



HAL
open science

Fixed-domain asymptotic properties of composite likelihood estimators for Gaussian processes

François Bachoc, Agnès Lagnoux

► **To cite this version:**

François Bachoc, Agnès Lagnoux. Fixed-domain asymptotic properties of composite likelihood estimators for Gaussian processes. *Journal of Statistical Planning and Inference*, 2020, 209, pp.62-75. hal-02079975

HAL Id: hal-02079975

<https://hal.science/hal-02079975>

Submitted on 26 Mar 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.

Fixed-domain asymptotic properties of composite likelihood estimators for Gaussian processes

F. Bachoc¹ and A. Lagnoux²

¹Institut de Mathématiques de Toulouse; UMR5219. Université de Toulouse; CNRS. UT3, F-31062 Toulouse, France.

²Institut de Mathématiques de Toulouse; UMR5219. Université de Toulouse; CNRS. UT2J, F-31058 Toulouse, France.

March 26, 2019

Abstract

We consider the estimation of the variance and spatial scale parameters of the covariance function of a one-dimensional Gaussian process with fixed smoothness parameter s . We study the fixed-domain asymptotic properties of composite likelihood estimators. As an improvement of previous references, we allow for any fixed number of neighbor observation points, both on the left and on the right sides, for the composite likelihood. First, we examine the case where only the variance parameter is unknown. We prove that for small values of s , the composite likelihood estimator converges at a sub-optimal rate and we provide its non-Gaussian asymptotic distribution. For large values of s , the estimator converges at the optimal rate. Second, we consider the case where the variance and the spatial scale are jointly estimated. We obtain the same conclusions as for the first case for the estimation of the microergodic parameter. The theoretical results are confirmed in numerical simulations.

Keywords: Gaussian process, composite likelihood, microergodicity, consistency, convergence rate, non-Gaussian limit, fixed-domain asymptotics.

1 Introduction

Gaussian processes are widely used in statistical science to model spatial data. When fitting a Gaussian field, one has to deal with the issue of the estimation of its covariance function. In many cases, it is assumed that this function belongs to a given parametric model or family of covariance functions, which turns the problem into a parametric estimation problem. Within this framework, the maximum likelihood estimator (MLE) [29, 37]

of the covariance parameters has been deeply studied in the last years in the two following asymptotic frameworks. The fixed-domain asymptotic framework, sometimes called infill asymptotics [9, 37], corresponds to the case where more and more data are observed in some fixed bounded sampling domain in \mathbb{R}^d ; while the increasing-domain asymptotic framework corresponds to the case where the sampling domain, also in \mathbb{R}^d , increases with the number of observed data and the distance between any two sampling locations is bounded away from 0. The asymptotic behavior of the MLE of the covariance parameters can be quite different under these two frameworks [46].

Under increasing-domain asymptotics, generally speaking, for all (identifiable) covariance parameters, the MLE is consistent and asymptotically normal under some mild regularity conditions. The asymptotic covariance matrix is equal to the inverse of the (asymptotic) Fisher information matrix [3, 10, 26, 32].

The situation is significantly different under fixed-domain asymptotics. Indeed, two types of covariance parameters can be distinguished: microergodic and non-microergodic parameters. A covariance parameter is microergodic if, for two different values of it, the two corresponding Gaussian measures are orthogonal, see [19, 37]. It is non-microergodic if, even for two different values of it, the two corresponding Gaussian measures are equivalent. Non-microergodic parameters cannot be estimated consistently, but misspecifying them asymptotically results in the same statistical inference as specifying them correctly [34, 35, 36, 45]. In the case of isotropic Matérn covariance functions with $d \leq 3$, [45] shows that only a reparametrized quantity obtained from the variance and the spatial scale parameters is microergodic. The asymptotic normality of the MLE of this microergodic parameter is then obtained in [21]. Similar results for the special case of the exponential covariance function were obtained previously in [44].

The maximum likelihood method is generally considered as the best option for estimating the covariance parameters of a Gaussian process (at least in the framework of the present paper, where the true covariance function does belong to the parametric model, see also [2, 4]). Nevertheless, the evaluation of the likelihood function requires to solve a system of linear equations and to compute a determinant. For a data set of n observations, the computational burden is $O(n^3)$, making this method computationally untractable for large data sets. This fact motivates the search for estimation methods with a good balance between computational complexity and statistical efficiency. Among these methods, we can mention low rank approximation (see [38] and the references therein for a review), sparse approximation [16], covariance tapering [13, 22], Gaussian Markov random fields approximation [11, 30], submodel aggregation [7, 12, 17, 31, 40, 41] and composite likelihood. With composite likelihood we indicate a general class of objective functions based on the likelihood of marginal or conditional events [42]. This kind of estimation method has two important benefits: it is generally appealing when dealing with large data sets and it can be helpful when it is difficult to specify the full likelihood.

Consider the observations $y_1 = Y(x_1), \dots, y_n = Y(x_n)$ of a Gaussian process Y corresponding to the observation points x_1, \dots, x_n . In this work, we focus on composite likelihood estimators (CLEs) of the covariance parameters that maximize the sum, over $i = 1, \dots, n$, of the conditional log likelihood of y_i given a subset of $\{y_1, \dots, y_n\} \setminus \{y_i\}$ that corresponds

to observation points that are nearby x_i . These estimators have been considered in several references, including [27, 28, 39, 43]. More generally, the principle of conditioning based on neighbor observation points rather than on the full set of observation points is widely applied for Gaussian processes [14, 15].

Despite their popularity in practice, no general fixed-domain asymptotic results exist for the above CLEs. The existing results address the exponential covariance function in dimension one. In this case, letting $x_1 \leq \dots \leq x_n$ be the observation points, the CLE coincides with the MLE due to the Markov property when the likelihood of each y_i is evaluated conditionally to the previous observations y_{i-1}, \dots, y_{i-K} for any arbitrary value of $K \geq 1$ (see [44]). The CLE of the microergodic parameter is asymptotically Gaussian in this special case. When each y_i is evaluated conditionally to its two neighbor observations y_{i-1}, y_{i+1} , then the CLE of the microergodic parameter is also asymptotically Gaussian [6]. Finally, we remark that also pairwise likelihood estimators have been analyzed recently, in the case of the exponential covariance function in dimension one [5].

In this work, we provide a fixed-domain asymptotic analysis of composite likelihood, for Gaussian processes in dimension one, that extends the previous references considerably, in terms of generality. Indeed, we allow for covariance functions $\sigma^2 k_\alpha$ where $k_\alpha(t) = 1 - |\alpha t|^s + r(\alpha t)$ where the remainder $r(\alpha t)$ is negligible compared to $|\alpha t|^s$ as $t \rightarrow 0$. Here σ^2 is the variance parameter, α is the spatial scale parameter and s is the fixed smoothness parameter, with $0 < s < 3/2$. In contrast, only the special case with $s = 1$ corresponding to exponential covariance functions is considered in [6, 44]. In particular, we allow for general Matérn covariance functions with parameter ν between 0 and 0.75, while in [6, 44], only the case $\nu = 0.5$ is considered. Furthermore, we allow for any fixed number of neighbor observation points, both on the left and on the right, for the composite likelihood, as opposed to two neighbor points or only points on the left in [6, 44].

First we consider the case where only the variance parameter σ_0^2 is estimated. We show that if $0 < s < 1/2$, then the CLE converges at the sub-optimal rate n^s , with an explicit asymptotic variance and is not asymptotically Gaussian, regardless of the number of neighbors used. Furthermore, we provide its non-Gaussian asymptotic distribution. This result is somehow surprising since, in this setting, quadratic variation estimators, also having a small computational cost compared to the MLE, would converge at rate $n^{1/2}$ [1]. This could motivate practical adjustments of composite likelihood for Gaussian processes with small smoothness. For $1/2 \leq s < 3/2$, the CLE converges at the optimal rate $n^{1/2}$.

Second, we consider the case where the variance σ_0^2 and the spatial scale α_0 are jointly estimated, in which case $\sigma_0^2 \alpha_0^s$ is microergodic. We obtain the same conclusions as above. For $0 < s < 1/2$, the CLE has sub-optimal rate n^s and we provide its non-Gaussian asymptotic approximation. Furthermore, the CLE has rate $n^{1/2}$ for $1/2 \leq s < 3/2$.

Many of the proof techniques we suggest are original, notably to take into account several neighbor points, on the left and on the right, for the composite likelihood. This situation was not explored theoretically in the references [5, 6]. In particular, we approximate the conditional expectations and variances, given a fixed number of neighbor observations under fixed-domain asymptotics, see (26) and (27) in the proofs. Furthermore, we apply some concepts from the literature of quadratic variation estimators [1, 20] to composite

likelihood, such as finite sequences applied to functions with given orders of differentiability. In numerical simulations in the case $0 < s < 1/2$, we confirm the rate n^s and the expression of the asymptotic variance. We observe that the number of observations n may need to be very large for the asymptotic results to provide an accurate approximation of the finite sample results.

The rest of the paper is organized as follows. In Section 2, we introduce the model and the CLE. In Section 3, we provide the results for the estimation of the variance parameter σ_0^2 while Section 4 is dedicated to the estimation of the microergodic parameter $\sigma_0^2 \alpha_0^s$. Section 5 presents the numerical results. Concluding remarks are given in Section 6. All the proofs are given in the appendix.

2 The context and notation

We consider a centered Gaussian process Y defined on $[0, 1]$, real-valued. We consider a parametric model of stationary covariance functions of the form $\{\sigma^2 k_\alpha; \sigma^2 \geq 0, \alpha \in A\}$, with $A \subset \mathbb{R}^p$ and where k_α is a correlation function for $\alpha \in \mathbb{R}^p$. We let Y have covariance function $\sigma_0^2 k_{\alpha_0}$ for some fixed $\sigma_0^2 > 0$ and $\alpha_0 \in \mathbb{R}^p$. We consider the observation points $0 \leq x_1 \leq \dots \leq x_n \leq 1$ and the corresponding observed values $y_1 = Y(x_1), \dots, y_n = Y(x_n)$. Classically, the covariance parameters σ_0^2 and α_0 are estimated by maximum likelihood [29, 37]. The MLE is given by

$$(\hat{\sigma}_{ML}^2, \hat{\alpha}_{ML}) \in \underset{\sigma^2 \geq 0, \alpha \in A}{\operatorname{argmin}} (n \log(\sigma^2) + \log(\det(R_\alpha)) + y^\top R_\alpha^{-1} y), \quad (1)$$

where R_α is the $n \times n$ matrix $[K_\alpha(x_i, x_j)]_{1 \leq i, j \leq n}$ and where $y = (y_1, \dots, y_n)^\top$. The computation cost of the likelihood criterion in (1) is $O(n^3)$ and is prohibitive when n becomes larger than, say, 10^4 .

To tackle this problem, several references [27, 28, 39, 43] have studied and used composite likelihood, that we now present. The principle is to sum the conditional log likelihood of each observation, given the K (resp. L) observations corresponding to the left (resp. right) nearest neighbor observation points.

We let $K \in \mathbb{N}$ and $L \in \mathbb{N}$ be fixed. For any $i \in \{K + 1, \dots, n - L\}$, we define the vector $r_{\alpha, K, L; i}$ by

$$r_{\alpha, K, L; i} = (k_\alpha(x_{i-K}, x_i), \dots, k_\alpha(x_{i-1}, x_i), k_\alpha(x_{i+1}, x_i), \dots, k_\alpha(x_{i+L}, x_i))^\top,$$

the vector of local observations $y_{K, L; i}$ by

$$y_{K, L; i} = (y_{i-K}, \dots, y_{i-1}, y_{i+1}, \dots, y_{i+L})^\top$$

and the submatrix $R_{\alpha,K,L;i}$ by

$$R_{\alpha,K,L;i} = \begin{pmatrix} k_{\alpha}(x_{i-K}, x_{i-K}) & \dots & k_{\alpha}(x_{i-K}, x_{i-1}) & k_{\alpha}(x_{i-K}, x_{i+1}) & \dots & k_{\alpha}(x_{i-K}, x_{i+L}) \\ \vdots & & & & & \vdots \\ k_{\alpha}(x_{i-1}, x_{i-K}) & \dots & k_{\alpha}(x_{i-1}, x_{i-1}) & k_{\alpha}(x_{i-1}, x_{i+1}) & \dots & k_{\alpha}(x_{i-1}, x_{i+L}) \\ k_{\alpha}(x_{i+1}, x_{i-K}) & \dots & k_{\alpha}(x_{i+1}, x_{i-1}) & k_{\alpha}(x_{i+1}, x_{i+1}) & \dots & k_{\alpha}(x_{i+1}, x_{i+L}) \\ \vdots & & & & & \vdots \\ k_{\alpha}(x_{i+L}, x_{i-K}) & \dots & k_{\alpha}(x_{i+L}, x_{i-1}) & k_{\alpha}(x_{i+L}, x_{i+1}) & \dots & k_{\alpha}(x_{i+L}, x_{i+L}) \end{pmatrix}.$$

The CLE estimator is then given by

$$(\hat{\sigma}^2, \hat{\alpha}) \in \underset{\sigma^2 \geq 0, \alpha \in A}{\operatorname{argmin}} \sum_{i=K+1}^{n-L} \mathcal{L}_{\sigma^2, \alpha}(y_i | y_{i-K}, \dots, y_{i-1}, y_{i+1}, \dots, y_{i+L}), \quad (2)$$

where $\mathcal{L}_{\sigma^2, \alpha}(y_i | y_{i-K}, \dots, y_{i-1}, y_{i+1}, \dots, y_{i+L})$ is defined as (-2) times the logarithm of the conditional probability density function of y_i given $y_{i-K}, \dots, y_{i-1}, y_{i+1}, \dots, y_{i+L}$ under the covariance parameters σ^2, α . We remark that, by Gaussian conditioning (see e.g. [29, Appendix A]), for $i = K + 1, \dots, n - L$, we have

$$\begin{aligned} & \mathcal{L}_{\sigma^2, \alpha}(y_i | y_{i-K}, \dots, y_{i-1}, y_{i+1}, \dots, y_{i+L}) \\ &= \log \left(\sigma^2 \left(1 - r_{\alpha, K, L; i}^{\top} R_{\alpha, K, L; i}^{-1} r_{\alpha, K, L; i} \right) \right) + \frac{\left(y_i - r_{\alpha, K, L; i}^{\top} R_{\alpha, K, L; i}^{-1} y_{K, L; i} \right)^2}{\sigma^2 \left(1 - r_{\alpha, K, L; i}^{\top} R_{\alpha, K, L; i}^{-1} r_{\alpha, K, L; i} \right)}. \end{aligned}$$

The computational cost of computing the composite likelihood criterion in (2) is $O(n)$ if K and L are fixed, as opposed to $O(n^3)$ for the likelihood criterion in (1).

In the rest of the paper, to lighten notation, we write r_i and R_i for $r_{\alpha, K, L; i}$ and $R_{\alpha, K, L; i}$.

Remark 2.1. *We have defined the CLE for Gaussian processes in dimension one. Indeed, all the asymptotic results in the paper hold for one-dimensional Gaussian processes. It should however be remarked that the CLE is well defined and used for Gaussian processes in dimension larger than one [27, 28, 39, 43]. In fact, the principle of composite likelihood is applicable whenever a relevant distance can be considered on the input space of the Gaussian process. This generality and flexibility is an asset of composite likelihood.*

3 Estimation of a variance parameter

3.1 Main assumptions and expression of the estimator

In the rest of the paper, we consider the regular design of observation points given by $\{x_1 = 1/n, \dots, x_n = 1\}$. Although the proofs in the paper could be extended to other types of designs of observation points, we state and prove our theoretical results in the case of the regular design, for a better readability. Furthermore, we let A be a compact subset of $(0, \infty)$ and we let $k_{\alpha}(t) = k(\alpha t)$, where k is a fixed stationary correlation function. Hence,

in the parametric model considered in this paper, the parameters are the variance σ^2 and the spatial scale α .

In this section, we let A be reduced to the singleton $\{1\}$, we let $\alpha_0 = 1$ (that is the correlation function is known) and we aim at deriving the asymptotic properties of the CLE of the unknown variance σ_0^2 . We then remark that $k = k_{\alpha_0}$.

Recall that the observation vector is $(Y(x_1), \dots, Y(x_n))^T$ which is the vector of observations at times x_1, \dots, x_n and that $y_i = Y(x_i)$ for $i = 1, \dots, n$. Let us introduce the process Z such that $Y = \sigma_0 Z$. The process Z is centered and Gaussian with covariance function k leading to the reduced observations $z_i = y_i/\sigma_0$. Analogously to $y_{K,L;i}$, we define $z_{K,L;i}$. For $i = K + 1, \dots, n - L$, let the prediction be given by

$$\hat{z}_i := \mathbb{E}[z_i | z_{K,L;i}] = r_i^\top R_i^{-1} z_{K,L;i}$$

and the prediction variance be given by

$$\hat{\sigma}_i^2 := \text{Var}(z_i | z_{K,L;i}) = 1 - r_i^\top R_i^{-1} r_i.$$

Then, after simple computations, one may derive the value of $\hat{\sigma}^2$ in (2):

$$\frac{\hat{\sigma}^2}{\sigma_0^2} = \frac{1}{n - L - K} \sum_{i=K+1}^{n-L} \frac{(z_i - \hat{z}_i)^2}{\hat{\sigma}_i^2}. \quad (3)$$

Obviously, this estimator is unbiased and its variance is given by

$$\text{Var} \left(\frac{\hat{\sigma}^2}{\sigma_0^2} \right) = \frac{2}{(n - L - K)^2} \sum_{i,j=K+1}^{n-L} \frac{\text{Cov}(z_i - \hat{z}_i, z_j - \hat{z}_j)^2}{\hat{\sigma}_i^2 \hat{\sigma}_j^2} \quad (4)$$

after application of Mehler's formula.

In the rest of the paper, we assume that the correlation function k satisfies the following condition.

Condition 3.1. *The correlation function k has the following expansion*

$$k(t) = 1 - |t|^s + r(t), \quad (5)$$

where r is twice differentiable, $r(0) = 0$ and $0 < s < 3/2$. Furthermore, for $1/2 \leq s < 3/2$, r'' is bounded.

Condition 3.1 means that the Gaussian process Y is continuous but not differentiable. The quantity s is interpreted as a smoothness parameter and, for instance, $s = 1$ enables to recover the exponential covariance function $\exp\{-|t|\}$, as considered in [5, 6, 44]. Condition 3.1 holds for the Matérn covariance function with parameter ν when $s = 2\nu$, see Section 3.4. Remark that in Condition 3.1, $r(t) = O(t)$ for any $0 < s < 3/2$ and $r(t) = O(t^2)$ for $1/2 \leq s < 3/2$. We also note that one may easily extend the results of this section to any $\alpha_0 > 0$. In fact, we have considered the case $\alpha_0 = 1$ for the ease of the readability of the proofs.

3.2 Main results

Let us define the $(K+L) \times (K+L)$ matrix b as the inverse of the $(K+L) \times (K+L)$ matrix B given by $B_{i,j} = i^s + j^s - |i-j|^s$. It has been proved in [1, Proposition 2.4] that B is invertible. Let us define additionally the $(K+L) \times (K+L)$ matrix C given by $C_{i,j} = i^s j^s$. In the case $0 < s < 1/2$, the additional condition will also be needed.

Condition 3.2. *We assume that $K \in \mathbb{N}$, $L \in \mathbb{N}$ and $0 < s < 1/2$ are such that*

$$\sum_{k=1}^{K+L} b_{K,k} k^s \neq 0.$$

In fact, we can show that Condition 3.2 holds when $K = 0$ or $L = 0$ (see Lemma A.5 in the appendix). When both K and L are non-zero, we are not able to prove that Condition 3.2 holds for all values of $K \in \mathbb{N}$, $L \in \mathbb{N}$ and $0 < s < 1/2$. Anyway, we have seen numerically that Condition 3.2 holds for all the numerous values of K , L and s that we have tried.

We remark that we have the following identity, after some simple computations,

$$\left(\sum_{k=1}^{K+L} b_{K,k} k^s \right)^2 = (bCb)_{K,K}. \quad (6)$$

In the next theorem, we provide the asymptotic order of magnitude of the variance of the CLE.

Theorem 3.3. *We assume that Condition 3.1 holds and we let $K \geq 0$ and $L \geq 0$ be fixed such that $K + L \geq 2$.*

(i) *If $0 < s < 1/2$, then*

$$\text{Var} \left(\frac{\hat{\sigma}^2}{\sigma_0^2} \right) = O \left(\frac{1}{n^{2s}} \right). \quad (7)$$

Furthermore if Condition 3.2 is fulfilled,

$$\text{Var} \left(\frac{\hat{\sigma}^2}{\sigma_0^2} \right) \sim \frac{4}{n^{2s}} \frac{(bCb)_{K,K}^2}{b_{K,K}^2} \int_0^1 (1-t)(1-t^s+r(t))^2 dt. \quad (8)$$

(ii) *If $1/2 \leq s < 3/2$, then*

$$\text{Var} \left(\frac{\hat{\sigma}^2}{\sigma_0^2} \right) = O \left(\frac{1}{n} \right). \quad (9)$$

Obviously, one may now derive the consistency of $\hat{\sigma}^2$ as soon as $0 < s < 3/2$ since its variance goes to zero.

Corollary 3.4 (Consistency). *We assume that Condition 3.1 holds. Then, for $0 < s < 3/2$ and for all $K \geq 0$ and $L \geq 0$ such that $K + L \geq 2$, $\hat{\sigma}^2$ is consistent:*

$$\hat{\sigma}^2 \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \sigma_0^2.$$

Theorem 3.3 shows that for $0 < s < 1/2$, the CLE converges at a sub-optimal rate $n^s < n^{1/2}$, while there exist estimators of σ_0^2 with optimal rate $n^{1/2}$, for instance the MLE in (1). This may be considered as a drawback of the CLE, since other estimators with small computational cost $O(n)$ exist that converge at the optimal rate $n^{1/2}$ for any $0 < s < 3/2$, for instance quadratic variation estimators (see Section 3.3 below). We also remark that the values of the number K and L of neighbors has no impact on the rate of convergence, but has an impact on the asymptotic variance in the case $0 < s < 1/2$.

The next corollary provides the asymptotic variance when $0 < s < 1/2$ in two particular cases.

Corollary 3.5 (Particular cases). *We assume that Condition 3.1 holds and that $0 < s < 1/2$.*

(i) *For $K = 1$ and $L = 1$, one gets*

$$\text{Var} \left(\frac{\hat{\sigma}^2}{\sigma_0^2} \right) \sim \frac{1}{n^{2s}} \frac{(1 - 2^{s-1})^4}{(1 - 2^{s-2})^2} \int_0^1 (1-t)(1-t^s + r(t))^2 dt.$$

(i) *For $K = 2$ and $L = 0$, one gets*

$$\text{Var} \left(\frac{\hat{\sigma}^2}{\sigma_0^2} \right) \sim \frac{1}{n^{2s}} \frac{2^{2s-4}}{(1 - 2^{s-2})^2} \int_0^1 (1-t)(1-t^s + r(t))^2 dt.$$

When $K + L = 1$, the proof of Theorem 3.3 can not be applied. Anyway, one may easily prove that (8) and (9) still hold.

Proposition 3.6. *We consider the case $K = 1$ and $L = 0$ (or by symmetry $K = 0$ and $L = 1$). We assume that Condition 3.1 holds and that $0 < s < 1/2$. Then one gets*

$$\text{Var} \left(\frac{\hat{\sigma}^2}{\sigma_0^2} \right) \sim \frac{1}{n^{2s}} \int_0^1 (1-t)(1-t^s + r(t))^2 dt. \quad (10)$$

When $1/2 \leq s < 3/2$, (9) still holds.

In the next proposition, we show that the CLE $\hat{\sigma}^2$ converges to a non-Gaussian random variable when $0 < s < 1/2$.

Proposition 3.7. *We assume that Conditions 3.1 and 3.2 hold. Then for $0 < s < 1/2$, for $K \geq 0$ and $L \geq 0$ such that $K + L \geq 1$, the random variable $n^s (\hat{\sigma}^2/\sigma_0^2 - 1)$ does not converge in distribution to a Gaussian random variable. In fact, we prove that*

$$n^s \left(\frac{\hat{\sigma}^2}{\sigma_0^2} - 1 \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \frac{(bCb)_{K,K}}{b_{K,K}} \left(\int_0^1 Z(t)^2 dt - 1 \right).$$

This proposition is a direct consequence of Theorem 4.1 below, with $\hat{\alpha} = \alpha_0 = 1$, see also Remark 4.3.

3.3 Quadratic a -variations

The aim of this section is to compare the previous asymptotic results to the ones obtained with the estimator based on quadratic variations in [1]. In that view, we consider a non-zero finite support sequence $a = (a_m)_{m \in \mathbb{N}}$ of real numbers with zero sum. Let $L(a)$ be the length of a : the indices of first and last non-zero elements of a have difference $L(a)$. Since the starting point of the sequence plays no particular role, we will assume when possible that the first non-zero element of a is a_0 . Hence, the last non-zero element is $a_{L(a)-1}$. We define the order $M(a)$ as the first non-zero moment of the sequence a :

$$\sum_{m=0}^{L(a)-1} a_m m^k = 0, \quad \text{for } 0 \leq k < M(a) \quad \text{and} \quad \sum_{m=0}^{L(a)-1} a_m m^{M(a)} \neq 0.$$

To any finite sequence a , with length $L(a)$, we define the quadratic a -variation of Y by

$$V_{a,n} = \sum_{i=1}^{n-L(a)+1} \left(\sum_{j=0}^{L(a)-1} a_j y_{i+j} \right)^2. \quad (11)$$

Guided by the moment method, the authors of [1] define an estimator $C_{a,n}$ of σ_0^2 from $V_{a,n}$. They prove that $C_{a,n}$ is asymptotically unbiased (see Section 3.3 in [1]) and such that, for $0 < s < 3/2$,

$$\mathbb{E}[(C_{a,n} - \sigma_0^2)^2] = O\left(\frac{1}{n}\right). \quad (12)$$

In the simplest case where a has non-zero elements $a_0 = 1$ and $a_1 = -1$ only, (12) means that basing the estimator $C_{a,n}$ on the differences $y_i - y_{i-1}$ yields the optimal rate $n^{1/2}$. In contrast, using the differences $y_i - \mathbb{E}[y_i|y_{i-1}]$ yields the sub-optimal rate n^s when $0 < s < 1/2$ from Proposition 3.6. Hence, while it could be intuitive that using $y_i - \mathbb{E}[y_i|y_{i-1}]$ would be more efficient than using $y_i - y_{i-1}$, this intuition is infirmed by our asymptotic results. Similarly, our results show that using $y_i - \mathbb{E}[y_i|y_{i-K}, \dots, y_{i-1}, y_{i+1}, \dots, y_{i+L}]$ is less efficient than using $\sum_{j=0}^{L(a)-1} a_j y_{i+j}$ when $0 < s < 1/2$, which is not obvious to anticipate.

3.4 Application to the Matérn covariance functions

Let, for $0 < \sigma^2 < \infty$, $0 < \alpha < \infty$ and $0 < \nu < \infty$, $k_{\sigma,\alpha,\nu} : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ be defined by

$$k_{\sigma,\alpha,\nu}(t) = \frac{\sigma^2(\alpha t)^\nu}{2^{\nu-1}\Gamma(\nu)} K_\nu(\alpha t)$$

where Γ is the Gamma function and K_ν is the modified Bessel function of the second kind. The function $(x, y) \mapsto k_{\sigma,\alpha,\nu}(|x - y|)$ is the Matérn covariance function [37, 24]. When ν is not an integer, we have [24]

$$k_{\sigma,\alpha,\nu}(t) = \sigma^2 \sum_{k=0}^{\infty} \frac{\alpha^{2k} t^{2k}}{2^{2k} k! \prod_{i=1}^k (i - \nu)} - \frac{\pi \sigma^2}{\Gamma(\nu) \sin(\nu\pi)} \sum_{k=0}^{\infty} \frac{\alpha^{2k+2\nu} t^{2k+2\nu}}{2^{2k+2\nu} k! \Gamma(k+1+\nu)}. \quad (13)$$

To shorten notation, we simply write k_ν for $k_{1,1,\nu}$. In view of Condition 3.1, we prove the following lemma.

Lemma 3.8. *When $0 < \nu < 1$, we have with a fixed finite non-zero A_ν*

$$k_\nu(t) = 1 - A_\nu t^{2\nu} + g_\nu(t)t^2. \quad (14)$$

Furthermore, g_ν is C^2 -differentiable on \mathbb{R}^+ .

Consequently, Condition 3.1 holds for the Matérn covariance function \tilde{k}_ν defined by $\tilde{k}_\nu(t) = k_\nu(A_\nu^{-1/2\nu}t)$ with $s = 2\nu$ and $r(t) = O(t^2)$.

4 Joint estimation of the variance and spatial scale parameters

We now let A be a general compact subset of $(0, \infty)$, that is we consider the joint estimation of the variance and the spatial scale parameters. As in Section 3, we assume that k satisfies Condition 3.1. In the case where k is a Matérn covariance function, it is well-known from [45] that the parameters σ^2 and α are non-microergodic but that the parameter $\sigma^2\alpha^{2\nu}$ is microergodic. Hence, recalling that $s = 2\nu$ for the correspondence between our assumptions and the Matérn model, we will provide asymptotic results for the estimation of $\sigma_0^2\alpha_0^s$ only. We assume that the true σ_0^2 and α_0 are fixed in $(0, \infty)$, but we do not need to assume that α_0 belongs to A . This is because σ^2 is unrestricted in (2), thus there always exists $(\sigma^2, \alpha) \in (0, \infty) \times A$ such that $\sigma^2\alpha^s = \sigma_0^2\alpha_0^s$.

The next theorem provides the rate of convergence of the CLE $\hat{\sigma}^2\hat{\alpha}^s$ of the microergodic parameter.

Theorem 4.1. *We assume that k satisfies Condition 3.1 and we let $K \geq 0$ and $L \geq 0$ be such that $K + L \geq 2$.*

(i) *If $0 < s < 1/2$, then*

$$\frac{\hat{\sigma}^2\hat{\alpha}^s}{\sigma_0^2\alpha_0^s} - 1 = \frac{\hat{\alpha}^s}{n^s} \frac{(bCb)_{K,K}}{b_{K,K}} \left(\frac{\hat{\alpha}^s}{\alpha_0^s} \int_0^1 Z(t)^2 dt - 1 \right) + o_{\mathbb{P}} \left(\frac{1}{n^s} \right). \quad (15)$$

(ii) *If $1/2 \leq s < 3/2$, then*

$$\frac{\hat{\sigma}^2\hat{\alpha}^s}{\sigma_0^2\alpha_0^s} - 1 = O \left(\frac{1}{\sqrt{n}} \right). \quad (16)$$

The interpretation of Theorem 4.1 is the same as the one of Theorem 3.3 in Section 3 for the estimation of the variance parameter. The CLE converges at the sub-optimal rate n^s for $0 < s < 1/2$. For $1/2 \leq s < 3/2$, the CLE converges at the optimal rate $n^{1/2}$. We remark that in Section 3, we state our results in terms of the asymptotic order of magnitude of the

variance of $\hat{\sigma}^2$. In this section, we can not analyze the variance of $\hat{\sigma}^2\hat{\alpha}^s$, because $\hat{\alpha}^s$ does not have an explicit expression. This is why Theorem 4.1 is stated in terms of convergence in distribution and probability rather than with variances or mean square errors.

We remark that in Theorem 4.1, when $0 < s < 1/2$, the asymptotic approximation of $\hat{\sigma}^2\hat{\alpha}^s/\sigma_0^2\alpha_0^s - 1$ depends on the distribution of $\hat{\alpha}^s$, for which little is known ([46] considers the exponential covariance function for which $s = 1$). Nevertheless, the random variable

$$\hat{\alpha}^s \frac{(bCb)_{K,K}}{b_{K,K}} \left(\frac{\hat{\alpha}^s}{\alpha_0^s} \int_0^1 Z(t)^2 dt - 1 \right) \quad (17)$$

is non-Gaussian, because its minimum is $-A_{\text{sup}}^s (bCb)_{K,K}/b_{K,K} > -\infty$ where A_{sup} is the supremum of the compact set A . Furthermore, this random variable is non-constant because $\hat{\alpha}^{2s} \int_0^1 Z(t)^2 dt$ is non-constant. Indeed, $\int_0^1 Z(t)^2 dt$ has a non-zero variance and has a non-zero probability to belong to any set $[0, \epsilon]$ with arbitrarily small $\epsilon > 0$ (see [23, 25]) and $\hat{\alpha}^{2s}$ is bounded by A_{sup}^{2s} .

With this argument, we see that the random variable (17) can not converge to a Gaussian distribution (including a constant) as $n \rightarrow \infty$. Hence the CLE is asymptotically non-Gaussian for $0 < s < 1/2$. Finally, if $A = \{\alpha_1\}$ with a fixed $\alpha_1 \in (0, \infty)$, then $n^s(\hat{\sigma}^2\hat{\alpha}^s/\sigma_0^2\alpha_0^s - 1)$ converges to a fixed non-Gaussian random variable, which variance is proportional to α_1^{4s} .

As in Section 3, the proof of Theorem 4.1 does not apply when $K+L = 1$, but its conclusion still holds.

Proposition 4.2. *We consider the case $K = 1$ and $L = 0$ (or by symmetry $K = 0$ and $L = 1$). We assume that k satisfies Condition 3.1 and that $0 < s < 1/2$. Then one gets*

$$\frac{\hat{\sigma}^2\hat{\alpha}^s}{\sigma_0^2\alpha_0^s} - 1 \sim \frac{\hat{\alpha}^s}{2n^s} \left(\frac{\hat{\alpha}^s}{\alpha_0^s} \int_0^1 Z(t)^2 dt - 1 \right). \quad (18)$$

When $1/2 \leq s < 3/2$, (16) still holds.

The next remark shows the correspondence between the results of Section 3 and Theorem 3.3.

Remark 4.3. *If $A = \{1\}$ and $\alpha_0 = 1$, then $\hat{\alpha} = 1$ and Theorem 4.1 reduces to*

$$\frac{\hat{\sigma}^2}{\sigma_0^2} - 1 \sim \frac{1}{n^s} \frac{(bCb)_{K,K}}{b_{K,K}} \left(\int_0^1 Z(t)^2 dt - 1 \right), \quad (19)$$

from which we deduce that $(\hat{\sigma}^2/\sigma_0^2 - 1)$ is asymptotically unbiased. Moreover, computing the variance of $\int_0^1 Z(t)^2 dt$ leads to the same expression as in (10) in Proposition 3.6.

5 Numerical experiments

Now we compare the asymptotic variance of the CLE with its exact finite sample variance. We consider the setting of Section 3, where a single variance parameter is estimated and the correlation function k is known and satisfies Condition 3.1. Since the CLE $\hat{\sigma}^2$ has an explicit expression, its variance can be written explicitly as

$$\frac{2}{(n - L - K)^2} \frac{1}{\hat{\sigma}_{K+1}^4} \sum_{i,j=K+1}^{n-L} (k(x_i, x_j) - r_i^\top R_i^{-1} r_{i,j} - r_j^\top R_j^{-1} r_{j,i} + r_i^\top R_i^{-1} R_{i,j} R_j^{-1} r_j)^2, \quad (20)$$

with the notation of Section 2, with $r_{a,\ell}$ being the column covariance vector between $y_{K,L;a}$ and y_ℓ under covariance function k and $R_{a,\ell}$ being the covariance matrix between $y_{K,L;a}$ and $y_{K,L;\ell}$ under covariance function k . Equation (20) directly follows from (4).

We consider the case where $0 < s < 1/2$ in Condition 3.1, so that the asymptotic variance of $\hat{\sigma}^2$ is given explicitly in Theorem 3.3. As correlation functions, we consider the generalized Slepian function [33] given by $k(t) = (1 - |t|^s)^+$ and the Matérn covariance function given by $k(t) = k_{1,1,\nu}(A_\nu^{-1/2\nu} t)$ with $s = 2\nu$ and with the notation of Section 3.4.

In Figure 5, for $s = 0.15$ and $s = 0.30$, for these two correlation functions and for various values of K and L , we plot the ratios of the exact finite sample variance in (20) over the asymptotic variance in Theorem 3.3. We observe that these ratios do converge to one as n increases, which confirms Theorem 3.3. We also observe that n may need to be very large for the ratios to be close to one. Hence, the asymptotic approximation given by Theorem 3.3 may become accurate only for very large n . For moderate values of n , the ratios are larger than one, so that the asymptotic variance underestimates the finite sample variance. We also observe that the ratios are larger when K and L are larger. In other words, when the CLE is based on more neighbors, n needs to be larger for the asymptotic variance to be close to the finite sample one. Furthermore, $s = 0.3$ leads to larger ratios than $s = 0.15$ and we have observed in other (unreported) experiments that for fixed n , the ratios increase with s . Hence, for fixed n , the asymptotic variance provides a more accurate approximation of the exact variance when s is small. Finally, the ratios are similar between the generalized Slepian and Matérn covariance functions.

6 Concluding remarks

We have provided a fixed-domain asymptotic analysis of the CLE, for one-dimensional Gaussian processes that are non-differentiable and are characterized by a general smoothness parameter $0 < s < 3/2$. Our analysis improves the previous references [5, 6, 44] by allowing for general covariance functions and general numbers of neighbors for the composite likelihood. A conclusion that we obtain, which was not obvious to anticipate, is that the CLE converges at a sub-optimal rate, for $0 < s < 1/2$, independently of the number of neighbors used for the composite likelihood.

There are some possible extensions of our results that, we believe, could be obtained by following the same proof structures. These extensions include obtaining the asymptotic

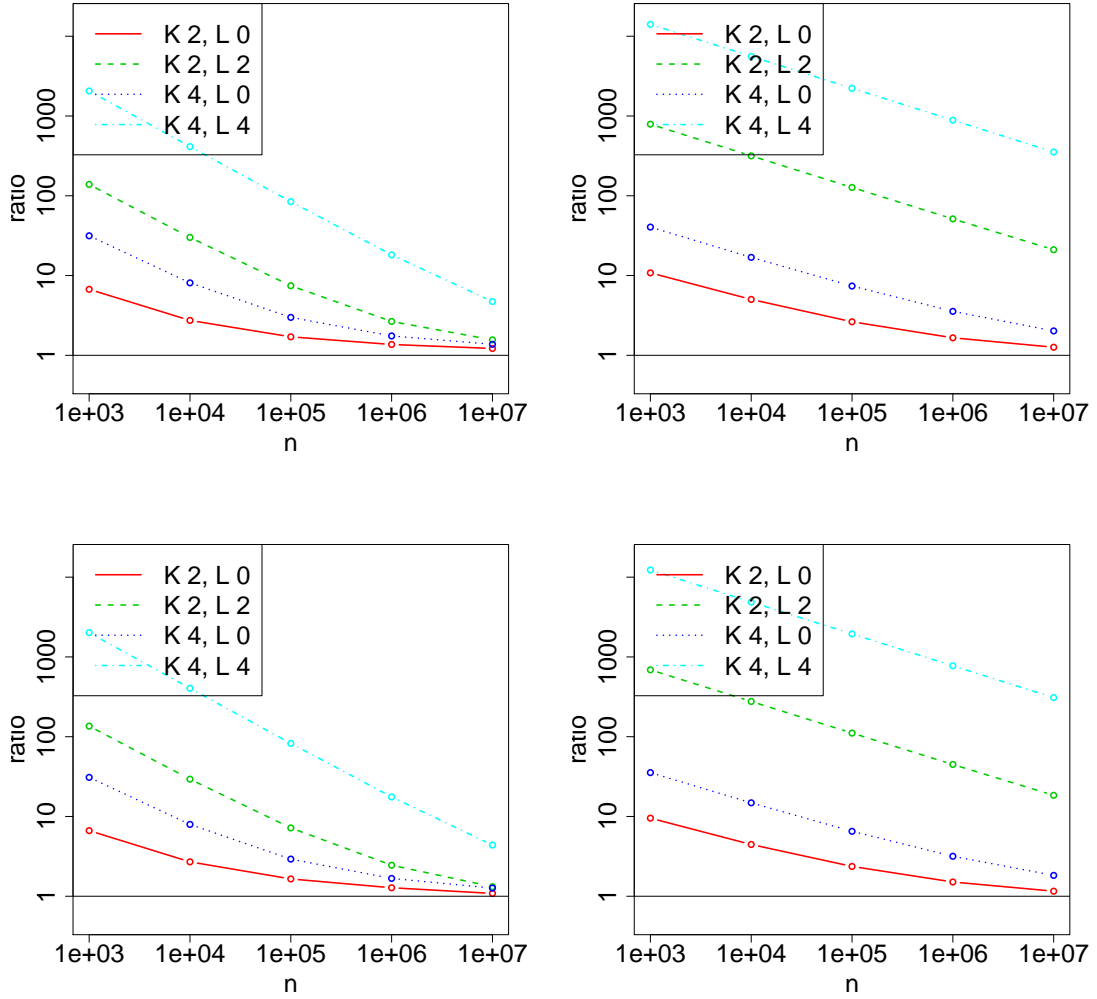


Figure 1: Plot of the ratios of the finite sample variance in (20) over the asymptotic variance in Theorem 3.3. Top: generalized Slepian correlation function. Bottom: Matérn covariance function. Left: $s = 0.15$. Right: $s = 0.30$.

distribution of the CLE when $1/2 \leq s < 3/2$, treating the case $3/2 \leq s \leq 2$, and allowing for irregularly spaced observation points.

Other open problems remain, that would potentially require more work and new approaches. For instance, it would be interesting to obtain asymptotic results for differentiable processes and in the multi-dimensional case.

A Proofs

In the paper, $(Const)$ stands for a generic constant that may differ from one line to another.

A.1 Discrete a -differences

Some of the proofs of this paper rely on the notion of discrete a -differences already introduced in Section 3.3 and used in the literature of quadratic variation estimators [1, 20]. Some technical results are collected in the following lemma.

Lemma A.1. *Let $(a_m)_m$ be a sequence of order M .*

(i) *As $n \rightarrow \pm\infty$, one has*

$$\sum_{m=-\infty}^{+\infty} a_m |n+m|^s = O\left(|n|^{s-M}\right). \quad (21)$$

(ii) *Let f be a M -times continuously differentiable function. Then, as $n \rightarrow +\infty$,*

$$\sup_{t \in [0,1]} \left| \sum_{m=-\infty}^{+\infty} a_m f\left(t + \frac{m}{n}\right) \right| = O\left(\frac{1}{n^M}\right). \quad (22)$$

Proof. (i) By a Taylor expansion with integral reminder of $|n+m|^s$ near n at order $(M-1)$, one gets, for n large enough,

$$\begin{aligned} |n+m|^s &= |n|^s + ms|n|^{s-1} + \dots + \frac{m^{M-1}}{(M-1)!} s(s-1)\dots(s-M+2)|n|^{s-M+1} \\ &\quad + \frac{m^M}{(M-1)!} s(s-1)\dots(s-M+1) \int_0^1 (1-\eta)^{M-1} |n+m\eta|^{s-M} d\eta. \end{aligned} \quad (23)$$

Hence, using the vanishing moments of the sequence $a = (a_m)_m$, we have

$$\begin{aligned} &\left| \sum_{m=-\infty}^{+\infty} a_m |n+m|^s \right| \\ &= \frac{|s(s-1)\dots(s-M+1)|}{(M-1)!} \sum_{m=-\infty}^{+\infty} |a_m| |m|^M \int_0^1 (1-\eta)^{M-1} |n+m\eta|^{s-M} d\eta \\ &\leq \frac{|s(s-1)\dots(s-M+1)|}{(M-1)!} |n|^{s-M} \sum_{m=-\infty}^{+\infty} |a_m| |m|^M \int_0^1 (1-\eta)^{M-1} \left|1 + \frac{m}{n}\eta\right|^{s-M} d\eta. \end{aligned}$$

For a fixed value of m , the term $|1 + m\eta/n|^{s-M}$ is bounded by 2 for n large enough. Then the proof is complete.

(ii) We follow the same lines as in (i). By a Taylor expansion with integral reminder of $f(t + m/n)$ near t at order $M - 1$, one gets

$$f\left(t + \frac{m}{n}\right) = \sum_{k=0}^{M-1} \frac{1}{k!} \left(\frac{m}{n}\right)^k f^{(k)}(t) + \int_t^{t+m/n} \frac{f^{(M)}(\eta)}{(M-1)!} \left(t + \frac{m}{n} - \eta\right)^{M-1} d\eta.$$

Hence, using the vanishing moments of the sequence $a = (a_m)_m$, one gets

$$\begin{aligned} & \left| \sum_{m=-\infty}^{+\infty} a_m f\left(t + \frac{m}{n}\right) \right| \\ & \leq \left| \sum_{k=0}^{M-1} \frac{1}{k!} \frac{f^{(k)}(t)}{n^k} \sum_{m=-\infty}^{+\infty} a_m m^k \right| + \sum_{m=-\infty}^{+\infty} |a_m| \int_t^{t+m/n} \frac{|f^{(M)}(\eta)|}{(M-1)!} \left|t + \frac{m}{n} - \eta\right|^{M-1} d\eta \\ & \leq \frac{(Const)}{(M-1)!} \sum_{m=-\infty}^{+\infty} |a_m| \left|\frac{m}{n}\right|^M \\ & \leq \frac{(Const)}{n^M}. \end{aligned}$$

Then the proof is complete. □

A.2 Proof of Theorem 3.3(i)

We recall the expression of the variance of the estimator

$$\text{Var}\left(\frac{\hat{\sigma}^2}{\sigma_0^2}\right) = \frac{2}{(n-L-K)^2} \sum_{i,j=K+1}^{n-L} \frac{\text{Cov}(z_i - \hat{z}_i, z_j - \hat{z}_j)^2}{\hat{\sigma}_i^2 \hat{\sigma}_j^2}. \quad (24)$$

By the well-known virtual Leave-One-Out (LOO) formulas (see [2, Proposition 3.1]), the prediction variance is given by

$$\begin{aligned} \hat{\sigma}_i^2 &= \text{Var}(z_i | z_{K,L;i}) \\ &= \text{Var}(z_{K+1} | z_{K,L;K+1}) \\ &= \text{Var}(z_{K+1} | z_1, \dots, z_K, z_{K+2}, \dots, z_{K+L+1}) \\ &= \text{Var}_{|1}(z_{K+1} | z_2, \dots, z_K, z_{K+2}, \dots, z_{K+L+1}) \\ &= \frac{1}{(\text{Cov}_{|1}(z_2, \dots, z_K, z_{K+1}, z_{K+2}, \dots, z_{K+L+1})^{-1})_{K,K}} \end{aligned}$$

where we recall that $z_{K,L;i} = (z_{i-K}, \dots, z_{i-1}, z_{i+1}, \dots, z_{i+L})^\top$ and where $\text{Var}_{|1}$ and $\text{Cov}_{|1}$ stand for the variance and the covariance matrices conditionally to z_1 . To shorten notation,

we denote $r(|i - j|/n)$ by $r_n(i, j)$. Now, for $2 \leq i, j \leq K + L + 1$, one has, by Gaussian conditioning,

$$\begin{aligned} \text{Cov}_{|1}(z_i, z_j) &= \text{Cov}(z_i, z_j) - \text{Cov}(z_i, z_1) \text{Var}(z_1)^{-1} \text{Cov}(z_1, z_j) \\ &= \left(1 - \frac{|i - j|^s}{n^s} + r_n(i, j)\right) - \left(1 - \frac{|i - 1|^s}{n^s} + r_n(i - 1, 0)\right) \left(1 - \frac{|j - 1|^s}{n^s} + r_n(0, j - 1)\right) \\ &=: \frac{1}{n^s} B_{i-1, j-1} - \frac{1}{n^{2s}} C_{i-1, j-1} + R_{i-1, j-1}^{(n)} \end{aligned}$$

using $\text{Var}(z_1) = 1$, where B, C and $R^{(n)} \in \mathcal{M}_{K+L, K+L}(\mathbb{R})$ are given by

$$\begin{aligned} B_{i,j} &= i^s + j^s - |i - j|^s \\ C_{i,j} &= i^s j^s \\ R_{i,j}^{(n)} &= r_n(i, j) - r_n(i, 0) \left(1 - \frac{j^s}{n^s}\right) - r_n(0, j) \left(1 - \frac{i^s}{n^s}\right) - r_n(i, 0) r_n(0, j). \end{aligned}$$

In addition, we introduce the following notation $b^{(n)} := (B - n^{-s}C + n^s R^{(n)})^{-1}$ and we recall that $b = B^{-1}$. Finally, the prediction variance writes as

$$\hat{\sigma}_i^2 = \frac{1}{n^s} \frac{1}{b_{K,K}^{(n)}}, \quad \text{for } i = K + 1, \dots, n - L.$$

Since the matrix B is invertible, by continuity and the assumption on r , we obtain

$$b^{(n)} = b + o(1) \tag{25}$$

leading to

$$\hat{\sigma}_i^2 \sim \frac{1}{n^s} \frac{1}{b_{K,K}}. \tag{26}$$

Let us turn to the computation of the covariances in (24). Using again the well-known

virtual LOO formulas (see [2, Proposition 3.1]), the prediction error is given by

$$\begin{aligned}
z_i - \hat{z}_i &= z_i - \mathbb{E}[z_i | z_{i-K}, \dots, z_{i-1}, z_{i+1}, \dots, z_{i+L}] \\
&= (z_i - \mathbb{E}_{|i-K}[z_i]) - \mathbb{E}_{|i-K} \left[(z_i - \mathbb{E}_{|i-K}[z_i]) | z_{i-K+1}, \dots, z_{i-1}, z_{i+1}, \dots, z_{i+L} \right] \\
&= \frac{1}{(\text{Cov}_{|1}(z_2, \dots, z_K, z_{K+1}, z_{K+2}, \dots, z_{K+L+1})^{-1})_{K,K}} \\
&\quad \times \left(\text{Cov}_{|1}(z_2, \dots, z_K, z_{K+1}, z_{K+2}, \dots, z_{K+L+1})^{-1} \begin{pmatrix} z_{i-K+1} - \mathbb{E}[z_{i-K+1} | z_{i-K}] \\ \vdots \\ z_{i-1} - \mathbb{E}[z_{i-1} | z_{i-K}] \\ z_i - \mathbb{E}[z_i | z_{i-K}] \\ z_{i+1} - \mathbb{E}[z_{i+1} | z_{i-K}] \\ \vdots \\ z_{i+L} - \mathbb{E}[z_{i+L} | z_{i-K}] \end{pmatrix} \right)_K \\
&= \frac{1}{b_{K,K}^{(n)}} \sum_{k=1}^{K+L} b_{K,k}^{(n)} (z_{i-K+k} - \mathbb{E}[z_{i-K+k} | z_{i-K}]) \\
&= \frac{1}{b_{K,K}^{(n)}} \sum_{k=1}^{K+L} b_{K,k}^{(n)} \left(z_{i-K+k} - \left(1 - \frac{k^s}{n^s} + r_n(k, 0) \right) z_{i-K} \right), \tag{27}
\end{aligned}$$

where $\mathbb{E}_{|l}$ stands for the expectation conditionally to z_l . Now,

$$\sum_{i,j=K+1}^{n-L} \text{Cov}(z_i - \hat{z}_i, z_j - \hat{z}_j)^2 = \frac{1}{(b_{K,K}^{(n)})^4} \sum_{i,j=K+1}^{n-L} \left(\sum_{k,l=1}^{K+L} b_{K,k}^{(n)} b_{K,l}^{(n)} \right) \tag{28}$$

$$\begin{aligned}
&\times \text{Cov} \left(z_{i-K+k} - \left(1 - \frac{k^s}{n^s} + r_n(k, 0) \right) z_{i-K}, z_{j-K+l} - \left(1 - \frac{l^s}{n^s} + r_n(0, l) \right) z_{j-K} \right)^2 \\
&= \frac{1}{(b_{K,K}^{(n)})^4} \sum_{i,j=K+1}^{n-L} (\alpha_{i,j} + \beta_{i,j})^2, \tag{29}
\end{aligned}$$

where

$$\alpha_{i,j} = \frac{1}{n^{2s}} \left(\sum_{k=1}^{K+L} b_{K,k}^{(n)} k^s \right)^2 \left(1 - \frac{|i-j|^s}{n^s} + r_n(i, j) \right)$$

and $\beta_{i,j}$ is self explanatory. Now, we establish two lemmas in order to complete the proof of Theorem 3.3(i).

Lemma A.2. *Assume that*

$$\sum_{i,j=K+1}^{n-L} \beta_{i,j}^2 = o \left(\sum_{i,j=K+1}^{n-L} \alpha_{i,j}^2 \right), \tag{30}$$

then

$$\sum_{i,j=K+1}^{n-L} (\alpha_{i,j} + \beta_{i,j})^2 \sim \sum_{i,j=K+1}^{n-L} \alpha_{i,j}^2.$$

Proof. The result directly comes from the use of Cauchy-Schwarz inequality. \square

Lemma A.3. *We assume that $0 < s < 1/2$.*

(i) *One has $\sum_{i,j=K+1}^{n-L} \alpha_{i,j}^2 = o(n^{2-4s})$. Moreover, if Condition 3.2 holds, then*

$$\sum_{i,j=K+1}^{n-L} \alpha_{i,j}^2 \sim 2n^{2-4s} (bCb)_{K,K}^2 \int_0^1 (1-t)(1-t^s+r(t))^2 dt.$$

(ii) *The array $(\beta_{i,j})_{i,j=K+1,\dots,n}$ is such that $\sum_{i,j=K+1}^{n-L} \beta_{i,j}^2 = o(n^{2-4s})$.*

Applying Lemmas A.2 and A.3 together with (24), (26) and (29) completes the proof of Theorem 3.3(i). \square

Proof of Lemma A.3. (i) Recalling identity (6) and the definition of $\alpha_{i,j}$, one has

$$\begin{aligned} \sum_{i,j=K+1}^{n-L} \alpha_{i,j}^2 &= \frac{1}{n^{4s}} \left(\sum_{k=1}^{K+L} b_{K,k}^{(n)} k^s \right)^4 \sum_{i,j=K+1}^{n-L} \left(1 - \frac{|i-j|^s}{n^s} + r_n(i,j) \right)^2 \\ &= \frac{2}{n^{4s}} (bCb)_{K,K}^2 (1+o(1)) \sum_{m=0}^{n-L-(K+1)} (n-L-K-1-m) \left(1 - \frac{m^s}{n^s} + r\left(\frac{m}{n}\right) \right)^2 \\ &= 2n^{2-4s} (bCb)_{K,K}^2 (1+o(1)) \left(\int_0^1 (1-t)(1-t^s+r(t))^2 dt \right) \end{aligned} \quad (31)$$

by the convergence theorem of Riemann sums.

(ii) For fixed k and l in $\{1, \dots, K+L\}$, one can show that

$$\begin{aligned} b_{K,k}^{(n)} b_{K,l}^{(n)} \text{Cov} \left(z_{i-K+k} - \left(1 - \frac{k^s}{n^s} + r_n(k,0) \right) z_{i-K}, z_{j-K+l} - \left(1 - \frac{l^s}{n^s} + r_n(0,l) \right) z_{j-K} \right) \\ = \alpha_{i,j}^{k,l} + \beta_{i,j}^{k,l}, \end{aligned}$$

with

$$\alpha_{i,j}^{k,l} = b_{K,k}^{(n)} b_{K,l}^{(n)} \frac{k^s l^s}{n^{2s}} \left(1 - \frac{|i-j|^s}{n^s} + r_n(i,j) \right)$$

and $\beta_{i,j}^{k,l}$ consists in the remaining terms:

$$\beta_{i,j}^{k,l} = b_{K,k}^{(n)} b_{K,l}^{(n)} \left(\gamma_{i,j}^{k,l} + \delta_{i,j}^{k,l} + \epsilon_{i,j}^{k,l} + \phi_{i,j}^{k,l} + \psi_{i,j}^{k,l} + \mu_{i,j}^{k,l} + \nu_{i,j}^{k,l} \right) \quad (32)$$

where

$$\begin{aligned}
\gamma_{i,j}^{k,l} &= \frac{1}{n^s} (|i-j-l|^s + |i+k-j|^s - |i+k-j-l|^s - |i-j|^s), \\
\delta_{i,j}^{k,l} &= \frac{k^s}{n^{2s}} (|i-j|^s - |i-j-l|^s) + \frac{l^s}{n^{2s}} (|i-j|^s - |i+k-j|^s), \\
\epsilon_{i,j}^{k,l} &= r_n(i+k, j+l) - r_n(i+k, j) - r_n(i, j+l) + r_n(i, j), \\
\phi_{i,j}^{k,l} &= -r_n(0, l) \left(1 - \frac{|i+k-j|^s}{n^s} - \left(1 - \frac{k^s}{n^s} \right) \left(1 - \frac{|i-j|^s}{n^s} \right) \right) \\
&\quad - r_n(k, 0) \left(1 - \frac{|i-j-l|^s}{n^s} - \left(1 - \frac{l^s}{n^s} \right) \left(1 - \frac{|i-j|^s}{n^s} \right) \right), \\
\psi_{i,j}^{k,l} &= \frac{k^s}{n^s} (r_n(i, j+l) - r_n(i, j)) + \frac{l^s}{n^s} (r_n(i+k, j) - r_n(i, j)), \\
\mu_{i,j}^{k,l} &= r_n(0, l) \left(r_n(i, j) \left(1 - \frac{k^s}{n^s} \right) - r_n(i+k, j) \right) + r_n(k, 0) \left(r_n(i, j) \left(1 - \frac{l^s}{n^s} \right) - r_n(i, j+l) \right), \\
\nu_{i,j}^{k,l} &= r_n(k, 0) r_n(0, l) r_n(i, j) + r_n(k, 0) r_n(0, l) \left(1 - \frac{|i-j|^s}{n^s} \right).
\end{aligned}$$

Since $(a+b)^2 \leq 2a^2 + 2b^2$, using (25) and since $b_{K,k}^{(n)} b_{K,l}^{(n)}$ does not depend of i and j , it suffices to prove that, for $t_{i,j}^{k,l}$ being any term in the sum (32),

$$\sum_{i,j=K+1}^{n-L} \left(t_{i,j}^{k,l} \right)^2 = o(n^{2-4s}) \tag{33}$$

for any fixed k and l .

• Term $\gamma_{i,j}^{k,l}$. We consider the sequence

$$a_0 = -1, a_k = 1, a_{-l} = 1, a_{k-l} = -1 \text{ and } a_m = 0 \text{ for } m \notin \{0, k, -l, k-l\}$$

if $k \neq l$ and we consider the sequence

$$a_0 = -2, a_k = 1, a_{-k} = 1 \text{ and } a_m = 0 \text{ for } m \notin \{0, k, -k\}$$

if $k = l$. Those sequences are of order 2. Hence, by Lemma A.1(i) with $r = i - j$, for $|i - j| \geq 2$,

$$\left| \gamma_{i,j}^{k,l} \right| \leq \frac{(Const)}{n^s} |i - j|^{s-2}.$$

Now,

$$\begin{aligned}
\sum_{i,j=K+1}^{n-L} \left(\gamma_{i,j}^{k,l} \right)^2 &\leq (\text{Const}) \left(n^{1-2s} + \frac{1}{n^{2s}} \sum_{\substack{i,j=(K+1),\dots,(n-L) \\ |i-j|\geq 2}} |i-j|^{2s-4} \right) \\
&\leq (\text{Const}) \left(n^{1-2s} + \frac{2}{n^{2s}} \sum_{m=2}^{n-L-(K+1)} (n-L-K-1-m)m^{2s-4} \right) \quad (34) \\
&\leq (\text{Const}) \left(n^{1-2s} + \frac{2}{n^{2s-1}} \sum_{m=2}^{+\infty} m^{2s-4} \right) \\
&\leq (\text{Const}) n^{1-2s},
\end{aligned}$$

which is $o(n^{2-4s})$ since $0 < s < 1/2$.

- Term $\delta_{i,j}^{k,l}$. Similarly to $\gamma_{i,j}^{k,l}$, we show that

$$\sum_{i,j=K+1}^{n-L} \left(\delta_{i,j}^{k,l} \right)^2 = O(n^{1-4s}) = o(n^{2-4s}).$$

- Term $\epsilon_{i,j}^{k,l}$. We want to use Lemma A.1(ii). If $k = l$,

$$\begin{aligned}
\epsilon_{i,j}^{k,k} &= 2r_n(i, j) - r_n(i+k, j) - r_n(i, j+k) \\
&= 2r \left(\frac{i-j}{n} \right) - r \left(\frac{i-j}{n} + \frac{k}{n} \right) - r \left(\frac{i-j}{n} - \frac{k}{n} \right).
\end{aligned}$$

Thus we can apply Lemma A.1(ii), with $t = (i-j)/n$ and the variation defined by $a_k = a_{-k} = -1$ and $a_0 = 2$ of order 2. This yields $\epsilon_{i,j}^{k,k} = O(1/n^2)$. Otherwise if $k \neq l$,

$$\begin{aligned}
\epsilon_{i,j}^{k,k} &= r_n(i+k, j+l) - r_n(i+k, j) - r_n(i, j+l) + r_n(i, j) \\
&= r \left(\frac{i-j}{n} + \frac{k-l}{n} \right) - r \left(\frac{i-j}{n} + \frac{k}{n} \right) - r \left(\frac{i-j}{n} - \frac{l}{n} \right) + r \left(\frac{i-j}{n} \right).
\end{aligned}$$

Thus we can apply Lemma A.1(ii), with $t = (i-j)/n$ and the variation defined by $a_{k-l} = 1$, $a_{-l} = a_k = -1$ and $a_0 = 1$ of order 2. This yields $\epsilon_{i,j}^{k,l} = O(1/n^2)$. Hence

$$\sum_{i,j=K+1}^{n-L} \left(\epsilon_{i,j}^{k,l} \right)^2 = O(n^{-2}) = o(n^{2-4s}).$$

- Term $\phi_{i,j}^{k,l}$. One has

$$\begin{aligned}
& |r_n(0, l)| \left| 1 - \frac{|i+k-j|^s}{n^s} - \left(1 - \frac{k^s}{n^s}\right) \left(1 - \frac{|i-j|^s}{n^s}\right) \right| \\
&= |r_n(0, l)| \left| \frac{|i-j|^s}{n^s} - \frac{|i+k-j|^s}{n^s} + \frac{k^s}{n^s} \left(1 - \frac{|i-j|^s}{n^s}\right) \right| \\
&\leq o(n^{-1}) \left((Const) \frac{|i-j|^{s-1}}{n^s} + \frac{1}{n^s} \right) \\
&\leq o(n^{-1-s})
\end{aligned}$$

using Lemma A.1(i) applied to the sequence of order 1: $a_0 = 1$ and $a_k = -1$. Analogously, the second term in $\phi_{i,j}^{k,l}$ is also $o(n^{-1-s})$. Since the $o(n^{-1-s})$ above do not depend on i, j , one gets that $\left(\phi_{i,j}^{k,l}\right)_{i,j}$ satisfies Condition (33) for any fixed k and l .

- Term $\psi_{i,j}^{k,l}$. Using Lemma A.1(ii), we derive that

$$\psi_{i,j}^{k,l} = O\left(\frac{1}{n^{s+1}}\right),$$

implying that $\left(\psi_{i,j}^{k,l}\right)_{i,j}$ satisfies Condition (33) for any fixed k and l .

- Term $\mu_{i,j}^{k,l}$. Using Lemma A.1(ii), one has $\mu_{i,j}^{k,l} = o(n^{-1}) O(n^{-1} + n^{-s}) = o(n^{-1-s})$. Thus, $\left(\mu_{i,j}^{k,l}\right)_{i,j}$ satisfies Condition (33) for any fixed k and l .
- Term $\nu_{i,j}^{k,l}$. Straightforwardly, one gets

$$\sum_{i,j=K+1}^{n-L} \left(\nu_{i,j}^{k,l}\right)^2 = O(n^{2-2}) = o(n^{2-4s}).$$

The proof is now complete. □

A.3 Proof of Theorem 3.3(ii)

The proof follows the same lines as the proof of Theorem 3.3(i) except that Lemma A.3 is updated to Lemma A.4.

Lemma A.4. *We assume that $1/2 \leq s < 3/2$.*

(i) *One has $\sum_{i,j=K+1}^{n-L} \alpha_{i,j}^2 = O(n^{1-2s})$.*

(ii) *The array $(\beta_{i,j})_{i,j=K+1,\dots,n-L}$ is also such that $\sum_{i,j=K+1}^{n-L} \beta_{i,j}^2 = O(n^{1-2s})$.*

Proof. The proof of (i) is straightforward from (31) in Lemma A.3. As in the proof of Lemma A.3, the proof of (ii) only requires to prove that, for $t_{i,j}^{k,l}$ being any term in the sum (32),

$$\sum_{i,j=K+1}^{n-L} \left(t_{i,j}^{k,l}\right)^2 = O\left(n^{1-2s}\right) \quad (35)$$

for any fixed k and l .

- Term $\gamma_{i,j}^{k,l}$. By (34), the sequence $(\gamma_{i,j}^{k,l})_{i,j}$ satisfies (35) since the sum at the line after (34) converges as soon as $s < 3/2$ for any fixed k and l .
- Term $\delta_{i,j}^{k,l}$. Analogously, one has

$$\begin{aligned} \sum_{i,j=K+1}^{n-L} \left(\delta_{i,j}^{k,l}\right)^2 &\leq \frac{(\text{Const})}{n^{4s}} \sum_{i,j=(K+1),\dots,(n-L)} |i-j|^{2s-2} \\ &\leq (\text{Const})n^{1-4s} \sum_{m=0}^{n-L-(K+1)} m^{2s-2} \\ &\leq (\text{Const})n^{1-4s} \left((\text{Const}) + \int_2^n t^{2s-2} dt \right) \\ &\leq (\text{Const})n^{1-4s} \left((\text{Const}) + n^{2s-1} \right) \\ &\leq (\text{Const})n^{-2s} = O\left(n^{1-2s}\right). \end{aligned}$$

- Term $\epsilon_{i,j}^{k,l}$. Using the proof of Lemma A.3, we still get that $\epsilon_{i,j}^{k,l} = O(1/n^2)$ that leads to the result as soon as $s < 3/2$.
- Term $\phi_{i,j}^{k,l}$. Since $r(t) = O(t^2)$ for $1/2 \leq s < 3/2$,

$$\sum_{i,j=K+1}^{n-L} \left(\phi_{i,j}^{k,l}\right)^2 = O\left(n^2 \cdot n^{-4}\right) = O\left(n^{1-2s}\right).$$

- Term $\psi_{i,j}^{k,l}$. Similarly, from Lemma A.3(ii), we derive that

$$\sum_{i,j=K+1}^{n-L} \left(\psi_{i,j}^{k,l}\right)^2 = O\left(\frac{n^2}{n^{2s+2}}\right) = O\left(n^{1-2s}\right).$$

- Term $\mu_{i,j}^{k,l}$. Analogously, the sequence $(\mu_{i,j}^{k,l})_{i,j}$ satisfies (35) as soon as $s < 3/2$ for any fixed k and l .
- Term $\nu_{i,j}^{k,l}$. Using once again $r(t) = O(t^2)$ for $1/2 \leq s < 3/2$, one gets that the sequence $(\nu_{i,j}^{k,l})_{i,j}$ satisfies (35).

The proof is now complete. □

The proof of Theorem 3.3(ii) then follows straightforwardly. □

A.4 Proof of Proposition 3.6

We only prove the result in the case $K = 1$ and $L = 0$. When $K = 0$ and $L = 1$, we get the result by symmetry. Here, r_i and R_i reduce to $r_i = k(x_{i-1}, x_i) = 1 - \frac{1}{n^s} + r_n(1, 0)$ and $R_i = k(x_{i-1}, x_{i-1}) = 1$. Hence, setting $i - j = a$, one gets

$$\text{Cov} \left(z_i - r_i^\top R_i^{-1} z_{1,0;i}, z_j - r_j^\top R_j^{-1} z_{1,0;j} \right) = \frac{g(a)}{n^s} + \frac{1 - g(a)}{n^{2s}} - \frac{|a|^s}{n^{3s}} + \frac{1}{n^{2s}} r_n(a, 0) + t_n(a)$$

where g is the quadratic variation of order 2 given by $g(a) = |a + 1|^s - 2|a|^s + |a - 1|^s$ and t_n is a remaining term involving the rest function r . Moreover, the variance term $\hat{\sigma}_i^2$ reduces to

$$\hat{\sigma}_i^2 = \frac{1}{n^s} \left(2 - \frac{1}{n^s} \right) - r_n(1, 0) \left(r_n(1, 0) + 2 \left(1 - \frac{1}{n^s} \right) \right) \sim \frac{2}{n^s}.$$

Finally,

$$\begin{aligned} \text{Var} \left(\frac{\hat{\sigma}^2}{\sigma_0^2} \right) &= \frac{2}{(n-1)^2} \sum_{i,j=2}^n \frac{\text{Cov} \left(z_i - r_i^\top R_i^{-1} z_{1,0;i}, z_j - r_j^\top R_j^{-1} z_{1,0;j} \right)^2}{(1 - r_i^\top R_i^{-1} r_i)(1 - r_j^\top R_j^{-1} r_j)} \\ &= \frac{4n}{(n-1)^2} \frac{1}{4 + o(1)} \sum_{a=0}^{n-2} \left(1 - \frac{a}{n} \right) \left(g(a) + \frac{1 - g(a)}{n^s} - \frac{a^s}{n^{2s}} + \frac{1}{n^s} r_n(a, 0) + t_n(a) \right)^2. \end{aligned}$$

The terms in $g(a)$, namely, the terms proportional to $g(a)^2$, $g(a)$ and $g(a)a^s$ are $O(1/n)$. Indeed, if we consider the term in $g(a)^2$ for example, one has, by Lemma A.1(i),

$$g(a)^2 = O(a^{2s-4}).$$

Now since

$$\frac{1}{n} \sum_{a=0}^{n-2} a^{2s-4} \leq \frac{1}{n} \sum_{a=0}^{+\infty} a^{2s-4} = O\left(\frac{1}{n}\right)$$

by Lemma A.1(i), one gets that

$$\frac{1}{n} \sum_{a=0}^{n-2} g(a)^2 = O\left(\frac{1}{n}\right)$$

as soon as $s < 3/2$. The same reasoning and conclusion hold for

$$\frac{1}{n^{s+1}} \sum_{a=0}^{n-2} g(a) \quad \text{and} \quad \frac{1}{n^{2s+1}} \sum_{a=0}^{n-2} g(a)a^s.$$

Let us turn now to the preponderant term: one has

$$\frac{1}{n} \sum_{a=0}^{n-2} \left(1 - \frac{a}{n}\right) \frac{1}{n^{2s}} \left(1 - \frac{a^s}{n^s} + r_n(a, 0)\right)^2 \sim \frac{1}{n^{2s}} \int_0^1 (1-t)(1-t^s + r(t))^2 dt$$

by the convergence theorem of Riemann sums.

Finally, the terms involving $t_n(a)$ can be treated as in the proof of Theorem 3.3(i) and using extensively Lemma A.1(ii). Thus, they are negligible with respect to the preponderant term. The proof of (10) is now complete. The proof for the case $1/2 \leq s < 3/2$ is carried out similarly. \square

A.5 On Condition 3.2

Lemma A.5. *Condition 3.2 holds for $K = 0$ or $L = 0$.*

Proof. Let M be the fractional Brownian motion (FBM) process, defined by $\text{Cov}(B(t), B(r)) = |r|^s + |t|^s - |t-r|^s$ for $r, t \geq 0$. For $i \geq 1$, we define $I_i^{(1)} = B(i) - B(i-1)$. From the well-known virtual LOO formulas (see [2, Proposition 3.1]), we have, for $n \geq 2$,

$$\begin{aligned} \sum_{i=1}^n b_{n,i} i^s &= b_{n,n} \left(|n|^s - \mathbb{E} \left(B(n) \mid (B(i) = |i|^s)_{i=1, \dots, n-1} \right) \right) \\ &= b_{n,n} \left(|n|^s - |n-1|^s - \mathbb{E} \left(I_n^{(1)} \mid (I_i^{(1)} = |i|^s - |i-1|^s)_{i=1, \dots, n-1} \right) \right). \end{aligned}$$

Now let M_I be the covariance matrix of $(I_1^{(1)}, \dots, I_n^{(1)})$. The matrix M_I is invertible since it is the covariance matrix of an invertible linear transformation of $(B(1), \dots, B(n))$, which has an invertible covariance matrix from [1]. Hence we obtain, with m_I the inverse of M_I , again from the virtual LOO formulas,

$$\sum_{i=1}^n b_{n,i} i^s = \frac{b_{n,n}}{m_{I,n,n}} \left(\sum_{i=1}^n m_{I,n,i} (|i|^s - |i-1|^s) \right). \quad (36)$$

We have, for $i \neq j$, $i, j \in \{1, \dots, n\}$,

$$\begin{aligned} \text{Cov}(I_i^{(1)}, I_j^{(1)}) &= \text{Cov}(B(i) - B(i-1), B(j) - B(j-1)) \\ &= -|i-j|^s + |i-j+1|^s + |i-j-1|^s - |i-j|^s \\ &< 0, \end{aligned}$$

by strict concavity of the function $|\cdot|^s$ on $[0, \infty)$. Hence the matrix M_I has positive diagonal elements, negative off-diagonal elements and is strictly positive definite. Hence, from Theorem 2.5.3 of [18], the elements of m_I are non-negative. This shows that $\sum_{i=1}^n b_{n,i} i^s > 0$ from (36). \square

A.6 Proof of Lemma 3.8

From (13), we have

$$A_\nu = \frac{\pi}{\Gamma(\nu) \sin(\nu\pi) 2^{2\nu} \Gamma(1+\nu)}$$

and

$$g_\nu(t) = \sum_{k=0}^{\infty} \frac{t^{2k}}{2^{2k+2} (k+1)! \prod_{i=1}^{k+1} (i-\nu)} - \frac{\pi}{\Gamma(\nu) \sin(\nu\pi)} \sum_{k=1}^{\infty} \frac{t^{2(k-1)+2\nu}}{2^{2k+2\nu} k! \Gamma(k+1+\nu)}.$$

Hence, the function g_ν is C^2 -differentiable on \mathbb{R}^+ by dominated convergence. \square

A.7 Proof of Theorem 4.1(i)

Since we have

$$(\hat{\sigma}^2, \hat{\alpha}) \in \operatorname{argmin}_{(\sigma^2, \alpha) \in (0, \infty) \times A} \sum_{i=K+1}^{n-L} \mathcal{L}(y_i | y_{K,L;i}),$$

the value of $\hat{\sigma}^2$ is given by, with $\hat{z}_{i,\hat{\alpha}} = r_{\hat{\alpha},K,L;i}^\top R_{\hat{\alpha},K,L;i}^{-1} z_{K,L;i}$ and $\hat{\sigma}_{i,\hat{\alpha}}^2 = 1 - r_{\hat{\alpha},K,L;i}^\top R_{\hat{\alpha},K,L;i}^{-1} r_{\hat{\alpha},K,L;i}$,

$$\frac{\hat{\sigma}^2}{\sigma_0^2} = \frac{1}{n-K-L} \sum_{i=K+1}^{n-L} \frac{(z_i - \hat{z}_{i,\hat{\alpha}})^2}{\hat{\sigma}_{i,\hat{\alpha}}^2}.$$

Let $\hat{r}_n(i, j) = r(\hat{\alpha} | i - j / n)$. Let us define the $(K+L) \times (K+L)$ matrix $b^{(n,\hat{\alpha})}$ as the inverse of the $(K+L) \times (K+L)$ matrix $(B - (\hat{\alpha}^s/n^s)C + (n^s/\hat{\alpha}^s)R^{(n,\hat{\alpha})})$, where $R^{(n,\hat{\alpha})}$ is given by

$$R_{i,j}^{(n,\hat{\alpha})} = \hat{r}_n(i, j) - \hat{r}_n(i, 0) \left(1 - \hat{\alpha}^s \frac{j^s}{n^s}\right) - \hat{r}_n(0, j) \left(1 - \hat{\alpha}^s \frac{i^s}{n^s}\right) - \hat{r}_n(i, 0) \hat{r}_n(0, j).$$

We remark that we have

$$\operatorname{Cov}_{1,\hat{\alpha}}(z_2, \dots, z_{K+L+1}) = \frac{\hat{\alpha}^s}{n^s} \left(B - \frac{\hat{\alpha}^s}{n^s} C + \frac{n^s}{\hat{\alpha}^s} R^{(n,\hat{\alpha})} \right).$$

Then, using the fact that $r(t) = O(t)$ for $0 < s < 1/2$ and $r(t) = O(t^2)$ for $1/2 \leq s < 3/2$, one gets

$$b^{(n,\hat{\alpha})} = b + \frac{\hat{\alpha}^s}{n^s} b C b \left(1 + o\left(\frac{1}{n^s}\right) \right) =: b + \frac{\hat{\alpha}^s}{n^s} \widetilde{b C b}.$$

The variance term $\hat{\sigma}_{i,\hat{\alpha}}^2$ reduces to

$$\hat{\sigma}_{i,\hat{\alpha}}^2 = \operatorname{Var}(z_i | z_{K,L;i}) = \frac{1}{(\operatorname{Cov}_{1,\hat{\alpha}}(z_2, \dots, z_K, z_{K+2}, \dots, z_{K+L+1})^{-1})_{K,K}} = \frac{\hat{\alpha}^s}{n^s} \frac{1}{b_{K,K}^{(n,\hat{\alpha})}},$$

from which we derive straightforwardly that

$$\hat{\sigma}_{i,\hat{\alpha}}^2 = \frac{\hat{\alpha}^s}{n^s} \frac{1}{b_{K,K}} \left(1 - \frac{\hat{\alpha}^s}{n^s} \frac{(\widetilde{bCb})_{K,K}}{b_{K,K}} + o\left(\frac{1}{n^s}\right) \right).$$

Moreover, similarly as in (27),

$$\begin{aligned} b_{K,K}^{(n,\hat{\alpha})}(z_i - \hat{z}_{i,\hat{\alpha}}) &= \sum_{k=1}^{K+L} b_{K,k}^{(n,\hat{\alpha})} \left(z_{i-K+k} - z_{i-K} + \frac{\hat{\alpha}^s}{n^s} k^s z_{i-K} - \hat{r}_n(k, 0) z_{i-K} \right) \\ &= \sum_{k=1}^{K+L} b_{K,k} (z_{i-K+k} - z_{i-K}) + \frac{\hat{\alpha}^s}{n^s} \sum_{k=1}^{K+L} ((bCb)_{K,k} (z_{i-K+k} - z_{i-K}) + b_{K,k} k^s z_{i-K}) \\ &\quad + \frac{\hat{\alpha}^{2s}}{n^{2s}} \left(\sum_{k=1}^{K+L} (\widetilde{bCb})_{K,k} k^s \right) z_{i-K} + a_{k,i,n}, \end{aligned}$$

where

$$\sup_{\substack{i=K+1,\dots,n-L \\ k=1,\dots,K+L}} |a_{k,i,n}| \leq \max \left\{ o\left(\frac{1}{n^{2s}}\right), O\left(\frac{1}{n^2}\right) \right\}.$$

Consequently, the ratio $\hat{\sigma}^2/\sigma_0^2$ equals

$$\begin{aligned} &\frac{1}{n-K-L} \frac{n^s}{\hat{\alpha}^s} \frac{1}{b_{K,K}} \left(1 - \frac{\hat{\alpha}^s}{n^s} \frac{(\widetilde{bCb})_{K,K}}{b_{K,K}} + o\left(\frac{1}{n^s}\right) \right) \sum_{i=K+1}^{n-L} \left[\sum_{k=1}^{K+L} b_{K,k} (z_{i-K+k} - z_{i-K}) + \right. \\ &\quad \left. \frac{\hat{\alpha}^s}{n^s} \sum_{k=1}^{K+L} ((bCb)_{K,k} (z_{i-K+k} - z_{i-K}) + b_{K,k} k^s z_{i-K}) + \frac{\hat{\alpha}^{2s}}{n^{2s}} \left(\sum_{k=1}^{K+L} (\widetilde{bCb})_{K,k} k^s \right) z_{i-K} + a_{k,i,n} \right]^2. \end{aligned} \tag{37}$$

- We start by computing the squares in (37).

First, we study the term $\sum_{i=K+1}^{n-L} \left(\sum_{k=1}^{K+L} b_{K,k} (z_{i-K+k} - z_{i-K}) \right)^2$. Its expectation is given by

$$\begin{aligned}
& \mathbb{E} \left[\sum_{i=K+1}^{n-L} \left(\sum_{k=1}^{K+L} b_{K,k} (z_{i-K+k} - z_{i-K}) \right)^2 \right] \\
&= \sum_{i=K+1}^{n-L} \sum_{k,l=1}^{K+L} b_{K,k} b_{K,l} \mathbb{E} [(z_{i-K+k} - z_{i-K}) (z_{i-K+l} - z_{i-K})] \\
&= \frac{\alpha_0^s}{n^s} \sum_{i=K+1}^{n-L} \sum_{k,l=1}^{K+L} b_{K,k} b_{K,l} B_{k,l} \\
&= \frac{\alpha_0^s}{n^s} \sum_{i=K+1}^{n-L} \sum_{k=1}^{K+L} b_{K,k} (bB)_{K,k} \\
&= \frac{\alpha_0^s}{n^s} \sum_{i=K+1}^{n-L} \sum_{k=1}^{K+L} b_{K,k} \delta_{K,k} \\
&\sim \alpha_0^s n^{1-s} b_{K,K}.
\end{aligned}$$

In addition, its variance is asymptotically in n^{1-2s} using Proposition 5 in [1] with a variation of order greater than 1 and $D = 0$. This yields a variation coefficient in $n^{-1/2}$ from which we conclude that

$$\sum_{i=K+1}^{n-L} \left(\sum_{k=1}^{K+L} b_{K,k} (z_{i-K+k} - z_{i-K}) \right)^2 = \alpha_0^s n^{1-s} b_{K,K} + O_{\mathbb{P}}(n^{1/2-s}).$$

Second, using (6), we study the term

$$\begin{aligned}
& \frac{\hat{\alpha}^{2s}}{n^{2s}} \sum_{i=K+1}^{n-L} \left(\sum_{k=1}^{K+L} \left((\widetilde{bCb})_{K,k} (z_{i-K+k} - z_{i-K}) + b_{K,k} k^s z_{i-K} \right) \right)^2 \\
&= \frac{\hat{\alpha}^{2s}}{n^{2s}} \sum_{i=K+1}^{n-L} \left(\sum_{k=1}^{K+L} (\widetilde{bCb})_{K,k} (z_{i-K+k} - z_{i-K}) \right)^2 + \frac{\hat{\alpha}^{2s}}{n^{2s}} (bCb)_{K,K} \left(\sum_{i=K+1}^{n-L} z_{i-K}^2 \right) \\
&\quad + 2 \frac{\hat{\alpha}^{2s}}{n^{2s}} \left(\sum_{k=1}^{K+L} b_{K,k} k^s \right) \sum_{i=K+1}^{n-L} z_{i-K} \left(\sum_{k=1}^{K+L} (\widetilde{bCb})_{K,k} (z_{i-K+k} - z_{i-K}) \right).
\end{aligned}$$

We have

$$\frac{\hat{\alpha}^{2s}}{n^{2s}} (bCb)_{K,K} \left(\sum_{i=K+1}^{n-L} z_{i-K}^2 \right) \sim \hat{\alpha}^{2s} n^{1-2s} (bCb)_{K,K} \int_0^1 Z(t)^2 dt.$$

Moreover, using the Hölder property of the process at order $\beta < 1/2$, we get, almost surely,

$$\begin{aligned} |z_{i-K+k} - z_{i-K}| &\leq \max_{i=K+1, \dots, n-L} |z_{i-K+k} - z_{i-K}| \\ &\leq (\text{Const}) \max_{k=1, \dots, K+L} \frac{k^\beta}{n^\beta} \\ &\leq (\text{Const}) \frac{(K+L)^\beta}{n^\beta} \\ &= O_{\mathbb{P}}(n^{-\beta}). \end{aligned}$$

Consequently,

$$\frac{\hat{\alpha}^{2s}}{n^{2s}} \sum_{i=K+1}^{n-L} \left(\sum_{k=1}^{K+L} (\widetilde{bCb})_{K,k} (z_{i-K+k} - z_{i-K}) \right)^2 = O_{\mathbb{P}}(n^{1-2s}) = o_{\mathbb{P}}(n^{1-2s}).$$

By Cauchy-Schwarz inequality, we conclude to

$$\begin{aligned} \frac{\hat{\alpha}^{2s}}{n^{2s}} \sum_{i=K+1}^{n-L} \left(\sum_{k=1}^{K+L} \left((\widetilde{bCb})_{K,k} (z_{i-K+k} - z_{i-K}) + b_{K,k} k^s z_{i-K} \right) \right)^2 \\ = \hat{\alpha}^{2s} n^{1-2s} (bCb)_{K,K} \int_0^1 Z(t)^2 dt + o_{\mathbb{P}}(n^{1-2s}). \end{aligned}$$

Third, we study the term

$$\frac{\hat{\alpha}^{4s}}{n^{4s}} \left(\sum_{k=1}^{K+L} (\widetilde{bCb})_{K,k} k^s \right)^2 \left(\sum_{i=K+1}^{n-L} z_{i-K}^2 \right) = \frac{\hat{\alpha}^{4s}}{n^{4s}} \left(\sum_{k=1}^{K+L} (bCb)_{K,k} k^s \right)^2 O_{\mathbb{P}}(n) = o_{\mathbb{P}}(n^{1-2s}).$$

• Now, let us turn to the double products in (37). First,

$$\begin{aligned} &2 \frac{\hat{\alpha}^s}{n^s} \sum_{i=K+1}^{n-L} \left(\sum_{k=1}^{K+L} b_{K,k} (z_{i-K+k} - z_{i-K}) \right) \left(\sum_{k=1}^{K+L} \left((\widetilde{bCb})_{K,k} (z_{i-K+k} - z_{i-K}) + b_{K,k} k^s z_{i-K} \right) \right) \\ &= 2 \frac{\hat{\alpha}^s}{n^s} \sum_{i=K+1}^{n-L} \left(\sum_{k=1}^{K+L} b_{K,k} (z_{i-K+k} - z_{i-K}) \right) \left(\sum_{l=1}^{K+L} (\widetilde{bCb})_{K,l} (z_{i-K+l} - z_{i-K}) \right) \\ &\quad + 2 \frac{\hat{\alpha}^s}{n^s} \left(\sum_{k=1}^{K+L} b_{K,k} k^s \right) \sum_{i=K+1}^{n-L} z_{i-K} \left(\sum_{k=1}^{K+L} b_{K,k} (z_{i-K+k} - z_{i-K}) \right). \end{aligned} \tag{38}$$

The first term in the right hand side of the previous equation is

$$\begin{aligned} &2 \frac{\hat{\alpha}^s}{n^s} \sum_{i=K+1}^{n-L} \left(\sum_{k=1}^{K+L} b_{K,k} (z_{i-K+k} - z_{i-K}) \right) \left(\sum_{l=1}^{K+L} (\widetilde{bCb})_{K,l} (z_{i-K+l} - z_{i-K}) \right) \\ &= 2 \frac{\hat{\alpha}^s}{n^s} \sum_{i=K+1}^{n-L} \sum_{k,l=1}^{K+L} b_{K,k} (\widetilde{bCb})_{K,l} (z_{i-K+k} - z_{i-K}) (z_{i-K+l} - z_{i-K}) \end{aligned}$$

where the expectation of the double sum provides

$$2\hat{\alpha}^s \alpha_0^s n^{1-2s} \sum_{k,l=1}^{K+L} b_{K,k} (\widetilde{bCb})_{K,l} B_{k,l} + o(n^{1-2s}) = 2\hat{\alpha}^s \alpha_0^s n^{1-2s} (\widetilde{bCb})_{K,K} + o(n^{1-2s})$$

while its variance is $O(n^{1-s})$ by similar arguments.

Taking the expectation of the sums in (38) gives

$$\begin{aligned} & 2 \frac{\hat{\alpha}^s}{n^s} \mathbb{E} \left[\left(\sum_{k=1}^{K+L} b_{K,k} k^s \right) \sum_{i=K+1}^{n-L} z_{i-K} \left(\sum_{k=1}^{K+L} b_{K,k} (z_{i-K+k} - z_{i-K}) \right) \right] \\ &= 2 \frac{\hat{\alpha}^s}{n^s} \left(\sum_{k=1}^{K+L} b_{K,k} k^s \right) \sum_{i=K+1}^{n-L} \sum_{k=1}^{K+L} b_{K,k} \mathbb{E} [z_{i-K} (z_{i-K+k} - z_{i-K})] \\ &= -2\hat{\alpha}^s \alpha_0^s n^{1-2s} \left(\sum_{k=1}^{K+L} b_{K,k} k^s \right)^2 + o(n^{1-2s}) \\ &= -2\hat{\alpha}^s \alpha_0^s n^{1-2s} (bCb)_{K,K} + o(n^{1-2s}). \end{aligned}$$

In addition, after tedious computations, taking the variance of the sums in (38) provides

$$4 \frac{\hat{\alpha}^{2s}}{n^{2s}} \frac{(\widetilde{bCb})_{K,K}}{b_{K,K}^4} \sum_{k,l=1}^{K+1} b_{K,k} b_{K,l} \sum_{i,j=K+1}^{n-L} \text{Cov}((z_{i-K+k} - z_{i-K})z_{i-K}, (z_{j-K+l} - z_{j-K})z_{j-K})$$

and is proven to be $O(n^{1-s})$. Indeed, for fixed k and l , we study

$$\begin{aligned} & \sum_{i,j=K+1}^{n-L} \text{Cov}((z_{i-K+k} - z_{i-K})z_{i-K}, (z_{j-K+l} - z_{j-K})z_{j-K}) \\ &= \sum_{i,j=K+1}^{n-L} \text{Cov}(z_{i-K+k}z_{i-K}, z_{j-K+l}z_{j-K}) - \text{Cov}(z_{i-K+k}z_{i-K}, z_{j-K}^2) \\ & \quad - \text{Cov}(z_{i-K}^2, z_{j-K+l}z_{j-K}) + \text{Cov}(z_{i-K}^2, z_{j-K}^2) \\ &= \frac{\alpha_0^s}{n^s} \sum_{i,j=K+1}^{n-L} (|i-j+k|^s + |i-j-l|^s - |i-j+k-l|^s - |i-j|^s) \\ & \quad + \frac{\alpha_0^{2s}}{n^{2s}} \sum_{i,j=K+1}^{n-L} [|i-j|^s (|i-j+k-l|^s - 2|i-j+k|^s - |i-j-l|^s + 2|i-j|^s) \\ & \quad + |i-j-l|^s (|i-j+k|^s - |i-j|^s)]. \end{aligned}$$

The term $(|i-j+k|^s + |i-j-l|^s - |i-j+k-l|^s - |i-j|^s)$ is a variation of order 2 implying that the first sum is $\alpha_0^s n^{1-s}$. The second sum is bounded by

$$\alpha_0^{2s} n^{1-2s} \left(C + \sum_{a=0}^{n-L-K-1} a^{s-1} \right) \leq \alpha_0^{2s} n^{1-2s} \left(C + \int_2^n t^{s-1} dt \right) \leq \alpha_0^{2s} n^{1-s}.$$

Finally, we conclude that the left hand side of (38) is given by

$$2\hat{\alpha}^s \alpha_0^s n^{1-2s} (\widetilde{bCb})_{K,K} - 2\hat{\alpha}^s \alpha_0^s n^{1-2s} (bCb)_{K,K} + o(n^{1-2s}).$$

Second, turning to the second double product in (37), we get, by Cauchy-Schwarz inequality,

$$\begin{aligned} & 2 \frac{\hat{\alpha}^{2s}}{n^{2s}} \sum_{i=K+1}^{n-L} \left(\sum_{k=1}^{K+L} b_{K,k} (z_{i-K+k} - z_{i-K}) \right) \left(\sum_{k=1}^{K+L} (\widetilde{bCb})_{K,k} k^s \right) z_{i-K} \\ & \leq 2 \frac{\hat{\alpha}^{2s}}{n^{2s}} \left(\sum_{i=K+1}^{n-L} \left(\sum_{k=1}^{K+L} b_{K,k} (z_{i-K+k} - z_{i-K}) \right)^2 \right)^{1/2} \left(\sum_{i=K+1}^{n-L} z_{i-K}^2 \right)^{1/2} \left(\sum_{k=1}^{K+L} (\widetilde{bCb})_{K,k} k^s \right) \\ & = n^{-2s} O_{\mathbb{P}}(n^{1/2-s/2}) O_{\mathbb{P}}(n^{1/2}) \\ & = O_{\mathbb{P}}(n^{1-5s/2}) \\ & = o_{\mathbb{P}}(n^{1-2s}). \end{aligned}$$

Third, for the last double product in (37), we proceed analogously:

$$\begin{aligned} & 2 \frac{\hat{\alpha}^{3s}}{n^{3s}} \left(\sum_{k=1}^{K+L} (\widetilde{bCb})_{K,k} k^s \right) \sum_{i=K+1}^{n-L} z_{i-K} \left(\sum_{k=1}^{K+L} \left((\widetilde{bCb})_{K,k} (z_{i-K+k} - z_{i-K}) + b_{K,k} k^s z_{i-K} \right) \right) \\ & \leq 2 \frac{\hat{\alpha}^{3s}}{n^{3s}} \left(\sum_{k=1}^{K+L} (\widetilde{bCb})_{K,k} k^s \right) \left(\sum_{i=K+1}^{n-L} z_{i-K}^2 \right)^{1/2} \\ & \quad \left(\sum_{i=K+1}^{n-L} \left(\sum_{k=1}^{K+L} \left((\widetilde{bCb})_{K,k} (z_{i-K+k} - z_{i-K}) + b_{K,k} k^s z_{i-K} \right) \right) \right)^{1/2} \\ & = n^{-3s} O_{\mathbb{P}}(n^{1/2}) O_{\mathbb{P}}(n^{1/2}) \\ & = O_{\mathbb{P}}(n^{1-3s}) \\ & = o_{\mathbb{P}}(n^{1-2s}). \end{aligned}$$

One can show straightforwardly that the effect of $a_{k,i,n}$ in (37) is negligible. Hence, plugging the asymptotic behaviors of the three squares and the three double products into (37) leads to (15) which concludes the proof. \square

A.8 Proof of Theorem 4.1(ii)

The proof directly comes following the same lines as the proofs of Theorems 3.3(ii) and 4.1(i). \square

A.9 Proof of Proposition 4.2

In the case $K = 1$ and $L = 0$, the value of $\hat{\sigma}^2$ is given by:

$$\frac{\hat{\sigma}^2}{\sigma_0^2} = \frac{1}{n-1} \sum_{i=2}^n \frac{(z_i - \hat{z}_{i,\hat{\alpha}})^2}{\hat{\sigma}_{i,\hat{\alpha}}^2}. \quad (39)$$

The variance term $\hat{\sigma}_{i,\hat{\alpha}}^2$ reduces to

$$\hat{\sigma}_{i,\hat{\alpha}}^2 = \text{Var}(z_i | z_{K,L;i}) = \text{Var}(z_i - r_i^\top R_i^{-1} z_{1,0;i}) = 1 - r_{i,\hat{\alpha}}^\top R_{i,\hat{\alpha}}^{-1} r_{i,\hat{\alpha}} = \frac{\hat{\alpha}^s}{n^s} \left(2 - \frac{\hat{\alpha}^s}{n^s} \right) + O\left(\frac{1}{n}\right). \quad (40)$$

Moreover,

$$\begin{aligned} (z_i - \hat{z}_{i,1})^2 &= \left(z_i - \left(1 - \frac{\hat{\alpha}^s}{n^s} + \hat{r}_n(1) \right) z_{i-1} \right)^2 \\ &= (z_i - z_{i-1})^2 + 2 \frac{\hat{\alpha}^s}{n^s} (z_i - z_{i-1}) z_{i-1} + \frac{\hat{\alpha}^{2s}}{n^{2s}} z_{i-1}^2 + a_{i,n} \end{aligned} \quad (41)$$

where $\sup_{i=2,\dots,n} |a_{i,n}| = O_{\mathbb{P}}(1/n)$.

• First, we study the variance of the term $\sum_{i=2}^n (z_i - z_{i-1})^2$. By Melher's formula and letting $a = |i - j|$, one gets

$$\begin{aligned} \text{Var} \left(\sum_{i=2}^n (z_i - \hat{z}_{i,\hat{\alpha}})^2 \right) &\leq (\text{Const}) \frac{\alpha_0^{2s}}{n^{2s-1}} \sum_{a=0}^{n-2} \left(1 - \frac{a}{n} \right) a^{s-2} \\ &\leq (\text{Const}) \frac{\alpha_0^{2s}}{n^{2s-1}} \sum_{a=0}^{n-2} a^{s-2} \\ &= O(n^{1-2s}). \end{aligned}$$

Consequently,

$$\sum_{i=2}^n \left((z_i - \hat{z}_{i,1})^2 - \mathbb{E}[(z_i - \hat{z}_{i,1})^2] \right) = O_{\mathbb{P}}(n^{1/2-s}).$$

But, since

$$\mathbb{E} \left[\sum_{i=2}^n (z_i - \hat{z}_{i,1})^2 \right] = \sum_{i=2}^n \text{Cov}(z_i - \hat{z}_{i,1}, z_i - \hat{z}_{i,1}) = 2n^{1-s} \alpha_0^s + O(1),$$

then

$$\sum_{i=2}^n (z_i - z_{i-1})^2 = 2n^{1-s} \alpha_0^s + O_{\mathbb{P}}(n^{1/2-s}). \quad (42)$$

- Second, $(1/n) \sum_{i=2}^n z_{i-1}^2$ is a sequence of random variables converging in the L^2 sense to $\int_0^1 Z(t)^2 dt$ (see, e.g. [37, 8]) and thus,

$$\sum_{i=2}^n \frac{\hat{\alpha}^{2s}}{n^{2s}} z_{i-1}^2 \sim_{\mathbb{P}} n^{1-2s} \hat{\alpha}^{2s} \int_0^1 Z(t)^2 dt. \quad (43)$$

- Third, let us turn to the double product in (41). We use the following result:

$$\text{Cov}(XY, ZT) = \text{Cov}(X, Z) \text{Cov}(Y, T) + \text{Cov}(X, T) \text{Cov}(Y, Z) \quad (44)$$

where (X, Y, Z, T) is a centered Gaussian vector. We have, after some tedious computations,

$$\begin{aligned} & 4 \frac{\hat{\alpha}^{2s}}{n^{2s}} \text{Var} \left(\sum_{i=2}^n (z_i - z_{i-1}) z_{i-1} \right) = 4 \frac{\hat{\alpha}^{2s}}{n^{2s}} \sum_{i,j=2}^n \text{Cov}((z_i - z_{i-1}) z_{i-1}, (z_j - z_{j-1}) z_{j-1}) \\ &= 4 \frac{\hat{\alpha}^{2s}}{n^{2s}} \sum_{i,j=2}^n \text{Cov}(z_i z_{i-1}, z_j z_{j-1}) - \text{Cov}(z_i z_{i-1}, z_{j-1}^2) - \text{Cov}(z_{i-1}^2, z_j z_{j-1}) + \text{Cov}(z_{i-1}^2, z_{j-1}^2) \\ &= 4 \frac{\hat{\alpha}^{2s} \alpha_0^s}{n^{3s}} \sum_{i,j=2}^n (|i-j+1|^s + |i-j-1|^s - 2|i-j|^s) + \frac{1}{n^{2s}} O_{\mathbb{P}} \left(n^2 \frac{1}{n} \right) + 4 \frac{\hat{\alpha}^{2s} \alpha_0^{2s}}{n^{4s}} \times \\ & \quad \sum_{i,j=2}^n (3|i-j|^{2s} + |i-j-1|^s |i-j+1|^s - 2|i-j|^s (|i-j+1|^s + |i-j-1|^s)). \end{aligned}$$

Using Lemma A.1, the first sum gives

$$\begin{aligned} & \sum_{i,j=2}^n (|i-j+1|^s + |i-j-1|^s - 2|i-j|^s) \\ & \leq 2 \sum_{a=0}^{n-2} (n-a) (|a+1|^s + |a-1|^s - 2a^s) \\ & \leq 2n \sum_{a=0}^{n-2} a^{s-2} = O(n). \end{aligned}$$

Analogously, the second sum gives

$$\begin{aligned}
& \sum_{i,j=2}^n (3|i-j|^{2s} + |i-j-1|^s |i-j+1|^s - 2|i-j|^s (|i-j+1|^s + |i-j-1|^s)) \\
& \sim 2 \sum_{a=0}^{n-2} (n-a) (3a^{2s} + |a-1|^s |a+1|^s - 2a^s (|a+1|^s + |a-1|^s)) \\
& = 2 \sum_{a=0}^{n-2} (n-a) (2a^s (a^s - |a+1|^s) + a^s (a^s - |a-1|^s) + |a-1|^s (|a+1|^s - a^s)) \\
& \leq 8n \left((Const) + \sum_{a=2}^{n-2} a^{2s-1} \right) \\
& \leq 8n \left((Const) + \int_1^{n-2} t^{2s-1} dt \right) \\
& \sim 8n^{1+2s} = O(n^{1+2s}).
\end{aligned}$$

Hence, the variance of the double product is $O(n^{1-2s})$ while the expectation provides

$$2 \frac{\hat{\alpha}^s}{n^s} \mathbb{E} \left[\sum_{i=2}^n (z_i - z_{i-1}) z_{i-1} \right] = 2 \frac{\hat{\alpha}^s}{n^s} \sum_{i=2}^n \left(\left(1 - \frac{\alpha_0^s}{n^s} + \hat{r}_n(1)\right) - 1 \right) = -2\alpha_0^s \hat{\alpha}^s n^{1-2s} + O_{\mathbb{P}}(n^{-s}).$$

Consequently,

$$2 \frac{\hat{\alpha}^s}{n^s} \sum_{i=2}^n (z_i - z_{i-1}) z_{i-1} = -2\alpha_0^s \hat{\alpha}^s n^{1-2s} + O(n^{1/2-s}). \quad (45)$$

Now, (42), (43) and (45) together with (40) lead to

$$\begin{aligned}
\frac{\hat{\sigma}^2}{\sigma_0^2} & \sim \frac{1}{2n^{1-s} \hat{\alpha}^s} \left(2n^{1-s} \alpha_0^s + O_{\mathbb{P}}(n^{1/2-s}) + \frac{\hat{\alpha}^s}{2n^s} 2n^{1-s} \alpha_0^s + n^{1-2s} \hat{\alpha}^{2s} \int_0^1 Z^2(t) dt \right. \\
& \quad \left. - 2n^{1-2s} \hat{\alpha}^s \alpha_0^s + O_{\mathbb{P}}(n^{-1/2}) \right).
\end{aligned}$$

Finally, one gets

$$\frac{\hat{\sigma}^2 \hat{\alpha}^s}{\sigma_0^2 \alpha_0^s} - 1 \sim \frac{\hat{\alpha}^s}{2n^s} \left(\frac{\hat{\alpha}^s}{\alpha_0^s} \int_0^1 Z^2(t) dt - 1 \right).$$

The proof of (18) is now complete. \square

Acknowledgments The authors warmly thank Guillaume Cebon for constructive and helpful discussions on the topic of this paper.

References

- [1] 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.
- [2] 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.
- [3] 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.
- [4] F. Bachoc. Asymptotic analysis of covariance parameter estimation for Gaussian processes in the misspecified case. *Bernoulli*, 24(2):1531–1575, 2018.
- [5] F. Bachoc, M. Bevilacqua, and D. Velandia. Composite likelihood estimation for a gaussian process under fixed domain asymptotics. *arXiv:1807.08988*, 2018.
- [6] 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.
- [7] Y. Cao and D. J. Fleet. Generalized product of experts for automatic and principled fusion of Gaussian process predictions. In *Modern Nonparametrics 3: Automating the Learning Pipeline workshop at NIPS, Montreal*, 2014. arXiv preprint arXiv:1410.7827.
- [8] H. Cramér and M. R. Leadbetter. *Stationary and related stochastic processes. Sample function properties and their applications*. John Wiley & Sons, Inc., New York-London-Sydney, 1967.
- [9] N. Cressie. *Statistics for spatial data*. J. Wiley, 1993.
- [10] N. Cressie and S. Lahiri. Asymptotics for REML estimation of spatial covariance parameters. *Journal of Statistical Planning and Inference*, 50:327–341, 1996.
- [11] A. Datta, S. Banerjee, A. O. Finley, and A. E. Gelfand. Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. *Journal of the American Statistical Association*, 111(514):800–812, 2016.
- [12] M. P. Deisenroth and J. W. Ng. Distributed Gaussian processes. *Proceedings of the 32nd International Conference on Machine Learning, Lille, France. JMLR: W&CP volume 37*, 2015.
- [13] R. Furrer, M. G. Genton, and D. Nychka. Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics*, 15(3):502–523, 2006.

- [14] R. B. Gramacy and D. W. Apley. Local Gaussian process approximation for large computer experiments. *Journal of Computational and Graphical Statistics*, 24(2):561–578, 2015.
- [15] R. B. Gramacy et al. laGP: large-scale spatial modeling via local approximate Gaussian processes in R. *Journal of Statistical Software*, 72(1):1–46, 2016.
- [16] J. Hensman and N. Fusi. Gaussian processes for big data. *Uncertainty in Artificial Intelligence*, pages 282–290, 2013.
- [17] G. E. Hinton. Training products of experts by minimizing contrastive divergence. *Neural computation*, 14(8):1771–1800, 2002.
- [18] R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge university press, 1991.
- [19] I. Ibragimov and Y. Rozanov. *Gaussian Random Processes*. Springer-Verlag, New York, 1978.
- [20] 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.
- [21] C. Kaufman and B. Shaby. The role of the range parameter for estimation and prediction in geostatistics. *Biometrika*, 100:473–484, 2013.
- [22] C. G. Kaufman, M. J. Schervish, and D. W. Nychka. Covariance tapering for likelihood-based estimation in large spatial data sets. *Journal of the American Statistical Association*, 103(484):1545–1555, 2008.
- [23] W. V. Li and W. Linde. Approximation, metric entropy and small ball estimates for gaussian measures. *Ann. Probab.*, 27:1556–1578, 1999.
- [24] W.-L. Loh. Estimating the smoothness of a Gaussian random field from irregularly spaced data via higher-order quadratic variations. *Ann. Statist.*, 43(6):2766–2794, 12 2015.
- [25] A. López-Lopera, F. Bachoc, N. Durrande, and O. Roustand. Finite-dimensional Gaussian approximation with linear inequality constraints. *SIAM/ASA Journal on Uncertainty Quantification*, 6(3):1224–1255, 2018.
- [26] K. Mardia and R. Marshall. Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika*, 71:135–146, 1984.
- [27] J. Mateu, E. Porcu, G. Christakos, and M. Bevilacqua. Fitting negative spatial covariances to geothermal field temperatures in Nea Kessani (Greece). *Environmetrics: The official journal of the International Environmetrics Society*, 18(7):759–773, 2007.

- [28] E. Pardo-Igúzquiza and P. A. Dowd. AMLE3D: A computer program for the inference of spatial covariance parameters by approximate maximum likelihood estimation. *Computers & Geosciences*, 23(7):793–805, 1997.
- [29] C. Rasmussen and C. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, Cambridge, 2006.
- [30] H. Rue and L. Held. *Gaussian Markov random fields, Theory and applications*. Chapman & Hall, 2005.
- [31] D. Rullière, N. Durrande, F. Bachoc, and C. Chevalier. Nested Kriging predictions for datasets with a large number of observations. *Statistics and Computing*, 28(4):849–867, 2018.
- [32] B. A. Shaby and D. Ruppert. Tapered covariance: Bayesian estimation and asymptotics. *Journal of Computational and Graphical Statistics*, 21(2):433–452, 2012.
- [33] D. Slepian. On the zeros of Gaussian noise. In *Proc. Sympos. Time Series Analysis (Brown Univ., 1962)*, pages 104–115. Wiley, New York, 1963.
- [34] M. Stein. Asymptotically efficient prediction of a random field with a misspecified covariance function. *The Annals of Statistics*, 16:55–63, 1988.
- [35] M. Stein. Bounds on the efficiency of linear predictions using an incorrect covariance function. *The Annals of Statistics*, 18:1116–1138, 1990.
- [36] M. Stein. Uniform asymptotic optimality of linear predictions of a random field using an incorrect second-order structure. *The Annals of Statistics*, 18:850–872, 1990.
- [37] M. Stein. *Interpolation of Spatial Data: Some Theory for Kriging*. Springer, New York, 1999.
- [38] M. L. Stein. Limitations on low rank approximations for covariance matrices of spatial data. *Spatial Statistics*, 8:1–19, 2014.
- [39] M. L. Stein, Z. Chi, and L. J. Welty. Approximating likelihoods for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(2):275–296, 2004.
- [40] V. Tresp. A Bayesian committee machine. *Neural Computation*, 12(11):2719–2741, 2000.
- [41] B. van Stein, H. Wang, W. Kowalczyk, T. Bäck, and M. Emmerich. Optimally weighted cluster Kriging for big data regression. In *International Symposium on Intelligent Data Analysis*, pages 310–321. Springer, 2015.
- [42] C. Varin, N. Reid, and D. Firth. An overview of composite likelihood methods. *Statistica Sinica*, 21:5–42, 2011.

- [43] A. V. Vecchia. Estimation and model identification for continuous spatial processes. *Journal of the Royal Statistical Society: Series B (Methodological)*, 50(2):297–312, 1988.
- [44] Z. Ying. Asymptotic properties of a maximum likelihood estimator with data from a Gaussian process. *Journal of Multivariate Analysis*, 36:280–296, 1991.
- [45] H. Zhang. Inconsistent estimation and asymptotically equivalent interpolations in model-based geostatistics. *Journal of the American Statistical Association*, 99:250–261, 2004.
- [46] H. Zhang and D. Zimmerman. Towards reconciling two asymptotic frameworks in spatial statistics. *Biometrika*, 92:921–936, 2005.