

## SEMI-PARAMETRIC ESTIMATION OF THE VARIOGRAM SCALE PARAMETER OF A GAUSSIAN PROCESS WITH STATIONARY INCREMENTS

JEAN-MARC AZAÏS<sup>1</sup>, FRANÇOIS BACHOC<sup>1</sup>, AGNÈS LAGNOUX<sup>2,\*</sup>  
AND THI MONG NGOC NGUYEN<sup>3</sup>

**Abstract.** We consider the semi-parametric estimation of the scale parameter of the variogram of a one-dimensional Gaussian process with known smoothness. We suggest an estimator based both on quadratic variations and the moment method. We provide asymptotic approximations of the mean and variance of this estimator, together with asymptotic normality results, for a large class of Gaussian processes. We allow for general mean functions, provide minimax upper bounds and study the aggregation of several estimators based on various variation sequences. In extensive simulation studies, we show that the asymptotic results accurately depict the finite-sample situations already for small to moderate sample sizes. We also compare various variation sequences and highlight the efficiency of the aggregation procedure.

**Mathematics Subject Classification.** 60G15, 62F12.

Received December 20, 2019. Accepted September 3, 2020.

### 1. INTRODUCTION

#### 1.1. General context and state of the art

Gaussian process models are widely used in statistics. For instance, they enable to interpolate observations by Kriging, notably in computer experiment designs to build a metamodel [44, 54]. A second type of application of Gaussian processes is the analysis of local characteristics of images [45] and one dimensional signals (*e.g.* in finance, see [26, 61] and the references therein). A central problem with Gaussian processes is the estimation of the covariance function or the variogram. In this paper, we consider a real-valued Gaussian process  $(X(t))_{t \in \mathbb{R}}$  with stationary increments. Its semi-variogram is well-defined and given by

$$V(h) := \frac{1}{2} \mathbb{E} \left[ (X(t+h) - X(t))^2 \right]. \quad (1.1)$$

---

*Keywords and phrases:* Gaussian processes, semi-parametric estimation, quadratic variations, scale covariance parameter, asymptotic normality, moment method, minimax upper bounds, aggregation of estimators.

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

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

<sup>3</sup> Faculty of Mathematics & Computer Science, University of Science, VNU-HCMC, Ho Chi Minh City, Viet Nam; Vietnam National University, Ho Chi Minh City, Viet Nam.

\* Corresponding author: [lagnoux@univ-tlse2.fr](mailto:lagnoux@univ-tlse2.fr)

Ideally, one aims at knowing perfectly the function  $V$  or at least estimate it precisely, either in a parametric setting or in a nonparametric setting. The parametric approach consists in assuming that the mean function of the Gaussian process (the drift) is a linear combination of known functions (often polynomials) and that the semi-variogram  $V$  belongs to a parametric family of semi-variograms  $\{V_\theta, \theta \in \Theta \subset \mathbb{R}^p\}$  for a given  $p$  in  $\mathbb{N}^*$ . Furthermore, in most practical cases, the semi-variogram is assumed to stem from a stationary autocovariance function  $k$  defined by  $k(h) = \text{Cov}(X(t), X(t+h))$ . In that case, the process is supposed to be stationary (assuming also the mean to be constant), and  $V$  can be rewritten in terms of the process autocovariance function  $k$ :  $V(h) = k(0) - k(h)$ . Moreover, a parametric set of stationary covariance functions is considered of the form  $\{k_\theta, \theta \in \Theta\}$  with  $\Theta \subset \mathbb{R}^p$ . In such a setting, several estimation procedures for  $\theta$  have been introduced and studied in the literature. Usually in practice, most of the software packages (like, *e.g.* `DiceKriging` [48]) use the maximum likelihood estimation method (MLE) to estimate  $\theta$  (see [44, 52, 54] for more details on MLE). Unfortunately, MLE necessitates to solve linear systems and compute determinants and is very often obtained via an iterative procedure for likelihood optimization. So, it is known to be computationally expensive and intractable for large data sets, and in addition, the iterative procedure may diverge (that is to say, in practice, the sequence of covariance parameters obtained during the iterative procedure for likelihood optimization may diverge or converge to parameters that are at the boundaries of the definition domain  $\Theta$ ) in some complicated situations (see Sect. 5.4). This has motivated the search for alternative estimation methods with a good balance between computational complexity and statistical efficiency. Among these methods, we can mention low rank approximation [55], sparse approximation [27], covariance tapering [20, 32], Gaussian Markov random fields approximation [16, 49], submodel aggregation [9, 19, 28, 50, 57, 58] and composite likelihood [6].

## 1.2. Framework and motivation

The approaches discussed above are parametric. In this paper, we consider a more general semi-parametric context. The Gaussian process  $X$  is only assumed to have stationary increments and no parametric assumption is made on the semi-variogram  $V$  in (1.1). Assume, to simplify, that the semi-variogram is a  $C^\infty$  function outside 0. This is the case for most of the models even if the sample paths are not regular, see the examples in Section 2.1. Let  $D$  be the order of differentiability in quadratic mean of  $(X(t))_{t \in \mathbb{R}}$ . This is equivalent to the fact that  $V$  is  $2D$  differentiable and not  $2D+2$  differentiable. Let us assume that the  $2D$ th derivative of  $V$  has the following expansion at the origin:

$$V^{(2D)}(h) = V^{(2D)}(0) + C(-1)^D |h|^s + r(h), \quad (1.2)$$

where  $C \geq 0$ ,  $0 < s < 2$ , and the remainder function  $r$  satisfies some hypothesis detailed further (see Sect. 2.1) and is a  $o(|h|^s)$  as  $h \rightarrow 0$ . Note that, since  $s < 2$ ,  $V$  is indeed not  $(2D+2)$  differentiable. The quantity  $s$  is the smoothness parameter and we call  $C$  the scale parameter. In this paper, we assume  $D$  and  $s$  to be known and we focus on the theoretical study of the semi-parametric estimation of  $C$  defined in (1.2) in dimension one. Remark that this makes it possible to test whether the Gaussian process stems from a white noise or not. Notice that we also perform some additional simulations in higher dimensions. The value of  $C$  may lead to significantly different behaviors of the process  $X$  as one can see in Figure 1 which represents several realizations of a Gaussian process with exponential covariance function for different values of  $C$ , *i.e.*  $V(h) = 1 - \exp(-C|h|)$  (which satisfies (1.2) with  $D = 0$ ).

In many cases, a large  $C$  is associated to a small dependence between the values of  $X$  and a small  $C$  is associated to an almost constant process  $X$ . For instance, in the above example  $V(h) = 1 - \exp(-C|h|)$ , the limit of  $V(h)$  is  $1 = \text{Var}(X(0))$  when  $C \rightarrow +\infty$  and the limit of  $V(h)$  is 0 when  $C \rightarrow 0$ . Hence, in this example, the case  $C = +\infty$  can be interpreted as independence between the values of  $X$  and the case  $C = 0$  can be interpreted as a constant process  $X$ .

As a motivating example, consider the following case where the estimation of  $C$  is beneficial. Assume that we observe a signal  $S$  depending on a vector of initial parameters  $x$  given by a computer code described by the

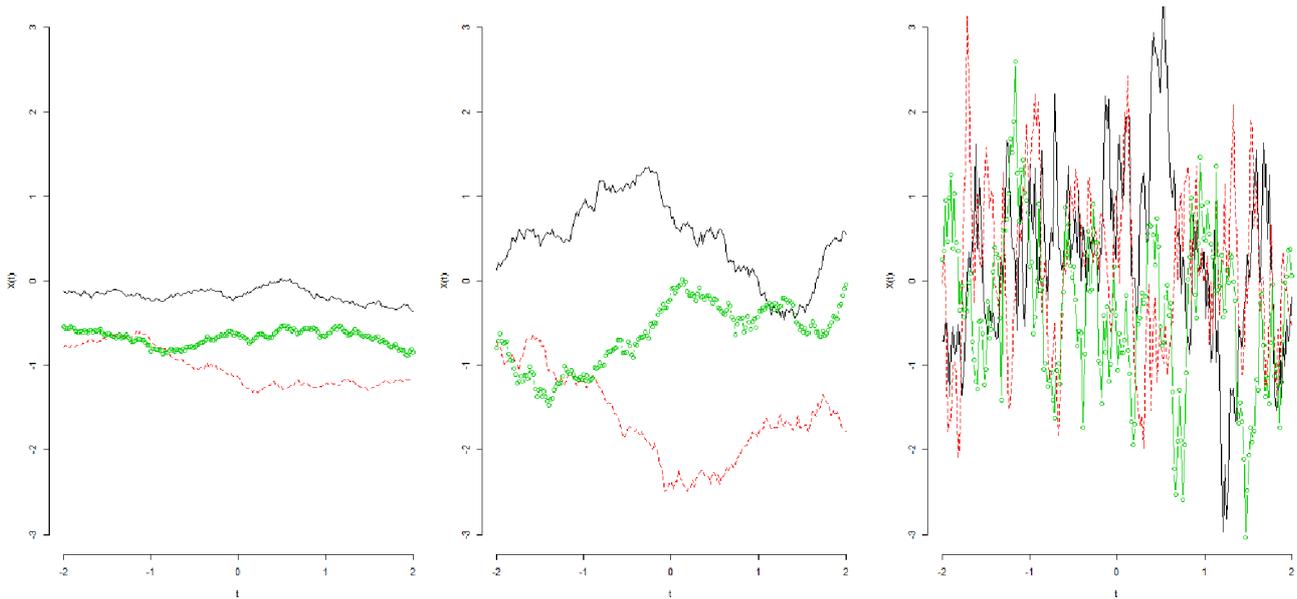


FIGURE 1. Several realizations of Gaussian processes with exponential covariance function with parameter  $C = 0.1$  (*left*), 10 (*middle*), and 100 (*right*).

application:

$$\begin{aligned}
 S : E &\rightarrow \mathbb{R}^{[0,1]} \\
 x &\mapsto (S_x(t), t \in [0, 1]),
 \end{aligned}
 \tag{1.3}$$

where  $E$  stands for the initial parameter space. In order to interpret the output curve  $t \mapsto S_x(t)$  for a given parameter  $x$ , it is useful to consider that this curve is the realization of a Gaussian process, with scale parameter  $C_x$ . It is then insightful to study the black-box  $x \mapsto C_x$ , for instance by means of a sensitivity analysis, or in the aim of finding which inputs  $x$  lead to  $C_x$  smaller than a given threshold. As discussed above, this can indicate output curves with a non-negligible dependence structure. A necessary step to such a study is the estimation of the value of  $C_x$ , given a discretized version of the curve  $t \mapsto S_x(t)$ .

More generally, estimating  $C$  can enable to assess if an observed signal is composed of almost independent components ( $C$  large) or not, and to quantify the level of dependence. We refer to the real data sets studied in Section 5.4 for further discussion.

### 1.3. State of the art on variogram estimation

Nonparametric estimation of the semi-variogram function is a difficult task since the resulting estimator must necessarily lead to a “valid” variogram (conditional negative definiteness property) ([13], p. 93). This requirement usually leads to complicated and computationally involved nonparametric estimators of the variogram [24, 25] that may need huge data sets to be meaningful. A simpler estimator based on the moment method has been proposed in [14, 40] but does not always conduce to a valid variogram. A classical approach to tackle this problem has been proposed in the geostatistics literature [10, 18, 31] and consists in fitting a parametric model of valid semi-variograms to a pointwise nonparametric semi-variogram estimator by minimizing a given distance between the nonparametric estimator and the semi-variograms at a finite number of lags. The reader is referred to [13], Chapter 2 for further details on semi-variogram model fitting and to [34] for least-squares methods.

### 1.4. State of the art on quadratic variations

In order to remedy the drawbacks of the MLE previously mentioned, we focus on an alternative estimation method using quadratic variations based on the observations of the process  $X$  at a triangular array of points  $(t_j)_j$ , where  $j = 1, \dots, n$ . This method is also an alternative to the existing methods mentioned in the previous paragraph. Quadratic variations have been first introduced by Levy in [37] to quantify the oscillations of the Brownian motion. Then a first result on the quadratic variations of a Gaussian non-differentiable process is due to Baxter (see *e.g.* [8], [22], Chap. 5 and [21]) that ensures (under some conditions) the almost sure convergence (as  $n$  tends to infinity) of

$$\sum_{j=1}^n (X(t_j) - X(t_{j-1}))^2, \quad (1.4)$$

where  $t_j = j/n$ , for  $j = 1, \dots, n$  (by convention  $t_0 = 0$  and  $X_0 = 0$ ). A generalization of the previous quadratic variations has been introduced in Guyon and Léon [23]: for a given real function  $H$ , the  $H$ -variation is given by

$$V_{H,n} := \sum_{j=1}^n H \left( \frac{X(t_j) - X(t_{j-1})}{(\text{Var}(X(t_j) - X(t_{j-1})))^{1/2}} \right). \quad (1.5)$$

In [23], technical conditions are assumed and a smoothness parameter  $0 < s < 2$ , similar to the one in (1.2) when  $D = 0$ , is considered. Then the most unexpected result of [23] is that  $(V_{H,n}/n)_n$  has a limiting normal distribution with convergence rate  $n^{1/2}$  when  $0 < s < 3/2$  whereas the limiting distribution is non normal and the convergence rate is reduced to  $n^{2-s}$  when  $3/2 < s < 2$ . Moreover, for statistical purposes, it has been proved by Coeurjolly that quadratic variations are optimal (details and precisions can be found in [11]). In [30], Istas and Lang generalized the results on quadratic variations. They allowed for observation points of the form  $t_j = j\delta_n$  for  $j = 1, \dots, n$ , with  $\delta_n$  depending on  $n$  and tending to 0 as  $n$  goes to infinity. They studied the generalized quadratic variations defined by:

$$V_{a,n} := \sum_{i=1}^{n-1} \left( \sum_{j \in \mathbb{Z}} a_j X(i + j\delta_n) \right)^2, \quad (1.6)$$

where the sequence  $a = (a_j)_j$  has a finite support and some vanishing moments. Then they built estimators of the smoothness parameter and the scale parameter  $C$  and showed that these estimators are almost surely consistent and asymptotically normal. In the more recent work of Lang and Roueff [35], the authors generalized the results of Istas and Lang [30] and Kent and Wood [33] on an increment-based estimator in a semi-parametric framework with different sets of hypothesis. Another generalization for non-stationary Gaussian processes and quadratic variations along curves is done in [1]. See also the studies of [42] and [11].

### 1.5. Contributions of the paper

Now let us present more precisely the framework considered in our paper. We assume that the Gaussian process  $X$  has stationary increments and is observed at times  $t_j = j\delta_n$  for  $j = 1, \dots, n$  with  $\delta_n$  tending to zero. Note that  $t_j$  also depends on  $n$  but we omit this dependence in the notation for simplicity. We will only consider  $\delta_n = n^{-\alpha}$  with  $0 < \alpha \leq 1$  throughout the article. Two cases are then considered:  $\alpha = 1$  ( $\delta_n = 1/n$ , infill asymptotic setting [13], that we call the infill situation throughout) and  $0 < \alpha < 1$  ( $\delta_n \rightarrow 0$  and  $n\delta_n \rightarrow \infty$ , mixed asymptotic setting [13], that we call the mixed situation throughout). The paper is devoted to the estimation of the scale parameter  $C$  from one or several generalized quadratic  $a$ -variations  $V_{a,n}$  defined in (1.6). Calculations show that the expectation of  $V_{a,n}$  is a function of  $C$  so that  $C$  can be estimated by the moment method.

Our study is related to the study of Istas and Lang [30] in which they estimate both the scale parameter  $C$  and the local Hölder index  $H = s + D/2$ . Our main motivation for focusing on the case where the local Hölder index is known is, on the one hand, to provide a simpler method to implement and analyze the estimator, and on the other hand to address more advanced statistical issues, such as efficiency, minimax upper bounds and aggregation of several estimators of  $C$ . In addition, our results hold under milder technical conditions than in [30], and in particular apply to most semi-variogram models commonly used in practice. We also show that a necessary condition in [30], namely the fact that the quantity in (2.3) is non-zero when the variation used has a large enough order, in fact always holds.

We establish asymptotic approximations of the expectation and the variance and a central limit theorem for the quadratic variations under consideration and for the estimators deduced from them. In particular, given a finite number of sequences  $a$ , we prove a joint central limit theorem (see Cor. 3.8). In addition, our method does not require a parametric specification of the drift (see Sect. 3.4); therefore it is more robust than MLE. Also, we obtain minimax upper bounds on the quadratic error of our method.

For a finite discrete sequence  $a$  with zero sum, we define its order as the largest integer  $M$  such that

$$\sum_j a_j j^\ell = 0 \quad \text{for } \ell = 1, \dots, M-1.$$

Roughly speaking  $\sum_j a_j f(j\delta_n)$  is an estimation of the  $M$ th derivative of the function  $f$  at zero. The order of the simplest sequence:  $(-1, 1)$  is  $M = 1$ . Natural questions then arise. What is the optimal sequence  $a$ ? In particular, what is the optimal order? Is it better to use the elementary sequence of order 1:  $(-1, 1)$  or the one of order 2:  $(-1, 2, -1)$ ? For a given order, for example  $M = 1$ , is it better to use the elementary sequence of order 1:  $(-1, 1)$  or a more general one, for example  $(-1, -2, 3)$  or even a sequence based on discrete wavelets? Can we efficiently combine the information of several quadratic  $a$ -variations associated to several sequences? As far as we know, these questions are not addressed yet in the literature. Unfortunately, the asymptotic variances we give in Proposition 3.1 or Theorem 3.7 do not allow either to address theoretically this issue. However, by Corollary 3.8, one may gather the information of different quadratic  $a$ -variations with different orders. In order to validate such a procedure, an important numerical study is performed. The main conclusion is that gathering the information of different quadratic  $a$ -variations with different orders  $M$  yields an aggregated estimator with asymptotic variance closer to the optimal Cramér–Rao bound computed in Section 4.2. This is illustrated in Figure 4. The benefit of combining different  $a$ -variations is also demonstrated in a Monte Carlo study in Section 5.3. We also illustrate (still in a Monte Carlo study) the convergence to the asymptotic distribution considering different models (namely, exponential and Matérn models).

Finally, we show that our suggested quadratic variation estimator can be easily extended to the two-dimensional case and we consider two real data sets in dimension two. When comparing our suggested estimator with MLE, we observe a very significant computational benefit for our estimator.

## 1.6. Organization of the paper

The paper is organized as follows. In Section 2, we detail the framework and present the assumptions on the process. In Section 3, we introduce our quadratic variation estimator and provide its asymptotic properties, together with a wide discussion. In this section, we consider also the addition of a drift and we give minimax upper bounds for the quadratic error of our method. Section 4 is devoted to the analysis of the statistical efficiency of our estimator. In Section 5, we provide the results of the numerical simulation and on the real data sets. A conclusion is provided in Section 6. In Appendix A, we discuss the assumptions made on the process. All the proofs have been postponed to Appendix B.

## 2. GENERAL SETTING AND ASSUMPTIONS

### 2.1. Assumptions on the process

In this paper, we consider a Gaussian process  $(X(t))_{t \in \mathbb{R}}$  which is not necessarily stationary but only has stationary increments. The process is observed at times  $j\delta_n$  for  $j = 1, \dots, n$  with  $\delta_n$  going to 0 as  $n$  goes to infinity. As mentioned in the introduction, we will only consider  $\delta_n = n^{-\alpha}$  with  $0 < \alpha \leq 1$  throughout the article. Two cases are then considered:  $\alpha = 1$  ( $\delta_n = 1/n$ , infill situation) and  $0 < \alpha < 1$  ( $\delta_n \rightarrow 0$  and  $n\delta_n \rightarrow \infty$ , mixed situation). The semi-variogram of  $X$  is defined by

$$V(h) := \frac{1}{2} \mathbb{E} \left[ (X(t+h) - X(t))^2 \right].$$

In the sequel, we denote by  $(Const)$  a positive constant which value may change from one occurrence to another. For the moment, we assume that  $X$  is centered, the case of non-zero expectation will be considered in Section 3.4. Now, we introduce the following assumptions. The form of  $(\mathcal{H}_1)$  and  $(\mathcal{H}_2)$  change following whether we are in the infill situation or in the particular mixed situation ( $\delta_n = n^{-\alpha}$  with  $0 < \alpha < 1$ ).

$(\mathcal{H}_0)$   $V$  is a  $C^\infty$  function on  $\mathbb{R} \setminus \{0\}$ .

**Infill situation:**  $\delta_n = 1/n$ .

$(\mathcal{H}_1)$  The semi-variogram is  $2D$ -times differentiable with  $D \geq 0$  and there exists  $C > 0$  and  $0 < s < 2$  such that for any  $h \in \mathbb{R}$ , we have

$$V^{(2D)}(h) = V^{(2D)}(0) + C(-1)^D |h|^s + r(h), \text{ with } r(h) = o(|h|^s) \text{ as } |h| \rightarrow 0. \quad (2.1)$$

In  $(\mathcal{H}_1)$ , the integer  $D$  is the greatest integer such that  $V$  is  $2D$ -times differentiable everywhere. We recall that, when  $X$  is assumed to be a stationary process, we have  $V(h) = k(0) - k(h)$ . If the covariance function  $k$  belongs to a parametric set of the form  $\{k_\theta; \theta \in \Theta\}$  with  $\Theta \subset \mathbb{R}^p$ , then  $C$  is a deterministic function of the parameter  $\theta$ .

$(\mathcal{H}_2)$

- when  $s < 3/2$ ,

$$\text{for some } \beta > s - 2, \text{ for } |h| \leq 1: |r^{(2)}(h)| \leq (Const) |h|^\beta;$$

- when  $s \geq 3/2$ ,

$$\text{for some } \beta > s - 3, \text{ for } |h| \leq 1: |r^{(3)}(h)| \leq (Const) |h|^\beta.$$

$(\mathcal{H}_3)$  As  $h \rightarrow 0$ ,

$$|r(h)| = o(|h|^{s+1/2}).$$

**Mixed situation:**  $\delta_n = n^{-\alpha}$  with  $0 < \alpha < 1$ .

We must add to  $(\mathcal{H}_1)$ :

$$|r(h)| \leq (Const) |h|^s \quad |h| > 1.$$

The new expression of  $(\mathcal{H}_2)$  is

- when  $s < 3/2$ , there exists  $\beta$  with  $s - 2 < \beta < -1/2$  such that, for all  $h \in \mathbb{R}$ ,

$$|r^{(2)}(h)| \leq (\text{Const}) |h|^\beta;$$

- when  $s \geq 3/2$ , there exists  $\beta$  with  $s - 3 < \beta < -1/2$  such that, for all  $h \in \mathbb{R}$ ,

$$|r^{(3)}(h)| \leq (\text{Const}) |h|^\beta.$$

Here  $(\mathcal{H}_3)$  writes

$$|r(h)| = o(|h|^{s+(1/2\alpha)}),$$

as  $h \rightarrow 0$ .

**Remark 2.1.**

- When  $D > 0$ , the  $D$ th derivative  $X^{(D)}$  in quadratic mean of  $X$  is a Gaussian stationary process with autocovariance function  $k$  given by  $k(h) = (-1)^{D+1} V^{(2D)}(h)$ . This implies that the Hölder exponent of the paths of  $X^{(D)}$  is  $s/2$ . Because  $s < 2$ ,  $D$  is exactly the order of differentiation of the paths of  $X$ .
- If we denote  $H = D + s/2$ ,  $H$  represents the local Hölder index of the process [29].

A discussion of the assumptions  $(\mathcal{H}_2)$  and  $(\mathcal{H}_3)$  is provided in Appendix A.

## 2.2. Examples of processes that satisfy our assumptions

We present a non exhaustive list of examples in dimension one that satisfy our hypotheses. In these examples, we provide their stationary covariance function  $k$ . Recall that this defines  $V$ , with  $V(h) = k(0) - k(h)$ .

- The exponential model:  $k(h) = \exp(-C|h|)$  ( $D = 0$ ,  $s = 1$ ). For this model,  $(\mathcal{H}_0)$  to  $(\mathcal{H}_2)$  always hold and  $(\mathcal{H}_3)$  holds when  $\alpha > 1/2$ , that is when the observation domain does not increase too fast.
- The generalized exponential model:  $k(h) = \exp(-C|h|^s)$ ,  $s \in (0, 2)$  ( $D = 0$ ). For this model,  $(\mathcal{H}_0)$  to  $(\mathcal{H}_2)$  always hold and  $(\mathcal{H}_3)$  holds when  $1/(2\alpha) < s$ . Hence, in the infill situation, we need  $s > 1/2$  and, in the mixed situation, the observation domain needs to increase slowly enough.
- The generalized Slepian model [53]:  $k(h) = (1 - C|h|^s)^+$ ,  $s \in (0, 1]$  ( $D = 0$ ). For this model, the condition  $s \in (0, 1]$  is needed for  $k$  to be a continuous non-negative definite function. In this case,  $(\mathcal{H}_0)$  to  $(\mathcal{H}_3)$  hold in the infill situation and when  $C < 1$ . For  $(\mathcal{H}_0)$ , we remark that, in this case,  $V$  is smooth on  $(0, 1]$  and not on  $(0, \infty)$ , but this is sufficient for all the results to hold.
- The Matérn model:

$$k(h) = \frac{2^{1-\nu}}{\Gamma(\nu)} (\sqrt{2\nu\theta}h)^\nu K_\nu(\sqrt{2\nu\theta}h),$$

where  $\nu > 0$  is the regularity parameter of the process. The function  $K_\nu$  is the modified Bessel function of the second kind of order  $\nu$ . See, e.g., [54] for more details on the model. In that case,  $D = \lfloor \nu \rfloor$  and  $s = 2\nu - 2D$ . Here, it requires tedious computations to express the scale parameter  $C$  as a function of  $\nu$  and  $\theta$ . However, in Section 5.1, we derive the value of  $C$  in two settings ( $\nu = 3/2$  and  $\nu = 5/2$ ). For this model,  $(\mathcal{H}_0)$  to  $(\mathcal{H}_2)$  always hold and  $(\mathcal{H}_3)$  holds when  $s < 2 - 1/(2\alpha)$ . Hence, in the infill situation we need  $s < 3/2$ , and in the mixed situation the observation domain needs to increase slowly enough.

All the previous examples are stationary (and thus have stationary increments). The following one is not stationary.

- The fractional Brownian motion (FBM) process denoted by  $(B_s(t))_{t \in \mathbb{R}}$  and defined by

$$\text{Cov}(B_s(u), B_s(t)) = C(|u|^s + |t|^s - |u - t|^s).$$

A reference on this subject is [12]. This process is classically indexed by its Hurst parameter  $H = s/2$ . Here,  $D = 0$  and  $s \in (0, 2)$  and  $C$  are left arbitrary. We call the FBM defined by  $C = 1$  the standard FBM.

From the previous list of examples, the conditions  $(\mathcal{H}_0)$ ,  $(\mathcal{H}_1)$  and  $(\mathcal{H}_2)$  do not seem restrictive, since they hold for all the examples and for any  $0 < s < 2$  ( $0 < s \leq 1$  for the generalized Slepian model). The condition  $(\mathcal{H}_3)$  is more restrictive, since for some of the examples, it holds only for a restrictive range of  $s$ . For these examples and for the values of  $s$  such that  $(\mathcal{H}_3)$  does not hold, we can still show a rate of convergence that is slower than  $n^{-1/2}$  for the estimator of  $C$ , see Section 3.5 and Appendix A.

We remark that the Gaussian model, or square-exponential, defined by  $k(h) = \sigma^2 e^{-h^2 \theta^2}$ , with  $(\sigma^2, \theta) \in (0, \infty)$ , does not satisfy our assumptions, because it is too regular (*i.e.* it is  $C^\infty$  everywhere and in particular at zero). Indeed, we have  $V(h) = \sigma^2 \theta^2 h^2 + o(h^2)$  as  $h \rightarrow 0$ , and hence this model can be interpreted similarly as in (2.1), with  $\sigma^2 \theta^2$  playing the role of the scale parameter  $C$  and with  $s$  in (2.1) taking the value 2. It is then a relevant statistical problem to estimate  $\sigma^2 \theta^2$ . However, the variation estimator developed here is not suitable for estimating  $\sigma^2 \theta^2$  in the Gaussian model. The main reason for this is that the variation estimator is designed for variograms that have only a finite order of differentiability at zero.

### 2.2.1. Detailed verification of the assumptions with the generalized exponential model

We consider the generalized exponential model, where  $k(h) = \exp(-C|h|^s)$  for some fixed  $s \in (0, 2)$ . Since we have  $V(h) = 1 - \exp(-C|h|^s)$  for  $h \in \mathbb{R}$ ,  $(\mathcal{H}_0)$  is trivially satisfied. Now we show that  $(\mathcal{H}_1)$  holds for  $D = 0$ . Indeed,  $V$  is a continuous function and we have

$$\begin{aligned} V(h) &= 0 + C|h|^s + 1 - C|h|^s - \exp(-C|h|^s) \\ &:= V(0) + C(-1)^0|h|^s + r(h), \end{aligned}$$

with

$$r(h) = 1 - C|h|^s - \exp(-C|h|^s).$$

As  $h \rightarrow 0$ ,  $r(h) = O(|h|^{2s}) = o(|h|^s)$  and thus  $(\mathcal{H}_1)$  holds in the infill situation. Furthermore, as  $|h| \rightarrow \infty$ ,  $r(h) = -C|h|^s + o(1) = O(|h|^s)$  and so  $(\mathcal{H}_1)$  holds also in the mixed situation. Let us now show that  $(\mathcal{H}_2)$  is also satisfied. First consider the case where  $s < 3/2$  and let

$$\beta = \begin{cases} \frac{1}{2}((s-2) + (-1/2)) & \text{if } 2s - 2 > -1/2 \quad (s > 3/4), \\ \frac{1}{2}((s-2) + (2s-2)) & \text{if } 2s - 2 \leq -1/2 \quad (s \leq 3/4). \end{cases}$$

One can check that  $s - 2 < \beta$  and that  $\beta < -1/2$  since  $s - 2 < -1/2$ . Consider the case where  $|h| \leq 1$ . Since

$$r(h) = \sum_{k=2}^{\infty} (-1)^{k+1} \frac{C^k}{k!} |h|^{sk},$$

one has

$$r^{(2)}(h) = |h|^{2s-2} \sum_{k=2}^{\infty} (-1)^{k+1} \frac{C^k}{k!} sk(sk-1) |h|^{s(k-2)} = |h|^{2s-2} g(h)$$

where  $g$  is a bounded function on  $[-1, 1]$ . Hence,  $|r^{(2)}(h)| \leq (Const)|h|^\beta$  since  $2s - 2 \geq \beta$ . Then the case  $s < 3/2$  is complete in the infill situation. Consider now the case where  $|h| \geq 1$ . Simply, one can show that

$$r^{(2)}(h) = -Cs(s - 1)|h|^{s-2} + (-Cs(s - 1)|h|^{s-2} + C^2s^2|h|^{2s-2}) \exp(-C|h|^s).$$

Hence

$$|r^{(2)}(h)| \leq (Const)|h|^{s-2} \leq (Const)|h|^\beta \quad \text{since } \beta > s - 2.$$

The case where  $s \geq 3/2$  can be treated analogously. Finally, it is simple to show that  $(\mathcal{H}_3)$  holds when  $1/(2\alpha) < s$ .

In the rest of the paper, we assume that  $0 < s < 2$ , unless specified otherwise.

### 2.3. Discrete $a$ -differences

Now, we consider a non-zero finite support sequence  $a$  of real numbers with zero sum. Let  $L(a)$  be its length. Since the starting point of the sequence plays no particular role, we will assume when possible that the first non-zero element is  $a_0$ . Hence, the last non-zero element is  $a_{L(a)-1}$ . We define the order  $M(a)$  of the sequence as the first non-zero moment of the sequence  $a$ :

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

To any sequence  $a$ , with length  $L(a)$  and any function  $f$ , we associate the discrete  $a$ -difference of  $f$  defined by

$$\Delta_{a,i}(f) = \sum_{j=0}^{L(a)-1} a_j f((i + j)\delta_n), \quad i = 1, \dots, n', \tag{2.2}$$

where  $n'$  stands for  $n - L(a) + 1$ . As a matter of fact, in the case of the simple quadratic  $a$ -variation given by  $a_0 = 0$  and  $a_1 = -1$ , the operator  $\Delta_a$  is a discrete differentiation operator of order one. More generally,  $\sum_{j=0}^{L(a)-1} a_j f(j\delta_n)$  is an approximation (up to some multiplicative coefficient) of the  $M(a)$ th derivative (when it exists) of the function  $f$  at zero.

We also define  $\Delta_a(X)$  as the Gaussian vector of size  $n'$  with entries  $\Delta_{a,i}(X)$  and  $\Sigma_a$  its variance-covariance matrix.

**Examples – elementary sequences.** The simplest case is the order one elementary sequence  $a^{(1)}$  defined by  $a_0^{(1)} = -1$  and  $a_1^{(1)} = 1$ . We have  $L(a^{(1)}) = 2$ ,  $M(a^{(1)}) = 1$ . More generally, we define the  $k$ th order elementary sequence  $a^{(k)}$  as the sequence with coefficients  $a_j^{(k)} = (-1)^{k-j} \binom{k}{j}$ ,  $j = 0, \dots, k$ . Its length is given by  $L(a^{(k)}) = k + 1$ .

For two sequences  $a$  and  $a'$ , we define their convolution  $b = a * a'$  as the sequence given by  $b_j = \sum_{k+l=j} a_k a'_l$ . In particular, we denote by  $a^{2*}$  the convolution  $a * a$ . Notice that the first non-zero element of  $b$  is not necessarily  $b_0$  but  $b_{-(L(a')-1)}$  as mentioned in the following properties.

**Properties 2.2.** The following properties of convolution of sequences are direct.

- (i) The support of  $a * a'$  (the indices of the non-zero elements) is included in the interval  $[-(L(a') - 1), (L(a) - 1)]$  while its order is  $M(a) + M(a')$ . In particular,  $a^{2*}$  has length  $2L(a) - 1$ , order  $2M(a)$  and is symmetrical.

(ii) The composition of two elementary sequences gives another elementary sequence.

The main result of this section is Proposition 2.6 that is required to quantify the asymptotic behavior of the two first moments of the quadratic  $a$ -variations defined in (3.1) (see Prop. 3.1). In order to prove (2.3), we establish two preliminary tools (Prop. 2.4 and Lemma 2.5). In that view, we need to define the integrated fractional Brownian motion (IFBM). We start from the FBM defined in Section 2.2 which has the following non anticipative representation:

$$B_s(u) = \int_{-\infty}^u f_s(t, u) dW(t),$$

where  $d(t)$  is a white noise defined on the whole real line and

$$f_s(t, u) = (Const) \left( ((u-t)^+)^{(s-1)/2} - ((-t)^+)^{(s-1)/2} \right).$$

For  $m \geq 0$  and  $t \geq 0$ , we define inductively the IFBM by

$$\begin{aligned} B_s^{(-0)}(u) &= B_s(u) \\ B_s^{(-m)}(u) &= \int_0^u B_s^{(-(m-1))}(t) dt. \end{aligned}$$

**Definition 2.3** (Non degenerated property). A process  $Z$  has the ND property if for every  $k > 0$  and every  $t_1 < t_2 < \dots < t_k$  belonging to the domain of definition of  $Z$ , the distribution of  $Z(t_1), \dots, Z(t_k)$  is non degenerated.

We have the following results.

**Proposition 2.4.** *The IFBM has the ND property.*

**Lemma 2.5.** *The variance function of the IFBM satisfies, for all  $m \in \mathbb{N}$ ,*

$$\text{Var}(B_s^{(-m)}(u) - B_s^{(-m)}(v)) = \sum_{i=1}^{N_m} (P^{m,i}(v)h_{m,i}(u) + P^{m,i}(u)h_{m,i}(v)) + (-1)^m \frac{2|u-v|^{s+2m}}{(s+1) \cdots (s+2m)},$$

where  $N_m \in \mathbb{N}$ , for  $i = 1, \dots, N_m$ ,  $P^{m,i}$  is a polynomial of degree less than or equal to  $m$  and  $h_{m,i}$  is some function.

**Proposition 2.6.** *If the sequence  $a$  has order  $M(a) > D$ , then*

$$\sum_j a_j^{2*} |j|^{2D+s} \neq 0 \quad \left( \text{i.e.} \quad (-1)^D \sum_j a_j^{2*} |j|^{2D+s} < 0 \right). \quad (2.3)$$

The case  $D = 0$  follows from [51], Lemma 2.10.8. The whole proof of Proposition 2.6 is postponed to Appendix B.1. Note that (2.3) is stated as an hypothesis in [30].

### 3. QUADRATIC $a$ -VARIATIONS

#### 3.1. Definition

Here, we consider the discrete  $a$ -difference applied to the process  $X$  and we define the quadratic  $a$ -variations by

$$V_{a,n} = \|\Delta_a(X)\|^2 = \sum_{i=1}^{n'} (\Delta_{a,i}(X))^2, \tag{3.1}$$

recalling that  $n' = n - L(a) + 1$ . When no confusion is possible, we will use the shorthand notation  $L$  and  $M$  for  $L(a)$  and  $M(a)$ .

#### 3.2. Main results on quadratic $a$ -variations

The basis of our computations of variances is the identity

$$\mathbb{E}[\Delta_{a,i}(X)\Delta_{a',i'}(X)] = -\Delta_{a*a',i-i'}(V), \tag{3.2}$$

for any sequences  $a$  and  $a'$ . A second main tool is the Taylor expansion with integral remainder (see, for example, (B.5)). So, we introduce another notation. For a sequence  $a$ , a scale  $\delta$ , an order  $q$  and a function  $f$ , we define

$$R(i, \delta, q, f, a) = - \sum_j a_j j^q \int_0^1 \frac{(1-\eta)^{q-1}}{(q-1)!} f((i+j\eta)\delta) d\eta. \tag{3.3}$$

By convention, we let  $R(i, \delta, 0, f, a) = -\Delta_{a,i}(f)$ , where in the expression of  $\Delta_{a,i}(f)$  in (2.2),  $\delta_n$  is replaced by  $\delta$ . Note that  $R(-i, \delta, 2q, |\cdot|^s, a * a') = R(i, \delta, 2q, |\cdot|^s, a' * a)$ . One of our main results is the following.

**Proposition 3.1** (Moments of  $V_{a,n}$ ). *Assume that  $V$  satisfies  $(\mathcal{H}_0)$  and  $(\mathcal{H}_1)$ .*

1) *If we choose a sequence  $a$  such that  $M > D$ , then*

$$\mathbb{E}[V_{a,n}] = nC(-1)^D \delta_n^{2D+s} [R(0, 1, 2D, |\cdot|^s, a^{2*})] (1 + o(1)), \tag{3.4}$$

*as  $n$  tends to infinity. Furthermore,  $(-1)^D R(0, 1, 2D, |\cdot|^s, a^{2*})$  is positive.*

2) *If  $V$  satisfies additionally  $(\mathcal{H}_2)$  and if we choose a sequence  $a$  such that  $M > D + s/2 + 1/4$ , then as  $n$  tends to infinity:*

$$\text{Var}(V_{a,n}) = 2nC^2 \delta_n^{4D+2s} \sum_{i \in \mathbb{Z}} R^2(i, 1, 2D, |\cdot|^s, a^{2*}) (1 + o(1)) \tag{3.5}$$

*and the series above is positive and finite.*

**Remark 3.2.** (i) Notice that (3.4) and (3.5) imply concentration in the sense that

$$\frac{V_{a,n}}{\mathbb{E}[V_{a,n}]} \xrightarrow[n \rightarrow +\infty]{L^2} 1.$$

(ii) In practice, since the parameters  $D$  and  $s$  are known, it suffices to choose  $M$  such that  $M \geq D + 1$  when  $s < 3/2$  and  $M \geq D + 2$  when  $3/2 \leq s < 2$ .

(iii) The expression of the asymptotic variance appears to be complicated. Anyway, in practice, it can be easily approximated. Some explicit examples are given in Section 5.

Following the same lines as in the proof of Proposition 3.1 and using the identities  $(a * a')_j = (a' * a)_{-j}$  and  $R(i, 1, 2D, |\cdot|^s, a * a') = R(-i, 1, 2D, |\cdot|^s, a' * a)$ , one may easily derive the corollary below. The proof is omitted.

**Corollary 3.3** (Covariance of  $V_{a,n}$  and  $V_{a',n}$ ). *Assume that  $V$  satisfies  $(\mathcal{H}_0)$ ,  $(\mathcal{H}_1)$ , and  $(\mathcal{H}_2)$ . Let us consider two sequences  $a$  and  $a'$  so that  $M(a) \wedge M(a') > D + s/2 + 1/4$ . Then, as  $n$  tends to infinity, one has*

$$\text{Cov}(V_{a,n}, V_{a',n}) = 2nC^2\delta_n^{4D+2s} \left[ \sum_{i \in \mathbb{Z}} R^2(i, 1, 2D, |\cdot|^s, a * a') \right] (1 + o(1)). \tag{3.6}$$

**Particular case –  $D = 0$ :**

- (i) We choose  $a$  as the first order elementary sequence ( $a_0 = -1, a_1 = 1$  and  $M = 1$ ). As  $n$  tends to infinity, one has

$$\begin{aligned} \mathbb{E}[V_{a,n}] &= nC\delta_n^s(2 + o(1)); \\ \text{Var}(V_{a,n}) &= 2nC^2\delta_n^{2s} \sum_{i \in \mathbb{Z}} (|i - 1|^s - 2|i|^s + |i + 1|^s)^2 (1 + o(1)), \quad s < 3/2. \end{aligned}$$

- (ii) General sequences. We choose two sequences  $a$  and  $a'$  so that  $M(a) \wedge M(a') > s/2 + 1/4$ . Then, as  $n$  tends to infinity, one has

$$\begin{aligned} \mathbb{E}[V_{a,n}] &= -nC\delta_n^s \left[ \sum_j a_j^{2*} |j|^s \right] (1 + o(1)); \\ \text{Var}(V_{a,n}) &= 2nC^2\delta_n^{2s} \sum_{i \in \mathbb{Z}} \left( \sum_j a_j^{2*} |i + j|^s \right)^2 (1 + o(1)); \\ \text{Cov}(V_{a,n}, V_{a',n}) &= 2nC^2\delta_n^{2s} \left( \sum_{|j| \leq L} a * a'_j |j|^s \right)^2 (1 + o(1)) \\ &\quad + nC^2\delta_n^{2s} \sum_{i \in \mathbb{Z}^*} \left( \left( \sum_{|j| \leq L} a * a'_j |i + j|^s \right)^2 + \left( \sum_{|j| \leq L} a' * a_j |i + j|^s \right)^2 \right) (1 + o(1)). \end{aligned}$$

Now we establish the central limit theorem.

**Theorem 3.4** (Central limit theorem for  $V_{a,n}$ ). *Assume  $(\mathcal{H}_0)$ ,  $(\mathcal{H}_1)$  and  $(\mathcal{H}_2)$  and  $M > D + s/2 + 1/4$ . Then  $V_{a,n}$  is asymptotically normal in the sense that*

$$\frac{V_{a,n} - \mathbb{E}[V_{a,n}]}{\sqrt{\text{Var}(V_{a,n})}} \xrightarrow[n \rightarrow +\infty]{\mathcal{D}} \mathcal{N}(0, 1). \tag{3.7}$$

**Remark 3.5.**

- If  $M = D + 1$ , the condition  $M > D + s/2 + 1/4$  in Proposition 3.1 implies  $s < 3/2$ . However, when  $M = D + 1$  and  $s \geq 3/2$ , it is still possible to compute the variance but the convergence is slower and the central limit theorem does not hold anymore. More precisely, we have the following.

◦ If  $s > 3/2$  and  $M = D + 1$  then, as  $n$  tends to infinity,

$$\text{Var}(V_{a,n}) = (\text{Const}) \times \delta_n^{4D+2s} \times n^{2s-4(M-D)+2} \times (1 + o(1)). \tag{3.8}$$

◦ If  $s = 3/2$  and  $M = D + 1$  then, as  $n$  tends to infinity

$$\text{Var}(V_{a,n}) = (\text{Const}) \times \delta_n^{4D+2s} \times n \log n \times (1 + o(1)). \tag{3.9}$$

We omit the proof. Analogous formula for the covariance of two variations can be derived similarly.

- Since the work of Guyon and León [23], it is a well-known fact that in the simplest case ( $D = 0, L = 2, M = 1$ ) and in the infill situation ( $\delta_n = 1/n, \alpha = 1$ ), the central limit theorem holds true for quadratic variations if and only if  $s < 3/2$ . Hence assumption  $M > D + s/2 + 1/4$  is minimal for  $D = 0$  in the following sense. When  $s < 3/2$ , the assumption becomes  $M \geq 1$ , which is not a restriction as we consider sequences  $a$  with zero sum throughout. When  $s \geq 3/2$ , the assumption becomes  $M \geq 2$ , and we have seen that the central limit theorem does not always hold for  $M = 1$ .

**Corollary 3.6** (Joint central limit theorem). *Assume that  $V$  satisfies  $(\mathcal{H}_0)$ ,  $(\mathcal{H}_1)$  and  $(\mathcal{H}_2)$ . Let  $a^{(1)}, \dots, a^{(k)}$  be  $k$  sequences with order greater than  $D + s/2 + 1/4$ . Assume also that, as  $n \rightarrow \infty$ , the  $k \times k$  matrix with term  $i, j$  equal to*

$$\frac{1}{n\delta_n^{4D+2s}} \text{Cov}(V_{a^{(i)},n}, V_{a^{(j)},n})$$

*converges to an invertible matrix  $\Lambda_\infty$ . Then,  $V_{a^{(1)}, \dots, a^{(k)},n} = (V_{a^{(1)},n}, \dots, V_{a^{(k)},n})^\top$  is asymptotically normal in the sense that*

$$\frac{V_{a^{(1)}, \dots, a^{(k)},n} - \mathbb{E}[V_{a^{(1)}, \dots, a^{(k)},n}]}{n^{1/2}\delta_n^{2D+s}} \xrightarrow[n \rightarrow +\infty]{\mathcal{D}} \mathcal{N}(0, \Lambda_\infty).$$

### 3.3. Estimators of C based on the quadratic a-variations

Guided by the moment method, we define

$$C_{a,n} := \frac{V_{a,n}}{n(-1)^D \delta_n^{2D+s} R(0, 1, 2D, |\cdot|^s, a^{2*})}. \tag{3.10}$$

Then  $C_{a,n}$  is an estimator of  $C$  which is asymptotically unbiased by Proposition 3.1. Now our aim is to establish its asymptotic behavior.

**Theorem 3.7** (Central limit theorem for  $C_{a,n}$ ). *Assume  $(\mathcal{H}_0)$  to  $(\mathcal{H}_3)$  and that  $M(a) > D + s/2 + 1/4$ . Then  $C_{a,n}$  is asymptotically normal. More precisely, we have*

$$\frac{C_{a,n} - C}{\sqrt{\text{Var}(C_{a,n})}} \xrightarrow[n \rightarrow +\infty]{\mathcal{D}} \mathcal{N}(0, 1), \tag{3.11}$$

with  $\text{Var}(C_{a,n}) = (\text{Const})n^{-1}(1 + o(1))$ .

The following corollary is of particular interest: it will give theoretical results when one aggregates the information of different quadratic  $a$ -variations with different orders. As one can see numerically in Section 5.2, such a procedure appears to be really promising and circumvents the problem of the determination of the optimal sequence  $a$ .

**Corollary 3.8.** *Under the assumptions of Theorem 3.7, consider  $k$  sequences  $a^{(1)}, \dots, a^{(k)}$  such that, for  $i = 1, \dots, k$ ,  $M(a^{(i)}) > D + s/2 + 1/4$ . Assume furthermore that the covariance matrix of  $(C_{a^{(i)},n}/\text{Var}(C_{a^{(i)},n})^{1/2})_{i=1,\dots,k}$  converges to an invertible matrix  $\Gamma_\infty$  as  $n \rightarrow \infty$ . Then,  $([C_{a^{(i)},n} - C]/\text{Var}(C_{a^{(i)},n})^{1/2})_{i=1,\dots,k}$  converges in distribution to the  $\mathcal{N}(0, \Gamma_\infty)$  distribution.*

### 3.4. Adding a drift

In this section, we do not assume anymore that the process  $X$  is centered and we set for  $t \geq 0$ ,

$$f(t) = \mathbb{E}[X(t)].$$

We write  $\bar{X}$  the corresponding centered process:  $\bar{X}(t) = X(t) - f(t)$ . As it is always the case in statistical applications, we assume that  $f$  is a  $C^\infty$  function. We emphasize on the fact that our purpose is not proposing an estimation of the mean function.

**Corollary 3.9.** *Assume the same assumptions as in Theorem 3.7, and recall that  $\delta_n = n^{-\alpha}$  for  $\alpha \in (0, 1]$ . Define*

$$K_{M,n}^\alpha = \sup_{t \in [0, n^{1-\alpha}]} |f^{(M)}(t)|.$$

and if we assume in addition that

$$K_{M,n}^\alpha = o\left(n^{-1/4} \delta_n^{D-M+s/2}\right), \quad (3.12)$$

then (3.11) still holds for  $X$ .

Note that in the infill situation ( $\delta_n = 1/n$ ,  $\alpha = 1$ ),  $K_{M,n}^1$  does not depend on  $n$ . Obviously, (3.12) is met if  $f$  is a polynomial up to an appropriate choice of the sequence  $a$  (and  $M$ ). In the infill situation, a sufficient condition for (3.12) is  $M > D + s/2 + 1/4$  which is always true. Moreover, it is worth noticing that we only assume regularity on the  $M$ th derivative of the drift. No parametric assumption on the model is required, unlike in the MLE procedure.

### 3.5. Minimax upper bound on the quadratic error

In this section, we provide a minimax upper bound on the quadratic error of the estimator  $C_{a,n}$ , over a class of possible variograms. Hence, whereas the variogram  $V$  is fixed in the other sections of this paper, we allow here for a varying variogram  $V$ , that is restricted to a function set. We consider here only the infill situation, with  $\delta_n = 1/n$ , for simplicity of exposition.

First we define this function set. We say that a function  $h : [-1, 1] \rightarrow \mathbb{R}$  is even if  $h(t) = h(-t)$  for  $t \in [-1, 1]$ . We say that it is conditionally negative definite if for any  $x_1, \dots, x_n \in [0, 1]$  and for any  $\alpha_1, \dots, \alpha_n \in \mathbb{R}$  such that  $\alpha_1 + \dots + \alpha_n = 0$ ,

$$\sum_{i,j=1}^n \alpha_i \alpha_j h(x_i - x_j) \leq 0.$$

We observe that any variogram  $V$  must be even and conditionally negative definite.

Consider  $D \in \mathbb{N}$ ,  $0 < s < 2$ ,  $0 < C < \infty$ ,  $0 < U < \infty$ ,  $0 < \mu < \infty$ ,  $0 < B < \infty$ . If  $s < 3/2$ , consider  $\beta > s - 2$  and if  $s \geq 3/2$ , consider  $\beta > s - 3$ . The function set to which the variogram  $V$  belongs is

$$\mathcal{C}_{D,s,C,U,\mu,B,\beta} = \left\{ V : [-1, 1] \rightarrow \mathbb{R}, \quad V \text{ is even conditionally negative definite, } V \text{ is } C^\infty \text{ on } [-1, 1] \setminus \{0\}, \right. \tag{3.13}$$

$$V \text{ is } 2D\text{-times differentiable at } 0 \text{ and for } h \in [-1, 1], \tag{3.14}$$

$$\left| V^{(2D)}(h) - V^{(2D)}(0) - C(-1)^D |h|^s \right| \leq U|h|^{s+\mu}, \tag{3.15}$$

$$\text{if } s < 3/2, \left| \left( V^{(2D)}(\cdot) - V^{(2D)}(0) - C(-1)^D |\cdot|^s \right)^{(2)}(h) \right| \leq B|h|^\beta \text{ for } h \in [-1, 1] \setminus \{0\}, \tag{3.16}$$

$$\left. \text{if } s \geq 3/2, \left| \left( V^{(2D)}(\cdot) - V^{(2D)}(0) - C(-1)^D |\cdot|^s \right)^{(3)}(h) \right| \leq B|h|^\beta \text{ for } h \in [-1, 1] \setminus \{0\} \right\}. \tag{3.17}$$

In the definition of  $\mathcal{C}_{D,s,C,U,\mu,B,\beta}$ , in (3.13), the condition that  $V$  is smooth outside of 0 corresponds to  $(\mathcal{H}_0)$ . In (3.15), it is assumed that  $V$  has the same approximation at 0 as in  $(\mathcal{H}_1)$ , where the remainder  $V^{(2D)}(h) - V^{(2D)}(0) - C(-1)^D |h|^s$ , that is written  $r(h)$  in  $(\mathcal{H}_1)$ , is bounded quantitatively by  $U|h|^{s+\mu}$ . Note that we only require  $\mu > 0$ , while in Theorem 3.7 we assume that  $(\mathcal{H}_3)$  holds, which corresponds to  $\mu > 1/2$ . The reason for this difference is that in Theorem 3.7, we prove a central limit theorem for the estimation of  $C$  with a convergence rate of order  $1/n^{1/2}$  and no asymptotic bias and that below, we will show that the minimax upper bound on the quadratic estimation error of  $C$  has order  $1/n$  only when  $\mu \geq 1/2$ . For  $0 < \mu < 1/2$ , the minimax upper bound is of order strictly larger than  $1/n$ . Finally, (3.16) and (3.17) correspond to  $(\mathcal{H}_2)$ .

We are interested in bounding the quadratic error

$$\mathbb{E}_V \left[ (C_{a,n} - C)^2 \right],$$

uniformly over all the possible variograms  $V \in \mathcal{C}_{D,s,C,U,\mu,B,\beta}$ . A uniform bound achieving this is a minimax upper bound. In the above expectation, we insist on the fact that  $C_{a,n} = C_{a,n}(X(1/n), \dots, X(1))$  and that  $X$  is a centered Gaussian process on  $[0, 1]$ , with variogram  $V \in \mathcal{C}_{D,s,C,U,\mu,B,\beta}$ . The next theorem provides the minimax upper bound on the quadratic error.

**Theorem 3.10.** *Consider  $D \in \mathbb{N}$ ,  $0 < s < 2$ ,  $0 < U < \infty$ ,  $0 < \mu < \infty$ ,  $0 < B < \infty$  to be fixed. If  $s < 3/2$ , consider  $\beta > s - 2$  to be fixed and if  $s \geq 3/2$ , consider  $\beta > s - 3$  to be fixed. Let  $0 < C_{\text{sup}} < \infty$  be fixed. Let the sequence  $a$  be fixed, with  $M(a) > D + s/2 + 1/4$ . Then, there exists a constant  $\mathcal{K}$ , depending only on  $D, s, U, \mu, B, \beta, C_{\text{sup}}, a$  such that, for  $n \geq L(a)$ ,*

$$\sup_{C \in (0, C_{\text{sup}}]} \sup_{V \in \mathcal{C}_{D,s,C,U,\mu,B,\beta}} \mathbb{E}_V \left[ (C_{a,n} - C)^2 \right] \leq \mathcal{K} n^{-\min(2\mu, 1)},$$

where in the above expectation the Gaussian process  $X$  has mean zero and variogram  $V$ .

In Theorem 3.10, the rate of decay of the quadratic error is  $n^{-\min(2\mu, 1)}$ . Let us interpret this. The variance of the estimator has order  $1/n$ , see Lemma B.4. The square bias has order  $n^{-\min(2\mu, 2)}$ , see Lemma B.3, and stems from the order of magnitude of the remainder term  $r(h) = V^{(2D)}(h) - V^{(2D)}(0) - C(-1)^D |h|^s$ , which is  $|h|^{s+\mu}$ . Hence, for  $\mu \in (0, 1/2)$ , the remainder term is too large and thus the square bias is too large, preventing the estimator  $C_{a,n}$  from achieving the parametric rate  $1/n$ . For  $\mu \geq 1/2$  the square bias becomes small enough

and the parametric rate  $1/n$  is achieved. As discussed above, the central limit theorem for  $C_{a,n}$  in Theorem 3.7, which provides an asymptotic variance of order  $1/n$  with no asymptotic bias, relies on  $(\mathcal{H}_3)$ , which corresponds to  $\mu > 1/2$  in the class  $\mathcal{C}_{D,s,C,U,\mu,B,\beta}$ . Note that our minimax upper bound is also uniform over the value of  $C$ , which is only upper bounded by  $C_{\text{sup}}$ . The expression of the constant  $\mathcal{K}$  can be found in the proof of Theorem 3.10.

**Remark 3.11.** The minimax upper bound in Theorem 3.10 provides an estimation error of order  $n^{-\min(\mu, 1/2)}$ . This order coincides with the one obtained in [35], where  $s$  rather than  $C$  is estimated and where a remainder term there, similar as  $r(h)$ , is also assumed of order  $|h|^{s+\mu}$ , similarly as in (3.15). We remark that [35] provides minimax upper and lower bounds, while we only consider upper bounds. In future work, it would be interesting to see if the lower bound techniques of [35], based in particular on the Van Trees inequality, could be extended to our setting.

**Remark 3.12.** It is important to observe here that the estimator  $C_{a,n}$  does not depend on the smoothness indices  $\mu$  and  $\beta$ , although the rate (partially) depends on  $\mu$  in Theorem 3.10. This rate could also possibly depend on  $\beta$  in other situations. In other semi-parametric statistical settings, the situation could be different, in that the estimators could depend on smoothness parameters (playing a similar role as  $\mu$  and  $\beta$  here). For instance, the optimal bandwidth in kernel smoothing typically depends on the underlying smoothness.

### 3.6. Elements of comparison with existing procedures

#### 3.6.1. Quadratic variations versus MLE

In this section, we compare our methodology to the very popular MLE method. For details on the MLE procedure, the reader is referred to, e.g. [44, 52].

#### 3.6.2. Model flexibility

As mentioned in the introduction, the MLE methodology is a parametric method and requires the covariance function to belong to a parametric family of the form  $\{k_\theta, \theta \in \Theta\}$ . In the procedure proposed in this paper, it is only assumed that the semi-variogram satisfies the conditions given in Section 2.1, and that  $D$  and  $s$  are known. In this latter case, the suggested variation estimator is feasible, while the MLE is not defined.

#### 3.6.3. Adding a drift

In order to use the MLE estimator, it is necessary to assume that the mean function of the process is a linear combination of known parametric functions:

$$f(t) = \sum_{i=1}^q \beta_i f_i(t),$$

with known  $f_1, \dots, f_q$  and where  $\beta_1, \dots, \beta_q$  need to be estimated. Our method is less restrictive and more robust. Indeed, we only assume the regularity of the  $M$ th derivative of the mean function in assumption (3.12). It does not require parametric assumptions neither on the semi-variogram nor the mean function. Furthermore, no estimation of the mean function is necessary.

#### 3.6.4. Computational cost

The cost of our method is only  $O(n)$  (the method only requires the computation of a sum) while the cost of the MLE procedure is known to be  $O(n^3)$ .

### 3.6.5. Practical issues

In some real data frameworks, it may occur that, for the MLE, the iterative procedure for likelihood optimization diverges, as can be seen in Section 5.4. Such a dead end cannot be possible with our procedure.

### 3.6.6. Quadratic variations versus other methods

### 3.6.7. Least-square estimators

In [34], the authors propose a Least Square Estimator (LSE). More precisely, given a parametric family of semi-variograms  $\{V_\theta, \theta \in \Theta\}$ , the Ordinary Least Square Estimation (OLSE) consists in minimizing in the parameter  $\theta$  the quantity

$$\sum_{h \in I} (V_n(h) - V_\theta(h))^2,$$

where  $I$  is a set of lags and  $V_n$  is a non-parametric estimator of the semi-variogram. Several variants as the Weighted Least Square Estimation and the General Least Square Estimation have been then introduced. Then the authors of [34] provide necessary and sufficient conditions for these estimators to be asymptotically efficient and they show that when the number of lags used to define the estimators is chosen to be equal to the number of variogram parameters to be estimated, the ordinary least squares estimator, the weighted least squares and the generalized least squares estimators are all asymptotically efficient. Similarly as for the MLE, least square estimators require a parametric family of variograms, while quadratic variation estimators do not.

### 3.6.8. Cross validation estimators

Cross validation estimators [3–5, 62] are based on minimizing scores based on the leave one out prediction errors, with respect to covariance parameters  $\theta$ , when a parametric family of semi-variograms  $\{V_\theta, \theta \in \Theta\}$  is considered. Hence, as the MLE, they require a parametric family of variograms. Furthermore, as the MLE, the computation cost is in  $O(n^3)$ , while this cost is  $O(n)$  for quadratic variation estimators.

### 3.6.9. Composite likelihood

Maximum composite likelihood estimators follow the principle of the MLE, with the aim of reducing its computational cost [39, 41, 56, 59, 60]. In this aim, they consist in optimizing, over  $\theta$ , the sum, over  $i = 1, \dots, n$  of the conditional likelihoods of the observation  $X_i = X(t_i)$ , given a small number of observations which observation locations are close to  $t_i$  when a parametric family of semi-variograms  $\{V_\theta, \theta \in \Theta\}$  is considered. The computation cost of an evaluation of this sum of conditional likelihood is  $O(n)$ , in contrast to  $O(n^3)$  for the full likelihood. Nevertheless, the composite likelihood estimation requires to perform a numerical optimization, while our suggested estimator does not. Furthermore, a parametric family of variograms is required for the composite likelihood but not for our estimator. Finally, [6] recently showed that the composite likelihood estimator has rate of convergence only  $n^s$  when  $D = 0$  and  $0 < s < 1/2$  in (1.2). Hence, the rate of convergence of quadratic variation estimators ( $n^{1/2}$ ) is larger in this case.

### 3.6.10. Already known results on quadratic $\alpha$ -variations

In [30], Istas and Lang consider a Gaussian process with stationary increments in the infill case and assume (1.2) as in our paper. Then they establish the asymptotic behavior of  $V_{\alpha,n}$  under more restrictive hypothesis of regularity on  $V$  than ours (in particular on  $r$  and on  $\delta_n$ ). Then they propose an estimation of both the local Hölder index  $H = D + s/2$  and the scale parameter  $C$ , based on quadratic  $\alpha$ -variations and study their asymptotic behavior. The expression of the estimation of  $C$  is much more complex than simply stems from the moment method. More precisely, they consider  $I$  sequences  $(a^j)_{j=1, \dots, I}$  with length  $p_j$  and the vector  $U$  of length  $I$  whose coordinate  $j$  is given by  $V_{a^j,n}$ . Noticing that the vector  $U/n$  converges to the product  $AZ$  where  $A$  is

a  $I \times p$  matrix derived from the sequences  $a$  with  $p = \max_j p_j$  and  $Z$  is the vector of  $(C(-1)^D(j\delta_n)^{2h})_{j=1,\dots,p}$ , they estimate  $Z$  by  $\hat{Z} = (A^\top A)^{-1}A^\top U$  and derive their estimators of  $H$  and  $C$  from  $\hat{Z}$ .

As explained in the introduction in Section 1, Lang and Roueff in [35] generalize the results of Istas and Lang in [30] and [33]. They consider the infill situation and use quadratic  $a$ -variations to estimate both the scale parameter  $C$  and smoothness parameter  $s$  under a similar hypothesis as in (1.2). Furthermore, they assume three types of regularity assumptions on  $V$ : Hölder regularity of the derivatives at the origin, Besov regularity and global Hölder regularity. Nevertheless, estimating both  $C$  and  $s$  leads to a more complex estimator of  $C$  and to proofs significantly different and more complicated.

To summarize, our contributions, additionally to the existing references [30, 33, 35], is to provide an estimation method for  $C$  which definition, implementation and asymptotic analysis are simpler. As a result, we need fewer technical assumptions. In fact, our assumptions can be easily shown to hold in many classical examples. This also enables us to study the aggregation of quadratic variation estimators from different sequences, see Section 5.2. Furthermore, our assumptions are easier to check than in [30] and less technical than in [35]. In addition, our proofs for the asymptotic analysis of quadratic variations are simpler than in the existing references.

#### 4. EFFICIENCY OF OUR ESTIMATION PROCEDURE

In this section, in order to decrease the asymptotic variance, we propose a procedure to combine several quadratic  $a$ -variations leading to aggregated estimators. Then our goal is to evaluate the quality of these proposed estimators. In that view, we compare their asymptotic variance with the theoretical Cramér–Rao bound in some particular cases in which this bound can be explicitly computed.

##### 4.1. Aggregation of estimators

Now in order to improve the estimation procedure, we suggest to aggregate a finite number of estimators:

$$\sum_{j=1}^k \lambda_j C_{a^{(j)},n}$$

based on  $k$  different sequences  $a^{(1)}, \dots, a^{(k)}$  with weights  $\lambda_1, \dots, \lambda_k$ . Ideally, one should provide an adaptive statistical procedure to choose the optimal number  $k^*$  of sequences, the optimal sequences and the optimal weights  $\lambda^*$ . Such a task is beyond the scope of this paper. Nevertheless, in this section, we consider a given number  $k$  of given sequences  $a^{(1)}, \dots, a^{(k)}$  leading to the estimators  $C_{a^{(1)},n}, \dots, C_{a^{(k)},n}$  defined by (3.10). Then we provide the optimal weights  $\lambda^*$ . Using [36] or [7], one can establish the following lemma.

**Lemma 4.1.** *We assume that for  $j = 1, \dots, k$ , the conditions of Corollary 3.8 are met. Let  $R$  be the  $k \times k$  asymptotic variance-covariance matrix of the vector of length  $k$  whose elements are given by  $(n^{1/2}/C)C_{a^{(j)},n}$ ,  $j = 1, \dots, k$ . Then for any  $\lambda_1, \dots, \lambda_k$  with sum one,*

$$(n^{1/2}/C) \left( \sum_{j=1}^k \lambda_j C_{a^{(j)},n} - C \right) \xrightarrow[n \rightarrow +\infty]{\mathcal{D}} \mathcal{N}(0, \lambda^T R \lambda).$$

Let  $\mathbf{1}_k$  be the “all one” column vector of size  $k$  and define

$$\lambda^* = \frac{R^{-1} \mathbf{1}_k}{\mathbf{1}_k^T R^{-1} \mathbf{1}_k}.$$

One has  $\sum_{j=1}^k \lambda_j^* = 1$  and

$$\lambda^{*T} R \lambda^* \leq \lambda^T R \lambda.$$

Let us now explain how the (optimally) aggregated estimator is computed in practice. The weights in Lemma 4.1 are functions of the matrix  $R$  which is the limit as  $n \rightarrow \infty$  of the covariance matrix  $R^{(n)}$  defined by, for  $i, j = 1, \dots, k$ ,

$$R_{i,j}^{(n)} = \frac{n}{C^2} \text{Cov} (C_{a^{(i)},n}, C_{a^{(j)},n}).$$

From Corollary 3.3 and (3.10), we have  $R^{(n)} - \tilde{R}^{(n)} \rightarrow 0$  as  $n \rightarrow \infty$ , where  $\tilde{R}^{(n)}$  is defined by, for  $i, j = 1, \dots, k$ ,

$$\tilde{R}_{i,j}^{(n)} = \frac{2 \sum_{\ell \in \mathbb{Z}} R^2(\ell, 1, 2D, |\cdot|^s, a^{(i)} * a^{(j)})}{R(0, 1, 2D, |\cdot|^s, (a^{(i)})^{2*}) R(0, 1, 2D, |\cdot|^s, (a^{(j)})^{2*})}. \tag{4.1}$$

Hence, in practice we can compute the weights given by

$$\tilde{\lambda}^{(n)} = \frac{(\tilde{R}^{(n)})^{-1} \mathbf{1}_k}{\mathbf{1}_k^T (\tilde{R}^{(n)})^{-1} \mathbf{1}_k}, \tag{4.2}$$

which only depend on  $a^{(1)}, \dots, a^{(k)}$ ,  $D$  and  $s$ , which are known quantities.

As will be shown with simulations in Section 5, the aggregated estimator considerably improves each of the original estimators  $C_{a^{(1)},n}, \dots, C_{a^{(k)},n}$ .

### 4.2. Cramér–Rao bound

To validate the aggregation procedure, we want to compare the obtained asymptotic variance with the theoretical Cramér–Rao bound. In that view, we compute in the following section the Cramér–Rao bound in two particular cases.

We consider a family  $Y_C$  ( $C \in \mathbb{R}^+$ ) of centered Gaussian processes. Let  $R_C$  be the  $(n-1) \times (n-1)$  variance-covariance matrix defined by

$$(R_C)_{i,j} = \text{Cov} (Y_C(i\delta_n) - Y_C((i-1)\delta_n), Y_C(j\delta_n) - Y_C((j-1)\delta_n)).$$

Assume that  $C \mapsto R_C$  is twice differentiable and  $R_C$  is invertible for all  $C \in \mathbb{R}^+$ . Then, let

$$I_C = \frac{1}{2} \text{Tr} \left( R_C^{-1} \left( \frac{\partial}{\partial C} R_C \right) R_C^{-1} \left( \frac{\partial}{\partial C} R_C \right) \right) \tag{4.3}$$

be the Fisher information. The quantity  $1/I_C$  is the Cramér–Rao lower bound for estimating  $C$  based on

$$\{Y_C(i\delta_n) - Y_C((i-1)\delta_n)\}_{i=2, \dots, n}$$

(see for instance [4, 15]). Now we give two examples of families of processes for which we can compute the Cramér–Rao lower bound explicitly. The first example is obtained from the IFBM defined in Section 2.2.

**Lemma 4.2.** *Let  $0 < s < 2$  and let  $X$  be equal to  $\sqrt{C}B_s^{(-D)}$  where  $B_s^{(-D)}$  is the IFBM. Then  $Y_C = X^{(D)}$  is a FBM whose semi-variogram  $V_C$  is given by*

$$V_C(h) = \frac{1}{2} \mathbb{E} \left[ (Y_C(t+h) - Y_C(t))^2 \right] = C|h|^s. \quad (4.4)$$

Hence in this case, we have  $1/I_C = 2C^2/(n-1)$ .

Now we consider a second example given by the generalized Slepian process defined in Section 2.2.

Let  $s \leq 1$  and  $Y_C$  with stationary autocovariance function  $k_C$  defined by

$$k_C(h) = (1 - (C/2)|h|^s)^+, \quad \text{for any } h \in \mathbb{R}. \quad (4.5)$$

This function is convex on  $\mathbb{R}$  and it follows from Pólya's theorem [43] that  $k_C$  is a valid autocovariance function. We thus easily obtain the following lemma whose proof is omitted.

**Lemma 4.3.** *Let  $X$  be the integration  $D$  times of  $Y_C$  defined via (4.5). Then, in the infill situation ( $\delta_n = 1/n$ ,  $\alpha = 1$ ) and for  $C < 2$ , the semi-variogram of  $Y_C$  is given by (4.4) and by consequence  $1/I_C = 2C^2/(n-1)$ .*

## 5. NUMERICAL RESULTS

In this section, we first study to which extent the asymptotic results of Proposition 3.1 and Theorem 3.7 are representative of the finite sample behavior of quadratic  $a$ -variations estimators. Then, we study the asymptotic variances of these estimators provided by Proposition 3.1 and that of the aggregated  $a$ -variations estimators of Section 4.1.

### 5.1. Simulation study of the convergence to the asymptotic distribution

We carry out a Monte Carlo study of the quadratic  $a$ -variations estimators in three different cases. In each of the three cases, we simulate  $N = 10,000$  realizations of a Gaussian process on  $[0, 1]$  with zero mean function and stationary autocovariance function  $k$ . In the case  $D = 0$ , we let  $k(h) = \exp(-C|h|)$ . Hence  $(\mathcal{H}_1)$  holds with  $D = 0$  and  $s = 1$ . In the case  $D = 1$ , we use the Matérn 3/2 autocovariance [48] :

$$k(h) = \left( 1 + \sqrt{3} \frac{|h|}{\theta} \right) e^{-\sqrt{3} \frac{|h|}{\theta}}.$$

One can show, by developing  $k$  into power series, that  $(\mathcal{H}_1)$  holds with  $D = 1$ ,  $s = 1$  and  $C = 6\sqrt{3}/\theta^3$ . Finally, in the case  $D = 2$ , we use the Matérn 5/2 autocovariance function:

$$k(h) = \left( 1 + \sqrt{5} \frac{|h|}{\theta} + \frac{5|h|^2}{3\theta^2} \right) e^{-\sqrt{5} \frac{|h|}{\theta}}.$$

Also  $(\mathcal{H}_1)$  holds true with  $D = 2$ ,  $s = 1$  and  $C = 200\sqrt{5}/3\theta^5$ .

In each of the three cases, we set  $C = 3$ . For  $n = 50$ ,  $n = 100$  and  $n = 200$ , we observe each generated process at  $n$  equispaced observation points on  $[0, 1]$  and compute the quadratic  $a$ -variations estimator  $C_{a,n}$  of Section 3.3. When  $D = i$ ,  $i = 0, 1, 2$ , we choose  $a$  to be the elementary sequence of order  $i + 1$ .

In Figure 2, we display the histograms of the 10,000 estimated values of  $C$  for the nine configurations of  $D$  and  $n$ . We also display the corresponding asymptotic Gaussian probability density functions provided by Proposition 3.1 and Theorem 3.7. We observe that there are few differences between the histograms and limit

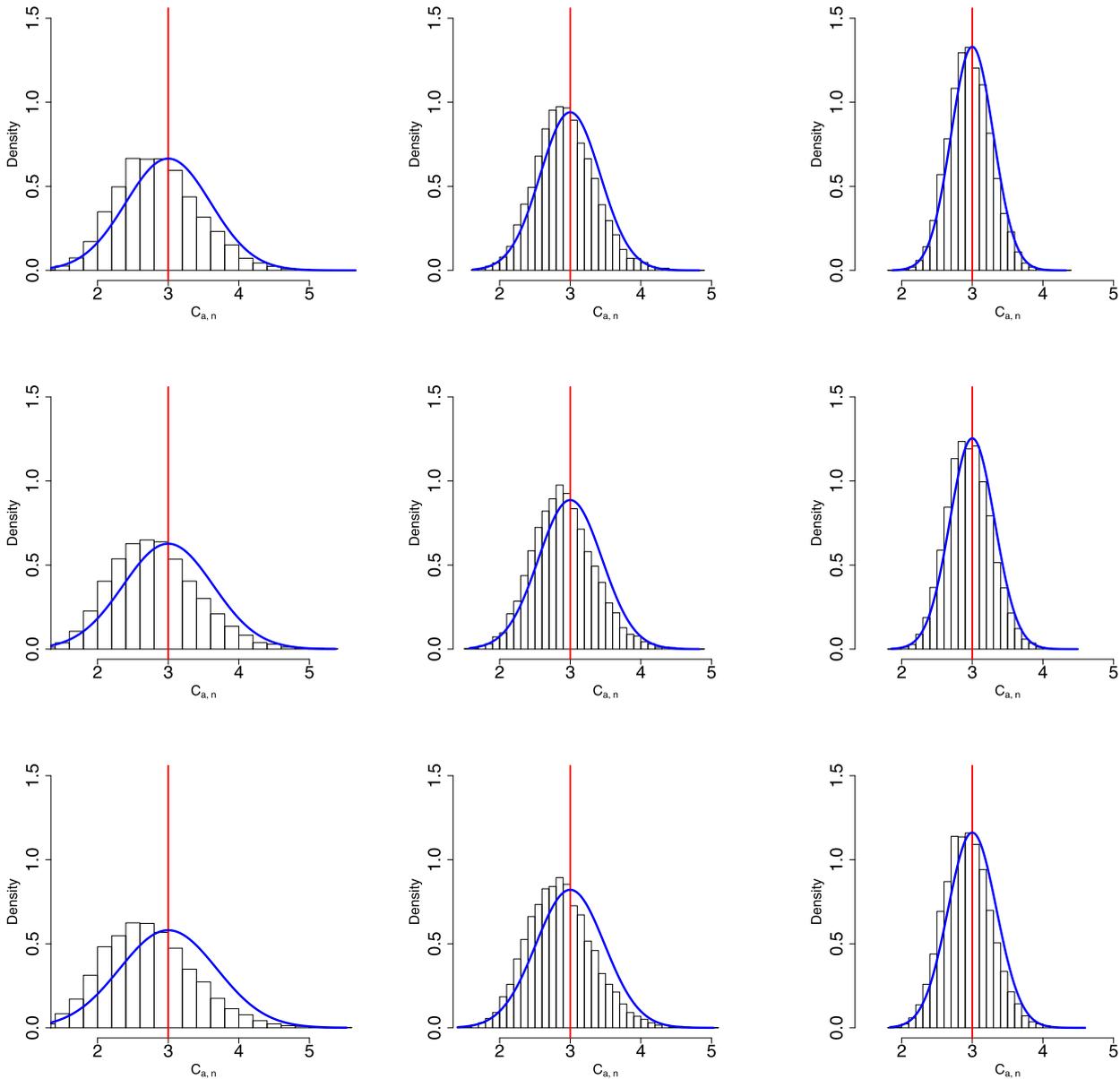


FIGURE 2. Comparison of the finite sample distribution of  $C_{a,n}$  (histograms) with the asymptotic Gaussian distribution provided by Proposition 3.1 and Theorem 3.7 (probability density function in blue line). The vertical red line denotes the true value of  $C = 3$ . From left to right,  $n = 50, 100, 200$ . From top to bottom,  $D = 0, 1, 2$ .

probability density functions between the cases ( $D = 0, 1, 2$ ). In these three cases, the limiting Gaussian distribution is already a reasonable approximation when  $n = 50$ . This approximation then improves for  $n = 100$  and becomes very accurate when  $n = 200$ . Naturally, we can also see the estimators' variances decrease as  $n$  increases. Finally, the figures suggest that the discrepancies between the finite sample and asymptotic distributions are slightly more pronounced with respect to the difference in mean values than to the difference in variances. As already pointed out, these discrepancies are mild in all the configurations.

## 5.2. Analysis of the asymptotic distributions

Now we consider the normalized asymptotic variance of  $C_{a,n}$  obtained from (3.5) in Proposition 3.1. We consider the infill situation ( $\delta_n = 1/n$ ,  $\alpha = 1$ ) and we let

$$\tilde{v}_{a,s} = \frac{2 \sum_{i \in \mathbb{Z}} R^2(i, 1, 2D, |\cdot|^s, a^{2*})}{R^2(0, 1, 2D, |\cdot|^s, a^{2*})}, \quad (5.1)$$

so that  $(n^{1/2}/C)(C_{a,n} - C)$  converges to a  $\mathcal{N}(0, \tilde{v}_{a,s})$  distribution as  $n \rightarrow \infty$ , where  $\tilde{v}_{a,s}$  does not depend on  $C$  (nor on  $n$ ).

First, we consider the case  $D = 0$  and we plot  $\tilde{v}_{a,s}$  as a function of  $s$  for various sequences  $a$  in Figure 3. The considered sequences are the following:

- the elementary sequence of order 1:  $a^{(1)}$  given by  $(-1, 1)$ ;
- the elementary sequence of order 2:  $a^{(2)}$  given by  $(1, -2, 1)$ ;
- the elementary sequence of order 3:  $a^{(3)}$  given by  $(-1, 3, -3, 1)$ ;
- the elementary sequence of order 4:  $a^{(4)}$  given by  $(1, -4, 6, -4, 1)$ ;
- a sequence of order 1 and with length 3:  $a^{(5)}$  given by  $(-1, -2, 3)$ ;
- a Daubechies wavelet sequence [17] with  $M = 2$  as in [30]:  $a^{(6)}$  given by  $(-0.1830127, -0.3169873, 1.1830127, -0.6830127)$ ;
- a second Daubechies wavelet sequence with  $M = 3$ :  $a^{(7)}$  given by  $(0.0498175, 0.12083221, -0.19093442, -0.650365, 1.14111692, -0.47046721)$ .

From Figure 3, we can draw several conclusions. First, the results of Section 4.2 suggest that 2 is a plausible lower bound for  $\tilde{v}_{a,s}$ . We shall call the value 2 the Cramér–Rao lower bound. Indeed, we observe numerically that  $\tilde{v}_{a,s} \geq 2$  for all the  $s$  and  $a$  considered here. Then we observe that, for any value of  $s$ , there is one of the  $\tilde{v}_{a,s}$  which is close to 2 (below 2.5). This suggests that quadratic variations can be approximately as efficient as maximum likelihood, for appropriate choices of the sequence  $a$ . We observe that, for  $s = 1$ , the elementary sequence of order 1 ( $a_0 = -1$ ,  $a_1 = 1$ ) satisfies  $\tilde{v}_{a,s} = 2$ . This is natural since for  $s = 1$ , this quadratic  $a$ -variations estimator coincides with the maximum likelihood estimator, when the observations stem from the standard Brownian motion. Except from this case  $s = 1$ , we could not find other quadratic  $a$ -variations estimators reaching exactly the Cramér–Rao lower bound 2 for other values of  $s$ .

Second, we observe that the normalized asymptotic variance  $\tilde{v}_{a,s}$  blows up for the two sequences  $a$  satisfying  $M = 1$  when  $s$  reaches 1.5. This comes from Remark 3.5: the variance of the quadratic  $a$ -variations estimators with  $M = 1$  is of order larger than  $1/n$  when  $s \geq 1.5$ . Consequently, we plot  $\tilde{v}_{a,s}$  for  $0.1 \leq s \leq 1.4$  for these two sequences. For the other sequences satisfying  $M \geq 2$ , we plot  $\tilde{v}_{a,s}$  for  $0.1 \leq s \leq 1.9$ .

Third, it is difficult to extract clear conclusions about the choice of the sequence: for  $s$  smaller than, say, 1.2 the two sequences with order  $M = 1$  have the smallest asymptotic variance. Similarly, the elementary sequence of order 2 has a smaller normalized variance than that of order 3 for all values of  $s$ . Also, the Daubechies sequence of order 2 has a smaller normalized variance than that of order 3 for all values of  $s$ . Hence, a conclusion of the study in Figure 3 is the following. When there is a sequence of a certain order for which the corresponding estimator reaches the rate  $1/n$  for the variance, there is usually no benefit in using a sequence of larger order. Finally, the Daubechies sequences appear to yield smaller asymptotic variances than the elementary sequences (the orders being equal). The sequence of order 1 given by  $(a_0, a_1, a_2) = (-1, -2, 3)$  can yield a smaller or larger asymptotic variance than the elementary sequence of order 1, depending on the value of  $s$ . For two sequences of the same order  $M$ , it seems nevertheless challenging to explain why one of the two provides a smaller asymptotic variance.

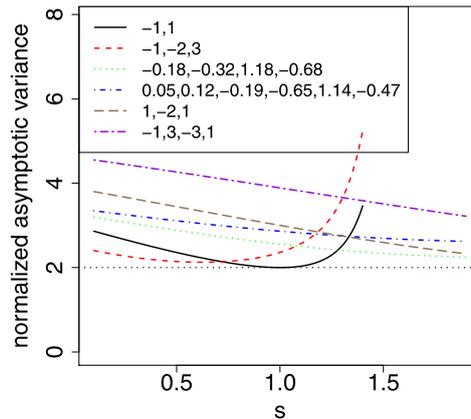


FIGURE 3. Case  $D = 0$ . Plot of the normalized asymptotic variance  $\tilde{v}_{a,s}$  of the quadratic  $a$ -variations estimator, as a function of  $s$ , for various sequences  $a$ . The legend shows the values  $a_0, \dots, a_l$  of these sequences (rounded to two digits). From top to bottom in the legend, the sequences are the elementary sequence of order 1, the sequence  $(-1, -2, 3)$  which has order 1, the Daubechies sequences of order 2 and 3 and the elementary sequences of orders 2 and 3. The horizontal line corresponds to the Cramér–Rao lower bound 2.

Now, we consider aggregated estimators, as presented in Section 4.1. A clear motivation for considering aggregation is that, in Figure 3, the smallest asymptotic variance  $\tilde{v}_{a,s}$  corresponds to different sequences  $a$ , depending on the values of  $s$ .

In Figure 4 left, we consider the case  $D = 0$  and we use four sequences:  $a^{(1)}$ ,  $a^{(5)}$ ,  $a^{(2)}$  and  $a^{(6)}$ . We plot their corresponding asymptotic variances  $\tilde{v}_{a^{(i)},s}$  as a function of  $s$ , for  $0.1 \leq s \leq 1.4$  as well as the variance of their aggregation. This variance of aggregation is computed from the covariance matrix given by (4.1) and the weights given by (4.2). Notice that the aggregated estimator, which asymptotic variance is plotted in Figure 4, only depends on known quantities and thus can be readily computed in practice.

It is then clear that aggregation drastically improves each of the four original estimators. The asymptotic variance of the aggregated estimator is very close to the Cramér–Rao lower bound 2 for all the values of  $s$ . In Figure 4 right, we perform the same analysis but with sequences of order larger than 1. The four considered sequences are now  $a^{(6)}$ ,  $a^{(2)}$ ,  $a^{(3)}$ , and  $a^{(4)}$ . The value of  $s$  varies from 0.1 to 1.9. Again, the aggregation is clearly the best.

Eventually, Figures 5 and 6 explore the case  $D = 1$ . Conclusions are similar.

### 5.3. Finite sample study of the aggregation of estimators

By definition, the asymptotic variances of aggregated estimators are smaller than the asymptotic variances obtained from their individual sequences. Numerically, in Section 5.2, we have seen that the benefit of the aggregation in terms of asymptotic variance can be relatively substantial.

In this section, we address the finite sample situation and quantify the benefit of the aggregation procedure. We study the infill situation ( $\delta_n = 1/n$ ) with  $C = 3$  and the sample sizes  $n = 50, 100, 200, 400$ . First, we consider the case  $D = 0$  and the generalized exponential covariance function  $k(h) = \exp(-C|h|^s)$  with  $s = 1.3$  and we make use of the elementary sequence of order 1:  $(-1, 1)$  and of order 2:  $(1, -2, 1)$ . As in Section 5.1, we simulate 10,000 observation vectors and compute 10,000 estimates of  $C$  obtained from the two estimators  $C_{a,n}$  corresponding to the two sequences. We also compute 10,000 estimates of  $C$  obtained from the aggregation of these two estimators, with the weights given by (4.2), which, we recall, only depend on known quantities. Then we compute the three empirical mean square errors, for estimating  $C$ , of the three estimators, from the two sequences and the aggregation. Second, we compute the same empirical mean square errors, but this time with

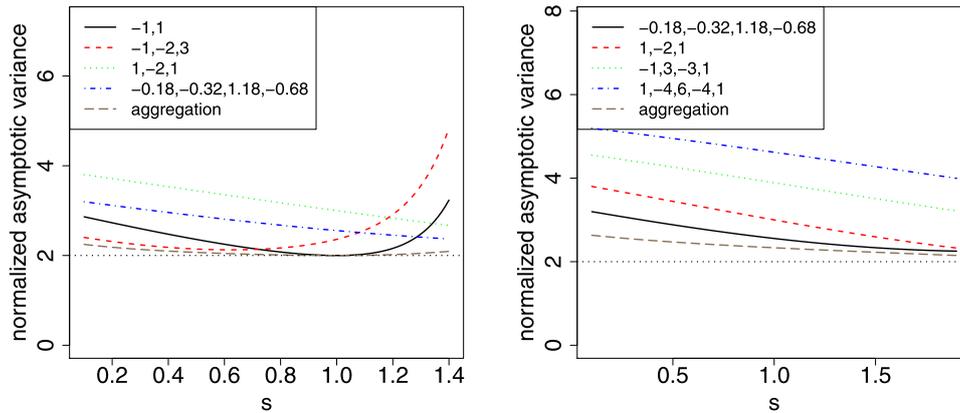


FIGURE 4. Case  $D = 0$ . Plot of the normalized asymptotic variance  $\tilde{v}_{a,s}$  of the quadratic  $a$ -variations estimator, as a function of  $s$ , for various sequences  $a$  and for their aggregation. On the left, including the order one elementary sequence, on the right without. The horizontal line corresponds to the Cramér–Rao lower bound 2.

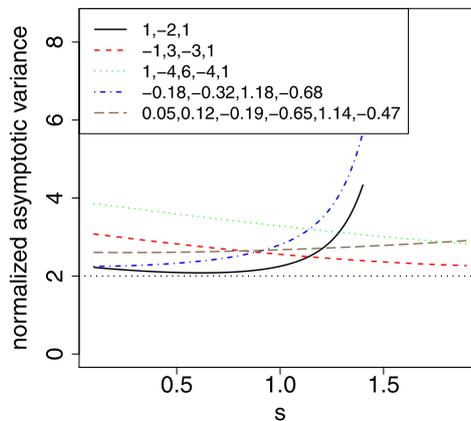


FIGURE 5. Same setting as in Figure 3 but for  $D = 1$ . From top to bottom in the legend, the sequences are the elementary sequences of order 2, 3 and 4 and the Daubechies sequences of order 2 and 3.

$D = 1, s = 1$ , the Matérn 3/2 covariance function and the elementary sequence of order 3:  $(-1, 3, -3, 1)$  and of order 4:  $(1, -4, 6, -4, 1)$ .

The empirical mean square errors are displayed in Table 1. We see that the mean square errors obtained from the aggregation procedure are smaller than those obtained for the two individual sequences, for  $D = 0$  and  $D = 1$  and for all sample sizes  $n = 50$  to  $n = 400$ . Hence, although the aggregation procedure is based on an asymptotic approximation of the variances and covariances, it is beneficial in finite sample, already for the sample size  $n = 50$ .

### 5.4. Real data examples

In this section, we consider real data of spatially distributed processes in dimension two. In this setting, we extend the estimation procedure based on the quadratic  $a$ -variations that is then compared to the MLE procedure.

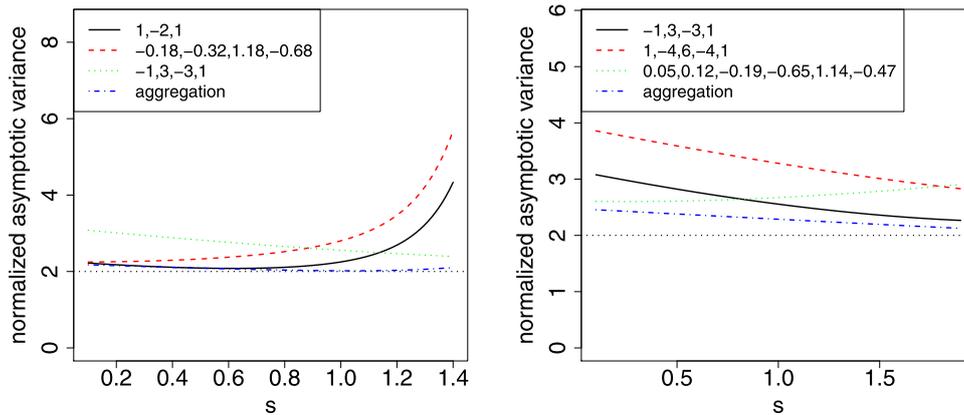


FIGURE 6. Same setting as in Figure 4 but for  $D = 1$ . On the left, from top to bottom in the legend, the sequences are the elementary sequence of order 2, the Daubechies sequence of order 2 and the elementary sequence of order 3. On the right, from top to bottom in the legend, the sequences are the elementary sequences of orders 3 and 4 and the Daubechies sequence of order 3.

TABLE 1. Top: empirical mean square errors (MSE) of estimators of  $C$  obtained from the sequences  $(-1, 1)$  and  $(1, -2, 1)$  and from their aggregation, for  $D = 0$ ,  $s = 1.3$ , and  $n = 50, 100, 200, 400$ . Bottom: same but with the two sequences  $(-1, 3, -3, 1)$  and  $(1, -4, 6, -4, 1)$  and with  $D = 1$ ,  $s = 1$ .

| $D$ | $s$ | $n$ | MSE $(-1, 1)$        | MSE $(1, -2, 1)$        | MSE aggregation |
|-----|-----|-----|----------------------|-------------------------|-----------------|
| 0   | 1.3 | 50  | 0.3990               | 0.4820                  | 0.3600          |
| 0   | 1.3 | 100 | 0.2040               | 0.2430                  | 0.1800          |
| 0   | 1.3 | 200 | 0.1050               | 0.1230                  | 0.0912          |
| 0   | 1.3 | 400 | 0.0553               | 0.0617                  | 0.047           |
| $D$ | $s$ | $n$ | MSE $(-1, 3, -3, 1)$ | MSE $(1, -4, 6, -4, 1)$ | MSE aggregation |
| 1   | 1   | 50  | 0.4570               | 0.5880                  | 0.4380          |
| 1   | 1   | 100 | 0.2270               | 0.2930                  | 0.2170          |
| 1   | 1   | 200 | 0.1140               | 0.1470                  | 0.1090          |
| 1   | 1   | 400 | 0.0576               | 0.0733                  | 0.0553          |

5.4.1. A moderate size data set

We compare two methods of estimation of the autocovariance function of a separable Gaussian model on a real data set of atomic force spectroscopy<sup>1</sup>. The data consist of observations taken on a grid of step  $1/15$  on  $[0, 1]^2$ , so they consist of 256 points of the form

$$X(i/15, j/15) \quad i = 0, \dots, 15, \quad j = 0, \dots, 15.$$

The first method is maximum likelihood estimation in a Kriging model, obtained from the function `km` of the R toolbox `DiceKriging` [48]. For this method, the mean and autocovariance functions are assumed to be  $\mathbb{E}(X(i/15, j/15)) = \mu$  and

$$\text{Cov}(X(i/15, j/15), X(i'/15, j'/15)) = \sigma^2 e^{-\theta_1|i-i'|/15} e^{-\theta_2|j-j'|/15}. \tag{5.2}$$

<sup>1</sup>Personal communication from C. Gales and J. M. Senard.

The parameters  $\mu, \sigma^2, \theta_1, \theta_2$  are estimated by maximum likelihood.

The second method assumes the same autocovariance model (5.2) and consists in the following steps.

- (1) Estimate  $\sigma^2$  by the sum of squares

$$\hat{\sigma}^2 = \frac{1}{256} \sum_{i,j=0}^{15} (X(i/15, j/15) - m)^2$$

with  $m = \sum_{i,j=0}^{15} X(i/15, j/15)$ .

- (2) For each column  $j$  of  $[X(i/15, j/15)]_{i,j=0,\dots,15}$ , the vector of 16 observations obey our model with  $s = 1$  and  $C_1 = \sigma^2\theta_1$ . Hence, we can estimate  $C_1$  by  $\hat{C}_{1,j}$  with the estimator (3.10), with the elementary sequence of order 1. We thus obtain an estimate  $\hat{C}_1$  by averaging the  $\hat{C}_{1,j}$  for  $j = 0, \dots, 15$ .
- (3) We perform the same analysis row by row to obtain an estimate  $\hat{C}_2$ .
- (4) For  $i = 1, 2$ ,  $\theta_i$  is estimated by  $\hat{\theta}_i = \hat{C}_i / \hat{\sigma}^2$ .

The first method, based on maximum likelihood, provides infinite values for  $\theta_1$  and  $\theta_2$ , so that it considers the 256 observed values as completely spatially independent. On the other hand, the second method provides the values  $\hat{\theta}_1 = 14.72$  and  $\hat{\theta}_2 = 15.73$ . This corresponds to a correlation of approximately  $1/e \approx 0.36$  between direct neighbors on the grid. Hence, the second method, based on our suggested quadratic variation estimator, is able to detect a weak correlation, unlike the maximum likelihood estimator. Actually, graphical observations of the data tend to suggest a small spatial dependence which is in accordance with the biological nature of the data.

We remark that, in general, the second method, based on the quadratic variation estimator, cannot provide infinite values of  $\theta_1$  and  $\theta_2$ , because the numerator in (3.10) is always finite and the denominator in (3.10) is always strictly positive. On the other hand, the first method, based on maximum likelihood, may provide infinite values of  $\hat{\theta}_1$  and  $\hat{\theta}_2$ , as is the case here. For instance, the likelihood can be maximum in the limit  $\theta_1 = +\infty$ , or, even if this is not the case, the iterative procedure for optimizing the likelihood can provide a sequence of values of  $\theta_1$  diverging to infinity.

**Remark 5.1.** In (5.2), we choose a separable model for its classicality (as done, for instance, in the R toolbox `DiceKriging` [48]). This separable model is characterized by having the basis vectors  $(1, 0)$  and  $(0, 1)$  as the two important directions. More precisely, in (5.2), the correlation function of  $i \mapsto X(i/15, j/15)$  is characterized by  $\theta_1$  only and the correlation function of  $j \mapsto X(i/15, j/15)$  is characterized by  $\theta_2$  only. It is then natural for the variation estimator to make, separately, first an analysis with the second component of  $X$  fixed (column by column) in order to estimate  $\theta_1$  and second an analysis with the first component of  $X$  fixed (row by row) in order to estimate  $\theta_2$ .

In addition, consider that (5.2) holds with 15 replaced by  $N - 1$  with  $N \rightarrow \infty$ . Then the consistency of the quadratic variation estimators  $\hat{C}_1$  and  $\hat{C}_2$  established in Lemma 5.2 below suggests that our analysis row by row and column by column is appropriate to the model (5.2).

Remark that the analysis row by row and column by column is also appropriate to other models where the directions  $(1, 0)$  and  $(0, 1)$  play a special role; for instance, one could consider the geometric anisotropic counterpart of (5.2), *i.e.*

$$\text{Cov}(X(i/15, j/15), X(i'/15, j'/15)) = \sigma^2 \exp\left\{-\left(\theta_1^2|i - i'|^2/15^2 + \theta_2^2|j - j'|^2/15^2\right)^{1/2}\right\}.$$

To conclude, our method of estimation row by row and column by column is perfectly coherent with the separable model (5.2), which is widely used in Kriging applications. Of course this method does not permit

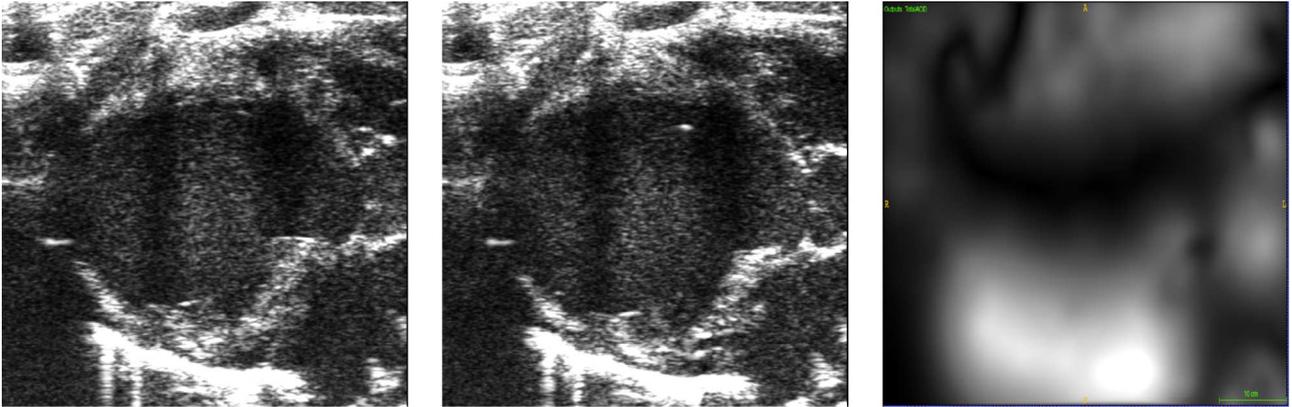


FIGURE 7. For the data set of Section 5.4.2: the two images to be registered (*left and middle*) and the field of deformation amplitude (*right*). On the right, light colors indicate large deformation amplitudes and dark colors indicate small deformation amplitudes.

to estimate anisotropy in other directions as in [46] or topothesy in the sense of [45] that correspond to more complex models including variations of the Hölder exponent.

Using the results of Section 3.5, we can prove straightforwardly the consistency of the quadratic variation estimators  $\hat{C}_1$  and  $\hat{C}_2$ , given by the second method with steps (2) and (3).

**Lemma 5.2.** *Assume that a centered Gaussian process  $X$  is observed at  $\{(i/(N-1), j/(N-1)); i, j = 0, \dots, N-1\}$  with a two-dimensional autocovariance function  $k$  having a separable expression given by*

$$k(x, y) = k_1(x)k_2(y),$$

where, for  $i = 1, 2$ , the variogram associated to  $k_i$  lies in  $\mathcal{C}_{D_i, s_i, C_i, U_i, \mu_i, B_i, \beta_i}$  introduced in Theorem 3.10, with the same setting and assumptions as in this theorem. Then the quadratic variation estimators  $\hat{C}_1$  and  $\hat{C}_2$ , described in the steps (2) and (3) above, are consistent; in other words,  $\hat{C}_i$  converges in probability to  $C_i$ , for  $i = 1, 2$ .

#### 5.4.2. A large size data set

The second data set consists in a two-dimensional field of deformation amplitude, corresponding to the registration of two real images. The deformation field is obtained from the software presented in [47]. Figure 7 displays the two images to be registered and the deformation field.

After a subsampling of the field of deformation amplitude, the data consist of observations taken on a rectangular grid of steps  $1/56$  and  $1/59$  on  $[0, 1]^2$ , so they consist of 3420 points of the form

$$X(i/56, j/59) \quad i = 0, \dots, 56, \quad j = 0, \dots, 59.$$

With these data, we consider the same autocovariance model as in Section 5.4.1. We estimate the parameters  $\theta_1$  and  $\theta_2$  from the same two methods as in Section 5.4.1. The first method provides  $\hat{\theta}_1 = 0.6547$  and  $\hat{\theta}_2 = 0.8770$  and takes about 22 minutes on a personal computer. The second method provides  $\hat{\theta}_1 = 0.107$  and  $\hat{\theta}_2 = 0.607$  and takes about 0.05 seconds on a personal computer. Hence, our suggested quadratic variation estimator provides a very significant computational benefit.

Both estimators conclude that the spatial correlation is more important along the  $x$ -axis than along the  $y$ -axis, which is graphically confirmed in Figure 7. As in Section 5.4.1, the maximum likelihood estimator provides less correlation than the quadratic variation estimator.

Finally, if the field of deformation is considered with no preliminary subsampling, its size is  $400 \times 600$ . In this case, the MLE cannot be directly implemented while the quadratic variation estimator can be.

## 6. CONCLUSION

We have provided an in-depth analysis of the estimation of the scale parameter of a one-dimensional Gaussian process by quadratic variations. Indeed, the knowledge of this scale parameter is essential when studying a Gaussian process, as it enables to quantify its dependence structure, or to test independence.

We have addressed a semi-parametric setting, where no parametric family of variograms needs to be assumed to contain the unknown variogram. We have suggested an estimator, based on previous references, which numerical implementation is straightforward. Our theoretical analysis follows the principles of previous references, but is significantly simpler and holds under mild and simple to check technical assumptions. Based on this theoretical analysis, we have been able to tackle more advanced statistical topics, such as the aggregation of estimators based on different sequences, in the aim of improving the statistical efficiency.

## APPENDIX A. DISCUSSION OF THE ASSUMPTIONS $(\mathcal{H}_2)$ AND $(\mathcal{H}_3)$

- The assumption  $(\mathcal{H}_2)$  may be difficult to interpret at first sight. This assumption is necessary in the proof of Proposition 3.1, where it is shown that the variance of the quadratic  $a$ -variation  $V_{a,n}$  in (3.1) is asymptotically equivalent to  $n\delta_n^{4D+2s}C^2\mathcal{V}_{D,s,a}$ , with an explicit constant  $\mathcal{V}_{D,s,a}$  not depending on  $r$ . In order to show that  $r$  has no impact on the asymptotic approximation of the variance, we need a condition that makes  $r$  negligible (compared to  $C(-1)^D|\cdot|^s$ ), in some sense. This is the role of  $(\mathcal{H}_2)$ , where we use the second or third derivative of  $r$ , from which we can use Taylor's theorem to a term involving  $r$  and exploit the order of the sequence  $a$ . We refer to (B.8) in the proof of Proposition 3.1 for more details. The condition  $\beta > s - 2$  or  $\beta > s - 3$  is exactly what we need to obtain, eventually, the rate  $n\delta_n^{4D+2s}$  and the explicit constant. The condition  $\beta < -1/2$ , that is additionally assumed in the mixed situation, is needed to guarantee that a series involving  $r$ , see also (B.8) in the proof of Proposition 3.1, is finite. If  $(\mathcal{H}_2)$  was not assumed, then the proof of Proposition 3.1 given here would not hold any longer. In addition, if  $(\mathcal{H}_2)$  was not assumed, one may ask whether there exist settings, for which the other assumptions hold, and where the variance of  $V_{a,n}$  has a larger order of magnitude than  $n\delta_n^{4D+2s}$ . We leave the question of an asymptotic analysis not involving  $(\mathcal{H}_2)$  open to future work. We remark that the condition  $(\mathcal{H}_2)$ , although technical, does not appear to be too restrictive, since it holds for all the common examples listed in Section 2.2, with no restriction on  $s$ . We finally remark that in [30], Theorem 1 (i), a bound on a derivative of  $r$  is also assumed, similarly as in  $(\mathcal{H}_2)$ .
- We have discussed above that the interpretation of  $(\mathcal{H}_2)$  is that, in the expression  $V^{(2D)}(h) - V^{(2D)}(0) = C(-1)^D|h|^s + r(h)$ , the remainder term  $r(h)$  is negligible compared to the main term  $C(-1)^D|h|^s$ . In the proof of Proposition 3.1, the useful notion of negligibly for us is the order of magnitude of the  $d$ th derivative as  $h \rightarrow 0$  (where  $d = 2$  or  $d = 3$ , depending on  $s$ ). The  $d$ th derivative of the main term is of order  $|h|^{s-d}$  and we assume that the  $d$ th derivative of the remainder term is of order  $|h|^\beta$  with  $\beta > s - d$ . Hence, we indeed assume that the  $d$ th derivative of the remainder term is negligible compared to that of the main term, as  $h \rightarrow 0$ . Furthermore, in the infill case, where we only assume  $\beta > s - d$ , we make a minimal assumption to guarantee this.
- Similarly as for  $(\mathcal{H}_2)$ , the interpretation of  $(\mathcal{H}_3)$  is that  $r$  is negligible. While  $(\mathcal{H}_2)$  is used to study the variance of the quadratic  $a$ -variation  $V_{a,n}$ ,  $(\mathcal{H}_3)$  is used to study its expectation. More precisely, an estimator  $C_{a,n}$  of  $C$  is constructed, by applying the moment method to  $V_{a,n}$ , see (3.10). The assumption  $(\mathcal{H}_3)$  then enables to show that  $\mathbb{E}(C_{a,n}) - C = o(n^{-1/2})$ , which allows for a central limit theorem with rate  $n^{-1/2}$  and no asymptotic bias, see Theorem 3.7. To prove the  $o(n^{-1/2})$ , we precisely need  $|r(h)| = o(|h|^{s+(1/2\alpha)})$  as  $h \rightarrow 0$ , see (B.11) in the proof of Theorem 3.7. We remark that when  $(\mathcal{H}_3)$  is not assumed and that only  $|r(h)| = o(|h|^s)$  is assumed, then we only have  $\mathbb{E}(C_{a,n}) - C = o(1)$ , from Proposition 3.1. In the infill situation, if we assume  $|r(h)| = o(|h|^{s+\mu})$  as  $h \rightarrow 0$ , with  $0 < \mu < 1/2$ , which is weaker than

$(\mathcal{H}_3)$ , then we obtain  $\mathbb{E}(C_{a,n}) - C = o(n^{-\mu})$ , see Section 3.5, that is a smaller rate of convergence of the expectation compared to  $(\mathcal{H}_3)$ .

### APPENDIX B. PROOFS

#### B.1 Proof of the results of Section 2.3

*Proof of Proposition 2.4.* By the stochastic Fubini theorem,

$$\begin{aligned} B_s^{(-m)}(u_1) &= \int_0^{u_1} du_2 \cdots \int_0^{u_m} du_{m+1} \int_{-\infty}^{u_{m+1}} d(t) f_s(t, u_{m+1}) \\ &= \int_{-\infty}^{u_1} d(t) \int_t^{u_1} du_2 \cdots \int_t^{u_m} du_{m+1} f_s(t, u_{m+1}) \\ &=: \int_0^{u_1} g_{m,s}(u_1, t) d(t). \end{aligned}$$

The positiveness of  $f_s(t, u)$  for  $u > 0$  implies that of  $g_{m,s}(t, u)$ . As a consequence, for  $0 < t_1 < \dots < t_k$ ,  $B_s^{(-m)}(t_k)$  includes a non-zero component:

$$\int_{t_{k-1}}^{t_k} g_{m,s}(u, t) d(t),$$

which is independent of  $(B_s^{(-m)}(t_1), \dots, B_s^{(-m)}(t_{k-1}))$  implying that  $B_s^{(-m)}(t_k)$  is not collinear to this set of variables. By induction, this implies in turn that  $B_s^{(-m)}(t_1), \dots, B_s^{(-m)}(t_k)$  are not collinear. □

*Proof of Lemma 2.5.* For  $m = 0$ , we have

$$\text{Var}(B_s^{(-0)}(u) - B_s^{(-0)}(v)) = 2|u - v|^s$$

so that the lemma holds with the convention  $(s + 1) \dots (s + 0) = 1$ . Thus we prove it by induction on  $m$  and assume that it holds for  $m \in \mathbb{N}$ . We have, with  $K^{(-r)}(u, v) = \mathbb{E}[B_s^{(-r)}(u)B_s^{(-r)}(v)]$ , for  $r \in \mathbb{N}$ ,

$$\begin{aligned} K^{(-m)}(u, v) &= \frac{1}{2} \left( \text{Var}(B_s^{(-m)}(u) - B_s^{(-m)}(0)) + \text{Var}(B_s^{(-m)}(v) - B_s^{(-m)}(0)) - \text{Var}(B_s^{(-m)}(u) - B_s^{(-m)}(v)) \right) \\ &= \psi(u) + \psi(v) - \frac{1}{2} \sum_{i=1}^{N_m} P^{m,i}(v) h_{m,i}(u) - \frac{1}{2} \sum_{i=1}^{N_m} P^{m,i}(u) h_{m,i}(v) - \frac{1}{2} (-1)^m \frac{2|u - v|^{s+2m}}{(s + 1) \dots (s + 2m)}, \end{aligned}$$

where  $\psi$  is some function. Since we have  $K^{(-(m+1))}(u, v) = \int_0^u \int_0^v K^{(-m)}(x, y) dx dy$ ,

$$\begin{aligned} K^{(-(m+1))}(u, v) &= \sum_{i=1}^{\tilde{N}_{m+1}} \tilde{P}^{m+1,i}(v) \tilde{h}_{m+1,i}(u) + \sum_{i=1}^{\tilde{N}_{m+1}} \tilde{P}^{m+1,i}(u) \tilde{h}_{m+1,i}(v) \\ &\quad + (-1)^{m+1} \frac{1}{(s + 1) \dots (s + 2m)} \int_0^v \left( \int_0^u |x - y|^{s+2m} dx \right) dy, \end{aligned} \tag{B.1}$$

where  $\tilde{N}_{m+1} \in \mathbb{N}$ , where for  $i = 1, \dots, \tilde{N}_{m+1}$ ,  $\tilde{P}^{m+1,i}$  is a polynomial of degree less than or equal to  $m + 1$  and  $\tilde{h}_{m+1,i}$  is some function. For  $v \leq u$ , we have

$$\begin{aligned} \int_0^v \left( \int_0^u |y-x|^{s+2m} dx \right) dy &= \int_0^v \left( \int_0^y (y-x)^{s+2m} dx + \int_y^u (x-y)^{s+2m} dx \right) dy \\ &= \int_0^v \left( \frac{y^{s+2m+1}}{2m+1} + \frac{(u-y)^{s+2m+1}}{2m+1} \right) dy \\ &= \frac{v^{s+2m+2}}{(2m+1)(2m+2)} - \frac{(u-v)^{s+2m+2}}{(2m+1)(2m+2)} + \frac{u^{s+2m+2}}{(2m+1)(2m+2)}. \end{aligned}$$

By symmetry, we obtain, for  $u, v \in \mathbb{N}$ ,

$$\int_0^u \left( \int_0^v |x-y|^{s+2m} dx \right) dy = \frac{u^{s+2m+2}}{(2m+1)(2m+2)} + \frac{v^{s+2m+2}}{(2m+1)(2m+2)} - \frac{|u-v|^{s+2m+2}}{(2m+1)(2m+2)}. \tag{B.2}$$

Hence, from the relation

$$\text{Var}(B_s^{(-m+1)}(u) - B_s^{(-m+1)}(v)) = K^{(-m+1)}(v, v) + K^{(-m+1)}(u, u) - 2K^{(-m+1)}(v, u),$$

(B.1), and (B.2), we conclude the proof of the lemma. □

*Proof of Proposition 2.6.* Using Lemma 2.5 (with  $m = D$ ) and the vanishing moments of  $a$  of order less than or equal to  $D$ , we have

$$\begin{aligned} \sum_{k,l} a_k a_l |k-l|^{2D+s} &= (Const)(-1)^D \sum_{k,l} a_k a_l \text{Var}(B_s^{(-D)}(k) - B_s^{(-D)}(l)) \\ &= (Const)(-1)^{D+1} \text{Var} \left( \sum_k a_k B_s^{(-D)}(k) \right). \end{aligned}$$

We conclude using the ND property of the IFBM stated in Proposition 2.4. □

### B.2 Preliminary results

**Lemma B.1.** *Let  $Z = (X, Y)$  be a centered Gaussian vector of dimension 2 then*

$$\text{Cov}(X^2, Y^2) = 2\text{Cov}^2(X, Y).$$

*Proof of Lemma B.1.* This Lemma is a consequence of the so called Mehler formula [2]. Its proof is immediate using the cumulant method. □

**Lemma B.2.** *Assume that  $V$  satisfies  $(\mathcal{H}_0)$ ,  $(\mathcal{H}_1)$  and  $(\mathcal{H}_2)$ . One has, when  $M > D + s + 1/4$ ,*

$$\max_{i=1, \dots, n'} \left( \sum_{i'=1, \dots, n'} |\Sigma_a(i, i')| \right) = o\left(\text{Var}(V_{a,n})^{1/2}\right).$$

*Proof of Lemma B.2.* Using the stationary increments of the process, one has

$$\max_{i=1, \dots, n'} \left( \sum_{i'=1, \dots, n'} |\Sigma_a(i, i')| \right) \leq 2 \sum_{i=0}^{n'-1} |\Sigma_a(1, 1+i)|. \tag{B.3}$$

Recall that

$$\Sigma_a(1, 1 + i) = \text{Cov}(\Delta_{a,1}(X), \Delta_{a,1+i}(X)) = -\Delta_{a^{2^*},i}(V) = \delta_n^{2D} R(i, \delta_n, 2D, V^{(2D)}, a^{2^*}).$$

We have seen in the proof of Proposition 3.1 ((B.7) and (B.9)) that for  $i$  sufficiently large

$$R(i, \delta_n, 2D, V^{(2D)}, a^{2^*}) \leq (\text{Const})(\delta_n^s i^{s-2(M-D)} + \delta_n^{d+\beta} i^\beta).$$

Thus, the sum in (B.3) is bounded by

$$(\text{Const})\delta_n^{2D+s}(n^{s-2(M-D)+1} + 1) + (\text{Const})\delta_n^{2D+d+\beta}(n^{1+\beta} + 1).$$

On the other hand, we have proved also in the proof of Proposition 3.1 that

$$\text{Var}(V_{a,n})^{1/2} = (\text{Const})n^{1/2}\delta_n^{2D+s}(1 + o(1))$$

giving the result. Thus, one has to check that

$$\delta_n^{2D+s}n^{s-2(M-D)+1}, \quad \delta_n^{2D+s}, \quad \delta_n^{2D+d+\beta}n^{1+\beta}, \quad \text{and} \quad \delta_n^{2D+d+\beta}$$

are  $o(n^{1/2}\delta_n^{2D+s})$  which is true by the assumptions made. We skip the details. □

### B.3 Proof of the main results

*Proof of Proposition 3.1.* 1) By definition of  $V_{a,n}$  in (3.1) and identity (3.2), we get

$$\mathbb{E}[V_{a,n}] = n' \mathbb{E}[\Delta_{a,i}(X)^2] = -n' \Delta_{a^{2^*},0}(V) = -n' \sum_j a_j^{2^*} V(j\delta_n). \tag{B.4}$$

Recall that  $n' = n - L + 1$  is the size of the vector  $\Delta_a(X)$ . In all the proof,  $j$  is assumed to vary from  $-L + 1$  to  $L - 1$ . We use a Taylor expansion of  $V((i + j)\delta_n)$  at  $(i\delta_n)$  and of order  $q \leq 2D$ :

$$V((i + j)\delta_n) = V(i\delta_n) + \dots + \frac{(j\delta_n)^{q-1}}{(q-1)!} V^{(q-1)}(i\delta_n) + (j\delta_n)^q \int_0^1 \frac{(1-\eta)^{q-1}}{(q-1)!} V^{(q)}((i + j\eta)\delta_n) d\eta. \tag{B.5}$$

Note that this expression is “telescopic” in the sense that if  $q < q' \leq 2D$ ,

$$\begin{aligned} & (j\delta_n)^q \int_0^1 \frac{(1-\eta)^{q-1}}{(q-1)!} V^{(q)}((i + j\eta)\delta_n) d\eta \\ &= \frac{(j\delta_n)^q}{(q)!} V^{(q)}(i\delta_n) + \dots + \frac{(j\delta_n)^{q'-1}}{(q'-1)!} V^{(q'-1)}(i\delta_n) + (j\delta_n)^{q'} \int_0^1 \frac{(1-\eta)^{q'-1}}{(q'-1)!} V^{(q')}((i + j\eta)\delta_n) d\eta. \end{aligned} \tag{B.6}$$

Combining (B.5) (with  $i = 0$  and  $q = 2D$ ), the vanishing moments of the sequence  $a^{2^*}$  and  $(\mathcal{H}_1)$  yields:

$$\begin{aligned} \mathbb{E}[V_{a,n}] &= n' \delta_n^{2D} R(0, \delta_n, 2D, V^{(2D)}, a^{2^*}) \\ &= n' C(-1)^D \delta_n^{2D+s} R(0, 1, 2D, |\cdot|^s, a^{2^*}) + n' \delta_n^{2D} R(0, \delta_n, 2D, r, a^{2^*}). \end{aligned}$$

The first term is non-zero by (2.3) in Proposition 2.6 and a dominated convergence argument together with  $(\mathcal{H}_1)$  shows that the last term is  $o(\delta_n^{2D+s})$  giving (3.4).

2) Using Lemma B.1, (B.5) with  $q = 2D$ , the fact that  $D \leq M$ , and the vanishing moments of the sequence  $a^{2*}$ , we obtain

$$\begin{aligned} \text{Var}(V_{a,n}) &= 2 \sum_{i,i'=1}^{n'} \text{Cov}^2(\Delta_{a,i}(X), \Delta_{a,i'}(X)) = 2 \sum_{i,i'=1}^{n'} (-\Delta_{a^{2*},i-i'}(V))^2 = 2 \sum_{i=-n'+1}^{n'-1} (n' - |i|) \Delta_{a^{2*},i}(V)^2 \\ &= 2\delta_n^{4D} \sum_{i=-n'+1}^{n'-1} (n' - |i|) R^2(i, \delta_n, 2D, V^{(2D)}, a^{2*}) \\ &= 2\delta_n^{4D} \sum_{i=-n'+1}^{n'-1} (n' - |i|) (C(-1)^D \delta_n^s R(i, 1, 2D, |\cdot|^s, a^{2*}) + R(i, \delta_n, 2D, r, a^{2*}))^2. \\ &=: A_n + B_n + C_n, \end{aligned}$$

where  $B_n$  comes from the double product.

(i) We show that  $A_n$  converges. Indeed,

$$A_n = 2C^2 \delta_n^{4D} \sum_{i=-n'+1}^{n'-1} (n' - |i|) \delta_n^{2s} R^2(i, 1, 2D, |\cdot|^s, a^{2*}) = 2C^2 n' \delta_n^{4D+2s} \sum_{i \in \mathbb{Z}} f_n(i),$$

with

$$f_n(i) := \frac{n' - |i|}{n'} R^2(i, 1, 2D, |\cdot|^s, a^{2*}) \mathbb{1}_{|i| \leq n'-1}.$$

Since  $f_n(i) \uparrow R^2(i, 1, 2D, |\cdot|^s, a^{2*})$  for fixed  $i$  and  $n'$  going to infinity, it suffices to study the series

$$\sum_{i \in \mathbb{Z}} R^2(i, 1, 2D, |\cdot|^s, a^{2*}).$$

Using (B.6), with  $q' = 2M$ ,  $|\cdot|^s$  instead of  $V^{(2D)}$  and  $\delta_n = 1$ , and using the vanishing moments of the sequence  $a^{2*}$ , we get, for  $i$  large enough so that  $i$  and  $i + j$  always have the same sign in the sum below,

$$R(i, 1, 2D, |\cdot|^s, a^{2*}) = R(i, 1, 2M, g, a^{2*}) = - \sum_j a_j^{2*} j^{2M} \int_0^1 \frac{(1 - \eta)^{2M-1}}{(2M - 1)!} g((i + j\eta)) d\eta,$$

where  $g$  is the  $2(M - D)$ th derivative of  $|\cdot|^s$  (defined on  $\mathbb{R} \setminus \{0\}$ ). For  $i$  sufficiently large,  $g(i + j\eta)$  is bounded by  $(Const)|i|^{s-2(M-D)}$  so that

$$R^2(i, 1, 2D, |\cdot|^s, a^{2*}) \text{ is bounded by } (Const)i^{2(s-2(M-D))}, \tag{B.7}$$

which is the general term of a convergent series.

(ii) Now we show that the term  $C_n$  is negligible compared to  $A_n$ . This will imply in turn that  $B_n$  is negligible compared to  $A_n$ , from the Cauchy-Schwarz inequality. We have to give bounds to the series with general term

$R^2(i, \delta_n, 2D, r, a^{2*})$  with

$$R(i, \delta_n, 2D, r, a^{2*}) = - \sum_j a_j^{2*} j^{2D} \int_0^1 \frac{(1-\eta)^{2D-1}}{(2D-1)!} r((i+j\eta)\delta_n) d\eta.$$

For fixed  $i$ , the assumptions (2.1) on  $r$  in  $(\mathcal{H}_1)$  are sufficient to build a dominated convergence argument to prove that  $R^2(i, \delta_n, 2D, r, a^{2*}) = o(\delta_n^{2s})$  which leads to the required result. So we concentrate our attention on indices  $i$  such that  $|i| > 2L$ . Now we use  $(\mathcal{H}_2)$  and the notation  $d = 2$  if  $s < 3/2$  and  $d = 3$  if  $s \geq 3/2$ . Consider  $\beta$  as in  $(\mathcal{H}_2)$  and remark that, in the mixed situation, we have  $s - d < \beta < -1/2$ . In the infill situation, it is assumed that for  $|h| < 1$ ,  $|r^{(d)}(h)| \leq (Const)|h|^\beta$  with  $\beta > s - d$ . Since  $h$  is restricted to  $[-1, 1]$ , we may also consider without loss of generality that  $\beta$  has been chosen such that  $s - d < \beta < -1/2$ . Using B.6 as in the proof of item 1), if  $2D + d \leq 2M$ , one gets

$$R(i, \delta_n, 2D, r, a^{2*}) = - \sum_j a_j^{2*} j^{2D+d} \delta_n^d \int_0^1 \frac{(1-\eta)^{2D+d-1}}{(2D+d-1)!} r^{(d)}((i+j\eta)\delta_n) d\eta. \tag{B.8}$$

The condition  $|i| > 2L$  ensures that the integral is always convergent. Then we have

$$R^2(i, \delta_n, 2D, r, a^{2*}) \leq (Const)\delta_n^{2d+2\beta} i^{2\beta}. \tag{B.9}$$

Since  $\beta < -1/2$ , the series in  $i$  converges and the contribution to  $C$  of the indices  $i$  such that  $|i| > 2L$  is bounded by  $(Const)\delta_n^{4D+2d+2\beta}$  which is negligible compared to  $\delta_n^{4D+2s}$  since  $d + \beta > s$ .  $\square$

*Proof of Theorem 3.4.* By a diagonalization argument,  $V_{a,n}$  can be written as

$$V_{a,n} = \sum_{i=1}^{n''} \lambda_i Z_i^2,$$

where  $\lambda_1, \dots, \lambda_{n''}$  are the non-zero eigenvalues of variance-covariance matrix  $\Sigma_a$  of  $\Delta_a(X)$  and the  $Z_i$  are independent and identically distributed standard Gaussian variables. Hence,

$$\frac{V_{a,n} - \mathbb{E}(V_{a,n})}{\sqrt{\text{Var}(V_{a,n})}} = \sum_{i=1}^{n''} \frac{\lambda_i}{\sqrt{\sum_{r=1}^{n''} \lambda_r^2}} (Z_i^2 - 1). \tag{B.10}$$

In such a situation, Lemma 2 in [30] implies that the Lindeberg condition is a sufficient condition required to prove the central limit theorem and is equivalent to

$$\max_{i=1, \dots, n''} |\lambda_i| = o\left(\sqrt{\text{Var}(V_{a,n})}\right).$$

From Lemma B.2, one has

$$\max_{i=1, \dots, n''} \left( \sum_{j=1}^{n''} |\Sigma_a(i, j)| \right) = o\left(\sqrt{\sum_{r=1}^{n''} \lambda_r^2}\right)$$

and the result follows using the following classical linear algebra result (see for instance [38, Ch. 6.2, p194])

$$\max_{i=1,\dots,n''} |\lambda_i| \leq \max_{i=1,\dots,n'} \left( \sum_{j=1}^{n'} |\Sigma_a(i, j)| \right).$$

□

*Proof of Corollary 3.6.* To prove the asymptotic joint normality it is sufficient to prove the asymptotic normality of any non-zero linear combination

$$LC(\gamma) = \sum_{j=1}^k \gamma_j V_{a^{(j)}, n},$$

where  $\gamma_j \in \mathbb{R}$  for  $j = 1, \dots, k$ . We have again the representation

$$LC(\gamma) = \sum_{i=1}^{n''} \lambda_i Z_i^2,$$

where the  $\lambda_i$ 's are now the non-zero eigenvalues of the variance-covariance matrix

$$\sigma' = \sum_{j=1}^k \gamma_j \Sigma_{a^{(j)}, n},$$

and the  $Z_i$ 's are as before. The Lindeberg condition has the same expression. On one hand, as  $n$  goes to infinity,

$$\frac{1}{n\delta_n^{4D+2s}} \sum_{i=1}^{n''} \lambda_i \rightarrow \gamma^\top \Lambda_\infty \gamma$$

where  $^\top$  stands for the transpose. On the other hand, by the triangular inequality for the operator norm (which is the maximum of the  $|\lambda_i|$ 's), one gets

$$\max_{i=1,\dots,n''} |\lambda_i| = \|\sigma'\|_{op} \leq \sum_{j=1}^k \gamma_j \|\Sigma_{a^{(j)}, n}\|_{op}.$$

In the proof of Theorem 3.4, we have established that  $\|\Sigma_{a^{(j)}, n}\|_{op} = o(n^{1/2}\delta_n^{2D+s})$  leading to the result. □

#### B.4 Proof of the remaining results

*Proof of Theorem 3.7.* We use the definition of  $C_{a,n}$  and the following decomposition:

$$\frac{C_{a,n} - C}{\sqrt{\text{Var}(C_{a,n})}} = \frac{C_{a,n} - \mathbb{E}[C_{a,n}]}{\sqrt{\text{Var}(C_{a,n})}} + \frac{\mathbb{E}[C_{a,n}] - C}{\sqrt{\text{Var}(C_{a,n})}} = \frac{V_{a,n} - \mathbb{E}[V_{a,n}]}{\sqrt{\text{Var}(V_{a,n})}} + \frac{\mathbb{E}[C_{a,n}] - C}{\sqrt{\text{Var}(C_{a,n})}}.$$

Following the proof of Proposition 3.1, the second term is proportional to

$$\sqrt{n}\delta_n^{-s} R(0, \delta_n, 2D, r, a^{2*}) = -\sqrt{n}\delta_n^{-s} \sum_i a_i^{2*} i^{2D} \int_0^1 \frac{(1-\eta)^{2D-1}}{(2D-1)!} r(i\eta\delta_n) d\eta \quad (\text{B.11})$$

which is negligible compared to  $(Const)n^{1/2+\alpha s-\alpha(s+1/2\alpha)}$  by  $(\mathcal{H}_3)$  and thus goes to 0 as  $n$  goes to infinity. Then Slutsky’s lemma and Theorem 3.4 lead straightforwardly to the required result.  $\square$

*Proof of Corollary 3.9.* Obviously, one has

$$V_{a,n}^X = \|\Delta_a(X)\|^2 = \|\Delta_a(f) + \Delta_a(\bar{X})\|^2.$$

Using the triangular inequality  $\|A + B\|^2 - \|A\|^2 \leq \|B\|^2 + 2\|A\|\|B\|$ , it suffices to have  $\|\Delta_a(f)\|^2 = o(\text{Var}(V_{a,n}(\bar{X})^{1/2}) = o(n^{1/2}\delta_n^{2D+s})$  to deduce the central limit theorem for  $X$  from that for  $\bar{X}$ . By application of the Taylor-Lagrange formula, one gets

$$\Delta_{a,i}(f) = (Const) \times \delta_n^M \times f^{(M)}(\xi),$$

with  $\xi \in [0, n^{1-\alpha}]$ . Then  $\|\Delta_a(f)\|^2 \leq n(K_{M,n}^\alpha)^2 \delta_n^{2M}$  and a sufficient condition is (3.12).  $\square$

In order to prove Theorem 3.10, we state and prove the two following lemmas.

**Lemma B.3.** *In the context of Theorem 3.10, there exists a constant  $\mathcal{K}'$ , depending only on  $D, s, U, \mu, B, \beta, C_{\text{sup}}$ ,  $a$  such that, for  $n \geq L(a)$ ,*

$$\sup_{C \in (0, C_{\text{sup}}]} \sup_{V \in \mathcal{C}_{D,s,C,U,\mu,B,\beta}} (\mathbb{E}_V [C_{a,n}] - C)^2 \leq \mathcal{K}' n^{-\min(2\mu, 2)}.$$

*Proof of Lemma B.3.* As in Section 2.1, we write for  $h \in [-1, 1]$ ,

$$r(h) = V^{(2D)}(h) - V^{(2D)}(0) - C(-1)^D |h|^s \tag{B.12}$$

and we remark that for  $V \in \mathcal{C}_{D,s,C,U,\mu,B,\beta}$ , we have  $|r(h)| \leq U|h|^{s+\mu}$  for  $h \in [-1, 1]$ . We observe that, since  $\delta_n = 1/n$ ,

$$C_{a,n} = \frac{V_{a,n}}{n(-1)^D n^{-2D-s} R(0, 1, 2D, |\cdot|^s, a^{2*})} \tag{B.13}$$

and that  $(-1)^D R(0, 1, 2D, |\cdot|^s, a^{2*})$  is a strictly positive constant, depending only on  $a, D$  and  $s$ . In the proof of Proposition 3.1, it is shown that, with  $L = L(a)$  and  $n' = n - L + 1$ ,

$$\mathbb{E}_V [V_{a,n}] = n' C (-1)^D n^{-2D-s} R(0, 1, 2D, |\cdot|^s, a^{2*}) + n' n^{-2D} R(0, 1/n, 2D, r, a^{2*}).$$

Hence we have

$$\begin{aligned} |\mathbb{E}_V [C_{a,n}] - C| &= \left| \mathbb{E}_V [C_{a,n}] - \frac{n'}{n} C + \left( \frac{n'}{n} - 1 \right) C \right| \\ &\leq \left| \frac{n' n^{-2D} R(0, 1/n, 2D, r, a^{2*})}{n(-1)^D n^{-2D-s} R(0, 1, 2D, |\cdot|^s, a^{2*})} \right| + \left( 1 - \frac{n'}{n} \right) C \\ &\leq n^s \left| \frac{R(0, 1/n, 2D, r, a^{2*})}{(-1)^D R(0, 1, 2D, |\cdot|^s, a^{2*})} \right| + \frac{L-1}{n} C. \end{aligned} \tag{B.14}$$

Let us bound the numerator in the above fraction. We have, for  $D = 0$ ,

$$\begin{aligned} |R(0, 1/n, 2D, r, a^{2*})| &\leq \left| \sum_i a_i^{2*} r(i/n) \right| \\ &\leq \left( \sum_i |a_i^{2*}| \right) U \left| \frac{L-1}{n} \right|^{s+\mu}, \end{aligned}$$

from the bound on  $r$  pointed out above, and observing that  $a_i^{2*} = 0$  for  $|i| \geq L$ .

When  $D \geq 1$ , we have

$$\begin{aligned} |R(0, 1/n, 2D, r, a^{2*})| &= \left| - \sum_i a_i^{2*} i^{2D} \int_0^1 \frac{(1-\eta)^{2D-1}}{(2D-1)!} r\left(\frac{i\eta}{n}\right) d\eta \right| \\ &\leq \left( \sum_i |a_i^{2*}| \right) (L-1)^{2D} \frac{1}{(2D-1)!} U \left| \frac{L-1}{n} \right|^{s+\mu}. \end{aligned}$$

Going back to (B.14), with

$$\tilde{\mathcal{N}} = \max \left( \left( \sum_i |a_i^{2*}| \right), \left( \sum_i |a_i^{2*}| \right) (L-1)^{2D} \frac{1}{(2D-1)!} \right),$$

we have that

$$|\mathbb{E}_V [C_{a,n}] - C| \leq n^{-\mu} \frac{\tilde{\mathcal{N}} U (L-1)^{s+\mu}}{(-1)^D R(0, 1, 2D, |\cdot|^s, a^{2*})} + \frac{L-1}{n} C.$$

This concludes the proof, for instance by taking

$$\mathcal{K}' = \left( \frac{\tilde{\mathcal{N}} U (L-1)^{s+\mu}}{(-1)^D R(0, 1, 2D, |\cdot|^s, a^{2*})} + (L-1) C_{\text{sup}} \right)^2.$$

□

**Lemma B.4.** *In the context of Theorem 3.10, there exists a constant  $\mathcal{K}''$ , depending only on  $D, s, U, \mu, B, \beta, C_{\text{sup}}, a$  such that*

$$\sup_{C \in (0, C_{\text{sup}}]} \sup_{V \in \mathcal{C}_{D, s, C, U, \mu, B, \beta}} \text{Var}_V(C_{a,n}) \leq \frac{\mathcal{K}''}{n}.$$

*Proof of Lemma B.4.* From (B.13), we have

$$\text{Var}_V(C_{a,n}) = \frac{\text{Var}_V(V_{a,n})}{n^{2-4D-2s} R(0, 1, 2D, |\cdot|^s, a^{2*})^2}.$$

With the notation of the proof of Proposition 3.1, and from the inequality  $(x + y)^2 \leq 2x^2 + 2y^2$ , we obtain

$$\text{Var}_V(V_{a,n}) \leq 2A_n + 2C_n.$$

From the proof of Proposition 3.1, recalling the notation  $n' = n - L + 1$ ,

$$\begin{aligned} A_n &= 2C^2 n' n^{-4D-2s} \sum_{i \in \mathbb{Z}} \frac{n' - |i|}{n'} R^2(i, 1, 2D, |\cdot|^s, a^{2*}) \mathbb{1}_{|i| \leq n'-1} \\ &\leq 2C^2 n^{1-4D-2s} \sum_{i \in \mathbb{Z}} R^2(i, 1, 2D, |\cdot|^s, a^{2*}). \end{aligned}$$

The above sum is a finite constant, depending only on  $D, a, s$ , from (B.7). Thus with

$$\tilde{\mathcal{K}}_1 = 2C_{\text{sup}}^2 \sum_{i \in \mathbb{Z}} R^2(i, 1, 2D, |\cdot|^s, a^{2*}),$$

we have

$$A_n \leq \tilde{\mathcal{K}}_1 n^{1-4D-2s}.$$

Next, from the proof of Proposition 3.1,

$$C_n = 2n^{-4D} \sum_{i=-n'+1}^{n'-1} (n' - |i|) R(i, 1/n, 2D, r, a^{2*})^2. \tag{B.15}$$

For  $|i| \leq 2L$ , when  $D = 0$ ,

$$|R(i, 1/n, 2D, r, a^{2*})| = \left| - \sum_j a_j^{2*} r \left( \frac{i+j}{n} \right) \right| \leq \left( \sum_j |a_j^{2*}| \right) U(3L-1)^{s+\mu} n^{-s-\mu}.$$

For  $|i| \leq 2L$ , when  $D > 0$ ,

$$\begin{aligned} |R(i, 1/n, 2D, r, a^{2*})| &= \left| - \sum_j a_j j^{2D} \int_0^1 \frac{(1-\eta)^{2D-1}}{(2D-1)!} r \left( \frac{i+j\eta}{n} \right) d\eta \right| \\ &\leq \left( \sum_j |a_j^{2*}| \right) (L-1)^{2D} \frac{1}{(2D-1)!} U(3L-1)^{s+\mu} n^{-s-\mu}. \end{aligned}$$

Hence, letting

$$\tilde{\mathcal{K}}_2 = \max \left( \left( \sum_j |a_j^{2*}| \right) U(3L-1)^{s+\mu}, \left( \sum_j |a_j^{2*}| \right) (L-1)^{2D} \frac{1}{(2D-1)!} U(3L-1)^{s+\mu} \right),$$

we obtain for  $|i| \leq 2L$ ,

$$|R(i, 1/n, 2D, r, a^{2*})| \leq \tilde{\mathcal{K}}_2 n^{-s-\mu}.$$

As in the proof of Proposition 3.1, we now write  $d = 2$  if  $s < 3/2$  and  $d = 3$  if  $s \geq 3/2$ . By definition we have  $\beta > s - d$ . As in the proof of Lemma B.3 we define  $r$  with (B.12). We recall that for  $|h| \leq 1$ ,  $|r^{(d)}(h)| \leq B|h|^\beta$ . Since  $|h| \leq 1$ , we can take  $\tilde{\beta}$ , only depending on  $\beta$ , such that  $\tilde{\beta} \leq \beta$  and  $s - d < \tilde{\beta} < -1/2$ . Then we have  $B|h|^\beta \leq B|h|^{\tilde{\beta}}$  for  $|h| \leq 1$  and thus  $|r^{(d)}(h)| \leq B|h|^{\tilde{\beta}}$  for  $|h| \leq 1$ .

Then, for  $|i| > 2L$ , from (B.8) in the proof of Proposition 3.1,

$$\begin{aligned} |R(i, 1/n, 2D, r, a^{2*})| &= \left| - \sum_j a_j^{2*} j^{2D+d} n^{-d} \int_0^1 \frac{(1-\eta)^{2D+d-1}}{(2D+d-1)!} r^{(d)}\left(\frac{i+j\eta}{n}\right) d\eta \right| \\ &\leq \left( \sum_j |a_j^{2*}| \right) (L-1)^{2D+d} n^{-d} \frac{1}{(2D+d-1)!} B \left( \frac{|i| - (L-1)}{n} \right)^{\tilde{\beta}}, \end{aligned} \quad (\text{B.16})$$

since the summand in the sum (B.16) is non-zero only for  $|j| \leq L-1$  and thus when  $|(i+j\eta)/n| \leq 1$  for  $0 \leq \eta \leq 1$ . Defining

$$\tilde{\mathcal{K}}_3 = \left( \sum_j |a_j^{2*}| \right) (L-1)^{2D+d} \frac{1}{(2D+d-1)!} B,$$

and observing that  $-d - \tilde{\beta} < -s$ , we obtain, for  $|i| > 2L$ ,

$$|R(i, 1/n, 2D, r, a^{2*})| \leq \tilde{\mathcal{K}}_3 n^{-s} (|i| - L + 1)^{\tilde{\beta}}.$$

Hence we obtain from (B.15)

$$\begin{aligned} C_n &\leq 2n^{1-4D} \left( (2L+1)\tilde{\mathcal{K}}_2^2 n^{-2s-2\mu} + \sum_{|i|>2L} \tilde{\mathcal{K}}_3^2 n^{-2s} (|i| - L + 1)^{2\tilde{\beta}} \right) \\ &\leq 2 \left( (2L+1)\tilde{\mathcal{K}}_2^2 + \sum_{|i|>2L} \tilde{\mathcal{K}}_3^2 (|i| - L + 1)^{2\tilde{\beta}} \right) n^{1-4D-2s}. \end{aligned}$$

We let

$$\tilde{\mathcal{K}}_4 = 2 \left( (2L+1)\tilde{\mathcal{K}}_2^2 + \sum_{|i|>2L} \tilde{\mathcal{K}}_3^2 (|i| - L + 1)^{2\tilde{\beta}} \right)$$

and observe that  $\tilde{\mathcal{K}}_4$  is finite because  $\tilde{\beta} < -1/2$ . Hence we have

$$C_n \leq \tilde{\mathcal{K}}_4 n^{1-4D-2s}$$

and thus

$$\text{Var}_V(C_{a,n}) \leq \frac{2(\tilde{\mathcal{K}}_1 + \tilde{\mathcal{K}}_4)n^{1-4D-2s}}{n^{2-4D-2s}R(0, 1, 2D, |\cdot|^s, a^{2*})^2}$$

and thus the proof is concluded with

$$\mathcal{K}'' = \frac{2(\tilde{\mathcal{K}}_1 + \tilde{\mathcal{K}}_4)}{R(0, 1, 2D, |\cdot|^s, a^{2*})^2}.$$

□

*Proof of Theorem 3.10.* The theorem follows directly from Lemmas B.3 and B.4, by taking  $\mathcal{K} = \mathcal{K}' + \mathcal{K}''$ . □

*Proof of Lemma 4.2.* (4.4) implies that  $\partial R_C / \partial C = R_1$  then (4.3) gives the result. □

*Proof of Lemma 5.2.* The estimator  $\hat{C}_{1,j}$  of step (2) obtained from any column  $j$  converges in the  $L^1$ -sense to  $C_1$ , with the  $L^1$ -difference going to zero uniformly over  $j = 1, \dots, N$ , from Theorem 3.10. Hence, after averaging,  $\hat{C}_1$  converges to  $C_1$  in the  $L^1$  sense. The result for  $\hat{C}_2$  is obtained similarly. □

*Acknowledgements.* This work has been partially supported by the French National Research Agency (ANR) through project PEPITO (no ANR-14-CE23-0011). The authors are grateful to Laurent Risser, for providing them the image registration data of Section 5.4.2. The authors warmly thank the reviewers, the comments of which greatly improved the manuscript.

## REFERENCES

- [1] R.J. Adler and R. Pyke, Uniform quadratic variation for Gaussian processes. *Stochastic Process. Appl.* **48** (1993) 191–209.
- [2] J.-M. Azaïs and M. Wschebor, Level sets and Extrema of Random Processes and Fields. John Wiley & Sons, Inc., Hoboken, NJ (2009).
- [3] F. Bachoc, Cross validation and maximum likelihood estimations of hyper-parameters of Gaussian processes with model misspecification. *Computat. Stat. Data Anal.* **66** (2013) 55–69.
- [4] F. Bachoc, Asymptotic analysis of the role of spatial sampling for covariance parameter estimation of Gaussian processes. *J. Multivariate Anal.* **125** (2014) 1–35.
- [5] F. Bachoc, Asymptotic analysis of covariance parameter estimation for Gaussian processes in the misspecified case. *Bernoulli* **24** (2018) 1531–1575.
- [6] F. Bachoc and A. Lagnoux, Fixed-domain asymptotic properties of maximum composite likelihood estimators for Gaussian processes. *J. Stat. Plann. Inference* **209** (2020) 62–75.
- [7] J.M. Bates and C.W. Granger, The combination of forecasts. *J. Oper. Res. Soc.* **20** (1969) 451–468.
- [8] G. Baxter, A strong limit theorem for Gaussian processes. *Proc. Am. Math. Soc.* **7** (1956) 522–527.
- [9] 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*. Preprint [arXiv:1410.7827](https://arxiv.org/abs/1410.7827) (2014).
- [10] I. Clark, Practical Geostatistics, Vol. 3. Applied Science Publishers, London (1979).
- [11] J.-F. Coeurjolly, Estimating the parameters of a fractional Brownian motion by discrete variations of its sample paths. *Stat. Inference Stoch. Process.* **4** (2001) 199–227.
- [12] S. Cohen and J. Istas, Fractional Fields and Applications, With a foreword by Stéphane Jaffard. Vol. 73 of *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Springer, Heidelberg (2013).
- [13] N. Cressie, Statistics for Spatial Data. John Wiley (1993).
- [14] N. Cressie and D.M. Hawkins, Robust estimation of the variogram: I. *J. Int. Assoc. Math. Geol.* **12** (1980) 115–125.
- [15] R. Dahlhaus, Efficient parameter estimation for self-similar processes. *Ann. Stat.* (1989) 1749–1766.
- [16] A. Datta, S. Banerjee, A.O. Finley, and A.E. Gelfand, Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. *J. Am. Stat. Assoc.* **111** (2016) 800–812.
- [17] I. Daubechies, Orthonormal bases of compactly supported wavelets. *Commun. Pure Appl. Math.* **41** (1988) 909–996.
- [18] M. David, Geostatistical Ore Reserve Estimation. Elsevier (2012).
- [19] M.P. Deisenroth and J.W. Ng, Distributed Gaussian processes, In *Proceedings of the 32nd International Conference on Machine Learning, Lille, France. JMLR: W&CP volume 37* (2015).
- [20] R. Furrer, M.G. Genton and D. Nychka, Covariance tapering for interpolation of large spatial datasets. *J. Comput. Graph. Stat.* **15** (2006) 502–523.

- [21] E.G. Gladyshev, A new limit theorem for stochastic processes with Gaussian increments. *Teor. Veroyatnost. Primenen.* **6** (1961) 57–66.
- [22] U. Grenander, Abstract inference. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York (1981).
- [23] X. Guyon and J. León, Convergence en loi des  $H$ -variations d'un processus gaussien stationnaire sur  $\mathbf{R}$ . *Ann. Inst. Henri Poincaré Probab. Statist.* **25** (1989) 265–282.
- [24] P. Hall, N.I. Fisher and B. Hoffmann, On the nonparametric estimation of covariance functions. *Ann. Stat.* **22** (1994) 2115–2134.
- [25] P. Hall and P. Patil, Properties of nonparametric estimators of autocovariance for stationary random fields. *Probab. Theory Related Fields* **99** (1994) 399–424.
- [26] J. Han and X.-P. Zhang, Financial time series volatility analysis using Gaussian process state-space models, in *2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE (2015) 358–362.
- [27] J. Hensman and N. Fusi, Gaussian processes for big data. *Uncertainty Artif. Intell.* (2013) 282–290.
- [28] G.E. Hinton, Training products of experts by minimizing contrastive divergence. *Neural Computat.* **14** (2002) 1771–1800.
- [29] I. Ibragimov and Y. Rozanov, Gaussian Random Processes. Springer-Verlag, New York (1978).
- [30] J. Istas and G. Lang, Quadratic variations and estimation of the local Hölder index of a Gaussian process. *Ann. Inst. Henri Poincaré Probab. Statist.* **33** (1997) 407–436.
- [31] A. Journel and C. Huijbregts, Mining geostatistics, in Bureau De Recherches Géologiques Et Minières, France Academic Pres Harcourt Brace & Company, Publishers, London, San Diego, New York, Boston, Sidney, Toronto (1978).
- [32] C.G. Kaufman, M.J. Schervish, and D.W. Nychka, Covariance tapering for likelihood-based estimation in large spatial data sets. *J. Am. Stat. Assoc.* **103** (2008) 1545–1555.
- [33] J.T. Kent and A.T.A. Wood, Estimating the fractal dimension of a locally self-similar Gaussian process by using increments. *J. Roy. Statist. Soc. Ser. B* **59** (1997) 679–699.
- [34] S.N. Lahiri, Y. Lee and N. Cressie, On asymptotic distribution and asymptotic efficiency of least squares estimators of spatial variogram parameters. *J. Stat. Plann. Inference* **103** (2002) 65–85.
- [35] G. Lang and F. Roueff, Semi-parametric estimation of the Hölder exponent of a stationary Gaussian process with minimax rates. *Stat. Inference Stoch. Process.* **4** (2001) 283–306.
- [36] F. Lavancier and P. Rochet, A general procedure to combine estimators. *Computat. Stat. Data Anal.* **94** (2016) 175–192.
- [37] P. Lévy, Le mouvement brownien plan. *Am. J. Math.* **62** (1940) 487–550.
- [38] D. Luenberger, Introduction to Dynamic Systems: Theory, Models, and Applications. Wiley (1979).
- [39] J. Mateu, E. Porcu, G. Christakos and M. Bevilacqua, Fitting negative spatial covariances to geothermal field temperatures in Nea Kessani (Greece). *Environmetrics* **18** (2007) 759–773.
- [40] G. Matheron, Traité de géostatistique appliquée, Tome I, Vol. 14 of *Editions Technip, Paris*. Mémoires du Bureau de Recherches Géologiques et Minières (1962).
- [41] E. Pardo-Igúzquiza and P.A. Dowd, AMLE3D: a computer program for the inference of spatial covariance parameters by approximate maximum likelihood estimation. *Comput. Geosci.* **23** (1997) 793–805.
- [42] O. Perrin, Quadratic variation for Gaussian processes and application to time deformation. *Stochastic Process. Appl.* **82** (1999) 293–305.
- [43] G. Pólya, Remarks on characteristic functions, in *Proceedings of the First Berkeley Symposium on Mathematical Statistics and Probability, August 13–18, 1945 and January 27–29, 1946*, edited by J. Neyman. Statistical Laboratory of the University of California, Berkeley. University of California Press, Berkeley, CA (1949) 115–123.
- [44] C. Rasmussen and C. Williams, Gaussian Processes for Machine Learning. The MIT Press, Cambridge (2006).
- [45] F. Richard, Anisotropy of Hölder Gaussian random fields: characterization, estimation, and application to image textures. *Stat. Comput.* **28** (2018) 1155–1168.
- [46] F. Richard and H. Biermé, Statistical tests of anisotropy for fractional Brownian textures. application to full-field digital mammography. *J. Math. Imaging Vis.* **36** (2010) 227–240.
- [47] L. Risser, F. Vialard, R. Wolz, M. Murgasova, D. Holm and D. Rueckert, ADNI: Simultaneous multiscale registration using large deformation diffeomorphic metric mapping. *IEEE Trans. Med. Imaging* **30** (2011) 1746–1759.
- [48] O. Roustant, D. Ginsbourger and Y. Deville, DiceKriging, DiceOptim: two R packages for the analysis of computer experiments by Kriging-based metamodeling and optimization. *J. Stat. Softw.* **51** (2012).
- [49] H. Rue and L. Held, Gaussian Markov Random Fields, Theory and Applications. Chapman & Hall (2005).
- [50] D. Rullière, N. Durrande, F. Bachoc and C. Chevalier, Nested Kriging predictions for datasets with a large number of observations. *Stat. Comput.* **28** (2018) 849–867.
- [51] G. Samorodnitsky and M. Taqqu, Non-Gaussian Stable Processes: Stochastic Models with Infinite Variance. Chapman ft Hall, London (1994).
- [52] T.J. Santner, B.J. Williams and W.I. Notz, The Design and Analysis of Computer Experiments. Springer Series in Statistics. Springer-Verlag, New York (2003).
- [53] D. Slepian, On the zeros of Gaussian noise, in *Proc. Sympos. Time Series Analysis (Brown Univ., 1962)*. Wiley, New York (1963) 104–115.
- [54] M. Stein, Interpolation of Spatial Data: Some Theory for Kriging. Springer, New York (1999).
- [55] M.L. Stein, Limitations on low rank approximations for covariance matrices of spatial data. *Spatial Stat.* **8** (2014) 1–19.
- [56] M.L. Stein, Z. Chi and L.J. Welty, Approximating likelihoods for large spatial data sets. *J. Roy. Stat. Soc.: Ser. B (Stat. Methodol.)* **66** (2004) 275–296.

- [57] V. Tresp, A Bayesian committee machine. *Neural Computat.* **12** (2000) 2719–2741.
- [58] 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*. Springer (2015) 310–321.
- [59] C. Varin, N. Reid and D. Firth, An overview of composite likelihood methods. *Stat. Sinica* **21** (2011) 5–42.
- [60] A.V. Vecchia, Estimation and model identification for continuous spatial processes. *J. Roy. Stat. Soc.: Ser. B (Methodological)* **50** (1988) 297–312.
- [61] Y. Wu, J.M. Hernández-Lobato and Z. Ghahramani, Gaussian process volatility model, in *Advances in Neural Information Processing Systems 27*, edited by Z. Ghahramani, M. Welling, C. Cortes, N.D. Lawrence and K.Q. Weinberger. Curran Associates, Inc. (2014) 1044–1052.
- [62] H. Zhang and Y. Wang, Kriging and cross-validation for massive spatial data. *Environmetrics* **21** (2010) 290–304.