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

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¶

June 8, 2018

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.

Keywords: quadratic variations, scale covariance parameter, asymptotic normality, moment method, aggregation of estimators.

1 Introduction

Gaussian processes models are widely used in spatial statistics and in particular to interpolate observations by Kriging. For example, this technique is used in computer experiment designs to build a meta-model [21, 24]. Usually the practitioner uses a model including a drift (often polynomial) and a stationary Gaussian model whose covariance belongs to some family, e.g. Matérn or exponential. In this paper, we limit the framework to unidimensional situations: we consider a real-valued process X on $\mathbb{R}$ for which

$$\mathrm{Cov}(X(s), X(t)) = f(C, s-t), \tag{1}$$

where the function f belongs to the prescribed class of covariance functions and the constant C is the unknown scaling parameter. In applications, the estimation of the parameters C is a crucial step since it constitutes a necessary preliminary to Kriging. Most of the software packages use the maximum likelihood method which is known to be computer intensive and may diverge in some complicated situations.

The aim of this paper is to propose another method of estimation based on quadratic variations. Quadratic variations have been first introduced by Levy in [17] that shows that,

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

where Z is the standard Gaussian process 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 $V_{1,n}$ defined by

$$V_{1,n} := \sum_{i=1}^{n-1} \left(X((i+1)\Delta_n) - X(i\Delta_n) \right)^2 \tag{2}$$

*Institut de Mathématiques de Toulouse; UMR5219. Université de Toulouse; CNRS. UT3, F-31062 Toulouse, France.
†Institut de Mathématiques de Toulouse; UMR5219. Université de Toulouse; CNRS. UT3, F-31062 Toulouse, France.
‡Institut de Mathématiques de Toulouse; UMR5219. Université de Toulouse; CNRS. ENAC, F-31400 Toulouse, France.
§Institut de Mathématiques de Toulouse; UMR5219. Université de Toulouse; CNRS. UT2J, F-31058 Toulouse, France.
¶VNUHCM - University of Science, Ho Chi Minh city, Viet Nam.

where the scale Δ_n tends to zero as $n \to +\infty$. A generalization of the quadratic variations $V_{1,n}$ has been introduced in Guyon and Léon [12]. For a given real function H, the H-variation is given by

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

where X is assumed to be a centered stationary Gaussian process in [12]. In fact for statistical purposes, it has been proved by Coeurjolly in [6] that quadratic variations are optimal. So we will limit our attention to this last case. The most unexpected result of [12] is that, if the local irregularity of the process defined by $\rho(h) = \mathrm{Cor}(X(t+h), X(t))$ 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, then

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

In [13], Istas and Lang generalize the results on quadratic variations of [12]. They consider a Gaussian process with stationary increments and observations of the process at times Δj for $j = 1, \ldots, n$, with Δ dependent on n and study a generalized quadratic variation:

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

where the sequence a has a finite support and some vanishing moments. Then they build estimators for the local Hölder index and the constant C and showed that they are almost surely consistent and asymptotic normal. In the more recent work of Lang and Roueff [15], the authors generalize 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 process X is observed at times $i\Delta_n$ for $i = 1, \ldots, n$ with Δ_n tending to zero. The paper is devoted to the estimation of the parameter C in (1) from one or several generalized quadratic a-variations of the type (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.

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 variations associated to several sequences? As long as we know, these questions are not addressed yet in the literature.

The first study of this paper in Section 3 is close to the study of Istas and Lang [13]. It establishes the expectation, the variance and a central limit theorem for the variations and the estimators deduced from them. Nevertheless, since we want to estimate the constant C only and not the local Hölder index, our hypotheses and proofs are simpler than thoseobtained in [13]. Moreover, one can easily check that our hypotheses are satisfied in all the commonly used Kriging models. In addition, we compute in Section4 the Cramér-Rao bound in this setting. Concerning the crucial choice of the sequence a, unfortunately the asymptotic variance given by Proposition 3.1 or Theorem 3.8 does not allow to address theoretically this issue. Thus, an important Monte-Carlo study is performed in Section 5. The main conclusion is that, if we aggregate the information of different a-variations with different orders, the results are close to the optimal Cramér-Rao bound as it can be seen in Figure 3. With this point of view, the choice of the sequence does not matter. In addition, our method does not require a parametric specification of the drift, see Section 3.4 and then is more robust than maximum likelihood.

2 General setting and assumptions

2.1 Assumptions on the process

In this paper, we consider actually a slightly more general framework than that in the introduction since we only assume that the Gaussian process $(X(t))_{t\in\mathbb{R}}$ has stationary increments. The process is observed at times $j\Delta$ for $j = 0, \ldots, n$ with $\Delta = \Delta_n$ going to 0 as n goes to infinity. Its variogram is then defined by

$$V(h) := \frac{1}{2}\mathbb{E}\left[(X(t+h) - X(t))^2\right]. \tag{5}$$

In the sequel, we let $\Delta = n^{-\alpha}$, $0 < \alpha \leqslant 1$ and we denote by $(Const)$ a positive constant which value may change from one occurrence to another. Note that the case $\alpha = 1$ then corresponds to the infill situation [24]. 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 $D \geqslant 0$ and there exists $C > 0$ and $0 < s < 2$ such that for any $h \in \mathbb{R}$, we have

$$V^{(2D)}(h) = V^{(2D)}(0) + C(-1)^D |h|^s + r(h), \text{ with } r(h) = o(|h|^s) \text{ and } |r(h)| \leqslant (Const)\,|h|^s. \tag{6}$$

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

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

Remark 2.1.

- 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 $\rho(h) = \mathrm{Cov}(X(t), X(t+h)) = (-1)^{D+1}V^{(2D)}(h)$. 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

All classical spatial models satisfy our hypotheses, except the Gaussian model which is too regular. Here is a non exhaustive list in dimension 1:

- the exponential model: $\rho(h) = \exp(-C|h|)$ $(D = 0,\ s = 1)$;
- the generalized exponential model: $\rho(h) = \exp(-(C|h|)^s), s \in (0, 2)$ $(D = 0,\ s = s)$;
- the generalized Slepian model [23]: $\rho(h) = (1 - (C|h|)^s)^+, s \in (0, 1]$ $(D = 0,\ s = s)$;
- the spherical model: $\rho(h) = (1 - C|h| + 0.5(\theta|h|)^3)^+$ $(D = 0,\ s = 1)$;
- the cubic model $\rho(h) = (1 - 3(\theta|h|)^2 + 2(\theta|h|)^3)^+$ $(D = 1,\ s = 1)$;
- the Matérn model:
$$\rho(h) = \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\sqrt{2\nu}\theta h\right)^\nu K_\nu(\sqrt{2\nu}\theta h),$$
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., [24] for more details on the model. In that case, $2D + s = \nu$;
- the fractional Brownian motion (FBM) process denoted by $(B_s(t))_{t\in\mathbb{R}}$ and defined by
$$\mathrm{Cov}(B_s(u), B_s(t)) = \left(|u|^s + |t|^s - |u-t|^s\right).$$
A reference on this subject is [7]. This process is classically indexed by its Hurst parameter $H = s/2$ and a multiplicative variance σ^2 is often introduced but we do not need it. Here, $D = 0$, $s = s$.

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 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:

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

To any sequence a, with length $L(a)$, we associate its discrete a-difference defined by

$$\Delta_{a,i}(f) = \sum_{j=0}^{L(a)-1} a_j f((i+j)\Delta), \quad i = 1, \dots n', \tag{7}$$

for a function f where n' stands for $n-L(a)+1$. As a matter of facts, $\sum_{j=0}^{L(a)-1} a_j f(j\Delta)$ 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 $\mathbf{\Delta_a}(X)$ as the Gaussian vector 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 a^{2*} the convolution $a * a$.

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

(i) The support of $a*b$ (the indices of the non-zero elements) is included in $-(L(b)-1), (L(a)-1)$ while its order is $M(a)+M(b)$. 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.

To state our next result, 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

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

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

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

Definition 2.3. A process Z has the ND property if for every $k > 0$ and every $t_1 < t_2 < \cdots < t_k$ 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 2.4. *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 positivity 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.

□

We also need the following lemma.

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

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

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

The proof of Lemma 2.5 is given in the appendix in Section 6.

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

$$\sum_j a_j^{2*} |j|^{2D+s} \neq 0 \quad (i.e. \quad (-1)^D \sum_j a_j^{2*} |i|^{2D+s} < 0). \tag{8}$$

Note that (8) is stated as an hypothesis in [13].

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

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

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

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

$$V_{a,n} = \|\boldsymbol{\Delta}_{\mathbf{a}}(X)\|^2 = \sum_{i=1}^{n'} (\Delta_{a,i}(X))^2, \tag{9}$$

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

$$\mathbb{E}[\Delta_{a,i}(X)\Delta_{a',i'}(X)] = -\Delta_{a*a',i-i'}(V), \tag{10}$$

for any sequences a and a'. A second main tool is the Taylor expansion with integral remainder (see, for example, (15)). So we introduce another notation. For a sequence a, a scale Δ, an order q and a function f, we define

$$R(i,\Delta,q,f,a) = -\sum_j a_j j^q \int_0^1 \frac{(1-\eta)^{q-1}}{(q-1)!} f((i+j\eta)\Delta)d\eta. \tag{11}$$

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 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*

$$\mathbb{E}[V_{a,n}] = nC(-1)^D\Delta^{2D+s}\left[R(0,1,2D,|\cdot|^s,a^{2*})\right](1+o(1)), \tag{12}$$

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:*

$$\mathrm{Var}(V_{a,n}) = 2nC^2\Delta^{4D+2s}\sum_{i\in\mathbb{Z}} R^2(i,1,2D,|\cdot|^s,a^{2*})(1+o(1)) \tag{13}$$

and the series above is positive and finite.

Remark 3.2. (i) Notice that (12) and (13) imply concentration in the sense that

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

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

Proof. **1)** By definition of $V_{a,n}$ in (9) and identity (10), we get

$$\mathbb{E}[V_{a,n}] = n'\mathbb{E}[\Delta_{a,i}(X)^2] = -n'\Delta_{a^{2*},0}(V) = -n'\sum_j a_j^{2*}V(j\Delta). \tag{14}$$

Recall that $n' = n - L + 1$ is the size of the vector $\boldsymbol{\Delta}_{\mathbf{a}}(X)$. 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 $q \leqslant 2D$:

$$V((i+j)\Delta) = V(i\Delta) + \cdots + \frac{(j\Delta)^{q-1}}{(q-1)!}V^{(q-1)}(i\Delta) + (j\Delta)^q\int_0^1 \frac{(1-\eta)^{q-1}}{(q-1)!}V^{(q)}((i+j\eta)\Delta)d\eta. \tag{15}$$

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

$$\begin{aligned}&(j\Delta)^q\int_0^1\frac{(1-\eta)^{q-1}}{(q-1)!}V^{(q)}((i+j\eta)\Delta)d\eta\\ &= \frac{(j\Delta)^q}{(q)!}V^{(q)}(i\Delta) + \cdots + \frac{(j\Delta)^{q'-1}}{(q'-1)!}V^{(q'-1)}(i\Delta) + (j\Delta)^{q'}\int_0^1\frac{(1-\eta)^{q'-1}}{(q'-1)!}V^{(q')}((i+j\eta)\Delta)d\eta. \end{aligned}\tag{16}$$

Combining (15) (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 Proposition 8 and a dominated convergence argument together with $(\mathcal{H}_1)$ shows that the last term is $o(\Delta^{2D+s})$ giving (12).

2) Using Lemma 6.1, (15) with $q = 2D$, the fact that $D \leqslant M$, and the vanishing moments of the sequence a^{2*}, we obtain

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

where B comes from the double product.

(i) We study the first term:

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

with

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

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 (16) , 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

$$
R^2(i, 1, 2D, |\cdot|^s, a^{2*}) \text{ is bounded by } (Const) i^{2(s-2(M-D))}, \tag{17}
$$

which is the general term of a convergent series.

(ii) Now we show that the term C is negligible compared to A. This will imply in turn that B is negligible compared to A, 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

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

For fixed i, the assumptions (6) 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 (16) as in the proof of item **1)**, if $2D + d \leqslant 2M$, one gets

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

The condition $|i| > 2L$ ensures that the integral is always convergent. Using (6),

$$R^2(i, \Delta, 2D, r, a^{2*}) \leqslant (Const)\Delta^{2d+2\beta} i^{2\beta}. \tag{18}$$

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 $d + \beta > s$. □

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*

$$\mathrm{Cov}(V_{a,n}, V_{a',n}) = 2nC^2\Delta^{4D+2s}\left[\sum_{i\in\mathbb{Z}} R^2(i, 1, 2D, |\cdot|^s, a * a')\right](1 + o(1)). \tag{19}$$

Particular case - $D = 0$:

(i) We choose a as the first order elementary sequence ($a_0 = -1$, $a_1 = 1$ and $M = 1$). One has

$$\mathbb{E}[V_{a,n}] = nC\Delta^s(2 + o(1));$$
$$\mathrm{Var}(V_{a,n}) = 2nC^2\Delta^{2s}\sum_{i\in\mathbb{Z}}\left(|i-1|^s - 2|i|^s + |i+1|^s\right)^2(1 + o(1)),\ s < 3/2$$

as n tends to infinity.

(ii) General sequence. We choose two sequences a and a' so that $M(a) \wedge M(a') > s/2 + 1/4$. Then

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

as n tends to infinity.

Remark 3.4. 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 $s \geqslant 3/2$, it is still possible to compute the variance but the speed is worse 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,

$$\mathrm{Var}(V_{a,n}) = (Const) \times \Delta^{4D+2s} \times n^{2s-4(M-D)+2} \times (1 + o(1)). \tag{20}$$

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

$$\mathrm{Var}(V_{a,n}) = (Const) \times \Delta^{4D+2s} \times n\log n \times (1 + o(1)). \tag{21}$$

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

Now we establish the central limit theorem.

Theorem 3.5 (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*

$$\frac{V_{a,n} - \mathbb{E}[V_{a,n}]}{\sqrt{\mathrm{Var}(V_{a,n})}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1). \tag{22}$$

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

$$\frac{V_{a,n} - \mathbb{E}(V_{a,n})}{\sqrt{\mathrm{Var}(V_{a,n})}} = \sum_{i=1}^{n''} \frac{\lambda_i}{\sqrt{\sum_{r=1}^{n''} \lambda_r^2}} (Z_i^2 - 1). \tag{23}$$

In such a situation, the Lindeberg condition is a sufficient condition required to prove the central limit theorem and is equivalent to

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

see Lemma 2 in [13]. From Lemma 6.2, one has

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

and the result follows using the following classical linear algebra result

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

See [18, Ch. 6.2, p194]. ☐

Remark 3.6. 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 $s < 3/2$.

Corollary 3.7 (Joint central limit theorem)**.** *Assume that* V *satisfies* $(\mathcal{H}_0)$*,* $(\mathcal{H}_1)$ *and* $(\mathcal{H}_2)$*. Let* $a^{(1)}, \ldots, a^{(k)}$ *be* k *sequences with order greater than* $D + s/2 + 1/4$*. Assume also that, as* $n \to \infty$*, the* $k \times k$ *matrix with term* i, j *equal to*

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

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

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

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

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

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

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

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,

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

with obvious notation. On the other hand, by the triangular inequality for the operator norm (which is the maximum of the $|\lambda_i|$'s), one gets

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

In the proof of Theorem 3.5, we have established that $\|\Sigma_{a^{(j)},n}\|_{op} = o(n^{1/2}\Delta^{2D+s})$ leading to the result. □

3.3 Estimators of C based on the quadratic a-variations

Guided by the moment method, we define

$$C_{a,n} := \frac{V_{a,n}}{n(-1)^D \Delta^{2D+s} R(0, 1, 2D, |\cdot|^s, a^{2*})}. \tag{24}$$

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 3.8 (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*

$$\frac{C_{a,n} - C}{\sqrt{\operatorname{Var}(C_{a,n})}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1). \tag{25}$$

By the definition (24) *of* $C_{a,n}$ *and Proposition 3.1,* $\operatorname{Var}(C_{a,n}) = (Const)n^{-1}(1 + o(1))$.

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

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

Following the proof of Proposition 3.1, 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 lemma and Theorem 3.5 lead straightforwardly to the required result. □

Corollary 3.9. *Under the assumptions of Theorem 3.8, consider* k *sequences* $a^{(1)}, \ldots, a^{(k)}$ *so that, for* $i = 1, \ldots, k$*,* $M(a^{(i)}) > D+s/2+1/4$*. Assume furthermore that the covariance matrix of* $(C_{a^{(i)},n}/\operatorname{Var}(C_{a^{(i)},n})^{1/2})_{i=1,\ldots,k}$ *converges to an invertible matrix* Γ_∞ *as* $n \to \infty$*. Then,* $([C_{a^{(i)},n} - C]/\operatorname{Var}(C_{a^{(i)},n})^{1/2})_{i=1,\ldots,k}$ *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 $t \geqslant 0$,

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

We write $X(t) = f(t) + \overline{X}(t)$. As it is always the case in Kriging applications, we assume that f is smooth.

Corollary 3.10. *Under the assumptions of Theorem 3.8. Define*

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

In addition, if

$$K^{\alpha}_{M,n} = o(n^{-1/4} \Delta^{D-M+s/2}), \tag{26}$$

then (25) still holds for X.

Note that in the infill situation ($\alpha = 1$), $K^1_{M,n}$ does not depend on n. Obviously, (26) 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 (26) is $M > D + s/2 + 1/4$ which is always true.

Proof. Obviously, one has

$$V^X_{a,n} = \|\mathbf{\Delta_a}(X)\|^2 = \|\mathbf{\Delta_a}(f) + \mathbf{\Delta_a}(\overline{X})\|^2.$$

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

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

with $\xi \in [0, n^{1-\alpha}]$. Then $\|\mathbf{\Delta_a}(f)\|^2 \leqslant n(K^{\alpha}_{M,n})^2 \Delta^{2M}$ and a sufficient condition is (26). □

3.5 Aggregation of estimators

Now we suggest a procedure to aggregate several quadratic a-variations estimators. We consider k sequences $a^{(1)}, ..., a^{(k)}$, with corresponding estimators $C_{a^{(1)},n}, \ldots, C_{a^{(k)},n}$ defined by (24). We assume that for $j = 1, ..., k$, the conditions of Corollary 3.9 are met. Let R be the $k \times k$ asymptotic variance-covariance matrix of the vector of length k whose elements are given by $(n^{1/2}/C)C_{a^{(j)},n}$, $j = 1, \ldots, k$. Let $\mathbf{1}_k$ be the "all one" column vector of size k and define

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

Elementary algebra shows that $\sum_{j=1}^k \lambda^*_j = 1$ and among all the possible linear combinations of the elements $(n^{1/2}/C)C_{a^{(j)},n}$ for $j = 1, \ldots, k$,

$$(n^{1/2}/C) \sum_{j=1}^{k} \lambda^*_j C_{a^{(j)},n}$$

is optimal in the sense that it has the smallest asymptotic variance (see e.g. [16] or [4]). Furthermore, it is a direct consequence of Corollary 3.7 that $(n^{1/2}/C)(\sum_{j=1}^k \lambda^*_j C_{a^{(j)},n} - C)$ converges to a $\mathcal{N}(0, \lambda^{*T} R \lambda^*)$ distribution as $n \to \infty$. Thus, the estimator $\sum_{j=1}^k \lambda^*_j C_{a^{(j)},n}$ is the optimal aggregation of $C_{a^{(1)},n}, ..., C_{a^{(k)},n}$. By construction, its asymptotic variance is not larger than any of the asymptotic variances of $C_{a^{(j)},n}$ denoted by $\widetilde{v}_{a^{(j)},s}$, for $j = 1, \ldots, k$. We call $\widetilde{v}_{a,s}$ the normalized asymptotic variance. Then Theorem 3.8 implies that $(n^{1/2}/C)(C_{a^{(i)},n} - C)$ converges to a $\mathcal{N}(0, \widetilde{v}_{a^{(i)},s})$ distribution as $n \to \infty$ (see (30) for the explicit expression of $\widetilde{v}_{a^{(j)},s}$ later on in the paper). As will be shown with simulations in Section 5, the aggregated estimator considerably improves each of the original estimators $C_{a^{(1)},n}, ..., C_{a^{(k)},n}$.

4 Cramér-Rao bound

In this section, we evaluate the quality of the proposed estimators. In that view, we compare their asymptotic variance with the theoretical Cramér-Rao bound in some ideal situations. More precisely, 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

$$(R_C)_{i,j} = \mathrm{Cov}\left(Y_C(i\Delta) - Y_C((i-1)\Delta), Y_C(i\Delta) - Y_C((i-1)\Delta)\right).$$

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

$$I_C = \frac{1}{2}\mathrm{Tr}\left(R_C^{-1}\left(\frac{\partial}{\partial C}R_C\right)R_C^{-1}\left(\frac{\partial}{\partial C}R_C\right)\right) \tag{27}$$

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

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

(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 4.1. *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*

$$V_C(h) = \frac{1}{2}\mathbb{E}\left[(Y_C(t+h) - Y_C(t))^2\right] = C|h|^s. \tag{28}$$

Hence in this case, we have $\frac{1}{I_C} = \frac{2C^2}{n-1}$.

Proof. (28) implies that $\partial R_C/\partial C = R_1$ then (27) gives the result. □

Now we consider a second example given by the generalized Slepian process defined in Section 2.2. Let $s \leqslant 1$ and Y_C with stationary covariance function ρ_C defined by

$$\rho_C(h) = (1 - (C/2)|h|^s)^+, \quad \text{for any} \quad h \in \mathbb{R}. \tag{29}$$

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.

Lemma 4.2. *Let X be the integration D times of Y_C defined by (29). Then we have, in the infill situation ($\alpha = 1$) and for $C < 2$, (28) 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 3.8 are representative of the finite sample behaviour 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 3.5.

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 [22] :

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

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 3.1 and Theorem 3.8. 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 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 (13) in Proposition 3.1. We let $\Delta = 1/n$ and

$$\widetilde{v}_{a,s} = \frac{2\sum_{i\in\mathbb{Z}} R^2(i, 1, 2D, |\cdot|^s, a^{2*})}{R^2(0, 1, 2D, |\cdot|^s, a^{2*}}, \tag{30}$$

so that $(n^{1/2}/C)(C_{a,n} - C)$ converges to a $\mathcal{N}(0, \widetilde{v}_{a,s})$ distribution as $n \to \infty$, where $\widetilde{v}_{a,s}$ already defined in Section 3.5 does not depend on C (nor on n).

First, we consider the case $D = 0$ and we plot $\widetilde{v}_{a,s}$ as a function of s for various sequences a in Figure 2. The conridered 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 Figure 2, we can draw several conclusions. First, the results of Section 4 suggest that 2 is a plausible lower bound for $\widetilde{v}_{a,s}$. We shall call the value 2 the Cramér-Rao lower bound. Indeed, we observe numerically that $\widetilde{v}_{a,s} \geqslant 2$ for all the s and a considered here. Then we observe that, for any value of s, there is one of the $\widetilde{v}_{a,s}$ 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 $\widetilde{v}_{a,s} = 2$. 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 $s = 1$,

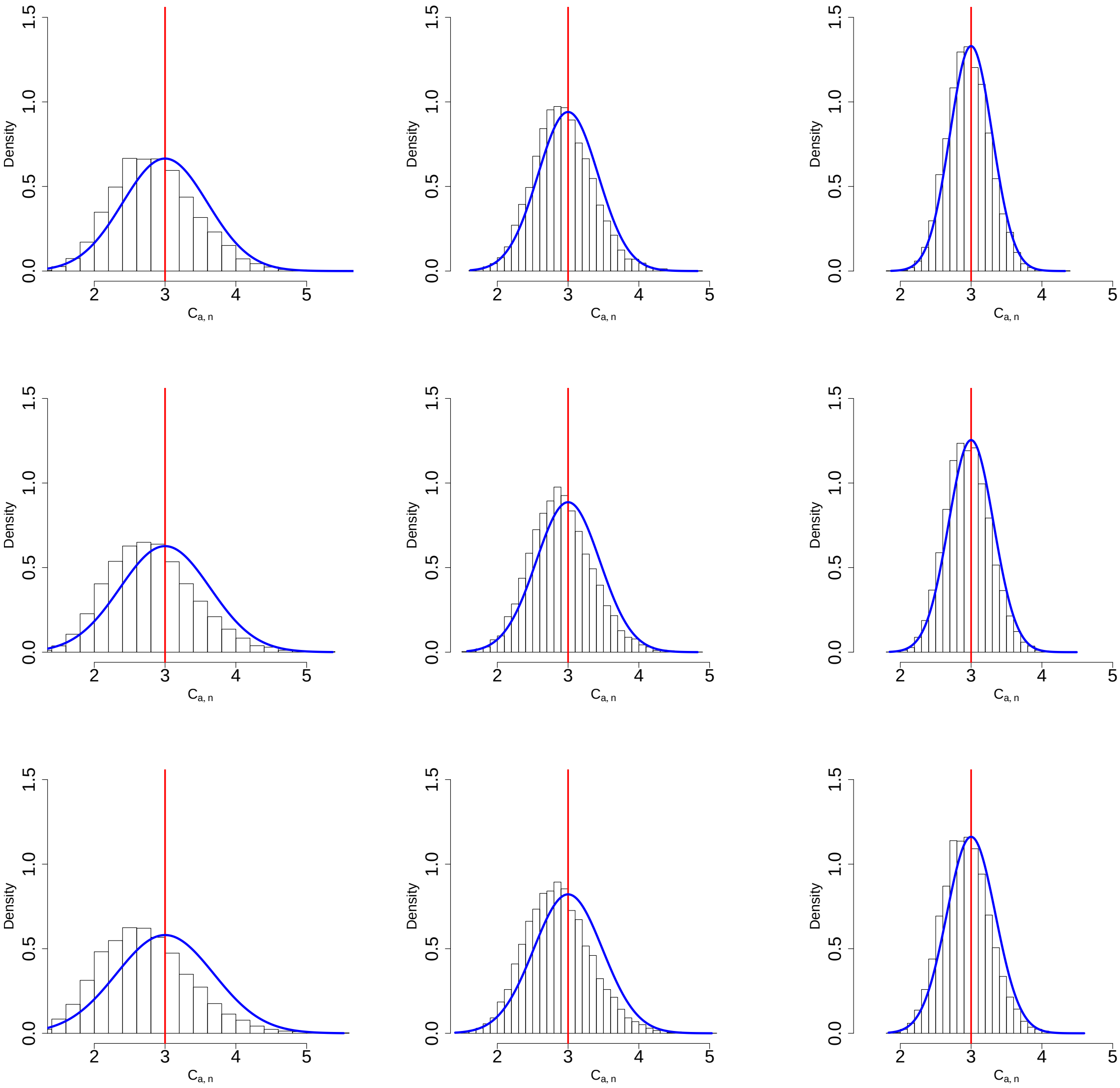

Figure 1: Comparison of the finite sample distribution of $C_{a,n}$ (histograms) with the asymptotic Gaussian distribution provided by Proposition 3.1 and Theorem 3.5 (probability density function in blue line). The vertical red line denotes the true value of $C = 3$. From left to right, $n = 50, 100, 200$. From top to bottom, $D = 0, 1, 2$.

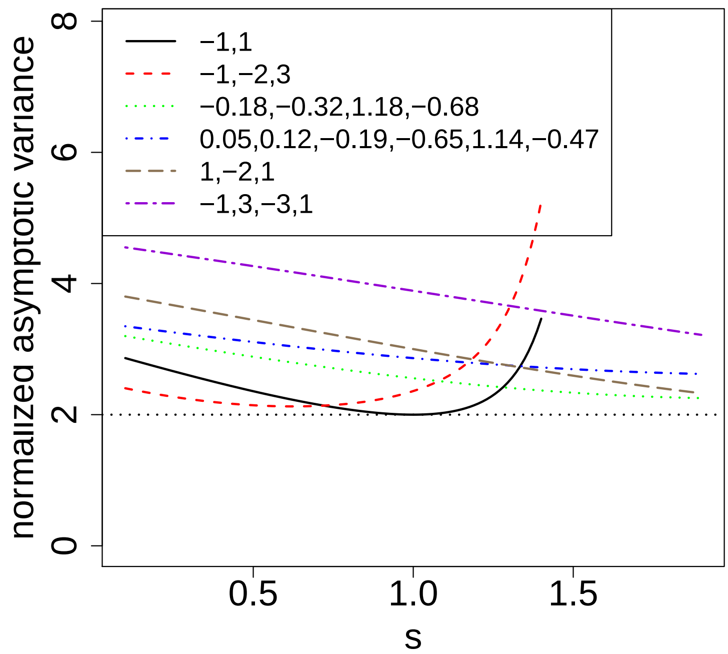

Figure 2: Case $D = 0$. Plot of the normalized asymptotic variance $\widetilde{v}_{a,s}$ of the quadratic a-variations estimator, as a function of s, for various sequences a. The legend shows the values $a_0, ..., a_l$ of these sequences (rounded to two digits). From top to bottom in the legend, the sequences are the elementary sequence of order 1, the sequence $(-1, -2, 3)$ which has order 1, the d 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.

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 $\widetilde{v}_{a,s}$ blows up for the two sequences a satisfying $M = 1$ when s reaches 1.5. This comes from Remark 3.4: the variance of the quadratic a-variations estimators with $M = 1$ is of order larger than $1/n$ when $s \geqslant 1.5$. Consequently, we plot $\widetilde{v}_{a,s}$ for $0.1 \leqslant s \leqslant 1.4$ for these two sequences. For the other sequences satisfying $M \geqslant 2$, we plot $\widetilde{v}_{a,s}$ for $0.1 \leqslant s \leqslant 1.9$.

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 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 3.5. A clear motivation for considering aggregation is that, in Figure 2, the smallest asymptotic variance $\widetilde{v}_{a,s}$ 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 $\widetilde{v}_{a^{(i)},s}$ as a function of s, for $0.1 \leqslant s \leqslant 1.4$ as well as

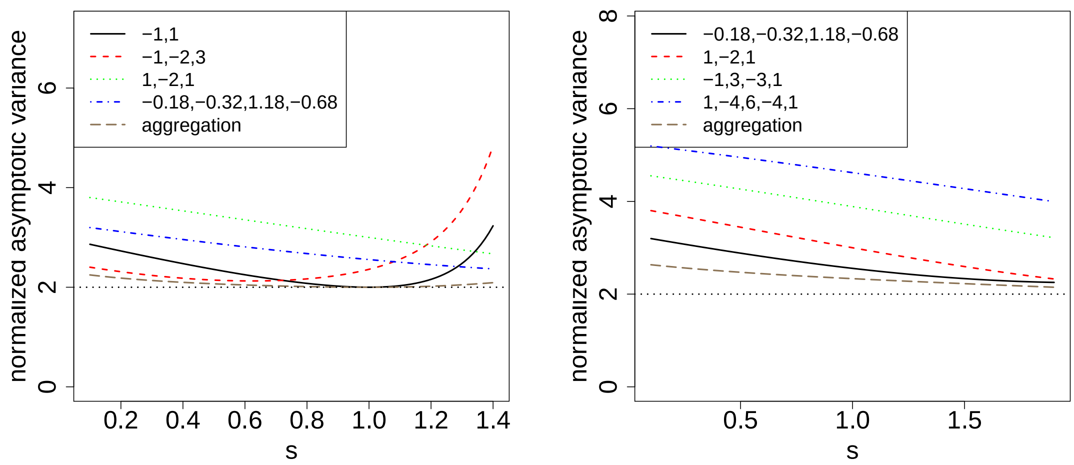

Figure 3: Case $D = 0$. Plot of the normalized asymptotic variance $\widetilde{v}_{a,s}$ of the quadratic a-variations estimator, as a function of s, for various sequences a and for their aggregation. On the left, including the order one elementary sequence, on the right without. The horizontal line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

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 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.

Eventually, Figures 4 and 5 explore the case $D = 1$. Conclusions are similar.

6 Appendix and technical results

Lemma 6.1. *Let $Z = (X, Y)$ be a centred Gaussian vector of dimension 2 then*

$$\mathrm{Cov}\left(X^2, Y^2\right) = 2\mathrm{Cov}^2\left(X, Y\right).$$

Proof. This Lemma is a consequence of the so called Mehler formula [2]. Let $H_m(x)$ be the Hermite polynomial of order m, i.e.,

$$H_m(x) = (-1)^m e^{x^2/2} \frac{d^m}{dx^m} e^{-x^2/2}.$$

Mehler formula states that if Z has for a variance-covariance matrix given by $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ then

$$\mathbb{E}\left[H_k(X) H_m(Y)\right] = \delta_{k,m} \rho^k k!.$$

We apply this formula with $k = m = 2$ (in that case $H_k(X) = X^2 - 1$) to get

$$2\mathrm{Cov}^2(X, Y) = \mathbb{E}\left[H_2(X) H_2(Y)\right] = \mathbb{E}\left[(X^2 - 1)(Y^2 - 1)\right] = \mathrm{Cov}(X^2, Y^2).$$

Eventually, we remark that this result can be generalized by homogeneity to the case of non-unit variance variables. □

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

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

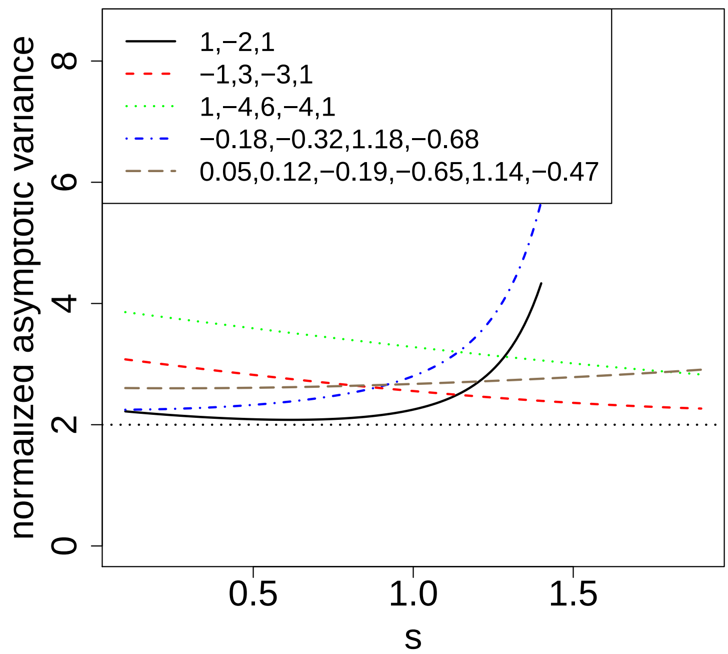

Figure 4: Same setting as in Figure 2 but for $D = 1$. From top to bottom in the legend, the sequences are the elementary sequences of order 2, 3 and 4 and the Daubechies sequences of order 2 and 3.

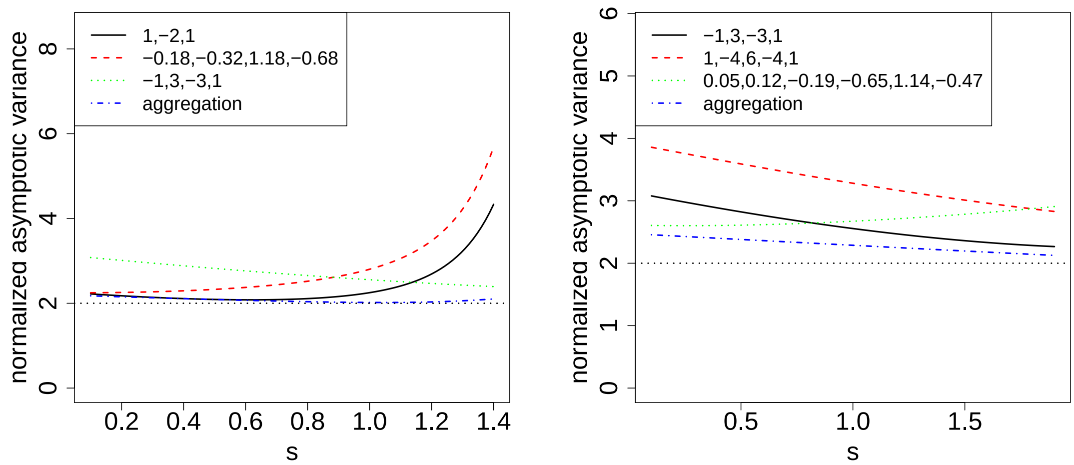

Figure 5: Same setting as in Figure 3 but for $D = 1$. On the left, from top to bottom in the legend, the sequences are the elementary sequence of order 2, the Daubechies sequence of order 2 and the elementary sequence of order 3. On the right, from top to bottom in the legend, the sequences are the elementary sequences of orders 3 and 4 and the Daubechies sequence of order 3.

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 $K^{(-r)}(u,v)=\mathbb{E}\big[B_s^{(-r)}(u)B_s^{(-r)}(v)\big]$, for $r\in\mathbb{N}$,

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

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

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

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

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

Hence, from the relation

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

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

Lemma 6.2. *Assume that V satisfies $(\mathcal{H}_0)$, $(\mathcal{H}_1)$, and $(\mathcal{H}_2)$. One has, when $M>D+s+1/4$,*

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

Proof. Using the stationarity of the increments of the process, one has

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

Recall that

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

We have seen in the proof of Proposition 3.1 ((17) and (18)) that for i sufficiently large

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

Thus the sum in (33) is bounded by

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

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

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

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. □

Acknowledgements This work has been partially supported by the French National Research Agency (ANR) through project PEPITO (no ANR-14-CE23-0011).

References

[1] Robert J. Adler and Ron Pyke. Uniform quadratic variation for Gaussian processes. *Stochastic Process. Appl.*, 48(2):191–209, 1993.

[2] Jean-Marc Azaïs and Mario Wschebor. *Level sets and extrema of random processes and fields*. John Wiley & Sons, Inc., Hoboken, NJ, 2009.

[3] François Bachoc. Asymptotic analysis of the role of spatial sampling for covariance parameter estimation of Gaussian processes. *Journal of Multivariate Analysis*, 125:1–35, 2014.

[4] John M Bates and Clive WJ Granger. The combination of forecasts. *Journal of the Operational Research Society*, 20(4):451–468, 1969.

[5] Glen Baxter. A strong limit theorem for Gaussian processes. *Proc. Amer. Math. Soc.*, 7:522–527, 1956.

[6] Jean-François Coeurjolly. Estimating the parameters of a fractional Brownian motion by discrete variations of its sample paths. *Stat. Inference Stoch. Process.*, 4(2):199–227, 2001.

[7] Serge Cohen and Jacques Istas. *Fractional fields and applications*, volume 73 of *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Springer, Heidelberg, 2013. With a foreword by Stéphane Jaffard.

[8] Rainer Dahlhaus. Efficient parameter estimation for self-similar processes. *The annals of Statistics*, pages 1749–1766, 1989.

[9] Ingrid Daubechies. Orthonormal bases of compactly supported wavelets. *Communications on pure and applied mathematics*, 41(7):909–996, 1988.

[10] E. G. Gladyšev. A new limit theorem for stochastic processes with Gaussian increments. *Teor. Verojatnost. i Primenen*, 6:57–66, 1961.

[11] Ulf Grenander. *Abstract inference*. John Wiley & Sons, Inc., New York, 1981. Wiley Series in Probability and Mathematical Statistics.

[12] Xavier Guyon and José León. Convergence en loi des H-variations d'un processus gaussien stationnaire sur **R**. *Ann. Inst. H. Poincaré Probab. Statist.*, 25(3):265–282, 1989.

[13] Jacques Istas and Gabriel Lang. Quadratic variations and estimation of the local Hölder index of a Gaussian process. *Ann. Inst. H. Poincaré Probab. Statist.*, 33(4):407–436, 1997.

[14] John T. Kent and Andrew T. A. Wood. Estimating the fractal dimension of a locally self-similar Gaussian process by using increments. *J. Roy. Statist. Soc. Ser. B*, 59(3):679–699, 1997.

[15] Gabriel Lang and François Roueff. Semi-parametric estimation of the Hölder exponent of a stationary Gaussian process with minimax rates. *Stat. Inference Stoch. Process.*, 4(3):283–306, 2001.

[16] Frédéric Lavancier and Paul Rochet. A general procedure to combine estimators. *Computational Statistics & Data Analysis*, 94:175–192, 2016.

[17] Paul Lévy. Le mouvement brownien plan. *Amer. J. Math.*, 62:487–550, 1940.

[18] D.G. Luenberger. *Introduction to Dynamic Systems: Theory, Models, and Applications*. Wiley, 1979.

[19] Olivier Perrin. Quadratic variation for Gaussian processes and application to time deformation. *Stochastic Process. Appl.*, 82(2):293–305, 1999.

[20] 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. Statistical Laboratory of the University of California, Berkeley. Berkeley, Calif.: University of California Press, 1949. 501 pp. Editor: Jerzy Neyman, p. 115-123*, pages 115–123, 1949.

[21] C.E. Rasmussen and C.K.I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, Cambridge, 2006.

[22] Olivier Roustant, David Ginsbourger, and Yves Deville. DiceKriging, DiceOptim: Two R packages for the analysis of computer experiments by Kriging-based metamodeling and optimization. *Journal of Statistical Software*, 51(1), 2012.

[23] David Slepian. On the zeros of Gaussian noise. In *Proc. Sympos. Time Series Analysis (Brown Univ., 1962)*, pages 104–115. Wiley, New York, 1963.

[24] M.L. Stein. *Interpolation of Spatial Data: Some Theory for Kriging*. Springer, New York, 1999.