Accuracy. *SIAM/ASA Journal of Uncertainty Quantification*, 10(4):1323–1349, 2022b.

Arno Solin and Manon Kok. Know your boundaries: Constraining Gaussian processes by variational harmonic features. In *AISTATS*, pages 2193–2202, 2019.

Michael L. Stein. A simple condition for asymptotic optimality of linear predictions of random fields. *Statistics and Probability Letters*, 17(5):399–404, 1993.

Michael L. Stein. *Interpolation of spatial data.* Springer Series in Statistics. Springer-Verlag, New York, 1999. Some theory for Kriging.

Wenpin Tang, Lu Zhang, and Sudipto Banerjee. On identifiability and consistency of the nugget in Gaussian spatial process models. *Journal of the Royal Statistical Society: Series B (Methodology)*, 83(5):1044–1070, 2021.

Daqing Wang. Fixed domain asymptotics and consistent estimation for Gaussian random field models in spatial statistics and computer experiments. *Technical Report: National University of Singapore*, 2010.

Daqing Wang and Wei-Liem Loh. On fixed-domain asymptotics and covariance tapering in Gaussian random field models. *Electronic Journal of Statistics*, 5:238–269, 2011.

P. Whittle. Stochastic processes in several dimensions. *Bulletin of the International Statistical Institute*, 40:974–994, 1963.

M. I. Yadrenko. *Spectral theory of random fields.* Optimization Software, Inc., Publications Division, New York, 1983.

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