



Asymptotic properties of the maximum likelihood and cross validation estimators for transformed Gaussian processes

François Bachoc, José Bétancourt, Reinhard Furrer, Thierry Klein

► To cite this version:

François Bachoc, José Bétancourt, Reinhard Furrer, Thierry Klein. Asymptotic properties of the maximum likelihood and cross validation estimators for transformed Gaussian processes. 2019. hal-02388266

HAL Id: hal-02388266

<https://hal-enac.archives-ouvertes.fr/hal-02388266>

Submitted on 1 Dec 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Asymptotic properties of the maximum likelihood and cross validation estimators for transformed Gaussian processes

Francois Bachoc¹, José Bétancourt², Reinhard Furrer³ and Thierry Klein⁴

¹*Institut de mathématique, UMR5219; Université de Toulouse; CNRS, UPS IMT, F-31062 Toulouse Cedex 9, France; e-mail: francois.bachoc@math.univ-toulouse.fr*

²*Institut de mathématique, UMR5219; Université de Toulouse; CNRS, UPS IMT, F-31062 Toulouse Cedex 9, France;
ENAC - Ecole Nationale de l'Aviation Civile, Université de Toulouse, France; e-mail: jose.betancourt@math.univ-toulouse.fr*

³*Department of Mathematics and Department of Computational Science, University of Zurich, CH-8057 Zurich; e-mail: reinhard.furrer@math.uzh.ch*

⁴*ENAC - Ecole Nationale de l'Aviation Civile, Université de Toulouse, France; Institut de mathématique, UMR5219; Université de Toulouse; e-mail: thierry.klein@math.univ-toulouse.fr*

Abstract: The asymptotic analysis of covariance parameter estimation of Gaussian processes has been subject to intensive investigation. However, this asymptotic analysis is very scarce for non-Gaussian processes. In this paper, we study a class of non-Gaussian processes obtained by regular non-linear transformations of Gaussian processes. We provide the increasing-domain asymptotic properties of the (Gaussian) maximum likelihood and cross validation estimators of the covariance parameters of a non-Gaussian process of this class. We show that these estimators are consistent and asymptotically normal, although they are defined as if the process was Gaussian. They do not need to model or estimate the non-linear transformation. Our results can thus be interpreted as a robustness of (Gaussian) maximum likelihood and cross validation towards non-Gaussianity. Our proofs rely on two technical results that are of independent interest for the increasing-domain asymptotic literature of spatial processes. First, we show that, under mild assumptions, coefficients of inverses of large covariance matrices decay at an inverse polynomial rate as a function of the corresponding observation location distances. Second, we provide a general central limit theorem for quadratic forms obtained from transformed Gaussian processes. Finally, our asymptotic results are illustrated by numerical simulations.

AMS 2000 subject classifications: Primary 62M30; secondary 62F12.

Keywords and phrases: covariance parameters, asymptotic normality, consistency, weak dependence, random fields, increasing-domain asymptotics.

1. Introduction

Kriging [41, 34] consists of inferring the values of a (Gaussian) random field given observations at a finite set of points. It has become a popular method for a large range of applications, such as geostatistics [31], numerical code approximation [35, 36, 9], calibration [33, 10], global optimization [27], and machine learning [34].

When considering a Gaussian process, one has to deal with the estimation of its covariance function. Usually, it is assumed that the covariance function belongs to a given parametric family (see [1] for a review of classical families). In this case, the estimation boils down to estimating the corresponding covariance parameters. Nowadays, the main estimation techniques are based on maximum likelihood [41, 34], cross-validation [48, 6, 7, 13] and variation estimators [25, 3, 4].

The asymptotic properties of estimators of the covariance parameters have been widely studied in the two following frameworks. The fixed-domain asymptotic framework, sometimes called infill asymptotics [41, 18], corresponds to the case where more and more data are observed in some fixed bounded sampling domain. The increasing-domain asymptotic framework corresponds to the case where the sampling domain increases with the number of observed data.

Under fixed-domain asymptotics, and particularly in low dimensional settings, not all covariance parameters can be estimated consistently (see [24, 41]). Hence, the distinction is made between microergodic and non-microergodic covariance parameters [24, 41]. Although non-microergodic parameters cannot be estimated consistently, they have an asymptotically negligible impact on prediction [38, 39, 40, 47]. There is, however, a fair amount of literature on the consistent estimation of microergodic parameters (see for instance [47, 28, 20, 42, 45, 46]).

This paper focuses on the increasing-domain asymptotic framework. Indeed, generally speaking, increasing-domain asymptotic results hold for significantly more general families of covariance functions than fixed-domain ones. Under increasing-domain asymptotics, the maximum likelihood and cross validation estimators of the covariance parameters are consistent and asymptotically normal under mild regularity conditions [30, 37, 7, 23].

All the asymptotic results discussed above are based on the assumption that the data come from a Gaussian random field. This assumption is indeed theoretically convenient but might be unrealistic for real applications. When the data stem from a non-Gaussian random field, it is still relevant to estimate the covariance function of this random field. Hence, it would be valuable to extend the asymptotic results discussed above to the problem of estimating the covariance parameters of a non-Gaussian random field.

In this paper, we provide such an extension, in the special case where the non-Gaussian random field is a deterministic (unknown) transformation of a Gaussian random field. Models of transformed Gaussian random fields have been used extensively in practice (for example in [17, 43, 2, 44]).

Under reasonable regularity assumptions, we prove that applying the (Gaussian) maximum likelihood estimator to data from a transformed Gaussian random field yields a consistent and asymptotically normal estimator of the covariance parameters of the transformed random field. This (Gaussian) maximum likelihood estimator corresponds to what would typically be done in practice when applying a Gaussian process model to a non-Gaussian spatial process. This estimator does not need to know the existence of the non-linear transformation function and is not based on the exact density of the non-Gaussian data. We refer to Remark 2 for further details and discussion on this point.

We then obtain the same consistency and asymptotic normality result when considering a cross validation estimator. In addition, we establish the joint asymptotic normality of both these estimators, which provides the asymptotic distribution of a large family of aggregated estimators. Our asymptotic results on maximum likelihood and cross validation are illustrated by numerical simulations.

To the best of our knowledge, our results (Theorems 3, 4, 5, 6 and 7) provide the first increasing-domain asymptotic analysis of Gaussian maximum likelihood and cross validation for non-Gaussian random fields. Our proofs intensively rely on Theorems 1 and 2. Theorem 1 shows that the components of inverse covariance matrices are bounded by inverse polynomial functions of the corresponding distance between observation locations. Theorem 2 provides a generic central limit theorem for quadratic forms constructed from transformed Gaussian processes. These two theorems have an interest in themselves.

The rest of the paper is organized as follows. In Section 2, general properties of transformed Gaussian processes are provided. In Section 3, Theorems 1 and 2 are stated. In Section 4, an application of these two theorems is given to the case of estimating a single variance parameter. In Section 5, the consistency and asymptotic normality results for general covariance parameters are given. The joint asymptotic normality result is also given in this section. The simulation results are provided in Section 6. All the proofs are provided in the appendix.

2. General properties of transformed Gaussian processes

In applications, the use of Gaussian process models may be too restrictive. One possibility for obtaining larger and more flexible classes of random fields is to consider transformations of Gaussian processes. In this section, we now introduce the family of transformed Gaussian processes that we will study asymptotically in this paper. This family is determined by regularity conditions on the covariance function of the original Gaussian process and on the transformation function.

Let us first introduce some notation. Throughout the paper, $C_{\text{inf}} > 0$ (resp. $C_{\text{sup}} < \infty$) denotes a generic strictly positive (resp. finite) constant. This constant never depends on the number of observations n , or on the covariance parameters (see Section 5), but is allowed to depend on other variables. We

mention these dependences explicitly in cases of ambiguity. The values of C_{inf} and C_{sup} may change across different occurrences.

For a vector x of dimension d we let $|x| = \max_{i=1,\dots,d} |x_i|$. Further, the Euclidean and operator norms are denoted by $\|x\|$ and by $\|M\|_{\text{op}} = \sup\{\|Mx\| : \|x\| \leq 1\}$, for any matrix M . We let $\lambda_1(B) \geq \dots \geq \lambda_r(B)$ be the r eigenvalues of a $r \times r$ symmetric matrix B . We let $\rho_1(B) \geq \dots \geq \rho_r(B) \geq 0$ be the r singular values of a $r \times r$ matrix B . We let \mathbb{N} be the set of non-zero natural numbers.

Further, we define the Fourier transform of a function $h : \mathbb{R}^d \rightarrow \mathbb{R}$ by $\hat{h}(f) = (2\pi)^{-d} \int_{\mathbb{R}^d} h(t) e^{-if^\top t} dt$, where $i^2 = -1$.

For a sequence of observation locations, the next condition ensures that a fixed distance between any two observation locations exists. This condition is classical [7, 11].

Condition 1. *We say that a sequence of observation locations, $(x_i)_{i \in \mathbb{N}}$, $x_i \in \mathbb{R}^d$, is asymptotically well-separated if we have $\inf_{i,j \in \mathbb{N}, i \neq j} |x_i - x_j| > 0$.*

The next condition on a stationary covariance function is classical under increasing-domain asymptotics. This condition provides asymptotic decorrelation for pairs of distant observation locations and implies that covariance matrices are asymptotically well-conditioned when a minimal distance between any two distinct observation locations exists [30, 7].

Condition 2. *We say that a stationary covariance function k on \mathbb{R}^d is sub-exponential and asymptotically positive if:*

i) C_{sup} and C_{inf} exist such that, for all $s \in \mathbb{R}^d$, we have

$$|k(s)| \leq C_{\text{sup}} \exp(-C_{\text{inf}}|s|); \quad (1)$$

ii) For any sequence $(x_i)_{i \in \mathbb{N}}$ satisfying Condition 1, we have $\inf_{n \in \mathbb{N}} \lambda_n(\Sigma) > 0$, where Σ is the $n \times n$ matrix $(k(x_i - x_j))_{i,j=1,\dots,n}$.

In Condition 2, we remark that $k : \mathbb{R}^d \rightarrow \mathbb{R}$ is called a stationary covariance function in the sense that $(x_1, x_2) \rightarrow k(x_1 - x_2)$ is a covariance function. We use this slight language abuse for convenience.

We also remark that, when non-transformed Gaussian processes are considered, a polynomial decay of the covariance function in Condition 2 i) is sufficient to obtain asymptotic results [7, 8]. Here an exponential decay is needed in the proofs to deal with the non-Gaussian case. Nevertheless, most classical families of covariance functions satisfy inequality (1).

When considering a transformed Gaussian process, we will consider a transformation satisfying the following regularity condition, which enables us to subsequently obtain regularity conditions on the covariance function of the transformed Gaussian process.

Condition 3. *Let $F : \mathbb{R} \rightarrow \mathbb{R}$ be a fixed non-constant continuously differentiable function, with derivative F' . We say that F is sub-exponential and non-decreasing if:*

i) For all $t \in \mathbb{R}$, we have $|F(t)| \leq C_{\text{sup}} \exp(C_{\text{sup}}|t|)$ and $|F'(t)| \leq C_{\text{sup}} \exp(C_{\text{sup}}|t|)$;

ii) The function F is non-decreasing on \mathbb{R} .

In the following lemma, we show that the covariance function of a transformed Gaussian process satisfies Condition 2, when Conditions 2 and 3 are satisfied, for the original process and for the transformation.

Lemma 1. *Assume that the stationary covariance function k satisfies Condition 2 and that the transformation F satisfies Condition 3. Let X be a zero-mean Gaussian process with covariance function k and let k' be the stationary covariance function of $F(X(\cdot))$. Then, k' satisfies Condition 2.*

In the next lemma, we show that we can replace the condition of an increasing-transformation by the condition of a monomial transformation of even degree (with an additive constant).

Lemma 2. *If a covariance function k satisfies Condition 2 (i) and if the Fourier transform \hat{k} of k is strictly positive on \mathbb{R}^d , then k satisfies Condition 2 (ii). Furthermore, in this case, in Lemma 1, Condition 3 (ii) can be replaced by the condition $F(x) = x^{2r} + u$ for $r \in \mathbb{N}$, $u \in \mathbb{R}$ and $x \in \mathbb{R}$.*

3. Two main technical results

3.1. Transformed Gaussian process framework

Throughout the paper, we will consider an unobserved latent Gaussian process Z on \mathbb{R}^d with $d \in \mathbb{N}$ fixed. Assume that Z has zero-mean and stationary covariance function k_Z . We assume throughout that k_Z satisfies Condition 2.

We consider a fixed transformation function T satisfying Condition 3. We assume that we observe the transformed Gaussian process Y , defined by $Y(s) = T(Z(s))$ for any $s \in \mathbb{R}^d$.

We assume throughout that the random field Y has zero-mean. We remark that, for a non-linear transformation $F : \mathbb{R} \rightarrow \mathbb{R}$, the random variable $F(X(s))$ does not necessarily have zero mean for $s \in \mathbb{R}^d$. Hence, we implicitly assume that T is of the form $F - \mathbb{E}[F(Z(x))]$, where F satisfies Condition 3. Note that $\mathbb{E}[F(Z(x))]$ is constant by stationarity and that, if F satisfies Condition 3 or the condition specified in Lemma 2, then $F - \mathbb{E}[F(Z(x))]$ also satisfies these conditions. Here, as in many references, the assumption of a zero-mean for Y is made by notational convenience and for the sake of brevity, and could be alleviated.

We let k_Y be the covariance function of Y . We remark that, from Lemma 1, k_Y also satisfies Condition 2.

We let $(s_i)_{i \in \mathbb{N}}$ be the sequence of observation locations, with $s_i \in \mathbb{R}^d$ for $i \in \mathbb{N}$. We assume that $(s_i)_{i \in \mathbb{N}}$ satisfies Condition 1.

For $n \in \mathbb{N}$, we let $y = (y_1, \dots, y_n)^\top = (Y(s_1), \dots, Y(s_n))^\top$ be the (non-Gaussian) observation vector and $R = (k_Y(s_i - s_j))_{i,j=1,\dots,n}$ be its covariance matrix.

The problem of estimating the covariance function k_Y from the observation vector y is crucial and has been extensively studied in the Gaussian case (when

T is a linear function). Classically, we assume that k_Y belongs to a parametric family of covariance functions. We will provide the asymptotic properties of two of the most popular estimators of the covariance parameters: the one based on the (Gaussian) maximum likelihood [34, 41] and the one based on cross validation [13, 6, 48]. To our knowledge, such properties are currently known only for Gaussian processes, and we will provide analogous properties in the transformed Gaussian framework.

3.2. Bounds on the elements of inverse covariance matrices

In the case of (non-transformed) Gaussian processes, one important argument for establishing the asymptotic properties of the maximum likelihood and cross validation estimators is to bound the largest eigenvalue of the inverse covariance matrix R^{-1} . Unfortunately, due to the non-linearity of the transformation T , such a bound on the largest eigenvalue is no longer sufficient in our setting.

To circumvent this issue, we obtain in the following theorem stronger control over the matrix R^{-1} : we show that its coefficients decrease polynomially quickly with respect to the corresponding distance between observation locations. This theorem may have an interest in itself.

Theorem 1. *Consider the setting of Section 3.1. For all fixed $0 < \tau < \infty$, we have, for all $n \in \mathbb{N}$ and $i, j = 1, \dots, n$*

$$\left| (R^{-1})_{i,j} \right| \leq \frac{C_{\text{sup}}}{1 + |s_i - s_j|^{d+\tau}},$$

where C_{sup} depends on τ but does not depend on n, i, j .

3.3. Central limit theorem for quadratic forms of transformed Gaussian processes

In the proofs on covariance parameter estimation of Gaussian processes, a central step is to show the asymptotic normality of quadratic forms of large Gaussian vectors. This asymptotic normality is established by diagonalizing the matrices of the quadratic forms. This diagonalization provides sums of squares of decorrelated Gaussian variables and thus sums of independent variables [25, 7].

In the transformed Gaussian case, one has to deal with quadratic forms involving transformations of Gaussian vectors. Hence, the previous arguments are not longer valid. To overcome this issue, we provide below a general central limit theorem for quadratic forms of transformed Gaussian vectors. This theorem may have an interest in itself.

This asymptotic normality result is established by considering a metric d_w generating the topology of weak convergence on the set of Borel probability measures on Euclidean spaces (see, e.g., [22] p. 393). We prove that the distance between the sequence of the standardized distributions of the quadratic forms and Gaussian distributions decreases to zero when n increases. The introduction

of the metric d_w enables us to formulate asymptotic normality results in cases when the sequence of standardized variances of the quadratic forms does not necessarily converge as $n \rightarrow \infty$.

Theorem 2. *Consider the setting of Section 3.1. Let $(A_n)_{n \in \mathbb{N}}$ be a sequence of matrices such that A_n has dimension $n \times n$ for any $n \in \mathbb{N}$. Let $A = A_n$ for concision. Assume that for all $n \in \mathbb{N}$ and for all $i, j = 1, \dots, n$,*

$$|A_{i,j}| \leq \frac{C_{\text{sup}}}{1 + |s_i - s_j|^{d+C_{\text{inf}}}},$$

where C_{sup} does not depend on i, j . Let

$$V_n = \frac{1}{n} y^\top A y. \tag{2}$$

Let \mathcal{L}_n be the distribution of $\sqrt{n}(V_n - \mathbb{E}[V_n])$. Then, as $n \rightarrow \infty$,

$$d_w(\mathcal{L}_n, \mathcal{N}[0, n \text{Var}(V_n)]) \rightarrow 0.$$

In addition, the sequence $(n \text{Var}(V_n))_{n \in \mathbb{N}}$ is bounded.

Remark 1. *In the case where limits E_∞ and σ_∞ exist, such that*

$$\mathbb{E}[V_n] - E_\infty = o(n^{-1/2})$$

and the sequence $(n \text{Var}(V_n))_{n \in \mathbb{N}}$ converges to a fixed variance σ_∞^2 , the result of Theorem 2 can be written in the classical form

$$\sqrt{n}(V_n - E_\infty) \xrightarrow{\mathcal{L}} \mathcal{N}[0, \sigma_\infty^2],$$

as $n \rightarrow \infty$.

4. Estimation of a single variance parameter

We let σ_0^2 be the marginal variance of Y , that is $\text{Var}(Y(s)) = \sigma_0^2$ for any $s \in \mathbb{R}^d$. We let $k_Y = \sigma_0^2 c_Y$ be the stationary covariance function of Y , where c_Y is a correlation function. We assume that the same conditions as in Section 3 hold. Then, the standard Gaussian maximum likelihood estimator of the variance parameter is

$$\hat{\sigma}_{\text{ML}}^2 = \frac{1}{n} y^\top C^{-1} y,$$

where $C = (c_Y(s_i - s_j))_{1 \leq i, j \leq n}$. One can simply show that $\mathbb{E}[\hat{\sigma}_{\text{ML}}^2] = \sigma_0^2$ even though y is not a Gaussian vector, since y has mean vector 0 and covariance matrix $\sigma_0^2 C$. Hence, a direct consequence of Theorems 1 and 2 is then that the maximum likelihood estimator is asymptotically Gaussian, with a $n^{1/2}$ rate of convergence, even though the transformed process Y is not a Gaussian process.

Corollary 1. Let \mathcal{L}_n be the distribution of $\sqrt{n}(\hat{\sigma}_{ML}^2 - \sigma_0^2)$. Then, as $n \rightarrow \infty$,

$$d_w(\mathcal{L}_n, \mathcal{N}[0, n \text{Var}(\hat{\sigma}_{ML}^2 - \sigma_0^2)]) \rightarrow 0.$$

In addition, the sequence $(n \text{Var}(\hat{\sigma}_{ML}^2 - \sigma_0^2))_{n \in \mathbb{N}}$ is bounded.

The proof of Theorem 2 actually allows us to study another estimator of the variance σ_0^2 of the form

$$\hat{\sigma}_{ML,K}^2 = \frac{1}{n} y^\top C_K^{-1} y,$$

where $(C_K^{-1})_{i,j} = (C^{-1})_{i,j} \mathbf{1}_{|s_i - s_j| \leq K}$. The proof of Theorem 2 then directly implies the following.

Corollary 2. Let K_n be any sequence of positive numbers tending to infinity. Let $\mathcal{L}_{K_n,n}$ be the distribution of $\sqrt{n}(\hat{\sigma}_{ML,K_n}^2 - \sigma_0^2)$. Then, as $n \rightarrow \infty$,

$$d_w(\mathcal{L}_{K_n,n}, \mathcal{N}[0, n \text{Var}(\hat{\sigma}_{ML,K_n}^2 - \sigma_0^2)]) \rightarrow 0.$$

In addition, we have

$$n \text{Var}(\hat{\sigma}_{ML,K_n}^2 - \sigma_0^2) - n \text{Var}(\hat{\sigma}_{ML}^2 - \sigma_0^2) \rightarrow 0.$$

The above corollary shows that one can taper the elements of C^{-1} when estimating the variance parameter, and obtain the same asymptotic distribution of the error, as long as the taper range goes to infinity, with no rate assumption. This result may have an interest in itself, in view of the existing literature on covariance tapering for Gaussian processes under increasing-domain asymptotics [23, 37]. We also remark that the computation costs of $\hat{\sigma}_{ML,K}^2$ and $\hat{\sigma}_{ML}^2$ have the same orders of magnitude because C^{-1} needs to be computed in both cases.

5. General covariance

5.1. Framework

As in Section 3.1, we consider a zero-mean Gaussian process Z defined on \mathbb{R}^d with covariance function k_Z satisfying Condition 2. Let Y be the random field defined for any $s \in \mathbb{R}^d$ by $Y(s) = T(Z(s))$, where T is a fixed function satisfying Condition 3. Furthermore we assume that Y has zero-mean function and we recall that from Lemma 1, its covariance function k_Y also satisfies Condition 2. Finally, the sequence of observation locations $(s_i)_{i \in \mathbb{N}}$ satisfies Condition 1.

Let $\{k_{Y,\theta}; \theta \in \Theta\}$ be a parametric set of stationary covariance functions on \mathbb{R}^d , with Θ a compact set of \mathbb{R}^p . We consider the following condition on this parametric set of covariance functions.

Condition 4. For all $s \in \mathbb{R}^d$, $k_{Y,\theta}(s)$ is three times continuously differentiable with respect to θ , and we have

$$\sup_{\theta \in \Theta} |k_{Y,\theta}(s)| \leq C_{\text{sup}} \exp(-C_{\text{inf}}|s|), \quad (3)$$

$$\sup_{\substack{\theta \in \Theta \\ \ell=1,2,3 \\ i_1, \dots, i_\ell=1, \dots, p}} \left| \frac{\partial k_{Y,\theta}(s)}{\partial \theta_{i_1}, \dots, \partial \theta_{i_\ell}} \right| \leq \frac{C_{\text{sup}}}{1 + |s|^{d+C_{\text{inf}}}}. \quad (4)$$

The smoothness condition in (4) is classical and is assumed for instance in [7]. As discussed after Condition 2, milder versions of (3) can be assumed for non-transformed Gaussian processes, but (3) is satisfied by most classical families of covariance functions nonetheless.

The next condition, on the Fourier transforms of the covariance functions in the model, is standard.

Condition 5. *We let $\hat{k}_{Y,\theta}$ be the Fourier transform of $k_{Y,\theta}$. Then $\hat{k}_{Y,\theta}(s)$ is jointly continuous with respect to θ and s and is strictly positive on $\Theta \times \mathbb{R}^d$.*

Finally, the next condition means that we address the well-specified case [6, 8], where the family of covariance functions does contain the true covariance function of Y . The well-specified case is considered in the majority of the literature on Gaussian processes.

Condition 6. *There exists θ_0 in the interior of Θ such that $k_Y = k_{Y,\theta_0}$.*

In the next two subsections, we study the asymptotic properties of two classical estimators (maximum likelihood and cross validation) for the covariance parameter θ_0 . The asymptotic properties of these estimators are already known for Gaussian processes and we extend them to the non-Gaussian process Y .

5.2. Maximum Likelihood

For $n \in \mathbb{N}$, let R_θ be the $n \times n$ matrix $(k_{Y,\theta}(s_i - s_j))_{i,j=1, \dots, n}$, and let

$$\hat{\theta}_{\text{ML}} \in \underset{\theta \in \Theta}{\operatorname{argmin}} L_\theta \quad (5)$$

with

$$L_\theta = \frac{1}{n} (\log(\det(R_\theta)) + y^\top R_\theta^{-1} y)$$

be a maximum likelihood estimator. We will provide its consistency under the following condition.

Condition 7. *For all $\alpha > 0$ we have*

$$\liminf_{n \rightarrow \infty} \inf_{\|\theta - \theta_0\| \geq \alpha} \frac{1}{n} \sum_{i,j=1}^n (k_{Y,\theta}(s_i - s_j) - k_{Y,\theta_0}(s_i - s_j))^2 > 0.$$

Condition 7 can be interpreted as a global indentifiability condition. It implies in particular that two different covariance parameters yield two different distributions for the observation vector. It is used in several studies, for instance [7].

Theorem 3. Consider the setting of Section 5.1 for Z , T , Y and $(s_i)_{i \in \mathbb{N}}$. Assume that Conditions 4, 5, 6 and 7 hold. Then, as $n \rightarrow \infty$,

$$\hat{\theta}_{ML} \xrightarrow{P} \theta_0.$$

Remark 2. The (Gaussian) maximum likelihood estimator $\hat{\theta}_{ML}$ in Theorem 3 is not, strictly speaking, a maximum likelihood estimator, because the distribution of the observation vector y is non-Gaussian. If the transformation T is known, and assuming that the covariance function of the Gaussian process Z belongs to a parametric set $\{k_{Z,\alpha}, \alpha \in \mathcal{A}\}$, a (non-Gaussian) maximum likelihood estimator $\hat{\alpha}_{Z,ML}(y)$ could be defined. When T is bijective, y is a fixed invertible transformation of $(Z(s_1), \dots, Z(s_n))$ and so this maximum likelihood estimator $\hat{\alpha}_{Z,ML}(y)$ coincides with the standard Gaussian maximum likelihood estimator based on Z . Asymptotically, it is known that $\hat{\alpha}_{Z,ML}(y)$ converges to the true parameter α_0 for which Z has covariance function k_{Z,α_0} . In Theorem 3, we show that $\hat{\theta}_{ML}$ similarly converges to the true parameter θ_0 for which Y has covariance function k_{Y,θ_0} . Notice that α_0 and θ_0 usually do not coincide. For instance, if the covariance models $\{k_{Y,\theta}; \theta \in \Theta\}$ and $\{k_{Z,\alpha}; \alpha \in \mathcal{A}\}$ are both $\{\sigma^2 e^{-\rho \|\cdot\|}\}$, then if $\alpha_0 = (\sigma_0^2, \rho_0)$, by Mehler's formula [5],

$$\text{Cov}(Y(s + \tau), Y(s)) = 2 \text{Cov}(Z(s + \tau), Z(s))^2 = 2\sigma_0^4 e^{-2\rho_0 \|\tau\|}$$

and thus $\theta_0 = (2\sigma_0^4, 2\rho_0)$. Hence, the exact (non-Gaussian) maximum likelihood estimator $\hat{\alpha}_{Z,ML}(y)$ based on the knowledge of T and modeling the covariance function of Z and the pseudo (Gaussian) maximum likelihood estimator $\hat{\theta}_{ML}$ that ignores T and models the covariance function of Y estimate covariance parameters of different natures and do not have the same limit.

Condition 8. For any $(\alpha_1, \dots, \alpha_p) \neq (0, \dots, 0)$, we have

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{i,j=1}^n \left(\sum_{\ell=1}^p \alpha_\ell \frac{\partial k_{Y,\theta_0}(s_i - s_j)}{\partial \theta_\ell} \right)^2 > 0.$$

Condition 8 can be interpreted as a regularity condition and as a local identifiability condition around θ_0 . In the next theorem, we provide the asymptotic normality of the maximum likelihood estimator. In this theorem, the matrices M_{θ_0} and Σ_{θ_0} depend on the number of observation locations.

Theorem 4. Consider the setting of Section 5.1 for Z , T , Y and $(s_i)_{i \in \mathbb{N}}$. Assume that Conditions 4, 5, 6, 7 and 8 hold. Let M_{θ_0} be the $p \times p$ matrix defined by

$$(M_{\theta_0})_{i,j} = \frac{1}{n} \text{tr} \left(R_{\theta_0}^{-1} \frac{\partial R_{\theta_0}}{\partial \theta_i} R_{\theta_0}^{-1} \frac{\partial R_{\theta_0}}{\partial \theta_j} \right).$$

Let Σ_{θ_0} be the $p \times p$ covariance matrix defined by $(\Sigma_{\theta_0})_{i,j} = \text{Cov}(n^{1/2} \partial L_{\theta_0} / \partial \theta_i, n^{1/2} \partial L_{\theta_0} / \partial \theta_j)$. Let $\mathcal{L}_{\theta_0,n}$ be the distribution of $\sqrt{n}(\hat{\theta}_{ML} - \theta_0)$. Then, as $n \rightarrow \infty$,

$$d_w(\mathcal{L}_{\theta_0,n}, \mathcal{N}[0, M_{\theta_0}^{-1} \Sigma_{\theta_0} M_{\theta_0}^{-1}]) \rightarrow 0. \quad (6)$$

In addition,

$$\limsup_{n \rightarrow \infty} \lambda_1(M_{\theta_0}^{-1} \Sigma_{\theta_0} M_{\theta_0}^{-1}) < +\infty. \quad (7)$$

Remark 3. *If the sequences of matrices M_{θ_0} and Σ_{θ_0} converge as $n \rightarrow \infty$, then $\sqrt{n}(\hat{\theta}_{ML} - \theta_0)$ converges in distribution to a fixed centered Gaussian distribution where the limiting covariance matrix is given by*

$$\lim_{n \rightarrow \infty} M_{\theta_0}^{-1} \Sigma_{\theta_0} M_{\theta_0}^{-1}.$$

Conditions 7 and 8 involve the model of covariance functions $\{k_{Y,\theta}; \theta \in \Theta\}$ and the sequence of observation locations $(s_i)_{i \in \mathbb{N}}$ but not the transformation T . They are further discussed, in a different context, in [12]. We believe that these conditions are mild. For instance, Conditions 7 and 8 hold when the sequence of observation locations $(s_i)_{i \in \mathbb{N}}$ is a randomly perturbed regular grid, as in [7].

Lemma 3 (see [7]). *For $i \in \mathbb{N}$, let $s_i = g_i + \epsilon_i$, where $(g_i)_{i \in \mathbb{N}}$ is a sequence with, for $N \in \mathbb{N}$, $\{g_1, \dots, g_{N^d}\} = \{(i_1, \dots, i_d); i_1 = 1, \dots, N, \dots, i_d = 1, \dots, N\}$ and where $(\epsilon_i)_{i \in \mathbb{N}}$ is a sequence of i.i.d. random variables with distribution on $[-1/2 + \delta, 1/2 - \delta]^d$ with $0 < \delta < 1/2$. Then, Condition 7 holds almost surely, provided that, for $\theta \neq \theta_0$, there exists $i \in \mathbb{Z}^d$ for which $k_{Y,\theta}(i + \epsilon_1 - \epsilon_2)$ and $k_{Y,\theta_0}(i + \epsilon_1 - \epsilon_2)$ are not almost surely equal.*

Furthermore, Condition 8 holds almost surely, provided that for $(\alpha_1, \dots, \alpha_p) \neq (0, \dots, 0)$, there exists $i \in \mathbb{Z}^d$ for which $\sum_{\ell=1}^p \alpha_\ell \partial k_{Y,\theta_0}(i + \epsilon_1 - \epsilon_2) / \partial \theta_\ell$ is not almost surely equal to zero.

5.3. Cross Validation

We consider the cross validation estimator consisting of minimizing the sum of leave-one-out square errors.

Since the leave-one-out errors do not depend on the variance $k_{Y,\theta}(0)$, we introduce some additional notation. In Subsections 5.3 and 5.4, we let $\Theta = [\sigma_{\text{inf}}^2, \sigma_{\text{sup}}^2] \times \mathcal{S}$ where $0 < \sigma_{\text{inf}}^2 < \sigma_{\text{sup}}^2 < \infty$ are fixed and where \mathcal{S} is compact in \mathbb{R}^{p-1} . We let $\theta = (\sigma^2, \psi)$ with $\sigma_{\text{inf}}^2 \leq \sigma^2 \leq \sigma_{\text{sup}}^2$ and $\psi \in \mathcal{S}$. We assume that for $\theta \in \Theta$, $k_{Y,\theta} = \sigma^2 c_{Y,\psi}$, with $c_{Y,\psi}$ a stationary correlation function. Similarly, we let $\theta_0 = (\sigma_0^2, \psi_0)$. For $\psi \in \mathcal{S}$, we let $C_\psi = (c_{Y,\psi}(s_i - s_j))_{i,j=1,\dots,n}$. Cross validation is defined for $n \in \mathbb{N}$ by

$$\hat{\psi}_{\text{CV}} \in \underset{\psi \in \mathcal{S}}{\operatorname{argmin}} CV_\psi, \quad (8)$$

with

$$CV_\psi = \frac{1}{n} y^\top C_\psi^{-1} \operatorname{diag}(C_\psi^{-1})^{-2} C_\psi^{-1} y,$$

where $\operatorname{diag}(M)$ is obtained by setting the off-diagonal elements of a square matrix M to zero. The criterion CV_ψ is the sum of leave-one-out square errors, as is shown for instance in [21, 6].

The asymptotic behaviour of $\hat{\psi}_{CV}$ was studied in the Gaussian framework in [7] and under increasing-domain asymptotics. We also remark that, in the Gaussian framework, a modified leave-one-out criterion was studied in [13] in the case of infill asymptotics.

The next identifiability condition is also made in [7].

Condition 9. For all $\alpha > 0$, we have

$$\liminf_{n \rightarrow \infty} \inf_{\|\psi - \psi_0\| \geq \alpha} \frac{1}{n} \sum_{i,j=1}^n (c_{Y,\psi}(s_i - s_j) - c_{Y,\psi_0}(s_i - s_j))^2 > 0.$$

The next theorem provides the consistency of the cross validation estimator.

Theorem 5. Consider the setting of Section 5.1 for Z , T , Y and $(s_i)_{i \in \mathbb{N}}$. Assume that Conditions 4, 5, 6 and 9 hold. Then, as $n \rightarrow \infty$,

$$\hat{\psi}_{CV} \xrightarrow{P} \psi_0.$$

The next condition is a local identifiability condition.

Condition 10. For any $(\alpha_1, \dots, \alpha_{p-1}) \neq (0, \dots, 0)$, we have

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{i,j=1}^n \left(\sum_{\ell=1}^{p-1} \alpha_\ell \frac{\partial}{\partial \psi_\ell} (c_{Y,\psi_0}(s_i - s_j)) \right)^2 > 0.$$

In the next theorem, we provide the asymptotic normality of the cross validation estimator. In this theorem, the matrices N_{ψ_0} and Γ_{ψ_0} depend on the number of observation locations.

Theorem 6. Consider the setting of Section 5.1 for Z , T , Y and $(s_i)_{i \in \mathbb{N}}$. Assume that Conditions 4, 5, 6, 9 and 10 hold. Let N_{ψ_0} be the $(p-1) \times (p-1)$ matrix defined by

$$\begin{aligned} (N_{\psi_0})_{i,j} &= -\frac{8}{n} \operatorname{tr} \left(\frac{\partial C_{\psi_0}}{\partial \psi_j} C_{\psi_0}^{-1} \operatorname{diag}(C_{\psi_0}^{-1})^{-3} \operatorname{diag} \left(C_{\psi_0}^{-1} \frac{\partial C_{\psi_0}}{\partial \psi_i} C_{\psi_0}^{-1} \right) C_{\psi_0}^{-1} \right) \\ &\quad + \frac{2}{n} \operatorname{tr} \left(\frac{\partial C_{\psi_0}}{\partial \psi_j} C_{\psi_0}^{-1} \operatorname{diag}(C_{\psi_0}^{-1})^{-2} C_{\psi_0}^{-1} \frac{\partial C_{\psi_0}}{\partial \psi_i} C_{\psi_0}^{-1} \right) \\ &\quad + \frac{6}{n} \operatorname{tr} \left(\operatorname{diag}(C_{\psi_0}^{-1})^{-4} \operatorname{diag} \left(C_{\psi_0}^{-1} \frac{\partial C_{\psi_0}}{\partial \psi_i} C_{\psi_0}^{-1} \right) \operatorname{diag} \left(C_{\psi_0}^{-1} \frac{\partial C_{\psi_0}}{\partial \psi_j} C_{\psi_0}^{-1} \right) C_{\psi_0}^{-1} \right). \end{aligned}$$

Let Γ_{ψ_0} be the $(p-1) \times (p-1)$ covariance matrix defined by $(\Gamma_{\psi_0})_{i,j} = \operatorname{Cov}(n^{1/2} \partial C V_{\psi_0} / \partial \psi_i, n^{1/2} \partial C V_{\psi_0} / \partial \psi_j)$. Let $\mathcal{Q}_{\psi_0, n}$ be the distribution of $\sqrt{n}(\hat{\psi}_{CV} - \psi_0)$. Then, as $n \rightarrow \infty$,

$$d_w \left(\mathcal{Q}_{\psi_0, n}, \mathcal{N}[0, N_{\psi_0}^{-1} \Gamma_{\psi_0} N_{\psi_0}^{-1}] \right) \rightarrow 0. \quad (9)$$

In addition,

$$\limsup_{n \rightarrow \infty} \lambda_1(N_{\psi_0}^{-1} \Gamma_{\psi_0} N_{\psi_0}^{-1}) < +\infty. \quad (10)$$

Similarly as for maximum likelihood, Condition 9 is a global identifiability condition for the correlation function. In the same way, Condition 10 is a local identifiability condition for the correlation function around ψ_0 . The sequence of observation locations presented in Lemma 3 also satisfies Conditions 9 and 10 by replacing $k_{Y,\theta}$ with $c_{Y,\psi}$.

We remark that the conditions for cross validation imply those for maximum likelihood.

Lemma 4. *Condition 9 implies Condition 7 and Condition 10 implies Condition 8.*

5.4. Joint asymptotic normality

From Theorems 4 and 6, both the maximum likelihood and cross validation estimators converge at the standard parametric rate $n^{1/2}$. Let us write $\hat{\theta}_{\text{ML}} = (\hat{\sigma}_{\text{ML}}^2, \hat{\psi}_{\text{ML}})$. In the case where T is the identity function (that is, where we observe Gaussian processes instead of transformed Gaussian processes), numerical experiments tend to show that $\hat{\psi}_{\text{ML}}$ is more accurate than $\hat{\psi}_{\text{CV}}$ [7]. Indeed, when T is the identity function, maximum likelihood is based on the Gaussian probability density function of the observation vector.

In contrast, when T is not the identity function, $\hat{\psi}_{\text{ML}}$ is an M-estimator based on a criterion which does not coincide with the observation probability density function anymore. Hence, it is conceivable that $\hat{\psi}_{\text{CV}}$ could become more accurate than $\hat{\psi}_{\text{ML}}$. Furthermore, it is possible that using linear combinations of these two estimators could result in a third one with improved accuracy [29, 14].

Motivated by this discussion, we now provide a joint central limit theorem for the maximum likelihood and cross validation estimators.

Theorem 7. *Consider the setting of Section 5.1 for Z, T, Y and $(s_i)_{i \in \mathbb{N}}$. Assume that Conditions 4, 5, 6, 9 and 10 hold. Let D_{θ_0} be the $(2p-1) \times (2p-1)$ block diagonal matrix with first $p \times p$ block equal to M_{θ_0} and second $(p-1) \times (p-1)$ block equal to N_{ψ_0} , with the notation of Theorems 4 and 6. Also let Ψ_{θ_0} be the $(2p-1) \times (2p-1)$ covariance matrix of the vector $n^{1/2}(\partial L_{\theta_0}/\partial \theta, \partial CV_{\psi_0}/\partial \psi)$.*

Let $\mathcal{Q}_{\theta_0, n}$ be the distribution of

$$\sqrt{n} \begin{pmatrix} \hat{\theta}_{\text{ML}} - \theta_0 \\ \hat{\psi}_{\text{CV}} - \psi_0 \end{pmatrix}.$$

Then, as $n \rightarrow \infty$,

$$d_w(\mathcal{Q}_{\theta_0, n}, \mathcal{N}[0, D_{\theta_0}^{-1} \Psi_{\theta_0} D_{\theta_0}^{-1}]) \rightarrow 0. \quad (11)$$

In addition,

$$\limsup_{n \rightarrow \infty} \lambda_1(D_{\theta_0}^{-1} \Psi_{\theta_0} D_{\theta_0}^{-1}) < +\infty. \quad (12)$$

Remark 4. *From Theorem 7, considering any C^1 function f from $\mathcal{S}^2 \rightarrow \mathcal{S}$ such that $f(\psi, \psi) = \psi$ for any $\psi \in \mathcal{S}$, and applying the classical delta method, we*

obtain the asymptotic normality of the new estimator $f(\hat{\psi}_{ML}, \hat{\psi}_{CV})$. A classical choice for f is $f(\psi_1, \psi_2) = \lambda\psi_1 + (1 - \lambda)\psi_2$, which leads to linear aggregation [29, 14]. We remark that selecting an optimal λ leading to the smallest asymptotic covariance matrix necessitates an estimation of the asymptotic covariance matrix in Theorem 7. We leave this as an open problem for further research.

6. Illustration

In this section we numerically illustrate the convergence of different estimators as stated in Corollary 1 and Theorems 4, 6 and 7. We use the following simulation setup in dimension $d = 2$. For $n = 4(m + 1/2)^2$ we define the grid $\{-m, \dots, m\}^2$ and add i.i.d. variables with uniform distribution on $[-0.4, 0.4]^2$ to obtain n observation points. Thus, we have a distance of at least $\Delta = 0.2$ between the individual observation locations. The zero-mean Gaussian process Z has stationary covariance function $k_Z(s) = \sigma_0^2 \exp(-\|s\|/\rho_0)$, $s \in \mathbb{R}^2$ and we will denote this reference case as the Gaussian case throughout. We define the zero-mean process $Y = T(Z) = Z^2 - \sigma_0^2$ and we will denote this case as the non-Gaussian case. Recall that $k_Y(s) = 2k_Z(2s) = 2\sigma_0^4 \exp(-2\|s\|/\rho_0)$ (see Remark 2). We set the marginal variance to $\sigma_0^2 = 1.5$ and the range to $\rho_0 = 2$. Hence, in the non-Gaussian case, the marginal variance of Y is $2\sigma_0^4 = 4.5$ with a range of $\rho_0/2$ which is equal to a half of that of Z .

To start, we consider the maximum likelihood estimates of the marginal variance parameters, when the range of Y or Z is assumed to be known, i.e., Corollary 1. Figure 1 illustrates the empirical densities of \mathcal{L}_n in Corollary 1 for $n = 100, 400$ and 900 observation locations based on $N = 2500$ replicates. For moderate n sizes and in the non-Gaussian case, the asymptotic variance $\sigma_\infty^2 := n \text{Var}(\hat{\sigma}_{ML}^2 - 2\sigma_0^2)$ (Corollary 1) can be calculated based on (26) and using

$$\begin{aligned} \text{Cov}(y_i y_j, y_k y_l) &= 4(k_{i,k}^2 k_{j,l}^2 + k_{i,l}^2 k_{j,k}^2) \\ &\quad + 16(k_{i,k} k_{i,l} k_{j,k} k_{j,l} + k_{i,j} k_{i,l} k_{j,k} k_{k,l} + k_{i,j} k_{i,k} k_{j,l} k_{k,l}), \end{aligned}$$

with $k_{i,j} = k_Z(s_i - s_j)$. The above display follows from tedious computations based on Isserlis' theorem. The densities in red in Figure 1 are based on the asymptotic distribution $\mathcal{N}[0, \sigma_\infty^2]$. In the non-Gaussian case and for $n \geq 400$, the calculation of σ_∞^2 is computationally prohibitive, so σ_∞^2 has instead been approximated by the empirical variance with the corresponding densities indicated in green. As expected, the convergence for the Gaussian case is faster than for the non Gaussian case. But in both situations, the results behave nicely.

We now turn to Theorem 4 and consider the bivariate variance and range maximum likelihood estimation. That is, we consider the two-dimensional maximum likelihood estimates of (σ_0^2, ρ_0) in the Gaussian case and of $(2\sigma_0^2, \rho_0/2)$ in the non-Gaussian case. Again, we do not observe many surprises. Skewness of the empirical distribution is slightly higher compared to the single variance parameter estimation, and convergence is slightly slower, as is illustrated in Figure 2.

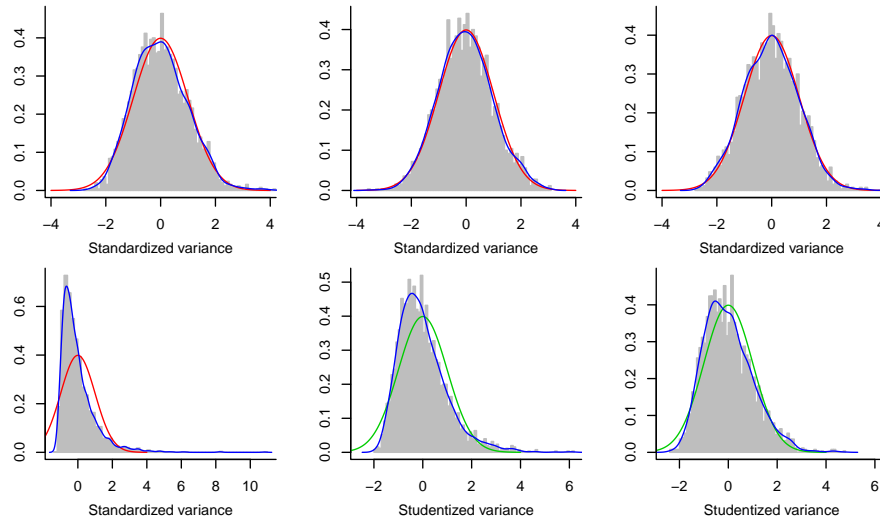


FIG 1. Histograms of standardized and studentized variance maximum likelihood estimates. Blue: empirical density (kernel density estimates), red: asymptotic density, green: asymptotic density based on the empirical variance. The top row shows results for the process Z (the Gaussian case), the bottom row for the transformed process $Y = Z^2 - \sigma_0^2$ (the non-Gaussian case). The columns are for $n = 100, 400, 900$. All panels are based on $N = 2500$ replicates.

For the general setting, when estimating jointly the variance and range parameter, the asymptotic bivariate covariance matrix is challenging to compute (see Theorem 4) and thus Figure 2 illustrates the empirical densities and densities based on the empirical bivariate covariance matrix.

We now consider not only maximum likelihood estimation of the variance and range, but also cross validation estimation of the range (see the beginning of Section 5.3). We have observed that the range estimates based on cross validation are much more variable, and in many situations the maximum was attained at the (imposed) boundary. Here we used the bound $[2/15, 12]$, i.e., smaller than the minimal distance between two observation locations and 6 (resp. 12) times the diameter of the observation points of Z (resp. Y). Estimates at or close to the boundary indicate convergence issues and would imply a second, possibly manual, inspection. For the reported results, we eliminated all cross validation cases that yielded estimates outside $[0.14, 11.4]$.

Figure 3 shows the mean squared error, squared bias and variance of $\hat{\sigma}_{\text{ML}}^2$, $\hat{\rho}_{\text{ML}}$ and $\hat{\rho}_{\text{CV}}$ under different settings. For maximum likelihood, we consider the univariate case (one parameter is estimated while the other is known) and the bivariate case (both parameters are jointly estimated). For cross validation, only the range parameter is estimated (see the beginning of Section 5.3), and thus only the univariate case is considered. The mean squared error is dominated by the variance component. Univariate maximum likelihood estimation for Gaus-

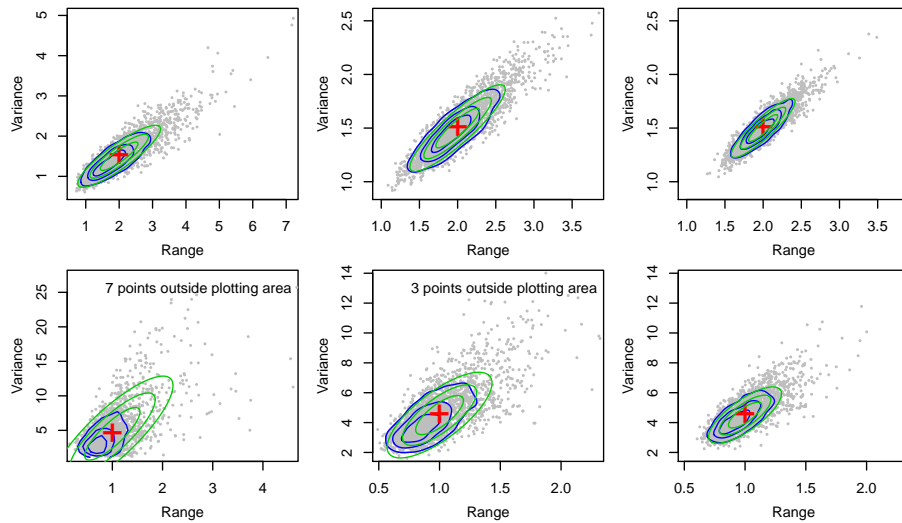


FIG 2. Scatter plots of variance and range maximum likelihood estimates. Blue: contour lines of kernel density estimates; green: contour lines of asymptotic density based on the empirical bivariate covariance matrix; red cross: true mean (true parameter values). The top row shows results for the process Z (Gaussian case), the bottom row for the transformed process $Y = Z^2 - \sigma_0^2$ (non-Gaussian case). The columns are for $n = 100, 400, 900$. All panels are based on $N = 2500$ replicates.

sian cases have low bias and the lowest variance (top left and right panel). Joint maximum likelihood estimation has a somewhat larger variance than individual estimation. Surprisingly, cross validation for Gaussian cases has a higher variance compared to cross validation for non-Gaussian cases.

Recall that Theorems 4, 6 and 7 show that, as n increases, the distribution of the standardized estimation error is close to a Gaussian distribution in terms of the metric d_w . In Figure 4, we illustrate this by computing one-dimensional Wasserstein distances between the empirical distribution of the standardized estimation errors and Gaussian distributions. The figure shows the Wasserstein distance ($p = 1$) as a function of the number of observation locations n for individual parameters and for specific bivariate settings (similarly as for Figure 3). In each case, the samples have been centered around the true mean (true parameter values) and standardized by an empirical standard deviation (n -weighted average over all the samples). Their empirical distribution is compared to the standardized Gaussian distribution. The top left panel shows that the densities of the cross validation parameters are converging slowest whereas their mean squared error is comparable (see Figure 3); the densities are highly skewed and thus lead to much larger Wasserstein distances compared to the distributions of the maximum likelihood derived parameters. For the bivariate maximum likelihood estimation the marginal distributions have very similar Wasserstein distances; in the center panels: the dashed and solid colored lines

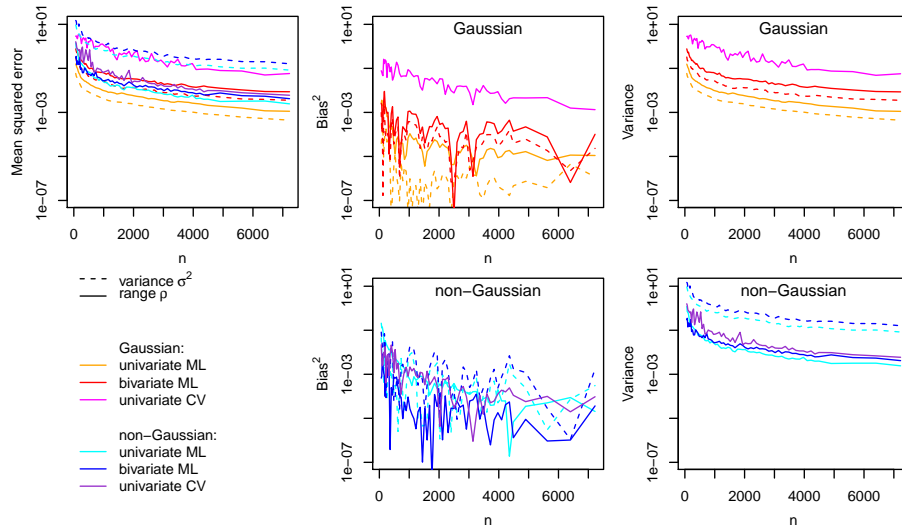


FIG 3. Mean squared error (top left), squared bias (middle column) and variance (right column) as a function of n for different settings in log-scale. The variance parameter is represented in dashed lines, the range parameter in solid lines. Process Z is shown by reddish colors (Gaussian case), transformed process $Y = Z^2 - \sigma_0^2$ in blueish colors (non-Gaussian case). The panels are based on $N = 250$ replicates that did not reveal any convergence issues.

are visually hardly separable. As suggested by the individual panels of Figures 1 and 2, convergence in the Gaussian case is much faster compared to the non-Gaussian case. The right column of Figure 4 illustrates the joint asymptotic normality of the range parameter estimators by maximum likelihood and cross validation. The gray lines there illustrate Theorem 7 and are Wasserstein distances for linear combinations of the range estimates by maximum likelihood and cross validation, i.e., $\lambda\hat{\rho}_{\text{ML}} + (1 - \lambda)\hat{\rho}_{\text{CV}}$ for $\lambda = j/10$, $j = 1, \dots, 9$. The highly skewed distribution of the cross validation-estimated range parameter for Gaussian processes is clearly visible. In the non Gaussian case, the effect of the skewness is less pronounced since the maximum likelihood is skewed as well.

7. Conclusion

We have shown that the covariance parameters of transformed Gaussian processes can be estimated by cross validation and Gaussian maximum likelihood, with the same rate of convergence as in the case of non-transformed Gaussian processes. In particular, Gaussian maximum likelihood works well asymptotically, despite the fact that the observations do not have a Gaussian distribution. Hence Gaussian maximum likelihood is here robust with respect to non-Gaussian data. This provides the first step of a theoretical validation of the use of Gaussian maximum likelihood in frequent cases where the data are non-Gaussian.

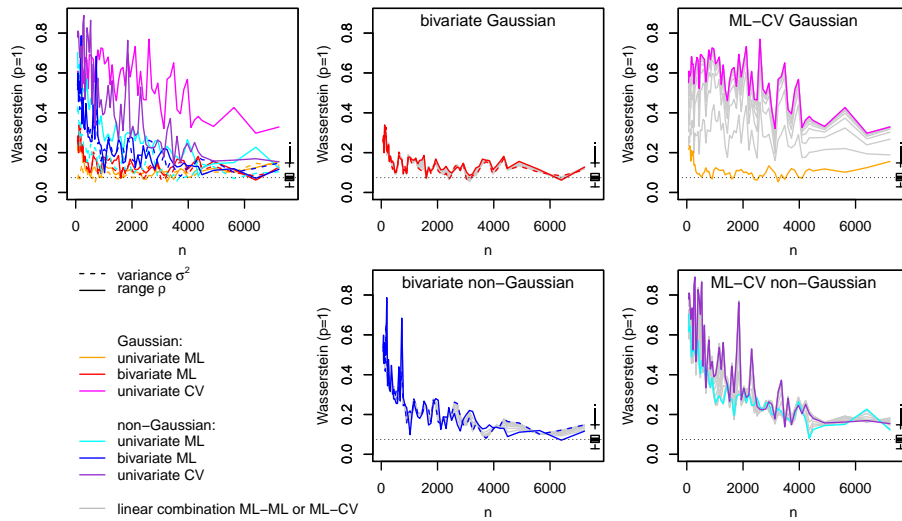


FIG 4. Wasserstein distance ($p = 1$). Top left: marginal for each parameter. Center panels: bivariate estimation of range and variance by maximum likelihood as in Theorem 4 and linear combinations thereof; right panels: univariate estimation of the range parameter by maximum likelihood and cross validation and with linear combinations of these two estimators as in Theorem 7. Top middle and right panels: Gaussian cases. Lower row panels: non-Gaussian case. The colors and line styles follow those in Figure 3. The gray lines are Wasserstein distances for estimates based on linear combinations of maximum likelihood (center column) and of maximum likelihood and cross validation (right column). The panels are based on $N = 250$ replicates that did not reveal any convergence issues. The small boxplot on the right in each panel shows the Wasserstein distance for sample size $n = 250$ of 10000 realizations of $\mathcal{N}[0, 1]$; the horizontal dotted line shows the median thereof.

In future research, it would be interesting to extend the results of this paper to other classes of non-Gaussian random fields rather than only transformed Gaussian processes. In addition, the asymptotic analysis of estimators of the transformation of transformed Gaussian processes is of great interest.

Appendix A: Proofs

A.1. Technical results

Lemma 5. Let $q \in \mathbb{N}$ be fixed. Let $g : \mathbb{R}^q \rightarrow \mathbb{R}^+$ be fixed and satisfy $g(x) \leq C_{\text{sup}} \exp(C_{\text{sup}}|x|)$. Let W be a Gaussian vector of dimension q . Then $\mathbb{E}[g(W)] < \infty$.

Proof. Without loss of generality, we can assume that W_i has variance 1 for

$i = 1, \dots, q$. We let w_i be the mean of W_i for $i = 1, \dots, q$. We have, for $t > 1$,

$$\begin{aligned} \mathbb{P}(g(W) \geq t) &\leq \mathbb{P}(C_{\text{sup}} \exp(C_{\text{sup}} \max_{i=1, \dots, q} |W_i|) \geq t) \\ &\leq \sum_{i=1}^q \mathbb{P}(C_{\text{sup}} \exp(C_{\text{sup}} |W_i|) \geq t) \\ &= \sum_{i=1}^q \mathbb{P}(|W_i| \geq (1/C_{\text{sup}}) \log(t/C_{\text{sup}})) \\ &\leq 2 \sum_{i=1}^q \mathbb{P}(W \geq (1/C_{\text{sup}}) \log(t/C_{\text{sup}}) - |w_i|), \end{aligned}$$

where $W \sim \mathcal{N}[0, 1]$. From the Gaussian tail inequality, we obtain, for

$$t \geq C_{\text{sup}} \exp(C_{\text{sup}} (\max_{i=1, \dots, q} |w_i| + 1)),$$

that

$$\mathbb{P}(g(W) \geq t) \leq \frac{2q}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left((1/C_{\text{sup}}) \log(t/C_{\text{sup}}) - \max_{i=1, \dots, q} |w_i| \right)^2\right).$$

The function of t above is clearly summable as $t \rightarrow +\infty$. Hence, we have $\mathbb{E}[g(W)] = \int_0^\infty \mathbb{P}(g(W) \geq t) < +\infty$. \square

Lemma 6. *Let X be centered Gaussian process with covariance function k_X satisfying Condition 2. Let F satisfy Condition 3. Let W be the spatial process $F(X(\cdot))$ and assume that W is centered. Let $(x_i)_{i \in \mathbb{N}}$ satisfy Condition 1.*

Then, we have, for any $r_1, r_2 \in \mathbb{N}$ and $\Delta \geq 0$,

$$\sup_{\substack{i_1, \dots, i_{r_1} \in \mathbb{N} \\ j_1, \dots, j_{r_2} \in \mathbb{N} \\ \min_{a=1, \dots, r_1, b=1, \dots, r_2} |x_{i_a} - x_{j_b}| \geq \Delta}} |\text{Cov}(W(x_{i_1}) \dots W(x_{i_{r_1}}), W(x_{j_1}) \dots W(x_{j_{r_2}}))| \leq C_{\text{sup}} e^{-C_{\text{inf}} \Delta},$$

where C_{sup} and C_{inf} depend on r_1, r_2 but not on Δ .

Proof. Let $\Delta \geq 0$, $i_1, \dots, i_{r_1} \in \mathbb{N}$ and $j_1, \dots, j_{r_2} \in \mathbb{N}$ such that $\min_{a=1, \dots, r_1, b=1, \dots, r_2} |x_{i_a} - x_{j_b}| \geq \Delta$.

Let $\epsilon_3 \sim \mathcal{N}[0, I_{r_1}]$ and $\epsilon_4 \sim \mathcal{N}[0, I_{r_2}]$, ϵ_3 and ϵ_4 being independent. Let R be the $r_2 \times r_1$ matrix $(k_X(x_{j_a}, x_{i_b}))_{a,b}$, let C_1 be the $r_1 \times r_1$ matrix $(k_X(x_{i_a}, x_{i_b}))_{a,b}$ and let C_2 be the $r_2 \times r_2$ matrix $(k_X(x_{j_a}, x_{j_b}))_{a,b}$. Let $M = RC_1^{-1}$. Let K be a matrix square root of $C_2 - RC_1^{-1}R^\top$. Let K_1 be the unique symmetric matrix square root of C_1 .

Then the vector $((K_1\epsilon_3)^\top, (MK_1\epsilon_3 + K\epsilon_4)^\top)$ has the same distribution as

$$(X(x_{i_1}), \dots, X(x_{i_{r_1}}), X(x_{j_1}), \dots, X(x_{j_{r_2}})).$$

For $i = 1, 2$, for $x = (x_1, \dots, x_{r_i}) \in \mathbb{R}^{r_i}$, let $f_i(x) = F(x_1) \cdots F(x_{r_i}) \in \mathbb{R}$. Then we have

$$\text{Cov}(W(x_{i_1}) \cdots W(x_{i_{r_1}}), W(x_{j_1}) \cdots W(x_{j_{r_2}})) = \text{Cov}(f_1(K_1\epsilon_3), f_2(MK_1\epsilon_3 + K\epsilon_4)).$$

By a Taylor expansion, there exists a random vector ϵ_5 belonging to the segment with endpoints $K\epsilon_4$ and $MK_1\epsilon_3 + K\epsilon_4$ such that, with $G_2(\epsilon_5)$ the gradient column vector of f_2 at ϵ_5 , we have

$$f_2(MK_1\epsilon_3 + K\epsilon_4) = f_2(K\epsilon_4) + (MK_1\epsilon_3)^\top G_2(\epsilon_5).$$

This yields

$$\begin{aligned} |\text{Cov}(W(x_{i_1}) \dots W(x_{i_{r_1}}), W(x_{j_1}) \dots W(x_{j_{r_2}}))| &= |\text{Cov}(f_1(K_1\epsilon_3), \epsilon_3^\top K_1^\top M^\top G_2(\epsilon_5))| \\ &\leq \sqrt{\mathbb{E}[f_1^2(K_1\epsilon_3)]} \sqrt{\mathbb{E}[(\epsilon_3^\top K_1^\top M^\top G_2(\epsilon_5))^2]}. \end{aligned}$$

From Condition 2 ii) and from the equivalence of norms, we obtain $\|M\|_{op} \leq C_{\text{sup}} e^{-C_{\text{inf}}\Delta}$ and $\|K_1\|_{op} \leq C_{\text{sup}}$, where C_{sup} and C_{inf} do not depend on $i_1, \dots, i_{r_1}, j_1, \dots, j_{r_2}, \Delta$. By equivalence of norms, we then obtain, with C_{sup} and C_{inf} not depending on $i_1, \dots, i_{r_1}, j_1, \dots, j_{r_2}, \Delta$,

$$|\text{Cov}(W(x_{i_1}) \dots W(x_{i_{r_1}}), W(x_{j_1}) \dots W(x_{j_{r_2}}))| \leq C_{\text{sup}} e^{-C_{\text{inf}}\Delta} \sqrt{\mathbb{E}[f_1^2(K_1\epsilon_3)] \mathbb{E}[(\|\epsilon_3\| \|G_2(\epsilon_5)\|)^2]}.$$

Now,

$$\|\epsilon_5\| \leq \|MK_1\epsilon_3\| + \|K\epsilon_4\| \leq C_{\text{sup}} (\|\epsilon_3\| + \|\epsilon_4\|),$$

from Condition 2 ii), where, again, C_{sup} does not depend on $i_1, \dots, i_{r_1}, j_1, \dots, j_{r_2}, \Delta$. Furthermore, $\|K_1\epsilon_3\| \leq C_{\text{sup}}\|\epsilon_3\|$. Eventually, we have

$$\begin{aligned} |\text{Cov}(W(x_{i_1}) \dots W(x_{i_{r_1}}), W(x_{j_1}) \dots W(x_{j_{r_2}}))| \\ \leq C_{\text{sup}} \exp(-C_{\text{inf}}\Delta) \sqrt{\mathbb{E}[f_1^2(K_1\epsilon_3)]} \sqrt{\mathbb{E}\left[\left(\|\epsilon_3\| \sup_{\|x\| \leq C_{\text{sup}}(\|\epsilon_3\| + \|\epsilon_4\|)} \|G_2(x)\|\right)^2\right]}. \end{aligned}$$

From Condition 3 i), we have $|f_1(K_1x)| \leq C_{\text{sup}} e^{C_{\text{sup}}\|x\|}$ and $\|G_2(x)\| \leq C_{\text{sup}} e^{C_{\text{sup}}\|x\|}$ where C_{sup} does not depend on x and $i_1, \dots, i_{r_1}, j_1, \dots, j_{r_2}, \Delta$. Hence the above square roots are finite from Lemma 5 and do not depend on $i_1, \dots, i_{r_1}, j_1, \dots, j_{r_2}$ and Δ . This concludes the proof. \square

Lemma 7. *Consider the setting of Section 3.1, that is k_Z satisfies Condition 2 and T satisfies Condition 3. For $n \in \mathbb{N}$ and $i, j = 1, \dots, n$ we have*

$$\sum_{k,l=1,\dots,n} |\text{Cov}(y_i y_j, y_k y_l)| \leq C_{\text{sup}},$$

where C_{sup} does not depend on n, i, j .

Proof. We let $d(a, (b, c)) = \min(|a - b|, |a - c|)$ for $a, b, c \in \mathbb{R}^d$. It is enough to show that

$$|\text{Cov}(y_i y_j, y_k y_l)| \leq C_{\text{sup}} \exp\left(-C_{\text{inf}} \max(d(s_k, (s_i, s_j)), d(s_l, (s_i, s_j)))\right). \quad (13)$$

Indeed, let, for $t \geq 0$, $N_{i,j,t}$ be the number of pairs (k, l) , with $1 \leq k, l \leq n$ such that

$$t \leq \max(d(s_k, (s_i, s_j)), d(s_l, (s_i, s_j))) \leq t + 1.$$

From Condition 1, we can show that we have $\sup_{n \in \mathbb{N}, i, j=1, \dots, n} N_{i,j,t} \leq C_{\text{sup}} t^{2d}$. Hence, we have

$$\sum_{i,j=1}^n \exp\left(-C_{\text{inf}} \max(d(s_k, (s_i, s_j)), d(s_l, (s_i, s_j)))\right) \leq \sum_{k=0}^{+\infty} C_{\text{sup}} (k+1)^{2d} \exp(-C_{\text{inf}} k) < +\infty.$$

Thus, (13) implies the result of the lemma and it suffices to prove (13).

Let $i, j, k, l \in \{1, \dots, n\}$ and let $\Delta = \max(d(s_k, (s_i, s_j)), d(s_l, (s_i, s_j)))$. By symmetry, we can consider that $d(s_k, (s_i, s_j)) = \Delta$.

If $|s_k - s_l| \leq |s_i - s_l|$ and $|s_k - s_l| \leq |s_j - s_l|$, then $|s_i - s_l| \geq \Delta/2$ and $|s_j - s_l| \geq \Delta/2$. Hence, we can apply Lemma 6 with distance $\Delta/2$ to obtain $|\text{Cov}(y_k y_l, y_i y_j)| \leq C_{\text{sup}} \exp(-C_{\text{inf}} \Delta/2)$, where C_{sup} and C_{inf} do not depend on n, i, j, k, l, Δ .

If $|s_i - s_l| \leq |s_k - s_l|$ and $|s_i - s_l| \leq |s_j - s_l|$, then $|s_k - s_l| \geq \Delta/2$. We then have

$$\begin{aligned} \text{Cov}(y_i y_j, y_k y_l) &= \mathbb{E}[y_i y_j y_k y_l] - \mathbb{E}[y_i y_j] \mathbb{E}[y_k y_l] \\ &= \text{Cov}(y_i y_j y_l, y_k) - \text{Cov}(y_i, y_j) \text{Cov}(y_k, y_l) \end{aligned} \quad (14)$$

since it is assumed that Y has zero-mean. In (14), the first and third covariances are bounded in absolute value by $C_{\text{sup}} \exp(-C_{\text{inf}} \Delta/2)$ from Lemma 6, because $|s_k - s_l| \geq \Delta/2$, $|s_k - s_i| \geq \Delta/2$ and $|s_k - s_j| \geq \Delta/2$. Hence we have $|\text{Cov}(y_k y_l, y_i y_j)| \leq C_{\text{sup}} \exp(-C_{\text{inf}} \Delta/2)$, where C_{sup} and C_{inf} do not depend on n, i, j, k, l, Δ .

If $|s_j - s_l| \leq |s_k - s_l|$ and $|s_j - s_l| \leq |s_i - s_l|$, we obtain the same bound by symmetry. We have thus considered all possible cases and the proof of (13) is concluded. \square

In the context of Theorem 2, the following lemma provides an approximation of V_n , based on replacing A by a sparse matrix. We remark that a similar approximation was shown in a time series context in [32]. Nevertheless, we find that our assumptions on the random field Y are more transparent and interpretable than the assumptions in [32], where cumulants are used. Because of these differences of assumptions, our proof of the following lemma differs from that in [32].

Lemma 8. *Let, for $K, n \in \mathbb{N}$, $A^{(K)}$ be the $n \times n$ matrix defined by*

$$A_{i,j}^{(K)} = A_{i,j} 1_{|s_i - s_j| \leq K}.$$

Then, under the same assumptions as in Lemma 7, we have

$$\sup_{n \in \mathbb{N}} n \text{Var}\left(\frac{1}{n} y^\top A y - \frac{1}{n} y^\top A^{(K)} y\right) \rightarrow_{K \rightarrow \infty} 0.$$

Proof. For any $K, n \in \mathbb{N}$ we have

$$n \text{Var}\left(\frac{1}{n} y^\top A y - \frac{1}{n} y^\top A^{(K)} y\right) = \frac{1}{n} \sum_{i,j,k,l=1}^n (A - A^{(K)})_{i,j} (A - A^{(K)})_{k,l} \text{Cov}(y_i y_j, y_k y_l).$$

We observe that $|(A - A^{(K)})_{k,l}|$ is equal to 0 or is smaller than $C_{\text{sup}}/(1 + K^{d+C_{\text{inf}}})$ by assumption. Hence we have

$$\begin{aligned} n \text{Var}\left(\frac{1}{n}y^\top Ay - \frac{1}{n}y^\top A^{(K)}y\right) &\leq C_{\text{sup}} \frac{1}{1 + K^{d+C_{\text{inf}}}} \frac{1}{n} \sum_{i,j,k,l=1}^n |(A - A^{(K)})_{i,j}| |\text{Cov}(y_i y_j, y_k y_l)| \\ &\leq C_{\text{sup}} \frac{1}{1 + K^{d+C_{\text{inf}}}} \frac{1}{n} \sum_{i,j=1}^n |(A - A^{(K)})_{i,j}| \\ &\leq C_{\text{sup}} \frac{1}{1 + K^{d+C_{\text{inf}}}} \max_{i=1,\dots,n} \sum_{j=1}^n \frac{1}{1 + |s_i - s_j|^{d+C_{\text{inf}}}} \\ &\leq C_{\text{sup}} \frac{1}{1 + K^{d+C_{\text{inf}}}}, \end{aligned}$$

where we have used Lemma 7 and where we have observed that $|(A - A^{(K)})_{i,j}|$ is equal to 0 or is smaller than $C_{\text{sup}}/(1 + |s_i - s_j|^{d+C_{\text{inf}}})$. We have also used Lemma 4 in [23] for the last inequality above. All the above constants C_{sup} and C_{inf} naturally do not depend on n , so the lemma is proved. \square

Lemma 9. *Consider the setting of Section 3.1, that is k_Z satisfies Condition 2 and T satisfies Condition 3. Let $a, b \in \mathbb{N}$. For $i \in \{1, \dots, a\}$, let $\alpha_i \in \mathbb{N}$ and let $I(i, 1), \dots, I(i, \alpha_i) \in \{1, \dots, n\}$. For $j \in \{1, \dots, b\}$, let $\beta_j \in \mathbb{N}$ and let $J(j, 1), \dots, J(j, \beta_j) \in \{1, \dots, n\}$. For $i = 1, \dots, a$, let f_i be a function from \mathbb{R}^{α_i} to \mathbb{R} . For $j = 1, \dots, b$, let g_j be a function from \mathbb{R}^{β_j} to \mathbb{R} . For $i = 1, \dots, a$, let $v^{(i)} = (f_i(Z(s_{I(i,1)}), \dots, Z(s_{I(i,\alpha_i)})))$. For $j = 1, \dots, b$, let $w^{(j)} = (g_j(Z(s_{J(j,1)}), \dots, Z(s_{J(j,\beta_j)})))$. Let*

$$\begin{aligned} &\alpha(\{v^{(1)}, \dots, v^{(a)}\}, \{w^{(1)}, \dots, w^{(b)}\}) \\ &= \sup \left\{ |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|; A \in \sigma(\{v^{(1)}, \dots, v^{(a)}\}), B \in \sigma(\{w^{(1)}, \dots, w^{(b)}\}) \right\}, \end{aligned} \tag{15}$$

where, for any set of random variables $\{\epsilon_1, \dots, \epsilon_r\}$, $\sigma(\{\epsilon_1, \dots, \epsilon_r\})$ is the sigma algebra generated by the random variables $\{\epsilon_1, \dots, \epsilon_r\}$. Let

$$\Delta = \inf_{\substack{i \in \{1, \dots, a\} \\ j \in \{1, \dots, b\} \\ \tilde{i} \in \{1, \dots, \alpha_i\} \\ \tilde{j} \in \{1, \dots, \beta_j\}}} |s_{I(i, \tilde{i})} - s_{J(j, \tilde{j})}|.$$

Then, we have

$$\alpha(\{v^{(1)}, \dots, v^{(a)}\}, \{w^{(1)}, \dots, w^{(b)}\}) \leq C_{\text{sup}} e^{-C_{\text{inf}} \Delta},$$

where C_{sup} and C_{inf} may depend on $a, \alpha_1, \dots, \alpha_a$ but do not depend on $b, (J(j, \tilde{j}))_{j=1, \dots, b, \tilde{j}=1, \dots, \beta_j}$ and Δ .

Proof. Let $\mathcal{I} = \{I(i, \tilde{i}); i = 1, \dots, a, \tilde{i} = 1, \dots, \alpha_i\}$ and let $\mathcal{J} = \{J(j, \tilde{j}); j = 1, \dots, b, \tilde{j} = 1, \dots, \beta_j\}$. In (15), any of the events A (resp. B) is an event defined on the set of random variables $\{Z(s_i)\}_{i \in \mathcal{I}}$ (resp. $\{Z(s_j)\}_{j \in \mathcal{J}}$). We thus obtain

$$\begin{aligned} & \alpha(\{v^{(1)}, \dots, v^{(a)}\}, \{w^{(1)}, \dots, w^{(b)}\}) \\ & \leq \alpha(\{Z(s_i)\}_{i \in \mathcal{I}}, \{Z(s_j)\}_{j \in \mathcal{J}}) \\ & := \sup \left\{ |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|; A \in \sigma(\{Z(s_i)\}_{i \in \mathcal{I}}), B \in \sigma(\{Z(s_j)\}_{j \in \mathcal{J}}) \right\}. \end{aligned}$$

Let $I_1 < \dots < I_{\bar{\alpha}}$ be such that $\{I_1, \dots, I_{\bar{\alpha}}\} = \mathcal{I}$ and let $v_Z = (Z(s_{I_1}), \dots, Z(s_{I_{\bar{\alpha}}}))^\top$. Let $J_1 < \dots < J_{\bar{\beta}}$ be such that $\{J_1, \dots, J_{\bar{\beta}}\} = \mathcal{J}$ and let $w_Z = (Z(s_{J_1}), \dots, Z(s_{J_{\bar{\beta}}}))^\top$. From Lemma 1 in Section 2.1 of [19], we have

$$\begin{aligned} & \alpha(\{Z(s_i)\}_{i \in \mathcal{I}}, \{Z(s_j)\}_{j \in \mathcal{J}}) \\ & \leq \sup \left\{ |\text{Cov}(v^\top v_Z, w^\top w_Z)|; \text{Var}(v^\top v_Z) = 1, \text{Var}(w^\top w_Z) = 1 \right\}. \end{aligned} \tag{16}$$

Let v and w be vectors belonging to the set in (16). The smallest eigenvalues of the covariance matrices of v_Z and w_Z are larger than a constant C_{inf} , not depending on \mathcal{I} and \mathcal{J} , since k_Z satisfies Condition 2. Thus we have

$$1 = \text{Var}(v^\top v_Z) = v^\top \text{Cov}(v_Z)v \geq C_{\text{inf}} \|v\|^2.$$

It follows that $\|v\|^2 \leq C_{\text{sup}}$, where C_{sup} does not depend on \mathcal{I} , \mathcal{J} and Δ . Similarly $\|w\|^2 \leq C_{\text{sup}}$.

We have

$$\begin{aligned} \text{Cov}^2(v^\top v_Z, w^\top w_Z) & \leq \|\text{Cov}(v^\top v_Z, w_Z)\|^2 \|w\|^2 \\ & \leq C_{\text{sup}} \sum_{j=1, \dots, \bar{\beta}} \text{Cov}(v^\top v_Z, Z(s_{J_j}))^2 \\ & \leq C_{\text{sup}} \|v\|^2 \sum_{i=1, \dots, \bar{\alpha}} \sum_{j=1, \dots, \bar{\beta}} \text{Cov}(Z(s_{I_i}), Z(s_{J_j}))^2 \\ & \leq C_{\text{sup}} \sum_{i=1, \dots, \bar{\alpha}} \sum_{\substack{j=1, \dots, n \\ |s_j - s_i| \geq \Delta}} e^{-C_{\text{inf}} |s_{I_i} - s_j|}, \end{aligned}$$

by definition of Δ , since k_Z satisfies Condition 2 and where C_{sup} and C_{inf} do not depend on b , \mathcal{J} and Δ . For any $i \in \{1, \dots, n\}$, the number of indices $j \in \{1, \dots, n\}$ such that $\tilde{\Delta} \leq |s_i - s_j| \leq \tilde{\Delta} + 1$ is smaller than $C_{\text{sup}} \tilde{\Delta}^d$, from Condition 1 and where C_{sup} only depends on d . This yields

$$\begin{aligned} \text{Cov}^2(v^\top v_Z, w^\top w_Z) & \leq C_{\text{sup}} (\alpha_1 + \dots + \alpha_a) \sum_{k=0}^{+\infty} C_{\text{sup}} (\Delta + k)^d e^{-C_{\text{inf}} |\Delta + k|} \\ & \leq C_{\text{sup}} e^{-C_{\text{inf}} |\Delta|/2} (\alpha_1 + \dots + \alpha_a) \sum_{k=1}^{+\infty} (\Delta + k)^d e^{-C_{\text{inf}} |\Delta + k|/2} \\ & \leq C_{\text{sup}} e^{-C_{\text{inf}} |\Delta|} \end{aligned}$$

where the different C_{sup} and C_{inf} do not depend on b , \mathcal{J} and Δ . This concludes the proof from (16). \square

Lemma 10. Consider a sequence $(x_i)_{i \in \mathbb{N}}$ of points in \mathbb{R}^d satisfying Condition 1. Let $\tau > 0$ be fixed. For $n \in \mathbb{N}$, let $(A_\theta)_{\theta \in \Theta}$ and $(B_\theta)_{\theta \in \Theta}$ be families of $n \times n$ matrices. Assume that for all $n \in \mathbb{N}$, $i, j = 1, \dots, n$ and $\theta \in \Theta$,

$$|(A_\theta)_{i,j}| \leq \frac{C_{\text{sup}}}{1 + |x_i - x_j|^{d+\tau}} \quad \text{and} \quad |(B_\theta)_{i,j}| \leq \frac{C_{\text{sup}}}{1 + |x_i - x_j|^{d+\tau}}$$

where C_{sup} does not depend on n, i, j, θ . Then we have for all $n \in \mathbb{N}$, $i, j = 1, \dots, n$ and $\theta \in \Theta$,

$$|(A_\theta B_\theta)_{i,j}| \leq \frac{C_{\text{sup}}}{1 + |x_i - x_j|^{d+\tau}}$$

where C_{sup} does not depend on n, i, j, θ .

Proof. We have,

$$\begin{aligned} |(A_\theta B_\theta)_{i,j}| &= \left| \sum_{\ell=1}^n (A_\theta)_{i,\ell} (B_\theta)_{\ell,j} \right| \\ &\leq \sum_{\ell=1}^n \frac{C_{\text{sup}}}{1 + |x_i - x_\ell|^{d+\tau}} \frac{C_{\text{sup}}}{1 + |x_j - x_\ell|^{d+\tau}} \\ &\leq \sum_{\substack{\ell=1, \dots, n \\ |x_i - x_\ell| \leq |x_j - x_\ell|}} \frac{C_{\text{sup}}}{1 + |x_i - x_\ell|^{d+\tau}} \frac{C_{\text{sup}}}{1 + (|x_i - x_j|/2)^{d+\tau}} \\ &\quad + \sum_{\substack{\ell=1, \dots, n \\ |x_j - x_\ell| \leq |x_i - x_\ell|}} \frac{C_{\text{sup}}}{1 + (|x_i - x_j|/2)^{d+\tau}} \frac{C_{\text{sup}}}{1 + |x_j - x_\ell|^{d+\tau}} \\ &\leq C_{\text{sup}} \frac{C_{\text{sup}}}{1 + |x_i - x_j|^{d+\tau}} \max_{a=1, \dots, n} \sum_{b=1}^n \frac{1}{1 + |x_a - x_b|^{d+\tau}} \\ &\leq \frac{C_{\text{sup}}}{1 + |x_i - x_j|^{d+\tau}} \end{aligned}$$

from Lemma 4 in [23]. \square

Lemma 11. Consider the setting of Section 5.1. Under Conditions 4 and 5, we have

$$\sup_{\theta \in \Theta} \lambda_1(R_\theta^{-1}) \leq C_{\text{sup}}, \quad (17)$$

$$\sup_{\theta \in \Theta} \lambda_1(R_\theta) \leq C_{\text{sup}}, \quad (18)$$

and

$$\sup_{\substack{\theta \in \Theta \\ \ell=1,2,3 \\ i_1, \dots, i_\ell=1, \dots, p}} \lambda_1 \left(\frac{\partial R_\theta}{\partial \theta_{i_1} \dots \partial \theta_{i_\ell}} \right) \leq C_{\text{sup}}. \quad (19)$$

Proof. Conditions 1, 4 and 5 imply (17) from Theorem 5 in [11]. Conditions 1 and 4 imply (18) and (19) from Lemma 6 in [23]. \square

Lemma 12. *Consider the setting of Section 5.1. Under Conditions 4 and 5, we have, for $n \in \mathbb{N}$ and $i, j \in \{1, \dots, n\}$,*

$$\sup_{\theta \in \Theta} |(R_\theta^{-1})_{i,j}| \leq \frac{C_{\text{sup}}}{1 + |s_i - s_j|^{d+C_{\text{inf}}}},$$

where C_{sup} and C_{inf} do not depend on n, i, j, θ .

Proof. One can show that the proof of Theorem 1 can be made uniform over $\theta \in \Theta$, thus yielding Lemma 12. \square

Lemma 13. *Consider the setting of Section 5.1. Under Conditions 4 and 5, we have,*

$$\inf_{\theta \in \Theta} \lambda_n(R_\theta) \geq C_{\text{inf}}, \quad (20)$$

$$\inf_{\theta \in \Theta} \lambda_n(\text{diag}(R_\theta^{-1})) \geq C_{\text{inf}}. \quad (21)$$

Proof. Equation (20) holds from (17). Then, (21) follows from (20) as in Lemma D.6 in [7]. \square

A.2. Proofs of the main results

Proof of Lemma 1. As a special case of Lemma 6, k' satisfies Condition 2 i).

Let us now show that k' satisfies Condition 2 ii). Let (x_i) satisfy Condition 1. Let $n \in \mathbb{N}$ be fixed and let R be the $n \times n$ covariance matrix $k(x_i - x_j)_{i,j=1,\dots,n}$. Let $a_1, \dots, a_n \in \mathbb{R}$. We have

$$\sum_{i,j=1}^n a_i a_j R_{i,j} = \text{Var} \left(\sum_{i=1}^n a_i F(X(x_i)) \right).$$

We now let $z = (X(x_1), \dots, X(x_n))^\top$ and $g : \mathbb{R}^n \rightarrow \mathbb{R}$ be defined by $g(t) = \sum_{i=1}^n a_i F(t_i)$. The gradient of g at t is $\nabla g(t) = (a_1 F'(t_1), \dots, a_n F'(t_n))^\top$. We use the inequality in Theorem 3.7 in [16]. This yields

$$\sum_{i,j=1}^n a_i a_j R_{i,j} \geq \mathbb{E} [\nabla g(z)]^\top \text{Cov}(z) \mathbb{E} [\nabla g(z)].$$

From Condition 2 ii), we have $\lambda_1(\text{Cov}(z)) \geq C_{\text{inf}}$. This yields

$$\begin{aligned} \sum_{i,j=1}^n a_i a_j R_{i,j} &\geq C_{\text{inf}} \sum_{i=1}^n \mathbb{E}^2 [(\nabla g(z))_i] \\ &= C_{\text{inf}} \sum_{i=1}^n a_i^2 \mathbb{E} [F'(z_i)]^2 \\ &= C_{\text{inf}} \left(\sum_{i=1}^n a_i^2 \right) \mathbb{E} [F'(z_1)]^2. \end{aligned}$$

From Condition 3, the above expectation is non-zero, which concludes the proof. \square

Proof of Lemma 2. The fact that k satisfies Condition 2 ii) follows from Theorem 4 in [11].

Let us now consider the case where F is defined by $F(x) = x^{2r} + u$. Since we consider a covariance function we can assume that $u = 0$ without loss of generality. Assume also that $k(0) = 1$ without loss of generality. From Lemma 6, k' satisfies Condition 2 i). Let us show that Condition 2 ii) is satisfied. Let $a, b \in \mathbb{R}^d$, let $c = k(a - b)$ and let $\lambda = (1 - c^2)^{1/2}$. With $(A_1, A_2) \sim \mathcal{N}[0, I_2]$, we have

$$\begin{aligned} k'(a - b) &= \text{Cov}(A_1^{2r}, (cA_1 + \lambda A_2)^{2r}) \\ &= \text{Cov}\left(A_1^{2r}, \sum_{i=0}^{2r} \binom{2r}{i} c^i \lambda^{2r-i} A_1^i A_2^{2r-i}\right) \\ &= \sum_{i=0}^{2r} \binom{2r}{i} c^i \lambda^{2r-i} \text{Cov}(A_1^{2r}, A_1^i A_2^{2r-i}). \end{aligned}$$

By independence of A_1 and A_2 , we obtain, for $i = 0, \dots, 2r$,

$$\text{Cov}(A_1^{2r}, A_1^i A_2^{2r-i}) = \mathbb{E}[A_2^{2r-i}] (\mathbb{E}[A_1^{2r+i}] - \mathbb{E}[A_1^{2r}] \mathbb{E}[A_1^i]). \quad (22)$$

From Isserlis' theorem, one can show that (22) is zero if i is odd and is strictly positive if i is even. As a consequence, we have

$$k'(a - b) = \sum_{i=0}^r \alpha_i k(a - b)^{2i}$$

with $\alpha_1, \dots, \alpha_r > 0$. Hence, the Fourier transform of k' is a linear combination of multiple convolutions of the Fourier transform of k , with strictly positive components. Since the Fourier transform of k is strictly positive everywhere, then also the Fourier transform of k' is strictly positive everywhere. Hence, from Theorem 4 in [11], k' satisfies Condition 2 ii). \square

Proof of Theorem 1. Condition 1 and Lemma 6 in [23] imply that the spectral norms of R^{-1} and R are bounded functions of n . Let $C_{\text{sup}} = \sup_n \lambda_1(R) < \infty$ and $C_{\text{inf}} = \inf_n \lambda_n(R) > 0$.

We write

$$R^{-1} = \frac{1}{C_{\text{sup}}} \left(I - \left(I - \frac{R}{C_{\text{sup}}} \right) \right)^{-1} = \frac{1}{C_{\text{sup}}} \sum_{\ell=0}^{\infty} \left(I - \frac{R}{C_{\text{sup}}} \right)^{\ell}. \quad (23)$$

We remark that the above sum is well-defined because the eigenvalues of $I - R/C_{\text{sup}}$ are between 0 and $1 - C_{\text{inf}}/C_{\text{sup}}$.

We denote $M = I - C_{\text{sup}}^{-1}R$ and $h_{i,j} = |s_i - s_j|$. Let $1 \leq A < \infty$ and $a > 0$ be fixed such that $M_{i,j} \leq Ae^{-2ah_{i,j}}$. Let $\delta = \inf_{i,j \in \mathbb{N}, i \neq j} |s_i - s_j| > 0$ (Condition 1).

Let $D < \infty$ be a constant such that $D \geq 1$ and, for any $L \geq \delta$ and $i \in \mathbb{N}$, the set $\{j \in \mathbb{N}; |s_i - s_j| \leq L\}$ has no more than $(D/2)L^d$ elements.

Let $0 < \mu < \infty$ be fixed. We show by induction over $\ell \in \mathbb{N}$ that there exists a constant $1 \leq \varphi < \infty$, depending on μ but not depending on ℓ, i, j , such that for $\ell \leq \mu \log(h_{i,j})$,

$$|(M^\ell)_{i,j}| \leq A^\ell \varphi^\ell D^\ell h_{i,j}^{d\ell-d} e^{-ah_{i,j}}. \quad (24)$$

In the case $\log(h_{i,j}) < 0$, there is nothing to prove in (24), so we consider i, j such that $\log(h_{i,j}) \geq 0$ when proving (24).

For $\ell = 1$, (24) holds. Assume that (24) holds for some $\ell \in \mathbb{N}$. We have

$$\begin{aligned} |(M^{\ell+1})_{i,j}| &= \sum_{r=1}^n (M^\ell)_{i,r} M_{j,r} \\ &\leq \sum_{r=1}^n A^\ell \varphi^\ell D^\ell h_{i,r}^{d\ell-d} e^{-ah_{i,r}} A e^{-2ah_{r,j}} \\ &= A^{\ell+1} \varphi^\ell D^\ell \sum_{r=1}^n h_{i,r}^{d\ell-d} e^{-ah_{i,r}} e^{-2ah_{r,j}}. \end{aligned}$$

Let now $B_i = \{x \in \mathbb{R}^d; |x - s_i| \leq |s_i - s_j|\}$ and $B_j = \{x \in \mathbb{R}^d; |x - s_j| \leq |s_i - s_j|\}$. From the triangle inequality we obtain

$$\begin{aligned} |(M^{\ell+1})_{i,j}| &\leq A^{\ell+1} \varphi^\ell D^\ell \sum_{r \in \mathbb{N}; s_r \in B_i \cup B_j} h_{i,r}^{d\ell-d} e^{-ah_{i,r}} e^{-2ah_{r,j}} \\ &\quad + A^{\ell+1} \varphi^\ell D^\ell \sum_{r \in \mathbb{N}; s_r \in B_i^c \cap B_j^c} h_{i,r}^{d\ell-d} e^{-ah_{i,r}} e^{-2ah_{r,j}} \\ &\leq A^{\ell+1} \varphi^\ell D^\ell 2(D/2) h_{i,j}^d h_{i,j}^{d\ell-d} e^{-ah_{i,j}} \\ &\quad + A^{\ell+1} \varphi^\ell D^\ell e^{-2ah_{i,j}} \sum_{r \in \mathbb{N}; s_r \in B_i^c} h_{i,r}^{d\ell-d} e^{-ah_{i,r}} \\ &\leq A^{\ell+1} \varphi^\ell D^{\ell+1} h_{i,j}^{d(\ell+1)-d} e^{-ah_{i,j}} \\ &\quad + A^{\ell+1} \varphi^\ell D^\ell e^{-ah_{i,j}} \left(e^{-ah_{i,j}} \sum_{r \in \mathbb{N}} Q r^{d\mu \log(h_{i,j})} e^{-a(r-1)} \right), \end{aligned}$$

where for the last inequality we let Qr^d be an upper bound on the cardinality of $\{b \in \mathbb{N}; |s_b - s_i| \in [r-1, r]\}$ for all $i \in \mathbb{N}$. The constant Q is finite and depends only on d and δ from Condition 1. We also let $\ell \leq \mu \log(h_{i,j})$ to show the last above inequality. Hence, in order to finish the proof of (24), it remains to show that the term (\cdot) in the above display is a bounded function of $h_{i,j}$, and to let $\varphi/2 \geq 1$ be a bound for the term (\cdot) .

We have, for $h_{i,j}$ large enough, with $\lceil \cdot \rceil$ the integer ceiling,

$$\begin{aligned}
 \left(e^{-ah_{i,j}} \sum_{r \in \mathbb{N}} Q r^{d\mu \log(h_{i,j})} e^{-a(r-1)} \right) &= Q e^a e^{-ah_{i,j}} \sum_{r \in \mathbb{N}} r^{\lceil d\mu \log(h_{i,j}) \rceil} e^{-ar} \\
 &= Q e^a e^{-ah_{i,j}} \left(\frac{2}{a} \right)^{\lceil d\mu \log(h_{i,j}) \rceil} \sum_{r \in \mathbb{N}} \left(\frac{ar}{2} \right)^{\lceil d\mu \log(h_{i,j}) \rceil} e^{-ar} \\
 &\leq Q e^a e^{-ah_{i,j}} \left(\frac{2}{a} \right)^{\lceil d\mu \log(h_{i,j}) \rceil} [d\mu \log(h_{i,j})]! \sum_{r \in \mathbb{N}} e^{ar/2} e^{-ar} \\
 &= Q e^a e^{-ah_{i,j}} \left(\frac{2}{a} \right)^{\lceil d\mu \log(h_{i,j}) \rceil} [d\mu \log(h_{i,j})]! \frac{1}{1 - e^{-ar/2}}.
 \end{aligned}$$

The above function of $h_{i,j}$ clearly goes to 0 as $h_{i,j}$ goes to ∞ . Thus, the above term (.) is bounded and thus (24) is proved.

Coming back to (23), using (24) and using the triangle inequality, we obtain, letting $\Delta = 1 - C_{\text{inf}}/C_{\text{sup}}$, and for $h_{i,j}$ large enough,

$$\begin{aligned}
 |(R^{-1})_{i,j}| &\leq \left(\sum_{1 \leq \ell \leq \mu \log(h_{i,j})} A^\ell \varphi^\ell D^\ell h_{i,j}^{d\ell-d} e^{-ah_{i,j}} \right) + \sum_{\mu \log(h_{i,j}) \leq \ell \leq \infty} \Delta^\ell \\
 &\leq \mu \log(h_{i,j}) (A\varphi D)^{\mu \log(h_{i,j})} h_{i,j}^{d\mu \log(h_{i,j})} e^{-ah_{i,j}} + \frac{\Delta^{\mu \log(h_{i,j})}}{1 - \Delta} \\
 &= \mu \log(h_{i,j}) (A\varphi D)^{\mu \log(h_{i,j})} h_{i,j}^{d\mu \log(h_{i,j})} e^{-ah_{i,j}} + \frac{h_{i,j}^{\mu \log(\Delta)}}{1 - \Delta}. \quad (25)
 \end{aligned}$$

In the above display, for any $\tau < \infty$ in the statement of Theorem 1, we can choose μ such that $\mu \log(\Delta) \leq -d - \tau$. Then, it is clear that the first summand in (25) is also smaller than a constant (depending on τ) time $h_{i,j}^{-d-\tau}$. This concludes the proof of Theorem 1, since also $\sup_{n \in \mathbb{N}} \max_{i,j=1,\dots,n} |(R^{-1})_{i,j}|$ is bounded by C_{sup} . \square

Proof of Theorem 2. We have

$$n \text{Var}(V_n) = \frac{1}{n} \sum_{i,j,k,l=1}^n A_{i,j} A_{k,l} \text{Cov}(y_i y_j, y_k y_l). \quad (26)$$

From Lemma 7, we obtain

$$\begin{aligned}
 n \text{Var}(V_n) &\leq \frac{1}{n} C_{\text{sup}} \sum_{i,j=1}^n |A_{i,j}| \\
 &\leq C_{\text{sup}} \max_{i=1,\dots,n} \sum_{j=1}^n \frac{1}{1 + |s_i - s_j|^{d+C_{\text{inf}}}} \\
 &\leq C_{\text{sup}}
 \end{aligned}$$

from Lemma 4 in [23]. Hence, $n \text{Var}(V_n)$ is bounded as $n \rightarrow \infty$.

Assume now that

$$d_w(\mathcal{L}_n, \mathcal{N}[0, n \operatorname{Var}(V_n)]) \not\rightarrow 0 \quad (27)$$

as $n \rightarrow \infty$. Because $n \operatorname{Var}(V_n)$ is bounded, there exists a subsequence $\phi(n)$ such that

$$d_w(\mathcal{L}_{\phi(n)}, \mathcal{N}[0, \phi(n) \operatorname{Var}(V_{\phi(n)})]) \not\rightarrow 0 \quad (28)$$

as $n \rightarrow \infty$ and $\phi(n) \operatorname{Var}(V_{\phi(n)}) \rightarrow V \in [0, \infty)$ as $n \rightarrow \infty$. It is then simple to show that this implies

$$d_w(\mathcal{L}_{\phi(n)}, \mathcal{N}[0, V]) \not\rightarrow 0 \quad (29)$$

as $n \rightarrow \infty$. If $V = 0$, then, from Chebyshev inequality, $\mathcal{L}_{\phi(n)}$ converges to a Dirac mass at zero and so (29) does not hold, yielding a contradiction.

Hence it remains to consider the case $\phi(n) \operatorname{Var}(V_{\phi(n)}) \rightarrow V \in (0, \infty)$ as $n \rightarrow \infty$ and where (29) holds.

To reach a contradiction, we will show that

$$\sqrt{\phi(n)} \left(\frac{V_{\phi(n)} - \mathbb{E}[V_{\phi(n)}]}{\sqrt{V}} \right) \rightarrow_{n \rightarrow \infty} \mathcal{N}[0, 1].$$

To simplify notations in the sequel, without loss of generality, we will consider that $\phi(n) = n$ and show that

$$\sqrt{n} \left(\frac{V_n - \mathbb{E}[V_n]}{\sqrt{V}} \right) \rightarrow_{n \rightarrow \infty} \mathcal{N}[0, 1], \quad (30)$$

where $n \operatorname{Var}(V_n) \rightarrow V \in (0, \infty)$ as $n \rightarrow \infty$. From Slutsky's lemma it is sufficient to show that

$$\sqrt{n} \left(\frac{V_n - \mathbb{E}[V_n]}{\sqrt{n \operatorname{Var}(V_n)}} \right) \rightarrow_{n \rightarrow \infty} \mathcal{N}[0, 1]. \quad (31)$$

For $K \geq 0$, let

$$V_n^{(K)} = \frac{1}{n} y^\top A^{(K)} y,$$

with the notation Lemma 8. We have

$$\begin{aligned} & \sup_{n \in \mathbb{N}} \mathbb{E} \left[\left(\sqrt{n} \left(\frac{V_n - \mathbb{E}[V_n]}{\sqrt{n \operatorname{Var}(V_n)}} \right) - \sqrt{n} \left(\frac{V_n^{(K)} - \mathbb{E}[V_n^{(K)}]}{\sqrt{n \operatorname{Var}(V_n^{(K)})}} \right) \right)^2 \right] \\ & \leq 2 \sup_{n \in \mathbb{N}} \operatorname{Var} \left(\sqrt{n} \left(\frac{V_n - V_n^{(K)}}{\sqrt{n \operatorname{Var}(V_n)}} \right) \right) + 2 \sup_{n \in \mathbb{N}} \operatorname{Var} \left(\sqrt{n} V_n^{(K)} \left(\frac{1}{\sqrt{n \operatorname{Var}(V_n)}} - \frac{1}{\sqrt{n \operatorname{Var}(V_n^{(K)})}} \right) \right) \\ & \rightarrow_{n \rightarrow \infty} 0 \end{aligned}$$

from Lemma 8 and because $V > 0$. Hence, from Theorem 4.2 in [15] (as in [32]), it is sufficient to show that there exists $L \in (0, \infty)$ such that for any fixed $K \geq L$, we have

$$\sqrt{n} \left(\frac{V_n^{(K)} - \mathbb{E}[V_n^{(K)}]}{\sqrt{n \operatorname{Var}(V_n^{(K)})}} \right) \rightarrow^{\mathcal{L}} \mathcal{N}[0, 1]$$

as $n \rightarrow \infty$, in order to prove (31) and thus to conclude the proof. We remark that, because of Lemma 8, we have $|\liminf_{n \rightarrow \infty} n \operatorname{Var}(V_n^{(K)}) - \liminf_{n \rightarrow \infty} n \operatorname{Var}(V_n)|$ goes to 0 as $K \rightarrow \infty$. Hence, we may take L such that $\liminf_{n \rightarrow \infty} n \operatorname{Var}(V_n^{(K)}) > 0$ for $K \geq L$. Hence, up to extracting a subsequence, it is sufficient to show

$$\sqrt{n}(V_n^{(K)} - \mathbb{E}[V_n^{(K)}]) \rightarrow^{\mathcal{L}} \mathcal{N}[0, V^{(K)}], \quad (32)$$

where $n \operatorname{Var}(V_n^{(K)}) \rightarrow V^{(K)} > 0$ as $n \rightarrow \infty$. We have

$$\begin{aligned} \sqrt{n}(V_n^{(K)} - \mathbb{E}[V_n^{(K)}]) &= \frac{1}{\sqrt{n}} \sum_{i,j=1}^n (y_i y_j - \mathbb{E}[y_i y_j]) A_{i,j} \mathbf{1}_{|s_i - s_j| \leq K} \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\sum_{\substack{j=1, \dots, n \\ |s_i - s_j| \leq K}} A_{i,j} (T(Z(s_i))T(Z(s_j)) - \mathbb{E}[T(Z(s_i))T(Z(s_j))]) \right) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n X_n(s_i), \end{aligned}$$

say, where X_n can be interpreted as a centered random field defined on $(s_i)_{i \in \mathbb{N}}$. We will now show that the sequence of random fields $(X_n)_{n \in \mathbb{N}}$ satisfies the conditions of Corollary 1 of [26].

We let, for $k, l \in \mathbb{N}$ and $r \geq 0$

$$\begin{aligned} \bar{\alpha}_{k,l}(r) &= \sup_{n \in \mathbb{N}} \sup \left\{ |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|; \right. \\ &\quad A \in \sigma(X_n(s_{I_1}), \dots, X_n(s_{I_{\bar{k}}}), B \in \sigma(X_n(s_{J_1}), \dots, X_n(s_{J_{\bar{l}}}), \\ &\quad \left. \bar{k} \leq k, \bar{l} \leq l, I_1, \dots, I_{\bar{k}}, J_1, \dots, J_{\bar{l}} \in \{1, \dots, n\}, \min_{\substack{\bar{k}=1, \dots, \bar{k}, \bar{l}=1, \dots, \bar{l}}} |s_{I_{\bar{k}}} - s_{J_{\bar{l}}}| \geq r \right\}. \end{aligned}$$

Let $N_K = \sup_{n \in \mathbb{N}} \max_{i=1, \dots, n} \sum_{j=1}^n \mathbf{1}_{|s_j - s_i| \leq K}$. Then $N_K \leq C_{\sup}$, where C_{\sup} depends only on K , d and δ from Condition 1. We remark that for $i = 1, \dots, n$, $X_n(s_i)$ is a function of the variables $Z(s_{I(n,i,1)}), \dots, Z(s_{I(n,i,\gamma(n,i))})$, with $\gamma(n,i) \leq N_K$ and with $|s_{I(n,i,\gamma)} - s_i| \leq K$ for $\gamma = 1, \dots, \gamma(n,i)$. Furthermore, for $|s_i - s_j| \geq r$ we have for $\gamma_i = 1, \dots, \gamma(n,i)$ and for $\gamma_j = 1, \dots, \gamma(n,j)$ that $|s_{I(n,i,\gamma_i)} - s_{I(n,j,\gamma_j)}| \geq r - 2K$. Hence, from Lemma 9, we have, for any $k \in \mathbb{N}$

$$\sup_{l \in \mathbb{N}} \bar{\alpha}_{k,l}(r) \leq C_{\sup} e^{-C_{\inf} r},$$

where C_{\sup} and C_{\inf} may depend on k and K .

We now let $D = (s_i)_{i \in \mathbb{N}}$, $D_n = (s_1, \dots, s_n)$, $Z_{i,n} = n^{-1/2} X_n(s_i)$ for $n \in \mathbb{N}$ and $i = 1, \dots, n$. We also let $c_{i,n} = n^{-1/2}$ for $n \in \mathbb{N}$ and $i = 1, \dots, n$. We remark that $Z_{i,n}/c_{i,n}$ can be written as $f(w)$ where w is a Gaussian vector of dimension less than N_K , with variances 1 and where $|f(x)| \leq C_{\sup} e^{C_{\sup} |x|}$, where C_{\sup} does not depend on $n \in \mathbb{N}$ and $i = 1, \dots, n$. One can thus show, from the Cauchy-Schwarz inequality and with the same techniques as in Lemma 5, that for any $q > 0$

$$\lim_{M \rightarrow +\infty} \sup_{n \in \mathbb{N}} \max_{i=1, \dots, n} \mathbb{E} \left[|Z_{i,n}/c_{i,n}|^{2+q} \mathbf{1}_{|Z_{i,n}/c_{i,n}| \geq M} \right] = 0. \quad (33)$$

With the previous notation and with (33), one can show that all the assumptions of Corollary 1 in [26] are satisfied. This shows (32) and thus concludes the proof. \square

Proof of Theorem 3. Let $\theta \in \Theta$ be fixed. We have

$$\text{Var}(L_\theta) = \frac{1}{n} \text{Var} \left(\frac{1}{n} (y^\top R_\theta^{-1} y) \right).$$

From Lemma 12 and Theorem 2, applied with $A_n = R_\theta^{-1}$, we obtain $\text{Var}(L_\theta) \rightarrow 0$ as $n \rightarrow \infty$. For $i = 1, \dots, p$,

$$\frac{\partial L_\theta}{\partial \theta_i} = \frac{1}{n} \text{tr} \left(R_\theta^{-1} \frac{\partial R_\theta}{\partial \theta_i} \right) + \frac{1}{n} \left(y^\top \left(-R_\theta^{-1} \frac{\partial R_\theta}{\partial \theta_i} R_\theta^{-1} \right) y \right),$$

which can be rewritten for convenience as

$$\frac{\partial L_\theta}{\partial \theta_i} = \frac{1}{n} \text{tr} (P_{\theta,i}) + \frac{1}{n} (y^\top Q_{\theta,i} y)$$

with

$$P_{\theta,i} = R_\theta^{-1} \frac{\partial R_\theta}{\partial \theta_i} \quad \text{and} \quad Q_{\theta,i} = -R_\theta^{-1} \frac{\partial R_\theta}{\partial \theta_i} R_\theta^{-1}.$$

The matrices R_θ^{-1} and $\partial R_\theta / \partial \theta_i$ are both valid choices for A_θ and B_θ in Lemma 10. From Gerschgorin Circle Theorem (GCT) and Lemma 4 in [23], we obtain $\sup_{\theta \in \Theta} \lambda_1(P_{\theta,i}^\top P_{\theta,i}) \leq C_{\text{sup}}$ and $\sup_{\theta \in \Theta} \lambda_1(Q_{\theta,i}) \leq C_{\text{sup}}$. This, in turn implies that $\sup_{\theta \in \Theta} \rho_1(P_{\theta,i}) \leq C_{\text{sup}}$. It follows that

$$\begin{aligned} \max_{i=1, \dots, p} \sup_{\theta \in \Theta} \left| \frac{\partial L_\theta}{\partial \theta_i} \right| &\leq \sup_{\theta \in \Theta} \left| \frac{1}{n} \text{tr} (P_{\theta,i}) + \frac{1}{n} (y^\top Q_{\theta,i} y) \right| \\ &\leq C_{\text{sup}} + C_{\text{sup}} \frac{\|y\|^2}{n} \\ &= O_p(1). \end{aligned}$$

Hence, Theorem 3 can be proved by proceeding as in the proof of Proposition 3.1 in [7]. \square

Proof of Theorem 4. From the proof of Theorem 3, we have for $i = 1, \dots, p$

$$\partial L_\theta / \partial \theta_i = \frac{1}{n} \text{tr}(P_{\theta,i}) + \frac{1}{n} (y^\top Q_{\theta,i} y),$$

where $P_{\theta,i}$ is a $n \times n$ matrix satisfying $\sup_{\theta \in \Theta} |(P_{\theta,i})_{a,b}| \leq C_{\text{sup}} / (1 + |s_a - s_b|^{d+C_{\text{inf}}})$ and $Q_{\theta,i}$ is a $n \times n$ symmetric matrix satisfying $\sup_{\theta \in \Theta} |(Q_{\theta,i})_{a,b}| \leq C_{\text{sup}} / (1 + |s_a - s_b|^{d+C_{\text{inf}}})$.

One can check that $\partial L_{\theta_0} / \partial \theta_i$ has mean zero for $i = 1, \dots, p$, since the mean value of $\partial L_{\theta_0} / \partial \theta_i$ is calculated as if Y were a Gaussian process with zero-mean and covariance function k_{Y, θ_0} . Let $\partial L_{\theta_0} / \partial \theta$ be the gradient column vector of L_θ at θ_0 . From Theorem 2, with $\mathcal{L}_{\Sigma, \theta_0, n}$ the distribution of $\sqrt{n} \partial L_{\theta_0} / \partial \theta$, as $n \rightarrow \infty$,

$$d_w(\mathcal{L}_{\Sigma, \theta_0, n}, \mathcal{N}[0, \Sigma_{\theta_0}]) \rightarrow 0. \quad (34)$$

In addition, for $i \in \{1, \dots, p\}$, the sequence $(n \text{Var}(\partial L_{\theta_0}/\partial \theta_i))$ is bounded, which implies that the elements of Σ_{θ_0} are bounded too.

One can check that the mean value of $\partial L_{\theta_0}/\partial \theta_i \partial \theta_j$ is $(M_{\theta_0})_{i,j}$ (also because this mean value is calculated as if Y were a Gaussian process with zero-mean and covariance function k_{Y, θ_0}).

Also, for $i, j = 1, \dots, p$, we have

$$\partial L_{\theta_0}/\partial \theta_i \partial \theta_j = \frac{1}{n} \text{tr}(C_{\theta, i, j}) + \frac{1}{n} (y^\top D_{\theta, i, j} y),$$

where $C_{\theta, i, j}$ and $D_{\theta, i, j}$ are sums of products of the matrices R_θ^{-1} , R_θ and the first and second derivative matrices of R_θ (see e.g., [7]). Hence, from Condition 4 and Lemma 12 used inside Lemma 10, we have $\sup_{\theta \in \Theta} |(C_{\theta, i, j})_{a, b}| \leq C_{\text{sup}}/(1 + |s_a - s_b|^{d+C_{\text{inf}}})$ and $\sup_{\theta \in \Theta} |(D_{\theta, i, j})_{a, b}| \leq C_{\text{sup}}/(1 + |s_a - s_b|^{d+C_{\text{inf}}})$.

Thus, the variance of $\partial L_{\theta_0}/\partial \theta_i \partial \theta_j$ goes to zero as $n \rightarrow \infty$ from Theorem 2. Hence

$$\partial L_{\theta_0}/\partial \theta_i \partial \theta_j \xrightarrow{p} (M_{\theta_0})_{i, j} \quad (35)$$

as $n \rightarrow \infty$.

It can be shown, similarly as in the proof of Proposition 3.3 in [7] that

$$\liminf_{n \rightarrow \infty} \lambda_p(M_{\theta_0}) > 0. \quad (36)$$

Hence, (7) follows.

Then, for $i, j, \ell \in \{1, \dots, p\}$, we have

$$\partial L_{\theta_0}/\partial \theta_i \partial \theta_j \partial \theta_\ell = \frac{1}{n} \text{tr}(E_{\theta, i, j, \ell}) + \frac{1}{n} (y^\top F_{\theta, i, j, \ell} y),$$

where $E_{\theta, i, j, \ell}$ and $F_{\theta, i, j, \ell}$ are sums of products of the matrices R_θ^{-1} , R_θ and the first, second and third derivative matrices of R_θ . Hence, from Condition 4 and Lemma 12 used inside Lemma 10, we have $\sup_{\theta \in \Theta} |(E_{\theta, i, j, \ell})_{a, b}| \leq C_{\text{sup}}/(1 + |s_a - s_b|^{d+C_{\text{inf}}})$ and $\sup_{\theta \in \Theta} |(F_{\theta, i, j, \ell})_{a, b}| \leq C_{\text{sup}}/(1 + |s_a - s_b|^{d+C_{\text{inf}}})$. Then, from GCT and Lemma 4 in [23], we have $\sup_{\theta \in \Theta} \rho_1(E_{\theta, i, j, \ell}) \leq C_{\text{sup}}$ and $\sup_{\theta \in \Theta} \rho_1(F_{\theta, i, j, \ell}) \leq C_{\text{sup}}$. Hence, as in the proof of Theorem 3, we can show

$$\sup_{\theta \in \Theta} \left| \frac{\partial L_{\theta_0}}{\partial \theta_i \partial \theta_j \partial \theta_\ell} \right| = O_p(1). \quad (37)$$

Also, $\lambda_1(M_{\theta_0})$ is clearly bounded as $n \rightarrow \infty$. From Theorem 2, $\lambda_1(\Sigma_{\theta_0})$ is bounded as $n \rightarrow \infty$. Hence, by considering subsequences along which M_{θ_0} and Σ_{θ_0} converge, and using (34), (35), (36) and (37), we can proceed as in the proof of Proposition D.10 in [7] and show (6). \square

Proof of Theorem 5. Let $\psi \in \mathcal{S}$ be fixed. We have

$$\text{Var}(CV_\psi) = \frac{1}{n} \text{Var} \left(\frac{1}{n} y^\top C_\psi^{-1} \text{diag}(C_\psi^{-1})^{-2} C_\psi^{-1} y \right).$$

From Lemmas 10, 12 and 13 (that can be trivially adapted by replacing θ by ψ), as well as Theorem 2, applied with $A_n = C_\psi^{-1} \text{diag}(C_\psi^{-1})^{-2} C_\psi^{-1}$, we obtain $\text{Var}(CV_\psi) \rightarrow 0$ as $n \rightarrow \infty$.

For $i = 1, \dots, p-1$,

$$\frac{\partial CV_\psi}{\partial \psi_i} = \frac{2}{n} y^\top A_{\psi,i} y$$

with

$$A_{\psi,i} = C_\psi^{-1} \text{diag}(C_\psi^{-1})^{-2} \left(\text{diag} \left(C_\psi^{-1} \frac{\partial C_\psi}{\partial \psi_i} C_\psi^{-1} \right) \text{diag}(C_\psi^{-1})^{-1} - C_\psi^{-1} \frac{\partial C_\psi}{\partial \psi_i} \right) C_\psi^{-1}.$$

As in the proof of Theorem 3, GCT and Lemma 4 in [23] lead us to $\sup_{\psi \in \mathcal{S}} \lambda_1(A_{\psi,i}^\top A_{\psi,i}) \leq C_{\text{sup}}$, which in turn implies $\sup_{\psi \in \mathcal{S}} \rho_1(A_{\psi,i}) \leq C_{\text{sup}}$. It follows that

$$\begin{aligned} \max_{i=1, \dots, p-1} \sup_{\psi \in \mathcal{S}} \left| \frac{\partial CV_\psi}{\partial \psi_i} \right| &\leq \sup_{\psi \in \mathcal{S}} \left| \frac{2}{n} (y^\top A_{\psi,i} y) \right| \\ &\leq C_{\text{sup}} \frac{\|y\|^2}{n} \\ &= O_p(1). \end{aligned}$$

Hence, Theorem 5 can also be proved by proceeding as in the proof of Proposition 3.4 in [7]. \square

Proof of Theorem 6. From Condition 4 and Lemma 12 used inside Lemma 10, we have for $i = 1, \dots, p-1$

$$\frac{\partial CV_\psi}{\partial \psi_i} = \frac{2}{n} y^\top A_{\psi,i} y,$$

where $A_{\psi,i}$ is a $n \times n$ matrix satisfying $\sup_{\psi \in \mathcal{S}} |(A_{\psi,i})_{a,b}| \leq C_{\text{sup}}/(1 + |s_a - s_b|^{d+C_{\text{inf}}})$. As in the proof of Theorem 4, one can check that $\partial CV_{\psi_0}/\partial \psi_i$ has mean zero for $i = 1, \dots, p-1$. Let $\partial CV_{\psi_0}/\partial \psi$ be the gradient column vector of CV_ψ at ψ_0 . From Theorem 2, with $\mathcal{L}_{\Gamma, \psi_0, n}$ the distribution of $\sqrt{n} \partial CV_{\psi_0}/\partial \psi$, as $n \rightarrow \infty$,

$$d_w(\mathcal{L}_{\Gamma, \psi_0, n}, \mathcal{N}[0, \Gamma_{\psi_0}]) \rightarrow 0. \quad (38)$$

In addition, for $i \in \{1, \dots, p-1\}$, the sequence $(n \text{Var}(\partial CV_{\psi_0}/\partial \psi_i))$ is bounded and thus, the elements of Γ_{ψ_0} are bounded too.

One can check that the mean value of $\partial CV_{\psi_0}/\partial \psi_i \partial \psi_j$ is $(N_{\psi_0})_{i,j}$. Furthermore, from Theorem 2, the variance of $\partial CV_{\psi_0}/\partial \psi_i \partial \psi_j$ goes to zero as $n \rightarrow \infty$. Hence, as $n \rightarrow \infty$,

$$\partial CV_{\psi_0}/\partial \psi_i \partial \psi_j \rightarrow^P (N_{\psi_0})_{i,j}. \quad (39)$$

It can be shown, similarly as in the proof of Proposition 3.7 in [7] that

$$\liminf_{n \rightarrow \infty} \lambda_{p-1}(N_{\psi_0}) > 0. \quad (40)$$

Hence, (10) follows. On the other hand, for $i, j = 1, \dots, p-1$, we have

$$\frac{\partial CV_\psi}{\partial \psi_i \partial \psi_j} = \frac{1}{n} y^\top D_{\psi, i, j} y,$$

where $D_{\psi, i, j}$ is computed as a sum of products of the matrices C_ψ^{-1} , C_ψ , the first and second derivative matrices of C_ψ and the diag operator (see e.g., [7]). Hence, from Condition 4 and Lemma 12 used inside Lemma 10, we have $\sup_{\psi \in \mathcal{S}} |(D_{\psi, i, j})_{a, b}| \leq C_{\text{sup}}/(1 + |s_a - s_b|^{d+C_{\text{inf}}})$.

Similarly, for $i, j, \ell \in \{1, \dots, p-1\}$, we have

$$\frac{\partial CV_\psi}{\partial \psi_i \partial \psi_j \partial \psi_\ell} = \frac{1}{n} y^\top E_{\psi, i, j, \ell} y,$$

where $E_{\psi, i, j, \ell}$ is a sum of products of the matrices C_ψ^{-1} , C_ψ , the first, second and third derivative matrices of C_ψ and the diag operator. Hence, from Condition 4 and Lemma 12 used inside Lemma 10, we have $\sup_{\psi \in \mathcal{S}} |(E_{\psi, i, j, \ell})_{a, b}| \leq C_{\text{sup}}/(1 + |s_a - s_b|^{d+C_{\text{inf}}})$. Then, from GCT and Lemma 4 in [23], we have $\sup_{\psi \in \mathcal{S}} \rho_1(E_{\psi, i, j, \ell}) \leq C_{\text{sup}}$. Hence, as in the proof of Theorem 3, we can show, for $i, j, \ell \in \{1, \dots, p-1\}$,

$$\sup_{\psi \in \mathcal{S}} \left| \frac{\partial CV_\psi}{\partial \psi_i \partial \psi_j \partial \psi_\ell} \right| = O_p(1). \quad (41)$$

Also, $\lambda_1(N_{\psi_0})$ is clearly bounded as $n \rightarrow \infty$. From Theorem 2, $\lambda_1(\Gamma_{\psi_0})$ is bounded as $n \rightarrow \infty$. Hence, by considering subsequences along which N_{ψ_0} and Γ_{ψ_0} converge, and using (38), (39), (40) and (41), we can proceed as in the proof of Proposition D.10 in [7] and show (9). \square

Proof of Lemma 4. We have, with $\theta = (\sigma^2, \psi) \in \Theta$ and $\theta_0 = (\sigma_0^2, \psi_0)$,

$$\begin{aligned} & \frac{1}{n} \sum_{i, j=1}^n (c_{Y, \psi}(s_i - s_j) - c_{Y, \psi_0}(s_i - s_j))^2 \\ &= \frac{1}{n} \sum_{i, j=1}^n \left(\frac{k_{Y, \theta}(s_i - s_j)}{k_{Y, \theta}(0)} - \frac{k_{Y, \theta_0}(s_i - s_j)}{k_{Y, \theta_0}(0)} \right)^2 \\ &\leq \frac{2}{n} \sum_{i, j=1}^n \left(\frac{k_{Y, \theta}(s_i - s_j)}{k_{Y, \theta}(0)} - \frac{k_{Y, \theta_0}(s_i - s_j)}{k_{Y, \theta}(0)} \right)^2 + \frac{2}{n} \sum_{i, j=1}^n \left(\frac{k_{Y, \theta_0}(s_i - s_j)}{k_{Y, \theta}(0)} - \frac{k_{Y, \theta_0}(s_i - s_j)}{k_{Y, \theta_0}(0)} \right)^2 \\ &\leq \frac{C_{\text{sup}}}{n} \sum_{i, j=1}^n (k_{Y, \theta}(s_i - s_j) - k_{Y, \theta_0}(s_i - s_j))^2 + \left(\frac{1}{k_{Y, \theta}(0)} - \frac{1}{k_{Y, \theta_0}(0)} \right)^2 C_{\text{sup}}, \end{aligned} \quad (42)$$

where the second C_{sup} comes from Lemma 11 and from the classical control of the square Frobenius norm by n times the largest square eigenvalue, for $n \times n$ symmetric matrices. If Condition 9 holds, then for all $\alpha > 0$,

$$\liminf_{n \rightarrow \infty} \inf_{\|\psi - \psi_0\| \geq \alpha} \frac{1}{n} \sum_{i, j=1}^n (c_{Y, \psi}(s_i - s_j) - c_{Y, \psi_0}(s_i - s_j))^2 > 0.$$

Consider a sequence $\theta_n = (\sigma_n^2, \psi_n) \in \Theta$ such that $\|\theta_n - \theta_0\| \geq \alpha$. If we can extract a subsequence n_m such that $\liminf_{m \rightarrow \infty} (\sigma_{n_m}^2 - \sigma_0^2)^2 > 0$, then clearly

$$\liminf_{m \rightarrow \infty} \frac{1}{n_m} \sum_{i,j=1}^{n_m} (k_{Y, \theta_{n_m}}(s_i - s_j) - k_{Y, \theta_0}(s_i - s_j))^2 > 0$$

by considering the diagonal terms in the above double sum. If we can not extract such a subsequence, then we can extract a subsequence n_m such that $\|\psi_{n_m} - \psi_0\| \geq \alpha/2$ and $\sigma_{n_m}^2 \rightarrow \sigma_0^2$ as $m \rightarrow \infty$. Along this subsequence

$$\liminf_{m \rightarrow \infty} \frac{1}{n_m} \sum_{i,j=1}^{n_m} (k_{Y, \theta_{n_m}}(s_i - s_j) - k_{Y, \theta_0}(s_i - s_j))^2 > 0$$

from (42). Hence, Condition 7 holds.

Let us now assume that Condition 8 does not hold. We have, with $\theta = (\sigma^2, \psi) \in \Theta$, with $\theta_0 = (\sigma_0^2, \psi_0)$ and with $(\beta_1, \dots, \beta_p) = (\beta_1, \alpha_1, \dots, \alpha_{p-1})$, where $\beta_1 \in \mathbb{R}$ is arbitrary,

$$\begin{aligned} & \frac{1}{n} \sum_{i,j=1}^n \left(\sum_{\ell=1}^{p-1} \alpha_\ell \frac{\partial}{\partial \psi_\ell} c_{Y, \psi_0}(s_i - s_j) \right)^2 \\ &= \frac{1}{n} \sum_{i,j=1}^n \left(\sum_{\ell=1}^p \beta_\ell \frac{\partial}{\partial \theta_\ell} \left(\frac{k_{Y, \theta_0}(s_i - s_j)}{k_{Y, \theta_0}(0)} \right) \right)^2 \\ &= \frac{1}{n} \sum_{i,j=1}^n \left(\sum_{\ell=1}^p \beta_\ell \frac{\frac{\partial}{\partial \theta_\ell} k_{Y, \theta_0}(s_i - s_j)}{k_{Y, \theta_0}(0)} - \sum_{\ell=1}^p \beta_\ell \frac{k_{Y, \theta_0}(s_i - s_j)}{k_{Y, \theta_0}(0)^2} \frac{\partial}{\partial \theta_\ell} k_{Y, \theta_0}(0) \right)^2 \\ &\leq \frac{2}{n} \sum_{i,j=1}^n \left(\sum_{\ell=1}^p \beta_\ell \frac{\frac{\partial}{\partial \theta_\ell} k_{Y, \theta_0}(s_i - s_j)}{k_{Y, \theta_0}(0)} \right)^2 + \frac{2}{n} \sum_{i,j=1}^n \left(\sum_{\ell=1}^p \beta_\ell \frac{k_{Y, \theta_0}(s_i - s_j)}{k_{Y, \theta_0}(0)^2} \frac{\partial}{\partial \theta_\ell} k_{Y, \theta_0}(0) \right)^2 \\ &\leq \frac{C_{\text{sup}}}{n} \sum_{i,j=1}^n \left(\sum_{\ell=1}^p \beta_\ell \frac{\partial}{\partial \theta_\ell} k_{Y, \theta_0}(s_i - s_j) \right)^2 + C_{\text{sup}} \left(\sum_{\ell=1}^p \beta_\ell \frac{\partial}{\partial \theta_\ell} k_{Y, \theta_0}(0) \right)^2, \end{aligned} \tag{43}$$

where the second C_{sup} comes from Lemma 11 and from the classical control of the square Frobenius norm by n times the largest square eigenvalue, for $n \times n$ symmetric matrices. If Condition 8 does not hold, there exists $(\beta_1^*, \dots, \beta_p^*) \neq (0, \dots, 0)$ and a subsequence n_m such that

$$\frac{1}{n_m} \sum_{i,j=1}^{n_m} \left(\sum_{\ell=1}^p \beta_\ell^* \frac{\partial}{\partial \theta_\ell} k_{Y, \theta_0}(s_i - s_j) \right)^2 \rightarrow_{m \rightarrow \infty} 0$$

and thus, considering the diagonal elements in the double sum above,

$$\beta_1^* = \sum_{\ell=1}^p \beta_\ell^* \frac{\partial}{\partial \theta_\ell} k_{Y, \theta_0}(0) = 0.$$

Hence, from (43), letting $(\beta_1^*, \dots, \beta_p^*) = (0, \alpha_1^*, \dots, \alpha_{p-1}^*)$, we have

$$\frac{1}{n_m} \sum_{i,j=1}^{n_m} \left(\sum_{\ell=1}^{p-1} \alpha_\ell^* \frac{\partial}{\partial \psi_\ell} c_{Y, \psi_0}(s_i - s_j) \right)^2 \rightarrow 0$$

and thus Condition 10 does not hold. \square

Proof of Theorem 7. Let λ and γ be two column vectors in \mathbb{R}^p and \mathbb{R}^{p-1} . Let also

$$W_n = \lambda^\top (\hat{\theta}_{\text{ML}} - \theta_0) + \gamma^\top (\hat{\psi}_{\text{CV}} - \psi_0),$$

for $n \in \mathbb{N}$.

From the proofs of Theorems 4 and 6 (see also the proof of Proposition D.10 in [7] that is referred to there), we know

$$\sqrt{n}(\hat{\theta}_{\text{ML}} - \theta_0) = \sqrt{n}M_{\theta_0}^{-1} \frac{\partial}{\partial \theta} L_{\theta_0} + o_p(1)$$

and

$$\sqrt{n}(\hat{\psi}_{\text{CV}} - \psi_0) = \sqrt{n}N_{\psi_0}^{-1} \frac{\partial}{\partial \psi} CV_{\psi_0} + o_p(1).$$

Also, from Condition 4 and Lemma 12 used inside Lemma 10, we have for $i = 1, \dots, p$,

$$\partial L_{\theta_0} / \partial \theta_i = \frac{1}{n} (y^\top A_{\theta_0, i} y) + c_{\theta_0},$$

where $c_{\theta_0} \in \mathbb{R}$ is deterministic and, for $i = 1, \dots, p-1$,

$$\partial CV_{\psi_0} / \partial \psi_i = \frac{1}{n} (y^\top B_{\psi_0, i} y),$$

where $A_{\theta, i}$ is a $n \times n$ symmetric matrix satisfying $\sup_{\theta \in \Theta} |(A_{\theta, i})_{a,b}| \leq C_{\text{sup}} / (1 + |s_a - s_b|^{d+C_{\text{inf}}})$ and $B_{\psi, i}$ is a $n \times n$ matrix satisfying $\sup_{\psi \in \mathcal{S}} |(B_{\psi, i})_{a,b}| \leq C_{\text{sup}} / (1 + |s_a - s_b|^{d+C_{\text{inf}}})$. As in the proofs of Theorems 4 and 6, one can check that $\partial L_{\theta_0} / \partial \theta_i$ has mean zero for $i = 1, \dots, p$ and $\partial CV_{\psi_0} / \partial \psi_i$ has mean zero for $i = 1, \dots, p-1$. Thus, we can rewrite

$$W_n = J_n - \mathbb{E}[J_n] + o_p(n^{-1/2})$$

with

$$J_n = \frac{1}{n} y^\top \left(\sum_{i=1}^p (\lambda^\top M_{\theta_0}^{-1})_i A_{\theta_0, i} + \sum_{i=1}^{p-1} (\gamma^\top N_{\psi_0}^{-1})_i B_{\psi_0, i} \right) y.$$

As the vectors λ and γ as well as the matrices $M_{\theta_0}^{-1}$ and $N_{\psi_0}^{-1}$ are fixed, the bound

$$\left| \left(\sum_{i=1}^p (\lambda^\top M_{\theta_0}^{-1})_i A_{\theta_0, i} + \sum_{i=1}^{p-1} (\gamma^\top N_{\psi_0}^{-1})_i B_{\psi_0, i} \right)_{k,l} \right| \leq \frac{C_{\text{sup}}}{1 + |s_k - s_l|^{d+C_{\text{inf}}}}$$

holds for all $k, l \in \{1, \dots, n\}$.

Let then $\mathcal{L}_{J, \theta_0, n}$ be the distribution of $n^{1/2}(J_n - \mathbb{E}[J_n])$. Then, from Theorem 2, as $n \rightarrow \infty$,

$$d_w(\mathcal{L}_{J, \theta_0, n}, \mathcal{N}[0, n \text{Var}(J_n)]) \rightarrow 0.$$

The variance can be written as

$$\begin{aligned} n \text{Var}(J_n) &= \sum_{i=1}^p \sum_{j=1}^p (\lambda^\top M_{\theta_0}^{-1})_i (\lambda^\top M_{\theta_0}^{-1})_j (\Sigma_{\theta_0})_{i,j} + \sum_{i=1}^{p-1} \sum_{j=1}^{p-1} (\gamma^\top N_{\psi_0}^{-1})_i (\gamma^\top N_{\psi_0}^{-1})_j (\Gamma_{\psi_0})_{i,j} \\ &\quad + 2 \sum_{i=1}^p \sum_{j=1}^{p-1} (\lambda^\top M_{\theta_0}^{-1})_i (\gamma^\top N_{\psi_0}^{-1})_j (\Omega_{\theta_0})_{i,j} \end{aligned}$$

with

$$\Omega_{i,j} = \text{Cov} \left(\sqrt{n} \frac{\partial}{\partial \theta_i} L_{\theta_0}, \sqrt{n} \frac{\partial}{\partial \psi_j} CV_{\psi_0} \right).$$

Hence, by applying product by blocks we get the matrix form expression

$$n \text{Var}(J_n) = (\lambda^\top, \gamma^\top) D_{\theta_0}^{-1} \Psi_{\theta_0} D_{\theta_0}^{-1} \begin{pmatrix} \lambda \\ \gamma \end{pmatrix}.$$

We conclude the proof by applying the Wald Theorem. \square

Acknowledgements

RF acknowledges the support of the Swiss National Science Foundation SNSF-175529. FB acknowledges the support of a PEPS from the French Centre national de la recherche scientifique. This research was partly undertaken within the RISCOPE ANR project.

References

- [1] P. Abrahamsen. A review of Gaussian random fields and correlation functions. Technical report, Norwegian Computing Center, 1997.
- [2] A. Alegría, S. Caro, M. Bevilacqua, E. Porcu, and J. Clarke. Estimating covariance functions of multivariate skew-Gaussian random fields on the sphere. *Spatial Statistics*, 22:388–402, 2017.
- [3] E. Anderes. On the consistent separation of scale and variance for Gaussian random fields. *The Annals of Statistics*, 38:870–893, 2010.
- [4] J.-M. Azaïs, F. Bachoc, T. Klein, A. Lagnoux, and T. M. N. Nguyen. Semi-parametric estimation of the variogram of a Gaussian process with stationary increments. *arXiv:1806.03135*, 2018.
- [5] J.-M. Azaïs and M. Wschebor. *Level Sets and Extrema of Random Processes and Fields*. John Wiley & Sons, 2009.
- [6] F. Bachoc. Cross validation and maximum likelihood estimations of hyper-parameters of Gaussian processes with model misspecification. *Computational Statistics and Data Analysis*, 66:55–69, 2013.

- [7] F. Bachoc. Asymptotic analysis of the role of spatial sampling for covariance parameter estimation of Gaussian processes. *Journal of Multivariate Analysis*, 125:1–35, 2014.
- [8] F. Bachoc. Asymptotic analysis of covariance parameter estimation for Gaussian processes in the misspecified case. *Bernoulli*, 24(2):1531–1575, 2018.
- [9] F. Bachoc, K. Ammar, and J. Martinez. Improvement of code behavior in a design of experiments by metamodeling. *Nuclear science and engineering*, 183(3):387–406, 2016.
- [10] F. Bachoc, G. Bois, J. Garnier, and J. Martinez. Calibration and improved prediction of computer models by universal Kriging. *Nuclear Science and Engineering*, 176(1):81–97, 2014.
- [11] F. Bachoc and R. Furrer. On the smallest eigenvalues of covariance matrices of multivariate spatial processes. *Stat*, 5(1):102–107, 2016.
- [12] F. Bachoc, F. Gamboa, J.-M. Loubes, and N. Venet. A Gaussian process regression model for distribution inputs. *IEEE Transactions on Information Theory*, 64(10):6620–6637, 2018.
- [13] F. Bachoc, A. Lagnoux, and T. M. N. Nguyen. Cross-validation estimation of covariance parameters under fixed-domain asymptotics. *Journal of Multivariate Analysis*, 160:42–67, 2017.
- [14] J. M. Bates and C. W. Granger. The combination of forecasts. *Journal of the Operational Research Society*, 20(4):451–468, 1969.
- [15] P. Billingsley. *Convergence of Probability Measures*. Wiley, New York, 1968.
- [16] T. Cacoullos. On upper and lower bounds for the variance of a function of a random variable. *The Annals of Probability*, pages 799–809, 1982.
- [17] J.-P. Chiles and P. Delfiner. *Geostatistics: Modeling Spatial Uncertainty*. John Wiley & Sons, 2009.
- [18] N. Cressie. *Statistics for Spatial Data*. John Wiley & Sons, 1993.
- [19] P. Doukhan. *Mixing, Properties and Examples*. Lecture Notes in Statistics. Springer-Verlag, New York, 1994.
- [20] J. Du, H. Zhang, and V. Mandrekar. Fixed-domain asymptotic properties of tapered maximum likelihood estimators. *The Annals of Statistics*, 37:3330–3361, 2009.
- [21] O. Dubrule. Cross validation of Kriging in a unique neighborhood. *Mathematical Geology*, 15:687–699, 1983.
- [22] R. M. Dudley. *Real Analysis and Probability*. Cambridge University Press, 2002.
- [23] R. Furrer, F. Bachoc, and J. Du. Asymptotic properties of multivariate tapering for estimation and prediction. *Journal of Multivariate Analysis*, 149:177–191, 2016.
- [24] I. Ibragimov and Y. Rozanov. *Gaussian Random Processes*. Springer-Verlag, New York, 1978.
- [25] J. Istas and G. Lang. Quadratic variations and estimation of the local Hölder index of a Gaussian process. *Annales de l’Institut Henri Poincaré*, 33:407–436, 1997.
- [26] N. Jenish and I. R. Prucha. Central limit theorems and uniform laws of large numbers for arrays of random fields. *Journal of econometrics*, 150(1):86–98,

- 2009.
- [27] D. Jones, M. Schonlau, and W. Welch. Efficient global optimization of expensive black box functions. *Journal of Global Optimization*, 13:455–492, 1998.
 - [28] C. Kaufman and B. Shaby. The role of the range parameter for estimation and prediction in geostatistics. *Biometrika*, 100:473–484, 2013.
 - [29] F. Lavancier and P. Rochet. A general procedure to combine estimators. *Computational Statistics & Data Analysis*, 94:175–192, 2016.
 - [30] K. Mardia and R. Marshall. Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika*, 71:135–146, 1984.
 - [31] G. Matheron. *La Théorie des Variables Régionalisées et ses Applications*. Fascicule 5 in Les Cahiers du Centre de Morphologie Mathématique de Fontainebleau. Ecole Nationale Supérieure des Mines de Paris, 1970.
 - [32] M. H. Neumann. A central limit theorem for triangular arrays of weakly dependent random variables, with applications in statistics. *ESAIM: Probability and Statistics*, 17:120–134, 2013.
 - [33] R. Paulo, G. Garcia-Donato, and J. Palomo. Calibration of computer models with multivariate output. *Computational Statistics and Data Analysis*, 56:3959–3974, 2012.
 - [34] C. Rasmussen and C. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, Cambridge, 2006.
 - [35] J. Sacks, W. Welch, T. Mitchell, and H. Wynn. Design and analysis of computer experiments. *Statistical Science*, 4:409–423, 1989.
 - [36] T. Santner, B. Williams, and W. Notz. *The Design and Analysis of Computer Experiments*. Springer, New York, 2003.
 - [37] B. A. Shaby and D. Ruppert. Tapered covariance: Bayesian estimation and asymptotics. *Journal of Computational and Graphical Statistics*, 21(2):433–452, 2012.
 - [38] M. Stein. Asymptotically efficient prediction of a random field with a misspecified covariance function. *Annals of Statistics*, 16:55–63, 1988.
 - [39] M. Stein. Bounds on the efficiency of linear predictions using an incorrect covariance function. *Annals of Statistics*, 18:1116–1138, 1990.
 - [40] M. Stein. Uniform asymptotic optimality of linear predictions of a random field using an incorrect second-order structure. *Annals of Statistics*, 18:850–872, 1990.
 - [41] M. Stein. *Interpolation of Spatial Data: Some Theory for Kriging*. Springer-Verlag, New York, 1999.
 - [42] D. Wang and W.-L. Loh. On fixed-domain asymptotics and covariance tapering in Gaussian random field models. *Electronic Journal of Statistics*, 5:238–269, 2011.
 - [43] G. Xu and M. G. Genton. Tukey g-and-h random fields. *Journal of the American Statistical Association*, 112(519):1236–1249, 2017.
 - [44] Y. Yan and M. G. Genton. Gaussian likelihood inference on data from trans-Gaussian random fields with Matérn covariance function. *Environmetrics*, 29(5-6):e2458, 2018.
 - [45] Z. Ying. Asymptotic properties of a maximum likelihood estimator with

- data from a Gaussian process. *Journal of Multivariate Analysis*, 36:280–296, 1991.
- [46] Z. Ying. Maximum likelihood estimation of parameters under a spatial sampling scheme. *Annals of Statistics*, 21:1567–1590, 1993.
- [47] H. Zhang. Inconsistent estimation and asymptotically equivalent interpolations in model-based geostatistics. *Journal of the American Statistical Association*, 99:250–261, 2004.
- [48] H. Zhang and Y. Wang. Kriging and cross validation for massive spatial data. *Environmetrics*, 21:290–304, 2010.