



HAL
open science

Semi-parametric estimation of the variogram of a Gaussian process with stationary increments

Jean-Marc Azaïs, François Bachoc, Thierry Klein, Agnès Lagnoux, Thi Mong
Ngoc Nguyen

► **To cite this version:**

Jean-Marc Azaïs, François Bachoc, Thierry Klein, Agnès Lagnoux, Thi Mong Ngoc Nguyen. Semi-parametric estimation of the variogram of a Gaussian process with stationary increments. 2019. hal-01802830v2

HAL Id: hal-01802830

<https://hal.science/hal-01802830v2>

Preprint submitted on 12 Feb 2019 (v2), last revised 8 Jan 2020 (v3)

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

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

TEST

Semi-parametric estimation of the variogram of a Gaussian process with stationary increments --Manuscript Draft--

Manuscript Number:	
Full Title:	Semi-parametric estimation of the variogram of a Gaussian process with stationary increments
Article Type:	Original research paper
Corresponding Author:	Agnès Lagnoux Institut de mathematiques de Toulouse FRANCE
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Institut de mathematiques de Toulouse
Corresponding Author's Secondary Institution:	
First Author:	Jean-Marc Azaïs
First Author Secondary Information:	
Order of Authors:	Jean-Marc Azaïs François Bachoc Thierry Klein Agnès Lagnoux Thi Mong Ngoc Nguyen
Order of Authors Secondary Information:	
Funding Information:	ANR-14-CE23-0011 (ANR-14-CE23-0011) PR Jean-Marc Azaïs
Abstract:	<p>We consider the semi-parametric estimation of a scale parameter of a one-dimensional Gaussian process with known smoothness. We suggest an estimator based on quadratic variations and on 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 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. Finally, we suggest an extension to the multi-dimensional case, and we apply it to two real data sets in dimension two.</p>
Suggested Reviewers:	

Semi-parametric estimation of the variogram of a Gaussian process with stationary increments

Jean-Marc Azaïs · François Bachoc · Thierry Klein ·
Agnès Lagnoux · Thi Mong Ngoc Nguyen

Received: date / Accepted: date

Abstract We consider the semi-parametric estimation of a scale parameter of a one-dimensional Gaussian process with known smoothness. We suggest an estimator based on quadratic variations and on 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 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. Finally, we suggest an extension to the multi-dimensional case, and we apply it to two real data sets in dimension two.

Keywords quadratic variations · scale covariance parameter · asymptotic normality · moment method · aggregation of estimators

1 Introduction

Gaussian process models are widely used in spatial statistics; in particular, in order to interpolate observations by Kriging. For example, this technique is used in computer experiment designs to build a metamodel [21, 27]. Usually, the practitioner uses a model for which the mean function of the Gaussian process (the drift) is assumed to be a linear combination of known functions (often polynomials), and the covariance function is assumed to belong to a parametric set of the form $\{\sigma^2 \rho_\theta; \sigma^2 \geq 0, \theta \in \Theta\}$ with $\Theta \subset \mathbb{R}^p$. Classical examples are the set of Matérn or exponential covariance functions [21, 27].

In this paper, we consider an intrinsically stationary one-dimensional Gaussian process X (i.e. with stationary increments). Its variogram is defined by

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

and has its $2D$ 'th derivative satisfying $V^{(2D)}(h) \approx V^{(2D)}(0) + C(-1)^D |h|^s$ as $h \rightarrow 0$ (see Section 2.1). Here D is the number of derivatives of the process, s is the smoothness parameter and we call C the scale parameter. In this paper, we assume D and s to be known and focus on the theoretical study of the estimation of C in dimension one. We also perform some additional simulations in higher dimensions.

In applications, the estimation of the scale parameter C is crucial. Only the knowledge of the constant C is involved in the covariance function and its estimation constitutes then a necessary preliminary step to Kriging. Indeed, for instance when $D = 1$, C provides the first order approximation of

J.-M. Azaïs and F. Bachoc

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

T. Klein

Institut de Mathématiques de Toulouse; UMR5219. Université de Toulouse; CNRS. ENAC, F-31400 Toulouse, France.

A. Lagnoux

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

T.M.N. Nguyen

VNUHCM - University of Science, Ho Chi Minh city, Viet Nam.

$\mathbb{E} \left[(X(t+h) - X(t))^2 \right]$ when h is small. When it is assumed that the covariance function of the Gaussian process belongs to a parametric set $\{\sigma^2 \rho_\theta; \sigma^2 \geq 0, \theta \in \Theta\}$ with $\Theta \subset \mathbb{R}^p$, C is a function of σ^2 and θ . In this case, most of the software packages (like, e.g. `DiceKriging` [23]) use the maximum likelihood estimation method (MLE) to estimate (σ^2, θ) and thus C (see [21, 25] for more details on MLE). Unfortunately, MLE is known to be computationally expensive. In addition, it may diverge in some complicated situations (see Section 5.3). Finally, MLE is applicable to estimate C only when the parametric set of covariance functions is given, while C is also relevant in the non-parametric case where no parametric assumptions are made on the variogram V in (1).

In order to overcome the drawbacks of the MLE, we propose an alternative method based on quadratic variations that estimates C with no parametric assumption. Quadratic variations have been first introduced by Levy in [17] to quantify the oscillations of the Brownian motion. Levy showed that,

$$\sum_{i=1}^{2^n} (Z(i/2^n) - Z((i-1)/2^n))^2 \xrightarrow[n \rightarrow +\infty]{\text{a.s.}} 1,$$

where Z is the standard Brownian motion on $[0, 1]$. A preliminary result on quadratic variations of a Gaussian non-differentiable process is Baxter's theorem (see e.g. [5], [11, Chap. 5] and [10]) that ensures (under some conditions) the almost sure convergence (as n tends to infinity) of

$$\sum_{i=1}^n (X(i/n) - X((i-1)/n))^2. \quad (2)$$

A generalization of the previous quadratic variations has been introduced in Guyon and Léon [12]: for a given real function H , the H -variation is given by

$$V_{H,n} := \sum_{i=1}^n H \left(\frac{X(i/n) - X((i-1)/n)}{\sqrt{\text{Var}(X(i/n) - X((i-1)/n))}} \right) \quad (3)$$

where X is assumed to be a centered stationary Gaussian process. The most unexpected result of [12] is that, if $\rho(h) = \text{Cov}(X(t+h), X(t))$ is such that $\rho(h) = 1 - |h|^s l(h)$ where s is a real number such that $0 < s < 2$ and l is a slowly varying function at zero and assuming some additional technical conditions, then

1. if $0 < s < 3/2$, $(V_{H,n}/n)$ has a limiting normal distribution with convergence rate $n^{1/2}$;
2. if $3/2 < s < 2$, $(V_{H,n}/n)$ has a limiting non normal distribution with convergence rate n^{2-s} .

In fact for statistical purposes, it has been proved by Coeurjolly that quadratic variations are optimal (details and precisions can be found in [6]). In [13], Istas and Lang generalized the results on quadratic variations. They considered a Gaussian process with stationary increments which is weaker than assuming the stationarity. The observations of the process are done at times $\Delta_n j$ for $j = 1, \dots, n$, with Δ_n depending on n . They studied the generalized quadratic variations defined by:

$$V_{a,n} := \sum_{i=1}^{n-1} \left(\sum_{k \in K} a_k X((i+k)\Delta_n) \right)^2, \quad (4)$$

where the sequence $a = (a_k)_{k \in K}$ has a finite support and some vanishing moments. Then they built estimators of the local Hölder index and the scale parameter C and showed that they are almost surely consistent and asymptotically normal. In the more recent work of Lang and Roueff [15], the authors generalized the results of Istas and Lang [13] and Kent and Wood [14] 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 [19] and [6].

Now let us present the framework considered in our paper. We assume that the intrinsically stationary Gaussian process X is observed at times $i\Delta_n$ for $i = 1, \dots, n$ with Δ_n tending to zero. The paper is devoted to the estimation of the scale parameter C from one or several generalized quadratic variations $V_{a,n}$ defined in (4). Calculations show that the expectation of $V_{a,n}$ is a function of C so that C can be estimated by the moment method.

The paper is organized as follows. In Section 2, we detail the framework and present the assumptions on the process. The study in Section 3 is close to the study of Istas and Lang [13] in which they estimate both the scale parameter C and the local Hölder index. Nevertheless, since we want to estimate the constant C only and not the local Hölder index, our assumptions and proofs are simpler than those in [13]. Moreover, one can easily check that our hypotheses are satisfied in all the commonly used Kriging models. We compute the expectation and the variance for the variations and we establish a central limit theorem for the variations and the estimators deduced from them. In particular, given a finite number of sequences a , we prove a joint central limit theorem (see Corollary 12) that allows the study of aggregated estimators. In addition, our method does not require a parametric specification of the drift (see Section 3.4); therefore it is more robust than MLE. Section 4 is dedicated to the efficiency of the estimators proposed. Indeed, natural questions then arise. What is the optimal sequence a ? In particular, what is the optimal order of the sequence, that is the number of zero moments (see Section 2.3)? Is it better to use the elementary sequence of order 1 $(-1, 1)$ or the one of order 2 $(-1, 2, -1)$? 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 long as we know, these questions are not addressed yet in the literature. Unfortunately, the asymptotic variance given by Proposition 5 or Theorem 11 does not allow either to address theoretically this issue. Anyway, we compute the optimal Cramér-Rao bound in two examples. Moreover, by Corollary 12, one may aggregate the information of different quadratic variations with different orders. In order to validate such a procedure, an important Monte Carlo study is performed in Section 5. The main conclusion is that aggregating the information of different quadratic variations with different orders produces closer results to the optimal Cramér-Rao bound computed in Section 4.2. The simulations are illustrated in Figure 3. In Section 5, together with the simulation based on aggregation, we illustrate numerically the convergence to the asymptotic distribution considering different models (exponential and Matérn models). Furthermore, in Section 5, 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 maximum likelihood, we observe a very significant computational benefit in favor of our estimator.

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 $j\Delta$ for $j = 1, \dots, n$ with $\Delta = \Delta_n$ going to 0 as n goes to infinity. Its variogram is defined by

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

In the sequel, we let $\Delta = n^{-\alpha}$, $0 < \alpha \leq 1$ and we denote by $(Const)$ a positive constant which value may change from one occurrence to another. Note that the case $\alpha = 1$ corresponds to the infill situation [27]. For the moment, we assume that X is centered, the case of non-zero expectation will be considered in Section 3.4. We introduce the following assumptions.

(\mathcal{H}_0) V is a smooth function on $(0, +\infty]$.

(\mathcal{H}_1) The 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

$$V^{(2D)}(h) = V^{(2D)}(0) + C(-1)^D |h|^s + r(h), \text{ with } r(h) = o(|h|^s) \text{ and } |r(h)| \leq (Const) |h|^s. \quad (5)$$

(\mathcal{H}_2) We assume that the remainder term r in (\mathcal{H}_1) is d -differentiable outside zero and $|r^{(d)}(h)| \leq (Const) |h|^\beta$ with $s - d < \beta < -1/2$. When $s < 3/2$, we set $d = 2$. When $s \geq 3/2$, we set $d = 3$.

(\mathcal{H}_3) $|r(h)| \leq (Const) |h|^{s+1/(2\alpha)}$.

If the covariance function belongs to a parametric set of the form $\{\sigma^2 \rho_\theta; \sigma^2 \geq 0, \theta \in \Theta\}$ with $\Theta \subset \mathbb{R}^p$, then C is a deterministic function of the parameters σ^2 and θ .

Remark 1

- When $D > 0$, the D -th derivative $X^{(D)}$ in quadratic mean of X is a Gaussian stationary process with covariance function ρ given by $\rho(h) = \text{Cov}(X(t), X(t+h)) = (-1)^{D+1}V^{(2D)}(h)$. 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 .
- Note that in the infill case ($\alpha = 1$), (\mathcal{H}_2) is almost minimal. Indeed, the condition $\beta < -1/2$ does not matter since the smaller β , the weaker the condition. And for example, when $s < 3/2$, the second derivative of the main term is of order $|h|^{s-2}$ and we only assume that $\beta > s - 2$.

2.2 Examples of processes that satisfy our assumptions

We present a non exhaustive list of examples in dimension 1 that satisfy our hypotheses.

- The exponential model: $\rho(h) = \exp(-C|h|)$ ($D = 0$, $s = 1$, $C = C$).
- The generalized exponential model: $\rho(h) = \exp(-(C|h|)^s)$, $s \in (0, 2)$ ($D = 0$, $s = s$, $C = C$).
- The generalized Slepian model [26]: $\rho(h) = (1 - (C|h|)^s)^+$, $s \in (0, 1]$ ($D = 0$, $s = s$, $C = C$).
- The spherical model: $\rho(h) = (1 - C|h| + 0.5(\theta|h|)^3)^+$ ($D = 0$, $s = 1$, $C = C$).
- The cubic model $\rho(h) = (1 - 3(\theta|h|)^2 + 2(\theta|h|)^3)^+$ ($D = 1$, $s = 1$, $C = 6\theta^2$).
- The Matérn model:

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

where $\nu > 0$ is the regularity parameter of the process. The function K_ν is the modified Bessel function of the second kind of order ν . See, e.g., [27] 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 ν and θ . However, in Section 5.1, we derive the value of C in two settings ($\nu = 3/2$ and $\nu = 5/2$).

All the previous examples are stationary (and thus intrinsically stationary). The following one is intrinsically stationary but 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 [7]. This process is classically indexed by its Hurst parameter $H = s/2$. Here, $D = 0$, $s = s$ and $C = C$. We call the FBM defined by $C = 1$ the standard FBM.

All classical spatial models satisfy our hypotheses, except the Gaussian model, or square-exponential, defined by $\rho(h) = \sigma^2 e^{-h^2\theta^2}$, with $(\sigma^2, \theta) \in (0, \infty)$, which is too regular.

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

$$\Delta_{a,i}(f) = \sum_{j=0}^{L(a)-1} a_j f((i+j)\Delta), \quad i = 1, \dots, n', \quad (6)$$

where n' stands for $n - L(a) + 1$. As a matter of fact, $\sum_{j=0}^{L(a)-1} a_j f(j\Delta)$ 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 $\Delta_{\mathbf{a}}(X)$ as the Gaussian vector of size n' with entries $\Delta_{a,i}(X)$ and Σ_a its variance-covariance matrix.

Examples - Elementary sequences. The simplest case is the order 1 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 *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 $-(L(a') - 1), (L(a) - 1)$ 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 4 that is required to quantify the asymptotic behaviors of the two first moments of the quadratic a -variations defined in (8) (see Proposition 5). In order to prove (7), we establish two preliminary tools (Proposition 3 and Lemma 1). 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 $dW(t)$ 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 1 (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.

Proposition 3 *The IFBM has the ND property.*

Proof By the stochastic Fubini theorem,

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

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:

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

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. \square

We also need the following lemma.

Lemma 1 *The variance function of the IFBM satisfies, for all $m \in \mathbb{N}$,*

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

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

The proof of Lemma 1 is given in the appendix in Section 6.

Proposition 4 *If the sequence a has order $M(a) > D$, then*

$$\sum_j a_j^{2*} |j|^{2D+s} \neq 0 \quad (\text{i.e.} \quad (-1)^D \sum_j a_j^{2*} |j|^{2D+s} < 0). \quad (7)$$

Note that (7) is stated as an hypothesis in [13].

Proof Using Lemma 1 (with $m = D$) and the vanishing moments of a of order less or equal than D , we have

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

We conclude using the ND property of the IFBM stated in Proposition 3. \square

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

$$V_{a,n} = \|\Delta_{\mathbf{a}}(X)\|^2 = \sum_{i=1}^{n'} (\Delta_{a,i}(X))^2, \quad (8)$$

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 Results on quadratic a -variations

The basis of our computations of variances is the identity

$$\mathbb{E}[\Delta_{a,i}(X)\Delta_{a',i'}(X)] = -\Delta_{a*a',i-i'}(V), \quad (9)$$

for any sequences a and a' . A second main tool is the Taylor expansion with integral remainder (see, for example, (14)). So we introduce another notation. For a sequence a , a scale Δ , an order q and a function f , we define

$$R(i, \Delta, q, f, a) = -\sum_j a_j j^q \int_0^1 \frac{(1-\eta)^{q-1}}{(q-1)!} f((i+j\eta)\Delta) d\eta. \quad (10)$$

By convention, we let $R(i, \Delta, 0, f, a) = -\Delta_{a,i}(f)$. 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 5 (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

$$\mathbb{E}[V_{a,n}] = nC(-1)^D \Delta^{2D+s} [R(0, 1, 2D, |\cdot|^s, a^{2*})] (1 + o(1)), \quad (11)$$

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 so that $M > D + s/2 + 1/4$, then as n tends to infinity:

$$\text{Var}(V_{a,n}) = 2nC^2 \Delta^{4D+2s} \sum_{i \in \mathbb{Z}} R^2(i, 1, 2D, |\cdot|^s, a^{2*}) (1 + o(1)) \quad (12)$$

and the series above is positive and finite.

Remark 6 (i) Notice that (11) and (12) 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.

Proof 1) By definition of $V_{a,n}$ in (8) and identity (9), we get

$$\mathbb{E}[V_{a,n}] = n' \mathbb{E}[\Delta_{a,i}(X)^2] = -n' \Delta_{a^{2*},0}(V) = -n' \sum_j a_j^{2*} V(j\Delta). \quad (13)$$

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

$$V((i + j)\Delta) = V(i\Delta) + \dots + \frac{(j\Delta)^{q-1}}{(q-1)!} V^{(q-1)}(i\Delta) + (j\Delta)^q \int_0^1 \frac{(1-\eta)^{q-1}}{(q-1)!} V^{(q)}((i + j\eta)\Delta) d\eta. \quad (14)$$

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

$$\begin{aligned} & (j\Delta)^q \int_0^1 \frac{(1-\eta)^{q-1}}{(q-1)!} V^{(q)}((i + j\eta)\Delta) d\eta \\ &= \frac{(j\Delta)^q}{(q)!} V^{(q)}(i\Delta) + \dots + \frac{(j\Delta)^{q'-1}}{(q'-1)!} V^{(q'-1)}(i\Delta) + (j\Delta)^{q'} \int_0^1 \frac{(1-\eta)^{q'-1}}{(q'-1)!} V^{(q')}((i + j\eta)\Delta) d\eta. \end{aligned} \quad (15)$$

Combining (14) (with $i = 0$ and $q = 2D$), the vanishing moments of the sequence a^{2*} and (\mathcal{H}_1) yields:

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

The first term is non-zero by (7) in Proposition 4 and a dominated convergence argument together with (\mathcal{H}_1) shows that the last term is $o(\Delta^{2D+s})$ giving (11).

2) Using Lemma 5, (14) with $q = 2D$, the fact that $D \leq M$, and the vanishing moments of the sequence a^{2*} , we obtain

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

where B_n comes from the double product.

(i) We show that A_n converges. Indeed,

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

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 (15), with $q' = 2M$, $|\cdot|^s$ instead of $V^{(2D)}$ and $\Delta = 1$, 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

$$R^2(i, 1, 2D, |\cdot|^s, a^{2*}) \text{ is bounded by } (Const)i^{2(s-2(M-D))}, \quad (16)$$

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

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

For fixed i , the assumptions (5) on r in (\mathcal{H}_1) are sufficient to build a dominated convergence argument to prove that $R^2(i, \Delta, 2D, r, a^{2*}) = o(\Delta^{2s})$ which leads to the required result. So we concentrate our attention on indices i such that $|i| > 2L$. Using (15) as in the proof of item 1), if $2D + d \leq 2M$, one gets

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

The condition $|i| > 2L$ ensures that the integral is always convergent. Using (5),

$$R^2(i, \Delta, 2D, r, a^{2*}) \leq (Const) \Delta^{2d+2\beta} i^{2\beta}. \quad (17)$$

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 $(Const) \Delta^{4D+2d+2\beta}$ which is negligible compared to Δ^{4D+2s} since $d + \beta > s$. \square

Following the same lines as in the proof of Proposition 5 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 7 (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*

$$\text{Cov}(V_{a,n}, V_{a',n}) = 2nC^2 \Delta^{4D+2s} \left[\sum_{i \in \mathbb{Z}} R^2(i, 1, 2D, |\cdot|^s, a * a') \right] (1 + o(1)). \quad (18)$$

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

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

(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

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

Now we establish the central limit theorem.

Theorem 8 (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

$$\frac{V_{a,n} - \mathbb{E}[V_{a,n}]}{\sqrt{\text{Var}(V_{a,n})}} \xrightarrow[n \rightarrow +\infty]{D} \mathcal{N}(0, 1). \quad (19)$$

Proof 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 Σ_a of $\Delta_a(X)$ and the Z_i are independent and identically distributed standard Gaussian variables. Hence,

$$\frac{V_{a,n} - \mathbb{E}(V_{a,n})}{\sqrt{\text{Var}(V_{a,n})}} = \sum_{i=1}^{n''} \frac{\lambda_i}{\sqrt{\sum_{r=1}^{n''} \lambda_r^2}} (Z_i^2 - 1). \quad (20)$$

In such a situation, Lemma 2 in [13] 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 6, 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 [18, Ch. 6.2, p194])

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

□

Remark 9

- If $M = D + 1$, the condition $M > D + s/2 + 1/4$ in Proposition 5 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,

$$\text{Var}(V_{a,n}) = (\text{Const}) \times \Delta^{4D+2s} \times n^{2s-4(M-D)+2} \times (1 + o(1)). \quad (21)$$

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

$$\text{Var}(V_{a,n}) = (\text{Const}) \times \Delta^{4D+2s} \times n \log n \times (1 + o(1)). \quad (22)$$

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

- Since the work of Guyon and León [12], it is a well known fact that in the simplest case ($D = 0, L = 2, M = 1$) and in the infill situation ($\alpha = 1$), 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.

Corollary 10 (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

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

converges to an invertible matrix Λ_∞ . 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 $n \rightarrow \infty$

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

Proof 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 λ_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,

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

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 8, we have established that $\|\Sigma_{a^{(j)},n}\|_{op} = o(n^{1/2}\Delta^{2D+s})$ leading to the result. \square

3.3 Estimators of C based on the quadratic a -variations

Guided by the moment method, we define

$$C_{a,n} := \frac{V_{a,n}}{n(-1)^D \Delta^{2D+s} R(0, 1, 2D, |\cdot|^s, a^{2*})}. \quad (23)$$

Then $C_{a,n}$ is an estimator of C which is asymptotically unbiased by Proposition 5. Now our aim is to establish its asymptotic behavior.

Theorem 11 (Central limit theorem for $C_{a,n}$) *Under the assumptions (\mathcal{H}_0) to (\mathcal{H}_3) , and if $M(a) > D + s/2 + 1/4$, then $C_{a,n}$ is asymptotically normal. More precisely, we have*

$$\frac{C_{a,n} - C}{\sqrt{\text{Var}(C_{a,n})}} \xrightarrow[n \rightarrow +\infty]{D} \mathcal{N}(0, 1), \quad (24)$$

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

Proof 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 5, the second term is proportional to

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

which converges to 0 as n goes to infinity by assumption (\mathcal{H}_3) . Then Slutsky's lemma and Theorem 8 lead straightforwardly to the required result. \square

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 12 *Under the assumptions of Theorem 11, consider k sequences $a^{(1)}, \dots, a^{(k)}$ so 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 Γ_∞ 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 Kriging applications, we assume that f is smooth.

Corollary 13 *Define*

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

Under the assumptions of Theorem 11 and if we assume in addition that

$$K_{M,n}^\alpha = o(n^{-1/4} \Delta^{D-M+s/2}), \quad (25)$$

then (24) still holds for X .

Note that in the infill situation ($\alpha = 1$), $K_{M,n}^1$ does not depend on n . Obviously, (25) 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 (25) 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.

Proof Obviously, one has

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

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

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

with $\xi \in [0, n^{1-\alpha}]$. Then $\|\Delta_{\mathbf{a}}(f)\|^2 \leq n(K_{M,n}^\alpha)^2 \Delta^{2M}$ and a sufficient condition is (25). \square

3.5 Elements of comparison with the MLE procedure

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. [21, 25].

Model flexibility As mentioned in the introduction, the MLE methodology is a parametric method and requires the kernel to belong to a parametric family of the form $\{\sigma^2 \rho_\theta; \sigma^2 \geq 0, \theta \in \Theta\}$. In the procedure proposed in this paper, it is only assumed that the 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.

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 β_1, \dots, β_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 (25).

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 $O(n^3)$.

Practical issues In some real data frameworks, it may occur that the MLE estimation diverges as can be seen in Section 5.3. Such a dead end can not be possible with our procedure.

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

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 $\lambda_1^*, \dots, \lambda_k^*$. 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 (23). Then we provide the optimal weights $\lambda_1^*, \dots, \lambda_k^*$. Using [16] or [4], one can establish the following lemma.

Lemma 2 *We assume that the conditions of Corollary 12 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 $\lambda_1, \dots, \lambda_k$ for which $\lambda_1 + \dots + \lambda_k = 1$,*

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

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

$$\lambda^* := (\lambda_1^*, \dots, \lambda_k^*)^\top = \frac{R^{-1} \mathbf{1}_k}{\mathbf{1}_k^\top R^{-1} \mathbf{1}_k}.$$

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

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

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 this 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

$$(R_C)_{i-1, j-1} = \text{Cov}(Y_C(i\Delta) - Y_C((i-1)\Delta), Y_C(j\Delta) - Y_C((j-1)\Delta)), \quad i, j = 2, \dots, n.$$

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

$$I_C = \frac{1}{2} \text{Tr} \left(R_C^{-1} \left(\frac{\partial}{\partial C} R_C \right) R_C^{-1} \left(\frac{\partial}{\partial C} R_C \right) \right) \quad (26)$$

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

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

(see for instance [3, 8]). 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 3 *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 variogram V_C is given by*

$$V_C(h) = \frac{1}{2} \mathbb{E} \left[(Y_C(t+h) - Y_C(t))^2 \right] = C|h|^s. \quad (27)$$

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

Proof (27) implies that $\partial R_C / \partial C = R_1$ then (26) gives the result. \square

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 covariance function ρ_C defined by

$$\rho_C(h) = (1 - (C/2)|h|^s)^+, \quad \text{for any } h \in \mathbb{R}. \quad (28)$$

This function is convex on \mathbb{R} and it follows from Pólya's theorem [20] that ρ_C is a valid covariance function. We thus easily obtain the following lemma whose proof is omitted.

Lemma 4 *Let X be the integration D times of Y_C defined via (28). Then, in the infill situation ($\alpha = 1$) and for $C < 2$, the variogram of Y_C is given by (27) 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 5 and Theorem 11 are representative of the finite sample behavior of quadratic a -variations estimators. Then, we study the asymptotic variances of these estimators provided by Proposition 5 and that of the aggregated a -variations estimators of Section 4.1. We conclude this section extending our empirical procedure to the multidimensional setting on two real data sets.

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 covariance function ρ . In the case $D = 0$, we let $\rho(h) = \exp(-C|h|)$. Hence (\mathcal{H}_1) holds with $D = 0$ and $s = 1$. In the case $D = 1$, we use the Matérn 3/2 covariance [24] :

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

One can show, by developing ρ 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 covariance function:

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

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 1, 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 5 and Theorem 11. We observe that there are few differences between the histograms and limit 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 that 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.

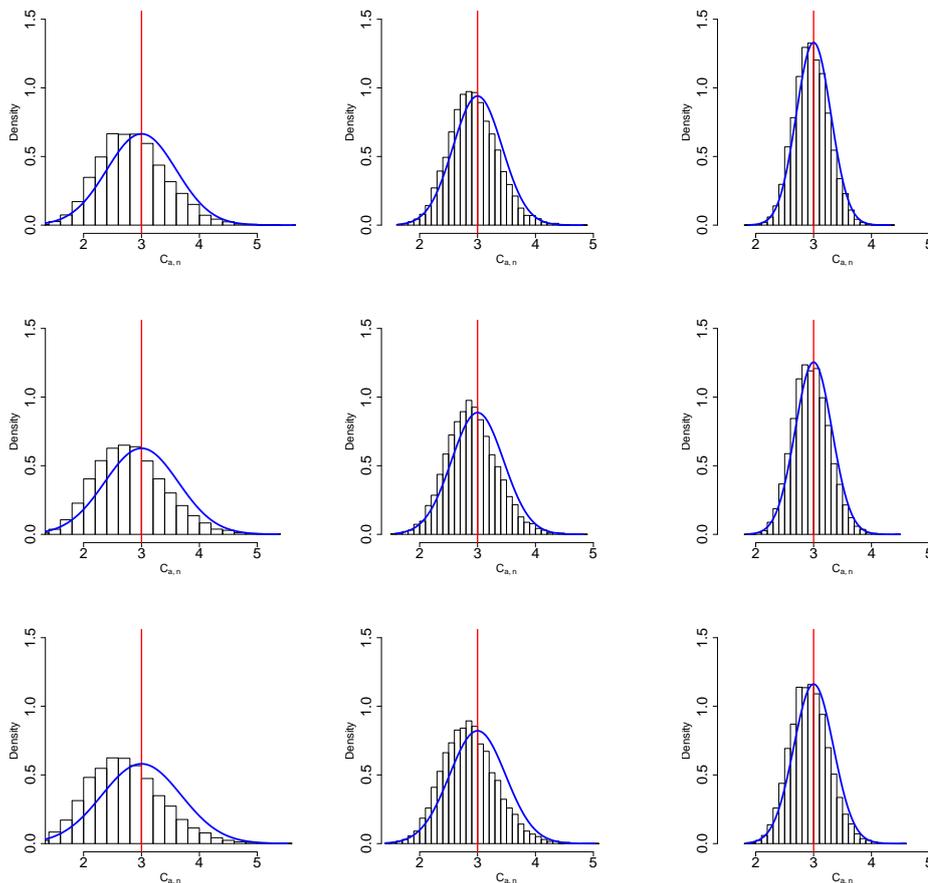


Fig. 1 Comparison of the finite sample distribution of $C_{a,n}$ (histograms) with the asymptotic Gaussian distribution provided by Proposition 5 and Theorem 8 (probability density function in blue line). The vertical red line denotes the true value of $C = 3$. From left to right, $n = 50, 100, 200$. From top to bottom, $D = 0, 1, 2$.

5.2 Analysis of the asymptotic distributions

Now we consider the normalized asymptotic variance of $C_{a,n}$ obtained from (12) in Proposition 5. We let $\Delta = 1/n$ and

$$\tilde{v}_{a,s} = \frac{2 \sum_{i \in \mathbb{Z}} R^2(i, 1, 2D, |\cdot|^s, a^{2*})}{R^2(0, 1, 2D, |\cdot|^s, a^{2*})}, \quad (29)$$

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 the left hand side of Figure 2. 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 with $M = 2$ [9] as in [13]: $a^{(6)}$ 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 the left hand side of Figure 2, 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.

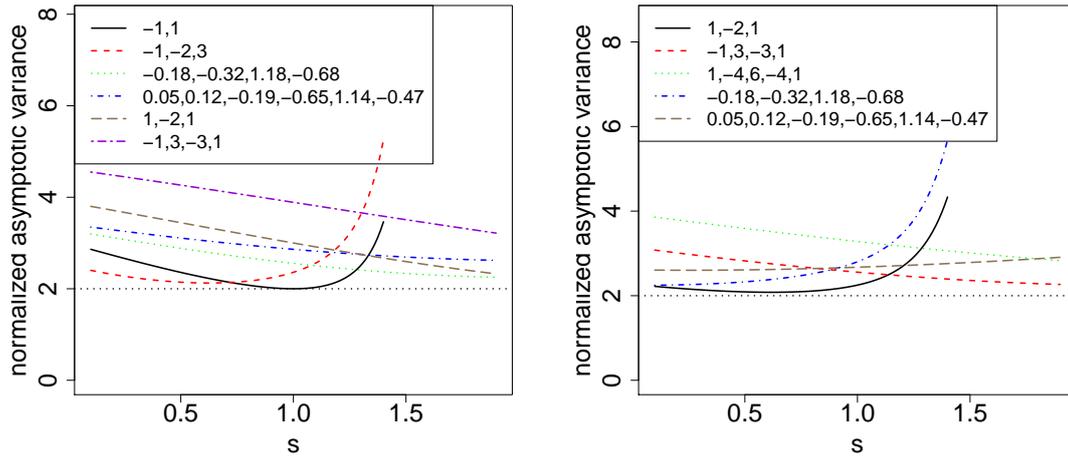


Fig. 2 Case $D = 0$ (left hand side) and $D = 1$ (right hand side). Plot of the normalized asymptotic variance $\tilde{v}_{a,s}$ of the quadratic a -variations estimator, as a function of s , for various sequences a . The legend shows the values a_0, \dots, a_l of these sequences (rounded to two digits). For $D = 0$, 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 d Daubechies sequences of order 2 and 3 and the elementary sequences of orders 2 and 3. 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. The horizontal line corresponds to the Cramér-Rao lower bound 2.

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 9: 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 the left hand side of Figure 2 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.

Now, we consider aggregated estimators, as presented in Section 4.1. A clear motivation for considering aggregation is that, in the left hand side of Figure 2, the smallest asymptotic variance $\tilde{v}_{a,s}$ corresponds to different sequences a , depending on the values of s .

In Figure 3 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. 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

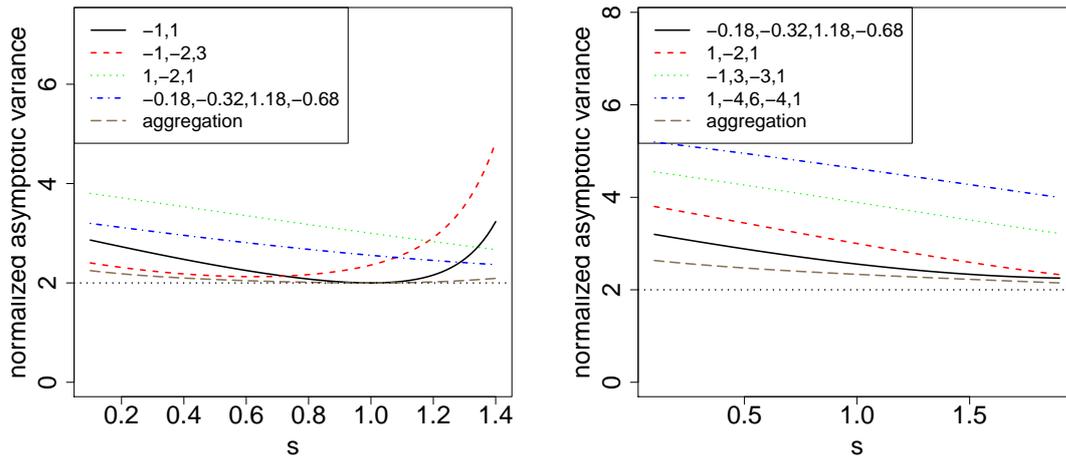


Fig. 3 Case $D = 0$. Plot of the normalized asymptotic variance $\hat{v}_{a,s}$ of the quadratic a -variations estimator, as a function of s , for various sequences a and for their aggregation. On the left, including the order one elementary sequence, on the right without. The horizontal line corresponds to the Cramér-Rao lower bound 2.

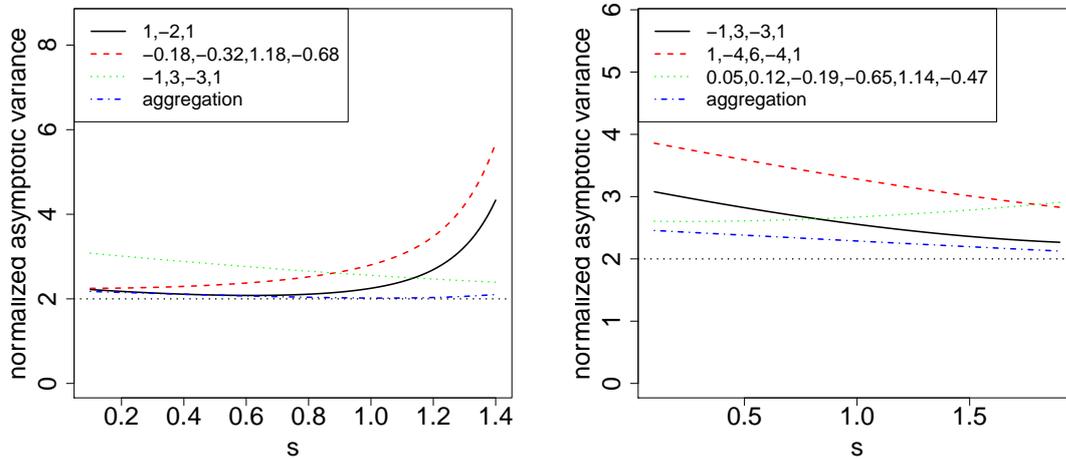


Fig. 4 Same setting as in Figure 3 but for $D = 1$. On the left, from top to bottom in the legend, the sequences are the elementary sequence of order 2, the Daubechies sequence of order 2 and the elementary sequence of order 3. On the right, from top to bottom in the legend, the sequences are the elementary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

lower bound 2 for all the values of s . In Figure 3 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.

Finally, the right hand side of Figure 2 and Figure 4 explore the case $D = 1$. Conclusions are similar.

5.3 Real data examples

In this section, we illustrate in two real data sets a possible extension of the estimation procedure in a two dimensional setting. Moreover, we show that our procedure based on quadratic a -variations outperforms the maximum likelihood procedure when the dimension is larger than 2.



Fig. 5 For the data set of Section 5.3.2: 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.

5.3.1 A moderate size data set

We compare two methods of estimation of the covariance function of a separable Gaussian model on a real data set of atomic force spectroscopy¹. 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` or the R toolbox `DiceKriging` [24]. For this method, the mean is assumed to be $\mathbb{E}(X(i/15, j/15)) = \mu$ and the covariance functions are assumed to be exponential (see Section 2.2):

$$\text{Cov}(X(i/15, j/15), X(i'/15, j'/15)) = \sigma^2 e^{-\theta_1|i-i'|/15} e^{-\theta_2|j-j'|/15}. \quad (30)$$

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

The second method assumes the same covariance model (30) and consists in the following steps.

- (1) Estimate σ^2 by the sum of squares

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

with $\hat{\mu} = (1/256) \sum_{i,j=0}^{15} X(i/15, j/15)$.

- (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 (23), with the elementary sequence of order 1. Thus we 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 $i = 1, 2$, θ_i is estimated by $\hat{\theta}_i = \hat{C}_i / \hat{\sigma}^2$.

The first method, based on maximum likelihood, provides infinite values for θ_1 and θ_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 a -variations estimator, is able to detect a weak correlation (that can be checked graphically) unlike the maximum likelihood estimator.

5.3.2 A large size data set

The second data set is 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 [22]. Figure 5 displays the two images to be registered and the deformation field.

¹ Personal communication from C. Gales and J. M. Senard.

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 covariance model as in Section 5.3.1 and we estimate the parameters θ_1 and θ_2 from the same two methods as in Section 5.3.1. The first method provides $\hat{\theta}_1 = 0.8770$ and $\hat{\theta}_2 = 0.6547$ and takes about 22 minutes on a personal computer, while the second method provides $\hat{\theta}_1 = 0.607$ and $\hat{\theta}_2 = 0.107$ in 0.05 seconds. Hence, our suggested quadratic a -variations 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 5. As in Section 5.3.1, the maximum likelihood estimator provides less correlation than the quadratic a -variations estimator.

Finally, if the field of deformation is considered with no preliminary subsampling, its size is 400×600 . In this case, the ML estimator can not be directly implemented while the quadratic a -variations estimator can be.

6 Appendix and technical results

Lemma 5 *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 This Lemma is a consequence of the so called Mehler formula [2]. Its proof is immediate using the cumulant method. \square

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

$$\begin{aligned} K^{(-(m+1))}(u, v) &= \sum_{i=1}^{\tilde{N}_{m+1}} \tilde{P}^{m+1,i}(v) \tilde{h}_{m+1,i}(u) + \sum_{i=1}^{\tilde{N}_{m+1}} \tilde{P}^{m+1,i}(u) \tilde{h}_{m+1,i}(v) \\ &\quad + (-1)^{m+1} \frac{1}{(s+1) \dots (s+2m)} \int_0^v \left(\int_0^u |x - y|^{s+2m} dx \right) dy, \end{aligned} \quad (31)$$

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 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 $u, v \in \mathbb{N}$,

$$\int_0^u \left(\int_0^v |x - y|^{s+2m} dx \right) dy = \frac{u^{s+2m+2}}{(2m+1)(2m+2)} + \frac{v^{s+2m+2}}{(2m+1)(2m+2)} - \frac{|u - v|^{s+2m+2}}{(2m+1)(2m+2)}. \quad (32)$$

Hence, from the relation

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

(31), and (32), we conclude the proof of the lemma. \square

Lemma 6 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 Using the stationary increments of the process, one has

$$\max_{i=1,\dots,n'} \left(\sum_{i'=1,\dots,n'} |\Sigma_a(i, i')| \right) \leq 2 \sum_{i=0}^{n'-1} |\Sigma_a(1, 1+i)|. \quad (33)$$

Recall that

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

We have seen in the proof of Proposition 5 ((16) and (17)) that for i sufficiently large

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

Thus the sum in (33) is bounded by

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

On the other hand, we have proved also in the proof of Proposition 5 that

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

giving the result. Thus, one has to check that

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

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

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 very grateful to Laurent Risser, for providing them the image registration data of Section 5.3.2.

References

1. Adler RJ, Pyke R (1993) Uniform quadratic variation for Gaussian processes. *Stochastic Process Appl* 48(2):191–209, DOI 10.1016/0304-4149(93)90044-5, URL [http://dx.doi.org/10.1016/0304-4149\(93\)90044-5](http://dx.doi.org/10.1016/0304-4149(93)90044-5)
2. Azaïs JM, Wschebor M (2009) Level sets and extrema of random processes and fields. John Wiley & Sons, Inc., Hoboken, NJ, URL <https://doi.org/10.1002/9780470434642>
3. Bachoc F (2014) Asymptotic analysis of the role of spatial sampling for covariance parameter estimation of Gaussian processes. *Journal of Multivariate Analysis* 125:1–35
4. Bates JM, Granger CW (1969) The combination of forecasts. *Journal of the Operational Research Society* 20(4):451–468
5. Baxter G (1956) A strong limit theorem for Gaussian processes. *Proc Amer Math Soc* 7:522–527, DOI 10.2307/2032765, URL <http://dx.doi.org/10.2307/2032765>
6. Coeurjolly JF (2001) Estimating the parameters of a fractional Brownian motion by discrete variations of its sample paths. *Stat Inference Stoch Process* 4(2):199–227, DOI 10.1023/A:1017507306245, URL <http://dx.doi.org/10.1023/A:1017507306245>
7. Cohen S, Istas J (2013) Fractional fields and applications, *Mathématiques & Applications (Berlin) [Mathematics & Applications]*, vol 73. Springer, Heidelberg, DOI 10.1007/978-3-642-36739-7, URL <https://doi.org/10.1007/978-3-642-36739-7>, with a foreword by Stéphane Jaffard
8. Dahlhaus R (1989) Efficient parameter estimation for self-similar processes. *The annals of Statistics* pp 1749–1766
9. Daubechies I (1988) Orthonormal bases of compactly supported wavelets. *Communications on pure and applied mathematics* 41(7):909–996

10. Gladýšev EG (1961) A new limit theorem for stochastic processes with Gaussian increments. *Teor Verojatnost i Primenen* 6:57–66
11. Grenander U (1981) *Abstract inference*. John Wiley & Sons, Inc., New York, wiley Series in Probability and Mathematical Statistics
12. Guyon X, León J (1989) Convergence en loi des H -variations d'un processus gaussien stationnaire sur \mathbf{R} . *Ann Inst H Poincaré Probab Statist* 25(3):265–282, URL http://www.numdam.org/item?id=AIHPB_1989__25_3_265_0
13. Istas J, Lang G (1997) Quadratic variations and estimation of the local Hölder index of a Gaussian process. *Ann Inst H Poincaré Probab Statist* 33(4):407–436, DOI 10.1016/S0246-0203(97)80099-4, URL [http://dx.doi.org/10.1016/S0246-0203\(97\)80099-4](http://dx.doi.org/10.1016/S0246-0203(97)80099-4)
14. Kent JT, Wood ATA (1997) Estimating the fractal dimension of a locally self-similar Gaussian process by using increments. *J Roy Statist Soc Ser B* 59(3):679–699
15. Lang G, Roueff F (2001) Semi-parametric estimation of the Hölder exponent of a stationary Gaussian process with minimax rates. *Stat Inference Stoch Process* 4(3):283–306, DOI 10.1023/A:1012227325436, URL <http://dx.doi.org/10.1023/A:1012227325436>
16. Lavancier F, Rochet P (2016) A general procedure to combine estimators. *Computational Statistics & Data Analysis* 94:175–192
17. Lévy P (1940) Le mouvement brownien plan. *Amer J Math* 62:487–550, DOI 10.2307/2371467, URL <http://dx.doi.org/10.2307/2371467>
18. Luenberger D (1979) *Introduction to Dynamic Systems: Theory, Models, and Applications*. Wiley, URL <https://books.google.fr/books?id=mvlQAAAAAAAJ>
19. Perrin O (1999) Quadratic variation for Gaussian processes and application to time deformation. *Stochastic Process Appl* 82(2):293–305, DOI 10.1016/S0304-4149(99)00037-X, URL [http://dx.doi.org/10.1016/S0304-4149\(99\)00037-X](http://dx.doi.org/10.1016/S0304-4149(99)00037-X)
20. Pólya G (1949) 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. Statistical Laboratory of the University of California, Berkeley. Berkeley, Calif.: University of California Press, 1949. 501 pp. Editor: Jerzy Neyman, p. 115–123, pp 115–123
21. Rasmussen C, Williams C (2006) *Gaussian Processes for Machine Learning*. The MIT Press, Cambridge
22. Risser L, Vialard F, Wolz R, Murgasova M, Holm D, Rueckert D (2011) ADNI: Simultaneous multiscale registration using large deformation diffeomorphic metric mapping. *IEEE Transactions on Medical Imaging* 30(10):1746–1759
23. Roustant O, Ginsbourger D, Deville Y (2012) Dicekriging, diceoptim: Two r packages for the analysis of computer experiments by kriging-based metamodeling and optimization. *Journal of Statistical Software*, Articles 51(1):1–55, DOI 10.18637/jss.v051.i01, URL <https://www.jstatsoft.org/v051/i01>
24. Roustant O, Ginsbourger D, Deville Y (2012) DiceKriging, DiceOptim: Two R packages for the analysis of computer experiments by Kriging-based metamodeling and optimization. *Journal of Statistical Software* 51(1)
25. Santner TJ, Williams BJ, Notz WI (2003) *The design and analysis of computer experiments*. Springer Series in Statistics, Springer-Verlag, New York, DOI 10.1007/978-1-4757-3799-8, URL <https://doi.org/10.1007/978-1-4757-3799-8>
26. Slepian D (1963) On the zeros of Gaussian noise. In: *Proc. Sympos. Time Series Analysis* (Brown Univ., 1962), Wiley, New York, pp 104–115
27. Stein M (1999) *Interpolation of Spatial Data: Some Theory for Kriging*. Springer, New York