



**HAL**  
open science

# Statistics for Gaussian random fields with unknown location and scale using Lipschitz-Killing curvatures

Elena Di Bernardino, Céline Duval

► **To cite this version:**

Elena Di Bernardino, Céline Duval. Statistics for Gaussian random fields with unknown location and scale using Lipschitz-Killing curvatures. *Scandinavian Journal of Statistics*, In press, 10.1111/sjos.12500 . hal-03347219

**HAL Id: hal-03347219**

**<https://hal.science/hal-03347219v1>**

Submitted on 17 Sep 2021

**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.

# Statistics for Gaussian Random Fields with Unknown Location and Scale using Lipschitz-Killing Curvatures

ELENA DI BERNARDINO\*, CÉLINE DUVAL†

## Abstract

In the present paper we study the average of Lipschitz-Killing (LK) curvatures of the excursion set of a stationary isotropic Gaussian field  $X$  on  $\mathbb{R}^2$ . The novelty is that the field can be non-standard, *i.e.*, with unknown mean and variance, which is more realistic from an applied viewpoint. To cope with the unknown location and scale parameters of  $X$ , we introduce novel fundamental quantities called *effective level* and *effective spectral moment*. We propose unbiased and asymptotically normal estimators of these parameters. From these asymptotic results, we build a test to determine if two images of excursion sets can be compared. This test is applied on both synthesized and real mammograms. Meanwhile, we establish the consistency of the empirical variance estimators of the third LK curvature under a weak condition on the correlation function of  $X$ .

**Key words:** Geometric inference, Excursion sets, Image analysis

**AMS Classification:** 62H11, 62M40, 60G60

## 1 Introduction

Lipschitz-Killing (LK) curvatures are geometrical tools which permit to analyse  $d$  dimensional objects. Considering a black and white image in dimension  $d = 2$ , there are three LK curvatures: the surface area, the half perimeter and the Euler characteristic. Each of them brings a distinct information on the geometry of the black (resp. white) zone. The surface area is related to its occupation density, the perimeter to its regularity and the Euler characteristic to its connectivity.

---

\*Laboratoire J.A. Dieudonné, UMR CNRS 7351, Université Côte d'Azur, Nice, France

†MAP5 UMR 8145, Université de Paris, Paris, France

In this paper, the images we consider are the excursion sets of the realization of a two-dimensional Gaussian stationary and isotropic random field  $X$  above a given level  $u$ , *i.e.* a black and white image indicating when the realization of  $X$  is above or below the level  $u$ . This work is part of a growing field of research, at the intersection of stochastic geometry and statistical analysis, which received an increasing attention in the recent years; see, for instance, [Lachièze-Rey \(2019\)](#), [Azaïs and Chassan \(2020\)](#), [Panigrahi et al. \(2019\)](#), [Cheng \(2016\)](#), [Azaïs and Pham \(2016\)](#) or [Pham \(2013\)](#). This literature investigates several geometrical stochastic objects and introduces consistent inference procedures to estimate them.

Using random fields in the modelization and analyzing their excursions with LK curvatures has been successfully exploited in many disciplines. In cosmology, to study the Cosmic Microwave Background radiation (see, *e.g.*, [Casaponsa et al. \(2016\)](#), [Schmalzing and Górski \(1998\)](#), [Gott et al. \(2007\)](#)) or to analyze the distribution of galaxies (see, *e.g.*, [Gott et al. \(2008\)](#)). The LK curvatures are also exploited in brain imaging (see [Adler and Taylor \(2011\)](#), Section 5, and the references therein) or to model sea waves (see, *e.g.*, [Longuet-Higgins \(1957\)](#), [Wschebor \(2006\)](#), [Lindgren \(2000\)](#)).

The average LK curvatures of the excursion set have been studied in a wide variety of contexts (see [Adler and Taylor \(2007, 2011\)](#), [Adler et al. \(2012\)](#) for a focus on the Gaussian kinematic formula and [Biermé and Desolneux \(2016\)](#) in case  $X$  is a shot noise). Empirical LK curvatures have also been studied and specific asymptotic results have been established. In the Gaussian framework, [Estrade and León \(2016\)](#) or [Di Bernardino et al. \(2017\)](#) studied the Euler characteristic whereas the area (also called sojourn times) is studied in [Bulinski et al. \(2012\)](#) or [Pham \(2013\)](#). See [Kratz and Vadlamani \(2018\)](#) and [Müller \(2017\)](#) for a joint analysis of LK curvatures.

These results permit to derive estimators and tests for the underlying field  $X$  (see [Di Bernardino et al. \(2017\)](#), [Biermé et al. \(2019\)](#), [Berzin \(2018\)](#)). Inference methods and tests using LK curvature devices only rely on the sparse observation of one excursion set and not on the covariance function nor on the marginal distribution of  $X$ , which require the observation of the entire field (see *e.g.* [Pantle et al. \(2010\)](#), [Nieto-Reyes et al. \(2014\)](#)).

**Removing the known mean and variance assumption** Most results concern the case where the field  $X$  is standard, *i.e.* centered and with unit variance. Furthermore the unit variance of its derivative is commonly required in the literature, assumption that is also not imposed in the present article. Imposing that the field is standard constrains the LK curvatures of the field. Indeed, from

the image of one excursion set if the field is symmetric, which is the case of Gaussian fields, the expected mean of the Euler characteristic of this set will be null at level  $u = 0$  and the expected mean of its area, will be  $1/2$  of the size of this image at this same level. More generally, the mean of the field indicates where the excursion set shows black and white zones in comparable proportions, whereas the variance conveys information on the range of levels where the excursion sets are non degenerate (all white or all black images). If one only observes one excursion set –and not the whole field  $X$ – it is often impossible to test whether the underlying field is centered or has unit variance. It is impossible to estimate the mean and variance of  $X$  from the sole excursion set. Then, from a statistical viewpoint, it is unrealistic to impose that the unobserved field  $X$  is standard.

The knowledge of the mean and variance of  $X$  is also crucial when it comes to compare the LK curvatures of several excursion sets. To compare the geometries of two given excursion sets, this information turns out to be crucial. Indeed, if the two excursion sets correspond to fields with distant mean, observing them at the same level will generate different pictures, but it does not mean that the underlying fields are necessarily different, *e.g.* one can be the translation of the other. This nonstandard context has deserved only a limited attention in the literature, we can mention, for instance, [David and Worsley \(1995\)](#) based on the Hadwiger characteristic.

In this article, in the case where  $X$  is a stationary and isotropic Gaussian random field, we introduce two quantities that we call *effective level* and *effective spectral moment* of the field (see Definition 1.2) that contain simultaneously the knowledge of the variance and mean of the field for every possible observation levels. Comparing excursion sets at the same *effective level* is meaningful as illustrated in Figure 1. In the first row we observe three different digital mammograms at the same fixed level. The obtained excursion sets look completely different; from this perspective the mammograms are dissimilar. But on the second row, we change the perspective and look at these objects at the same *effective level*, compared to the first row, the proportions of black and white zones are visually alike.

We propose consistent and asymptotically normal estimators for these *effective level* and *effective spectral moment* when  $X$  is a stationary and isotropic Gaussian random field (see Propositions 2.1 and 2.2). We relate these quantities to the model parameters and we show how these estimators permit to recover some model parameters and to build testing procedures.

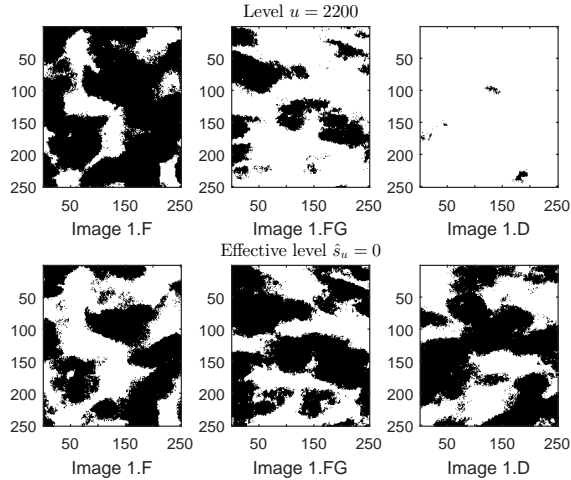


Figure 1: Excursion sets of 3 images of synthetic digital mammograms observed at the same level (first row) and at the same *effective* level (second row).

**Comparing images of exclusion sets** To compare two different fields from the observation of one excursion set of each using LK curvature devices, observing them at the same level  $u$  is irrelevant if the underlying fields have distant mean and/or variance. It is meaningful to compare them only if they are observed at the same *effective level* and *effective spectral moment*. This is why it is important to develop a consistent test to decide whether two images of excursion sets have the same *effective level* or not. This test is a mandatory preliminary step before putting in comparison two images of excursion sets. Its construction makes a full use of the asymptotically normal estimator of *effective level* as well as the consistent estimator of its limit variance, introduced in the present work (see Corollary 4.1).

**Asymptotic variance estimation** Frequently, when calibrating testing procedures or studying the deviations of the estimators, the proper calibration depends on the unknown asymptotic variances. Therefore, having consistent estimators of the asymptotic variances of the empirical LK curvatures is essential for the statistical study of these objects. If estimation strategies are known and used in practice in a wide range of contexts, there are few theoretical results concerning the consistency of these procedures. Some results exist in the case of the area if  $X$  is a Gaussian random field and under conditions on the correlation function of  $X$  that are difficult to check (see [Mattfeldt et al. \(2011\)](#), [Bulinski et al. \(2012\)](#)).

We establish in Theorem 3.1 the consistency of the empirical variance of the area when  $X$  is a stationary and isotropic Gaussian field whose correlation function decays polynomially (see Assumption (A1) below). For that we use the decomposition of Itô-Wiener chaos of the LK curvatures jointly with the diagram formula (see Taqqu (1977)). We also argue why this estimator should remain consistent to estimate the Euler characteristic variance (see Corollary 3.1). We point out that in this article, we are mainly interested in the Euler characteristic and area as they are numerically more stable to compute (see, *e.g.*, the discussion about the well known numerical instability of the perimeter estimation in Bierné et al. (2019)). Moreover, as we prove results on the asymptotic variance of the area, the numerical results in Section 4.1 rely only on this latter quantity.

**Outline of the paper** The paper is organized as follows. In the remaining of this section we define the three objects of interest, *i.e.*, the LK densities for the excursion sets of a two-dimensional Gaussian random field which is non centered and whose variance is not necessarily one. These formulae depend on two identifiable parameters, called *effective level* and *effective spectral moment*. In Section 2 we estimate these parameters and establish asymptotic normality results. The associated asymptotic variances are estimated in Section 3, using a subwindow technique. The consistency of the proposed subwindow empirical variance estimators is studied in Section 3.1 as well as some statistical implications in Section 3.2. In Section 4 we build a test based on the *effective level* to compare two images of excursion sets. This test is put into practice in Section 4.1 on synthesized 2D digital mammograms provided by GE Healthcare France (department Mammography) and in Section 4.2 on real digital mammograms provided by the Mammographic Image Analysis Society (MIAS). The proofs of the technical results are gathered in Appendix A.

## 1.1 Definitions and preliminary notions

In the whole paper,  $|\cdot|$  denotes equally the absolute value or the two dimensional Lebesgue measure and by  $|\cdot|_1$  its one-dimensional Hausdorff measure,  $\|\cdot\|$  denotes the Euclidian norm. We consider  $X$  being a Gaussian field defined on  $\mathbb{R}^2$  satisfying the following hypothesis.

(A0) The Gaussian field  $X$  is stationary, isotropic with mean  $\mathbb{E}[X(0)] = \mu$ , variance  $\mathbb{V}(X(0)) = \sigma^2$  and  $\mathbb{V}(X'(0)) = \lambda I_2$  for  $\lambda > 0$ , the second spectral moment,  $\sigma > 0$ ,  $\mu \in \mathbb{R}$  and  $I_2$  the  $2 \times 2$  identity matrix. Moreover, the trajectories of  $X$  are almost surely of class  $C^3$ .

Denote by  $r$  the covariance function of  $X$ , *i.e.*, for any fixed  $t$ ,  $r(t) = \text{Cov}(X(0), X(t))$  and by  $\rho$  the correlation function  $\rho(t) = \text{corr}(X(0), X(t))$ . Under Assumption [\(A0\)](#) the field is suppose to be  $C^3$ , then the covariance function is  $C^6$ . In some cases, *e.g.*, to get the asymptotic normality or consistency of our estimators, we impose additional assumptions on covariance and correlation functions of  $X$  gathered in the following.

$$(A1) \quad |\rho(t)| \leq (1 + \|t\|)^{-\gamma}, \quad \gamma > 2.$$

(A2) For any fixed  $t$  in  $\mathbb{R}^2$ , the covariance matrix of the random vector  $(X(t), X'(t), X''(t))$  has full rank and the covariance function  $r$  of  $X$  is such that,

$$\int_{\mathbb{R}^2} r(s) \, ds > 0, \quad M_r(t) \rightarrow 0 \quad \text{when} \quad \|t\| \rightarrow +\infty \quad \text{and} \quad M_r \in L^1(\mathbb{R}^2),$$

where

$$M_r(t) = \max \left( \left| \frac{\partial^{\mathbf{k}}}{\partial t^{\mathbf{k}}} r(t) \right|; \mathbf{k} = (i_1, \dots, i_\ell) \in \{1, 2\}^\ell, 0 \leq \ell \leq 4 \right)$$

$$\text{and} \quad \frac{\partial^{\mathbf{k}} r}{\partial t^{\mathbf{k}}}(t) = \frac{\partial^\ell}{\partial x_{i_1} \dots \partial x_{i_\ell}} r(t).$$

Assumptions [\(A0\)](#), [\(A1\)](#) and [\(A2\)](#) are standard when studying limit laws of non linear functionals of stationary Gaussian random fields. Assumptions [\(A0\)](#) and [\(A1\)](#) are used when working with the area, whereas for the Euler characteristic we rely on [\(A0\)](#) and [\(A2\)](#).

Rectangles, denoted by  $T$ , in  $\mathbb{R}^2$  are bounded and with non empty interior. In the following notation  $T \nearrow \mathbb{R}^2$  stands for the limit along any sequence of bounded rectangles that grows to  $\mathbb{R}^2$ . For that, let  $T$  be a bounded closed rectangle in  $\mathbb{R}^2$  with  $0 \in \overset{\circ}{T}$  and set  $N > 0$  and define

$$T^{(N)} := \{Nt : t \in T\}$$

the image of a fixed rectangle  $T$  by the dilatation  $t \mapsto Nt$  and then letting  $T \nearrow \mathbb{R}^2$  is equivalent to  $N \rightarrow \infty$ . Then,  $T^{(N)}$  is a Van Hove-growing sequence (VH-growing sequence, see Definition 6 in [Bulinski et al. \(2012\)](#)), *i.e.*,  $|\partial T^{(N)}|_1 / |T^{(N)}| \rightarrow 0$  as  $N \rightarrow \infty$ , where  $\partial T$  stands for the frontier of the set  $T$ . In the sequel, we sometimes drop the dependency in  $N$  of the rectangle  $T$  to soften notation.

## 1.2 Lipschitz-Killing curvatures of a given excursion set

Let  $u \in \mathbb{R}$ . For  $X$  a real-valued stationary random field defined on  $\mathbb{R}^2$ , we consider the excursion set within  $T$  above level  $u$ :

$$\{t \in T : X(t) \geq u\} = T \cap E_X(u), \quad \text{where} \quad E_X(u) := X^{-1}([u, +\infty)).$$

**Definition 1.1** (LK curvatures of  $E_X(u)$ ). Let  $X$  be a Gaussian field satisfying Assumption (A0). Define the following Lipschitz-Killing curvatures for the excursion set  $E_X(u)$ , for  $u \in \mathbb{R}$ ,

$$\mathcal{L}_2(X, u, T) := |T \cap E_X(u)| = \int_T \mathbf{1}_{\{E_X(u)\}}(t) dt, \quad (1)$$

$$\mathcal{L}_1(X, u, T) := |\partial(T \cap E_X(u))|_1 = \frac{1}{2} \lim_{\varepsilon \rightarrow 0} \int_T \delta_{]u-\varepsilon, u+\varepsilon[}(X(t)) \|\nabla X(t)\| dt, \quad (2)$$

$$\mathcal{L}_0(X, u, T) := \# \text{ connected components in } T \cap E_X(u) - \# \text{ holes in } T \cap E_X(u), \quad (3)$$

$$= \frac{1}{|T|} (\#(T \cap \mathcal{C}_{extr.}) - \#(T \cap \mathcal{C}_{saddle})), \quad (4)$$

where  $\mathcal{C}_{extr.} = \{t : X(t) \geq u, \nabla X(t) = 0, t \text{ is a local extremum}\}$ ,  $\mathcal{C}_{saddle} = \{t : X(t) \geq u, \nabla X(t) = 0, t \text{ is a saddle point}\}$ ,  $\delta_u$  the Dirac mass at  $u$  and  $\nabla X$  is the gradient field.

These three additive functionals,  $\mathcal{L}_j$  for  $j = 0, 1, 2$  in Definition 1.1, called Lipschitz-Killing curvatures, are also closely related to intrinsic volumes and Minkowski functionals. Roughly speaking, for  $A$  a Borelian set in  $\mathbb{R}^2$ ,  $\mathcal{L}_0(A)$  stands for the Euler characteristic of  $A$ ,  $\mathcal{L}_1(A)$  for the half perimeter of the boundary of  $A$  and  $\mathcal{L}_2(A)$  is equal to the area of  $A$ , *i.e.*, the two-dimensional Lebesgue measure.

In particular, when  $T$  is a bounded rectangle in  $\mathbb{R}^2$  with non empty interior,

$$\mathcal{L}_0(T) = 1, \quad \mathcal{L}_1(T) = \frac{1}{2} |\partial T|_1, \quad \mathcal{L}_2(T) = |T|. \quad (5)$$

Using the same formalism as in [Biermé et al. \(2019\)](#), the normalized LK curvatures are given by

$$C_i^{/T}(X, u) := \frac{\mathcal{L}_i(X, u, T)}{|T|}, \quad \text{for } i = 0, 1, 2,$$

and the associated LK densities by

$$C_i^*(X, u) := \lim_{T \nearrow \mathbb{R}^2} \mathbb{E}[C_i^{/T}(X, u)], \quad \text{for } i = 0, 1, 2. \quad (6)$$

In the case  $i = 2$ , one can easily write  $C_2^*(X, u) = \mathbb{E}[C_2^{/T}(X, u)] = \mathbb{P}(X(0) \geq u)$ . Computationally  $C_2^{/T}(X, u)$ ,  $C_1^{/T}(X, u)$  and  $C_0^{/T}(X, u)$  in (1), (2) and (3) can be evaluated in a given image by using the Matlab functions `bwarea`, `bwperim` and `bweuler` respectively. The numerical evaluation of (4) can be obtained for instance by finding local maxima, local minima and saddle points with the Matlab function `imextrema`.



**Definition 1.2.** Let  $X$  be a Gaussian random field satisfying Assumption (A0). We define the effective observation level (effective level in the sequel)

$$s_u := \frac{u - \mu}{\sigma}$$

and the effective second spectral moment (effective spectral moment in the sequel) of the field  $X$

$$w := \frac{\lambda}{\sigma^2}.$$

Notice that, if the underlying Gaussian field  $X$  is centered with unit variance, the effective level simply coincides with the observation level  $u$  and the effective spectral moment with the spectral moment of  $X$ . Using  $s_u$  and  $w$  as in Definition 1.2, we compute the expected value of these empirical LK curvatures of excursion sets of a Gaussian random field with unknown location and scale and derive the corresponding LK densities by using the following adapted Gaussian Kinematic Formula.

**Proposition 1.1.** Let  $X$  be a Gaussian random field satisfying Assumption (A0), denote  $\psi(x) = \mathbb{P}(\mathcal{N}(0, 1) \geq x)$ . Then, it holds that

$$\begin{aligned} \mathbb{E}[C_0^{/T}(X, u)] &= \frac{\psi(s_u)}{|T|} + \frac{\sqrt{w}}{2\pi} \exp\left\{-\frac{1}{2}s_u^2\right\} \frac{|\partial T|_1}{2|T|} + \frac{w}{(2\pi)^{3/2}} \exp\left\{-\frac{1}{2}s_u^2\right\} s_u, \\ \mathbb{E}[C_1^{/T}(X, u)] &= \psi(s_u) \frac{|\partial T|_1}{2|T|} + \frac{\sqrt{w}}{4} \exp\left\{-\frac{1}{2}s_u^2\right\}, \\ \mathbb{E}[C_2^{/T}(X, u)] &= \psi(s_u). \end{aligned}$$

Having  $T \nearrow \mathbb{R}^2$ , it follows that the LK densities defined in (6) are given by

$$C_0^*(X, u) = \frac{w}{(2\pi)^{3/2}} \exp\left\{-\frac{1}{2}s_u^2\right\} s_u, \quad C_1^*(X, u) = \frac{\sqrt{w}}{4} \exp\left\{-\frac{1}{2}s_u^2\right\}, \quad C_2^*(X, u) = \psi(s_u). \quad (7)$$

Proof of Proposition 1.1 is postponed to Section A.2. The LK densities given by (7) enable to identify  $s_u$  and  $w$ .

**Remark 1** (Unbiased estimators for LK densities). Let  $u \in \mathbb{R}$  and  $X$  as in Assumption (A0). Assume we observe  $T \cap E_X(u)$  for  $T$  a rectangle in  $\mathbb{R}^2$ . The following quantities are unbiased estimator of  $C_i^*(X, u)$  in (7) (see also [Biermé et al. \(2019\)](#)):

$$\widehat{C}_{2,T}(X, u) = C_2^{/T}(X, u), \quad (8)$$

$$\widehat{C}_{1,T}(X, u) = C_1^{/T}(X, u) - \frac{|\partial T|_1}{2|T|} C_2^{/T}(X, u), \quad (9)$$

$$\widehat{C}_{0,T}(X, u) = C_0^{/T}(X, u) - \frac{|\partial T|_1}{\pi|T|} C_1^{/T}(X, u) + \left( \frac{1}{2\pi} \left( \frac{|\partial T|_1}{|T|} \right)^2 - \frac{1}{|T|} \right) C_2^{/T}(X, u). \quad (10)$$

## 2 Estimators of the *effective level* and *effective spectral moment*

We focus on stationary Gaussian random fields with unknown location and scale parameters. In this section we build consistent estimators for  $s_u$  and  $w$  in Definition 1.2 relying on (8), (10) and Proposition 1.1. The importance of estimating these quantities lies in the fact that a comparison between two images of excursion sets of two fields (with unknown and eventually different location and scale parameters) can be proposed only if the considered observation levels make sense in terms of their respective mean and variance. These effective level and effective spectral moment can be used to determine if such a comparison is meaningful.

**Definition 2.1** (Estimators for  $s_u$  and  $w$ ). *Define the estimator of  $s_u$  built on the observation  $T \cap E_X(u)$ ,  $u$  being fixed, by  $\arg \min_{s \in \mathbb{R}} |\psi(s) - \widehat{C}_{2,T}(X, u)|$ , i.e.*

$$\widehat{s}_{u,T} := \psi^{-1}(\widehat{C}_{2,T}(X, u)). \quad (11)$$

Let  $T_1$  and  $T_2$  be two rectangles in  $\mathbb{R}^2$  such that  $\text{dist}(T_1, T_2) > 0$  and  $|T_1| = |T_2| > 0$ . Define the estimator of  $w$ ,  $u$  being fixed and  $u \neq \mu$ , by

$$\widehat{w}_{u,T_1,T_2} := \frac{\widehat{C}_{0,T_1}(X, u)(2\pi)^{3/2}}{\widehat{s}_{u,T_2} \exp\{-\frac{1}{2}(\widehat{s}_{u,T_2})^2\}}. \quad (12)$$

**Remark 2.** Notice that in Equation (12) we use two disjoint rectangles with same Lebesgue measure. To get the asymptotic normality of this quantity, we rely on a joint central limit theorem established for  $(\widehat{C}_{0,T_1}(X, u), \widehat{C}_{2,T_2}(X, u))$ . However, if  $T_1 = T_2$  such a result is not known, it can only be achieved for domains such that asymptotically  $\text{dist}(T_1, T_2) \rightarrow \infty$  (see the proof of Proposition 2.2). Indeed, we are not aware of any joint central limit theorem for the LK densities of excursion sets of stationary random fields on  $\mathbb{R}^2$ , except in the case of a Boolean model in Hug et al. (2016) or in the case of pixelated images in Reddy et al. (2018) and binary images in Ebner et al. (2018). Such a constraint on  $T_1$  and  $T_2$  is satisfied for instance taking  $T_1 \cup T_2 \subset T$  such that  $T_1 \cap T_2 = \emptyset$ ,  $|T_1| = |T_2|$ ,  $|T_1| > c|T|$ ,  $0 < c < 1/2$  and having  $T \uparrow \mathbb{R}^2$ . From a practical point of view, one only has one (large) image, the inference procedure can be implemented by partitioning it in sub-windows  $T_1$  and  $T_2$  separated by a “large” band (for instance of width  $\sqrt{\text{size of the image}}$ ) in order to build the estimator  $\widehat{w}_{u,T_1,T_2}$ .

**Remark 3** (Bias for  $\widehat{s}_{u,T}$  and  $\widehat{w}_{u,T_1,T_2}$ ). Estimator  $\widehat{s}_{u,T}$  in (11) is the quantile of the standard Gaussian distribution at random level  $1 - \widehat{C}_{2,T}(X, u)$ , where  $\widehat{C}_{2,T}(X, u)$  defined in (8) is unbiased.

It follows that  $\widehat{s}_{u,T}$  is biased, as the function  $\psi^{-1}$  is convex in  $(0, 0.5)$  (*resp.* concave in  $(0.5, 1)$ ), then it holds  $\mathbb{E}[\widehat{s}_{u,T}] > s_u$  if  $\widehat{C}_2^{/T}(X, u)$  in  $(0, 0.5)$  (*resp.*  $\mathbb{E}[\widehat{s}_{u,T}] < s_u$  if  $\widehat{C}_2^{/T}(X, u)$  in  $(0.5, 1)$ ). The bias of  $\widehat{s}_{u,T}$  certainly generates a bias also in the estimation of  $w$ . This bias for  $\widehat{s}_{u,T}$  can be assessed in Figure 2.

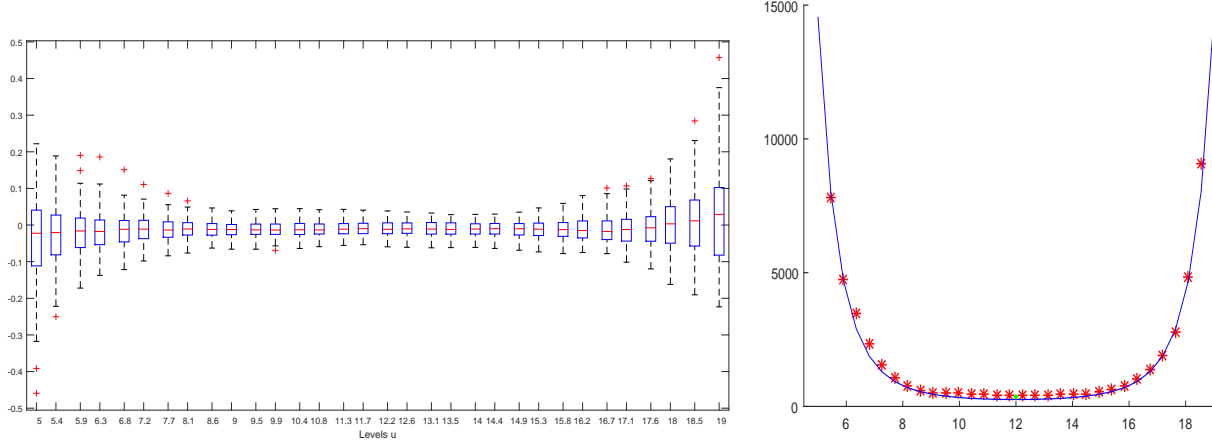


Figure 2: Synthetic digital mammograms study. Boxplot for  $\widehat{s}_{u,T} - s_u$  for different values of  $u$  (left panel). Empirical variance  $\sigma_{s_u}^2$  (red stars) and theoretical  $u \mapsto \sigma_{s_u}^2$  in (13) in blue line (right panel). These samples have been obtained with `Matlab` using circulant embedding matrix.

**Remark 4.** Let  $X$  be a stationary random field, non necessarily Gaussian, with finite variance. It holds  $\mathbb{E}[C_2^{/T}(X, u)] = \mathbb{P}(X(0) \geq u) = \mathbb{P}\left(\frac{X(0) - \mu}{\sigma} \geq s_u\right) =: F(s_u)$ . Then, given a family for the marginal distribution  $F$  of the field, known up to its first two moments, the *effective level* of the field can always be recovered using  $\widehat{s}_u = F^{-1}(C_2^{/T}(X, u))$ . This is not derived from the kinematic formula contrary to the equation leading to the computing of  $w$ .

**Proposition 2.1** (Asymptotic normality of  $(\widehat{s}_{u_1,T}, \dots, \widehat{s}_{u_m,T})$ ). *Let  $X$  be a Gaussian random field satisfying Assumptions (A0) and (A1). For a positive integer  $N$ , consider  $T^{(N)} = \{Nt : t \in T\}$  and  $\widehat{s}_{u_i, T^{(N)}}$  the estimator defined in (11) built on the observation  $T^{(N)} \cap E_X(u_i)$ , where  $u_1, \dots, u_m$  are fixed. Then,  $\sqrt{|T^{(N)}|}(\widehat{s}_{u_1, T^{(N)}} - s_{u_1}, \dots, \widehat{s}_{u_m, T^{(N)}} - s_{u_m})$  converges in distribution to a centered Gaussian vector with covariance matrix  $(\Sigma_{s_i, s_j}^2)_{1 \leq i, j \leq m}$ , with  $\Sigma_{s_i, s_j}^2 = \int_{\mathbb{R}^2} \int_0^{\rho(t)} h_{(u_i, u_j)}(r) dr dt \in (0, +\infty)$ , where*

$$h_{(u_i, u_j)}(r) = \frac{1}{\sqrt{1-r^2}} \exp \left\{ \frac{2rs_{u_i}s_{u_j} - r^2s_{u_i}^2 - r^2s_{u_j}^2}{2(1-r^2)} \right\}.$$

Proposition 2.1 is a consequence of Proposition A.2 and a multidimensional version of the delta method.

In the sequel we denote by  $\sigma_{s_u}^2 := \Sigma_{s, (u, u)}^2$  *i.e.*,

$$\sigma_{s_u}^2 = \int_{\mathbb{R}^2} \int_0^{\rho(t)} \frac{1}{\sqrt{1-r^2}} \exp \left\{ \frac{s_u^2 r}{1+r} \right\} dr dt \in (0, +\infty). \quad (13)$$

The function  $s_u^2 \mapsto \sigma_{s_u}^2$  is an increasing function, it is minimal when  $s_u^2$  is minimal *i.e.* at  $u = \mu$ . This leads to the following result.

**Corollary 2.1** (Asymptotic variance for  $\widehat{s}_{u=\mu, T}$ ). *Under assumptions of Proposition 2.1, the variance  $\sigma_{s_u}^2$  in (13) attains its minimum at  $u = \mu$ , *i.e.*,  $s_u = 0$ , and this minimum is*

$$\sigma_{s_\mu}^2 = \int_{\mathbb{R}^2} \arcsin(\rho(t)) dt.$$

Corollary 2.1 will be useful in Sections 4.1 and 4.2, where we frequently choose the observation level  $u$  such that the associated effective level  $s_u$  is null. This procedure will guarantee that the estimator has minimal variance. Corollary 2.1 is consistent with recent results on the so-called Berry's cancellation phenomenon, *i.e.*, the fact that the variance of the LK curvatures at level  $s_u$  reaches a minimum, because of the disappearance of a large number of chaotic terms (see Cammarota and Marinucci (2020)).

**Proposition 2.2** (Asymptotic normality of  $(\widehat{w}_{u_1}, \dots, \widehat{w}_{u_m})$ ). *Let  $X$  be a Gaussian random field satisfying Assumptions (A0), (A1) and (A2). Let  $T_1$  and  $T_2$  be two rectangles in  $\mathbb{R}^2$  such that  $\text{dist}(T_1, T_2) > 0$  and  $|T_1| = |T_2| > 0$ . For a positive integer  $N$ , let  $T_i^{(N)} = \{Nt : t \in T_i\}$ ,  $i \in \{1, 2\}$ . Consider  $\widehat{w}_{u_i, T_1^{(N)}, T_2^{(N)}}$  the estimator defined in (12) with  $u_i \neq \mu$  being fixed, for  $i = 1, \dots, m$ . Then,  $\sqrt{|T_1^{(N)}|} (\widehat{w}_{u_1, T_1^{(N)}, T_2^{(N)}} - w, \dots, \widehat{w}_{u_m, T_1^{(N)}, T_2^{(N)}} - w)$  converges in distribution to a centered Gaussian vector with covariance matrix  $(\Sigma_{w, (u_i, u_j)}^2)_{1 \leq i, j \leq m}$  given by*

$$\Sigma_{w, (u_i, u_j)}^2 = \frac{(2\pi)^3 \left( \Sigma_{C_0^*, (u_i, u_j)}^2 + \Sigma_{s, (u_i, u_j)}^2 (s_{u_i}^2 - 1)(s_{u_j}^2 - 1) C_0^*(X, u_i) C_0^*(X, u_j) \right)}{s_{u_i}^2 s_{u_j}^2 \exp \left\{ - \left( \frac{s_{u_i}^2}{2} + \frac{s_{u_j}^2}{2} \right) \right\}} < +\infty,$$

with  $\Sigma_{C_0^*, (u_i, u_j)}^2$  as in Proposition A.1 and  $\Sigma_{s, (u_i, u_j)}^2$  as in Proposition 2.1.

Proof of Proposition 2.2 is postponed to Section A.2.

**Remark 5.** The fact that  $\Sigma_{C_0^*(u,u)}^2$  is nonzero for all levels  $u$  still is an open problem. In Estrade and León (2016) the first Itô-Wiener chaos element of the series of  $\Sigma_{C_0^*(u,u)}^2$  has been calculated for all levels  $u$ . However, to guarantee that it is nonzero, higher order elements of this series need to be explored. Conversely, the first Itô-Wiener chaos element of the series of  $\sigma_{C_2^*,u}^2$  (see, e.g., Müller (2017) and Kratz and Vadlamani (2018)) guarantees it is non degenerate for all  $u \in \mathbb{R}$ .

**Numerical illustrations** In the following we provide an illustration of the finite sample performance of the proposed estimator  $\widehat{s}_{u,T}$  for several values of  $u$  (see Figure 2). We consider here  $M = 100$  sample simulations of Gaussian random fields as in Assumption (A0) with  $\mu = 12$ ,  $\sigma^2 = 4$  and covariance  $r(x) = e^{-\kappa^2 \|x\|^2}$ , for  $\kappa = 100/2^{10}$  in domains of size  $2^{10} \times 2^{10}$  pixels.

As expected from Proposition 2.1, the quality of the consistency can be evaluated in the displayed boxplots. Furthermore the associated asymptotic empirical variance is analysed and compared with the theoretical one in the case of  $u \mapsto \sigma_{s_u}^2$  in (13) (see right panel in Figure 2). Furthermore a bias in the estimation of  $s_u$  can be observed for values of level  $u$  far from  $\mu$  (see Remark 3). Similar results have been obtained for the estimator of  $\widehat{w}_{u,T_1,T_2}$  but they are omitted here for the sake of brevity.

### 3 Subwindow empirical variance estimation

#### 3.1 Consistent variance estimator

Having asymptotically normal estimators for  $s_u$  or  $w$  is insufficient if we cannot estimate their limit variance. Indeed, this variance plays a central role for instance to calibrate testing procedures, such as a test to determine if two images of excursion sets are comparable (see Section 4). Due to the specificity of our observations, which are excursion sets, proving that a variance estimator is consistent turns out to be a delicate, while fundamental, problem

**Definition of the subwindow estimators** We introduce a technique to estimate the asymptotic variances appearing in Propositions 2.1 and 2.2. Inspired by the cutting of  $T^{(N)}$  introduced in Pantle et al. (2010) (see also Bulinski et al. (2012), Section 5) we consider a classical empirical variance

estimator. To establish its consistency, we decompose this estimator on domains that are infinitely distant, mimicking the classical context of independent and identically distributed random variables.

Consider the following cutting of  $T^{(N)}$ : set  $M_N \in \mathbb{N}$  such that  $M_N \rightarrow \infty$  as  $N \rightarrow \infty$  and consider the grid pattern of  $T^{(N)}$  defined by  $(V^{(N,(i,j))})_{1 \leq i,j \leq M_N}$  such that  $\cup_{1 \leq i,j \leq M_N} V^{(N,(i,j))} = T^{(N)}$ , for  $(i,j) \neq (i',j')$ ,  $V^{(N,(i,j))} \cap V^{(N,(i',j'))} = \emptyset$  and  $|V^{(N,(i,j))}| := r_N$ ,  $\forall 1 \leq i,j \leq M_N$  where  $r_N \rightarrow \infty$  as  $N \rightarrow \infty$ . Observe that, whenever  $\max\{|i-i'|, |j-j'|\} \geq 2$  then  $\text{dist}(V^{(N,(i,j))}, V^{(N,(i',j'))}) \rightarrow \infty$  as  $N \rightarrow \infty$ . Then, by Propositions A.1 and A.2 the estimators  $\widehat{C}_\ell^{/V^{(N,(i,j))}}(X, u)$ , for  $\ell \in \{0, 2\}$ , are consistent and asymptotically independent. Therefore, we estimate the asymptotic variances appearing in Propositions A.1 and A.2 by

$$\widehat{\Sigma}_{C_\ell^*,(u,v)}^2 = \frac{1}{M_N^2 - 1} \sum_{i,j=1}^{M_N} \widehat{\xi}_N^{(i,j)}(u) \widehat{\xi}_N^{(i,j)}(v) - \left( \frac{1}{M_N^2 - 1} \sum_{i,j=1}^{M_N} \widehat{\xi}_N^{(i,j)}(u) \right) \left( \frac{1}{M_N^2 - 1} \sum_{i,j=1}^{M_N} \widehat{\xi}_N^{(i,j)}(v) \right), \quad (14)$$

where we define  $\widehat{\xi}_N^{(i,j)}(u) := \widehat{C}_\ell^{/V^{(N,(i,j))}}(X, u)$ , for all  $u \in \mathbb{R}$ ,  $\ell \in \{0, 2\}$ ,  $1 \leq i, j \leq M_N$ . The key point is that these estimators are identically distributed, as the field  $X$  is stationary, and are asymptotically independent whenever  $\max\{|i-i'|, |j-j'|\} \geq 2$  in view of Propositions A.1 and A.2 and the fact that  $\text{dist}(V^{(N,(i,j))}, V^{(N,(i',j'))}) \rightarrow \infty$ , as  $N \rightarrow \infty$ .

**Case of the area ( $\ell = 2$  in Equation (14))** To the knowledge of the authors, the only result establishing the consistency of (14), is Theorem 6 of Bulinski et al. (2012) in the case of the area ( $\ell = 2$ ). However, this results imposes a condition (39) on the fourth-order cumulant that is difficult to check and for which the example provided is “a random field  $X$  with finite dependence range”, which means that  $\rho = \rho \mathbf{1}_D$  for some finite domain  $D$ . In the following Theorem 3.1 we prove consistency of the estimator  $\widehat{\Sigma}_{C_2^*,(u,v)}^2$  defined in (14) under Assumption (A1).

**Theorem 3.1.** *Let  $X$  be a Gaussian random field satisfying Assumptions (A0) and (A1). Let  $\widehat{\Sigma}_{C_2^*,(u_q,u_k)}^2$  be as in (14) with  $q, k \in \{1, \dots, m\}$ . Then, it holds that  $\widehat{\Sigma}_{C_2^*,(u_q,u_k)}^2 \xrightarrow[N \rightarrow \infty]{\mathbb{P}} \Sigma_{C_2^*,(u_q,u_k)}^2$ .*

*Proof of Theorem 3.1.* We show the consistency of estimators  $\widehat{\Sigma}_{C_2^*,(u_q,u_k)}^2$  in (14), for all  $q, k \in \{1, \dots, m\}$ . First, note that  $\widehat{\Sigma}_{C_2^*,(u_q,u_k)}^2$  is given by the difference of two terms, write

$$\widehat{\Sigma}_{C_2^*,(u_q,u_k)}^2 := S_1(u, v) - S_2(u)S_2(v). \quad (15)$$

By Proposition 2.1 and the additivity of  $T \mapsto C_2^{/T}(X, u)$ , we get that for all  $u$ ,  $S_2(u) \xrightarrow[N \rightarrow \infty]{\mathbb{P}} C_2^*(X, u)$ .

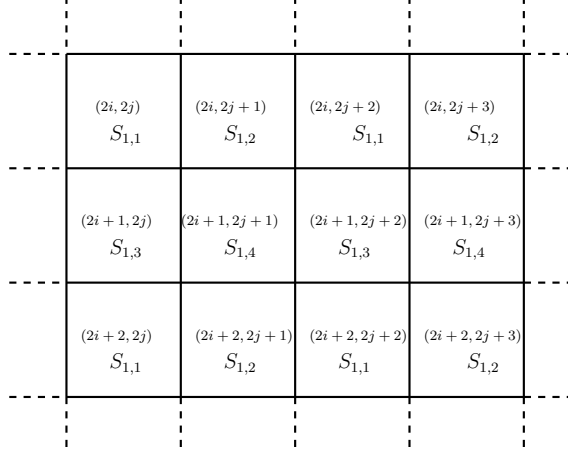


Figure 3: Decomposition of the sum  $S_1$  in (15) in four sums.

The remaining of the proof consists is noticing that the sum  $S_1$  can be rewritten as four sums each composed of identically distributed terms that are asymptotically independent. Decompose  $S_1$  as follows (see Figure 3) and without loss of generality assume that  $M_N$  is even, then

$$\begin{aligned}
S_1 &= \frac{1}{M_N^2 - 1} \sum_{i,j=1}^{M_N} \widehat{\xi}_N^{(2i,2j)}(u_q) \widehat{\xi}_N^{(2i,2j)}(u_k) + \frac{1}{M_N^2 - 1} \sum_{i,j=1}^{M_N} \widehat{\xi}_N^{(2i,2j+1)}(u_q) \widehat{\xi}_N^{(2i,2j+1)}(u_k) \\
&\quad + \frac{1}{M_N^2 - 1} \sum_{i,j=1}^{M_N} \widehat{\xi}_N^{(2i+1,2j)}(u_q) \widehat{\xi}_N^{(2i+1,2j)}(u_k) + \frac{1}{M_N^2 - 1} \sum_{i,j=1}^{M_N} \widehat{\xi}_N^{(2i+1,2j+1)}(u_q) \widehat{\xi}_N^{(2i+1,2j+1)}(u_k) \\
&:= S_{1,1} + S_{1,2} + S_{1,3} + S_{1,4}.
\end{aligned}$$

To show that  $S_{1,p} - \frac{1}{4} \mathbb{E}[\widehat{\xi}_N^{(1,1)}(u) \widehat{\xi}_N^{(1,1)}(v)] \xrightarrow[N \rightarrow \infty]{\mathbb{P}} 0$ ,  $1 \leq p \leq 4$ , we show that the convergence holds in  $L_2$ . We perform computations for  $S_{1,1}$ , other sums are treated similarly. It holds, using the stationarity of the field, that

$$\begin{aligned}
\mathbb{V}(S_{1,1}) &= \frac{1}{(M_N^2 - 1)^2} \sum_{i,i',j,j'=1}^{M_N} \text{Cov}(\widehat{\xi}_N^{(2i,2j)}(u_q) \widehat{\xi}_N^{(2i,2j)}(u_k), \widehat{\xi}_N^{(2i',2j')}(u_q) \widehat{\xi}_N^{(2i',2j')}(u_k)) \\
&= \frac{1}{(M_N^2 - 1)^2} \sum_{(i,j) \neq (i',j')=1}^{M_N} \text{Cov}(\widehat{\xi}_N^{(2i,2j)}(u_q) \widehat{\xi}_N^{(2i,2j)}(u_k), \widehat{\xi}_N^{(2i',2j')}(u_q) \widehat{\xi}_N^{(2i',2j')}(u_k)) \quad (16)
\end{aligned}$$

$$+ \frac{1}{(M_N^2 - 1)^2} \sum_{i,j=1}^{M_N} \text{Cov}(\widehat{\xi}_N^{(2i,2j)}(u_q) \widehat{\xi}_N^{(2i,2j)}(u_k), \widehat{\xi}_N^{(2i,2j)}(u_q) \widehat{\xi}_N^{(2i,2j)}(u_k)). \quad (17)$$

For all  $(i, j) \in \{1, \dots, M_N\}^2$ , if we set  $G = (X - \mu)/\sigma$  it holds that

$$\widehat{\xi}_N^{(i,j)}(u) = \widehat{C}_2^{V(N,(i,j))} (X, u) = \widehat{C}_2^{V(N,(i,j))} (G, s_u) = \frac{\mathcal{L}_2(G, s_u, V^{(N,(i,j))})}{|V^{(N,(i,j))}|}.$$

The remaining of the proof is a direct consequence of the following result.

**Proposition 3.1.** *Let  $X$  be a Gaussian random field satisfying Assumptions (A0) and (A1) and set  $G = (X - \mu)/\sigma$ . Let  $u_1, u_2, u_3, u_4$  be four fixed levels in  $\mathbb{R}$  and associated effective levels  $s_{u_i} = (u_i - \mu)/\sigma$ , for  $i \in \{1, 2, 3, 4\}$ . Let  $T$  and  $T'$  be such that  $|T| = |T'|$  and  $\text{dist}(T, T') \rightarrow \infty$ . Then, it holds that,*

$$\mathbb{E}[\mathcal{L}_2(G, s_{u_1}, T)\mathcal{L}_2(G, s_{u_2}, T)\mathcal{L}_2(G, s_{u_3}, T')\mathcal{L}_2(G, s_{u_4}, T')] = \psi(s_{u_1})\psi(s_{u_2})\psi(s_{u_3})\psi(s_{u_4})|T|^4 + o(|T|^3).$$

Proof of Proposition 3.1 is postponed to Section A.3. Proposition 3.1 ensures that (16) goes to 0. Let  $(i, j) \neq (i', j') \in \{1, \dots, M_N\}^4$  and  $T = V^{(N, (i, j))}$  and  $T' = V^{(N, (i', j'))}$ , for which for  $N$  large enough it holds  $\text{dist}(V^{(N, (i, j))}, V^{(N, (i', j'))}) > 1$  and  $|V^{(N, (i, j))}| = |V^{(N, (i', j'))}| \rightarrow \infty$  as  $N \rightarrow \infty$ , then we derive

$$\lim_{n \rightarrow \infty} \text{Cov}(\widehat{\xi}_N^{(2i, 2j)}(u_q)\widehat{\xi}_N^{(2i, 2j)}(u_k), \widehat{\xi}_N^{(2i', 2j')}(u_q)\widehat{\xi}_N^{(2i', 2j')}(u_k)) = 0.$$

This implies that (16) goes to 0 as  $N \rightarrow \infty$  using Cesàro Lemma. Moreover, Equation (17) vanishes using that  $|\widehat{\xi}_N| \leq 1$  almost surely, implying (17) is bounded by  $M_N^2/(M_N^2 - 1)^2 \rightarrow 0$ , as  $N \rightarrow \infty$ . This implies that  $\mathbb{V}(S_{1,1}) \rightarrow 0$  as  $N \rightarrow \infty$  using Cesàro Lemma. The same behaviour holds true for  $\mathbb{V}(S_{1,p})$  for all  $p \in \{1, 2, 3, 4\}$ , implying the desired result in Theorem 3.1.  $\square$

**Discussion** Theorem 3.1 establishes consistency of this subwindow variance estimator in the case of the area under Assumption (A1). This assumption seems weak as it is the same assumption as the one imposed to get the asymptotic normality of  $\widehat{C}_2(X, u)$  (see Theorem 4 of Bulinski et al. (2012)).

A key element in the proof of Theorem 3.1 is Proposition 3.1 which is similar to the constraint (39) of Bulinski et al. (2012). More precisely, in Proposition 3.1 we control covariance terms between distant domains. Proof of Proposition 3.1 relies on two ingredients: *i.* the Itô-Wiener chaos decomposition of  $\mathcal{L}_2(X, u, T)$  and *ii.* the diagram formula (see Taqqu (1977), Lemma 3.2) applied up to the dimension 4. This diagram formula in dimensions 3 and 4 leads to technical and lengthy computations where the constraint  $\text{dist}(V^{(N, (i, j))}, V^{(N, (i', j'))}) \rightarrow \infty$  as  $N \rightarrow \infty$ , and the cutting of  $T^{(N)}$ , plays a crucial part.

A key question from an applied point of view is the choice of  $M_N$  and  $|V^{(N, (1, 1))}|$  (see Section 5 of Bulinski et al. (2012)). Theoretically, it is only required is that  $M_N$  and  $|V^{(N, (1, 1))}|$  increase to



infinity, but in practice, we need a cutting of  $T$  in infinitely many rectangles such that in each sub-rectangle the excursion set is sufficiently rich to properly estimate  $\widehat{\xi}_N^{(i,j)}$ . A coarse comparison of the different order of magnitudes required to get a normalized central limit theorem using  $\widehat{\Sigma}_{C_2^*(u,v)}^2$  is to consider  $M_N \asymp \lfloor |T|^{1/4} \rfloor$  and  $|V^{(N,(1,1))}| \asymp \sqrt{|T|}$ .

The result below is a direct consequence of Theorem 3.1: the consistency of the estimator for  $\Sigma_{s,(u_i,u_j)}^2$ , introduced in Proposition 2.1.

**Corollary 3.1.** *Let  $X$  be a Gaussian random field satisfying Assumptions (A0) and (A1). Define, the estimator of  $\Sigma_{s,(u_i,u_k)}^2$ , where  $(i,k) \in \{1, \dots, m\}^2$ , by*

$$\widehat{\Sigma}_{s,(u_i,u_k)}^2 = 2\pi \exp \left\{ \frac{1}{2} (\widehat{s}_{u_i}^2 + \widehat{s}_{u_k}^2) \right\} \widehat{\Sigma}_{C_2^*(u_i,u_k)}^2, \quad (18)$$

with  $\widehat{s}_u := \widehat{s}_{u,T(N)}$  as in Proposition 2.1. Then, it holds that, for  $(i,k) \in \{1, \dots, m\}^2$ ,

$$\widehat{\Sigma}_{s,(u_i,u_k)}^2 \xrightarrow[N \rightarrow \infty]{\mathbb{P}} \Sigma_{s,(u_i,u_k)}^2.$$

**Case of the Euler characteristic ( $\ell = 0$  in Equation (14))** In the case of the Euler characteristic ( $\ell = 0$ ), the same proof should generalize under the following additional assumption.

(A3) Suppose that Assumption (A2) holds true with  $|M_r(t)| \leq C_M/(1 + \|t\|)^\tau$  where  $C_M$  is a positive constant and  $\tau > 2$ .

Indeed there is a Itô-Wiener chaos decomposition of  $\mathcal{L}_0(X, u, T)$  (see the Additional Supporting Information) with similar properties on the coefficients. However, computing the fourth moment with this chaos decomposition would imply to use the diagram formula in dimensions 18 and 24, which become extremely technical and burdensome. Therefore, we give in the Additional Supporting Information some ingredients to establish the following Conjecture 3.1 which is an equivalent of Proposition 3.1 in the case of the Euler characteristic having for all  $(i,j) \in \{1, \dots, M_N\}^2$ ,

$$\widehat{\xi}_N^{(i,j)}(u) = \widehat{C}_0^{V^{(N,(i,j))}}(X, u) = \frac{\mathcal{L}_0(X, u, V^{(N,(i,j))})}{|V^{(N,(i,j))}|}.$$

The missing step in the proof to have the result is the computation of the diagram formula in dimensions 18 and 24.

**Conjecture 3.1.** *Let  $X$  be a Gaussian random field satisfying Assumptions (A0) and (A3). Let  $T$  and  $T'$  be such that  $|T| = |T'|$  and  $\text{dist}(T, T') \rightarrow \infty$ . Let  $u_1, u_2, u_3, u_4$  be four fixed levels in  $\mathbb{R}$ . Then it holds that*

$$\begin{aligned} & \mathbb{E}[\mathcal{L}_0(X, u_1, T)\mathcal{L}_0(X, u_2, T)\mathcal{L}_0(X, u_3, T')\mathcal{L}_0(X, u_4, T')] \\ &= \frac{w^4}{(2\pi)^6} \exp\left\{-\frac{1}{2}(s_{u_1}^2 + s_{u_2}^2 + s_{u_3}^2 + s_{u_4}^2)\right\} s_{u_1} s_{u_2} s_{u_3} s_{u_4} |T|^4 + O(|T|^3), \end{aligned}$$

where  $w$  denotes the effective spectral moment of  $X$  in Definition 1.2.

Mimicking the proof of Theorem 3.1, we derive that a consequence of Conjecture 3.1 is that

$$\widehat{\Sigma}_{C_0^*, (u_q, u_k)}^2 \xrightarrow[N \rightarrow \infty]{\mathbb{P}} \Sigma_{C_0^*, (u_q, u_k)}^2. \quad (\text{Conjecture})$$

Then, using (Conjecture), one can conjecture the consistency of the estimator for  $\Sigma_{w, (u_i, u_j)}^2$ , introduced in Proposition 2.2.

**Conjecture 3.2.** *Let  $X$  be a Gaussian random field satisfying Assumptions (A0), (A1) and (A3).*

*Define the estimator of  $\Sigma_{w, (u_i, u_k)}^2$ , where  $(i, k) \in \{1, \dots, m\}^2$ ,*

$$\widehat{\Sigma}_{w, (u_i, u_k)}^2 = \frac{(2\pi)^3 \left( \widehat{\Sigma}_{C_0^*, (u_i, u_k)}^2 + \widehat{\Sigma}_{s, (u_i, u_k)}^2 (\widehat{s}_{u_i}^2 - 1)(\widehat{s}_{u_k}^2 - 1) \widehat{C}_{0, T(N)}(X, u_i) \widehat{C}_{0, T(N)}(X, u_k) \right)}{\widehat{s}_{u_i}^2 \widehat{s}_{u_k}^2 \exp\left\{-\left(\frac{\widehat{s}_{u_i}^2}{2} + \frac{\widehat{s}_{u_k}^2}{2}\right)\right\}},$$

with  $\widehat{s}_u := \widehat{s}_{u, T(N)}$  as in Proposition 2.1,  $\widehat{\Sigma}_{s, (u_i, u_k)}^2$  as in (18) and  $\widehat{C}_{0, T(N)}(X, u_k)$  as in (10). Then,

$$\widehat{\Sigma}_{w, (u_i, u_k)}^2 \xrightarrow[N \rightarrow \infty]{\mathbb{P}} \Sigma_{w, (u_i, u_k)}^2.$$

## 3.2 Statistical implications and numerical illustrations

In this section we consider several useful applications of theoretical results provided in Sections 2 and 3.1. Indeed, the asymptotically Gaussian estimators of the unknown location and scale parameters and the subwindow empirical variance estimators are used in the following to build a test of Gaussianity using LK densities (see Section 3.2.1). A procedure to build an interval for the unknown location  $\mu$  of the field is given in Section 3.2.2.

### 3.2.1 A Gaussianity test for non standard fields

Here we generalize the test in Section 3.1 of [Biermé et al. \(2019\)](#) in cases where the field is not supposed to be centered and with unit variance. In this section we are interested in testing

$$H_0 : X \text{ is a Gaussian field with unknown mean and unknown variance.}$$

Consider the quantities

$$R(u_1, u_2) := \frac{C_0^*(X, u_2)}{C_0^*(X, u_1)} \stackrel{\text{under } H_0}{=} \frac{s_{u_2}}{s_{u_1}} \exp \left\{ -\frac{1}{2}(s_{u_2}^2 - s_{u_1}^2) \right\} := R^{H_0}(u_1, u_2).$$

This ratio is empirically accessible provided the field is observed at two distinct levels  $u_1$  and  $u_2$ ,

$$\widehat{R}_T(u_1, u_2) := \frac{\widehat{C}_{0,T}(X, u_2)}{\widehat{C}_{0,T}(X, u_1)}. \quad (19)$$

We establish below a central limit theorem for  $\widehat{R}_T(u_1, u_2)$ . Notice that one can readily establish a CLT for  $\sqrt{|T|} \left( \widehat{R}_T(u_1, u_2) - R^{H_0}(u_1, u_2) \right)$  using, for instance, Proposition A.1, but conversely to Section 3.1 of [Biermé et al. \(2019\)](#), as  $R^{H_0}(u_1, u_2)$  is now unknown, this CLT does not allow to determine a rejection level. To overcome that problem, define

$$\widehat{R}_T^{H_0}(u_1, u_2) := \frac{\widehat{s}_{u_2}}{\widehat{s}_{u_1}} \exp \left\{ -\frac{1}{2}(\widehat{s}_{u_2}^2 - \widehat{s}_{u_1}^2) \right\}. \quad (20)$$

For technical reasons, we estimate  $\widehat{R}_{T_1}(u_1, u_2)$  and  $\widehat{R}_{T_2}^{H_0}(u_1, u_2)$  on rectangles  $T_1$  and  $T_2$  that are asymptotically infinitely distant so that both excursion sets are asymptotically independent and lead to independent estimators (see also Remark 2).

**Proposition 3.2.** *Assume that  $X$  is a Gaussian field satisfying Assumptions (A0), (A1) and (A2). Let  $u_1, u_2 \neq \mu$  being fixed. Let  $T_1$  and  $T_2$  be two rectangles in  $\mathbb{R}^2$  such that  $\text{dist}(T_1, T_2) > 0$  and  $|T_1| = |T_2| > 0$ . For any positive integer  $N$ , we define  $T_i^{(N)} = \{Nt : t \in T_i\}$ , for  $i = 1, 2$ . Then, under  $H_0$  it holds that*

$$\sqrt{|T_1^{(N)}|} \left( \widehat{R}_{T_1^{(N)}}(u_1, u_2) - \widehat{R}_{T_2^{(N)}}^{H_0}(u_1, u_2) \right) \xrightarrow[N \rightarrow \infty]{d} \mathcal{N}(0, \sigma_{R_{u_1, u_2}}^2),$$

where  $\sigma_{R_{u_1, u_2}}^2 < \infty$  and  $\widehat{R}_{T_1^{(N)}}(u_1, u_2)$  (resp.  $\widehat{R}_{T_2^{(N)}}^{H_0}(u_1, u_2)$ ) is defined as in (19) (resp. in (20)) built on the observation  $T_1^{(N)} \cap E_X(u_1)$  and  $T_1^{(N)} \cap E_X(u_2)$  (resp.  $T_2^{(N)} \cap E_X(u_1)$  and  $T_2^{(N)} \cap E_X(u_2)$ ).

Proof of Proposition 3.2 is postponed to Section A.2.

**Subwindow empirical estimation of variance**  $\sigma_{R_{u_1, u_2}}^2$  We provide a normalized version of Proposition 3.2, we build a consistent estimator for  $\sigma_{R_{u_1, u_2}}^2$ , i.e.,

$$\widehat{\sigma}_{R_{u_1, u_2}}^2 = \widehat{\sigma}_{g(C_0^*), (u_1, u_2)}^2 + \widehat{\sigma}_{h(s), (u_1, u_2)}^2 := \widehat{\mathbb{V}}(\widehat{R}_{T_1^{(N)}}(u_1, u_2)) |T_1^{(N)}| + \widehat{\mathbb{V}}(\widehat{R}_{T_2^{(N)}}^{H_0}(u_1, u_2)) |T_2^{(N)}|. \quad (21)$$

Consistent estimators for  $\mathbb{V}(\widehat{R}_{T^{(N)}}(u_1, u_2))$  and  $\mathbb{V}(\widehat{R}_{T^{(N)}}^{H_0}(u_1, u_2))$  are obtained using the subwindows technique of Section 3.1. Define  $\widehat{\sigma}_{h^{(s)},(u_1,u_2)}^2$  in (21) by

$$\widehat{\sigma}_{h^{(s)},(u_1,u_2)}^2 = e^{-(\widehat{s}_{u_2}^2 - \widehat{s}_{u_1}^2)} \left[ \widehat{\sigma}_{s_{u_2}}^2 \left( \frac{1 - \widehat{s}_{u_2}^2}{\widehat{s}_{u_1}} \right)^2 + 2 \widehat{\Sigma}_{s,(u_1,u_2)}^2 \left( \frac{1 - \widehat{s}_{u_2}^2}{\widehat{s}_{u_1}} \right) \left( \widehat{s}_{u_2} - \frac{\widehat{s}_{u_2}}{\widehat{s}_{u_1}^2} \right) + \widehat{\sigma}_{s_{u_1}}^2 \left( \widehat{s}_{u_2} - \frac{\widehat{s}_{u_2}}{\widehat{s}_{u_1}^2} \right)^2 \right],$$

with  $\widehat{s}_{u_i} := \widehat{s}_{u_i, T_2^{(N)}}^2$  defined in (11),  $\widehat{\Sigma}_{s,(u_1,u_2)}^2$  and  $\widehat{\sigma}_{s_{u_i}}^2$  as in Corollary 3.1. From Proposition 2.1 and Corollary 3.1, it holds that  $\widehat{\sigma}_{h^{(s)},(u_1,u_2)}^2 \xrightarrow[N \rightarrow \infty]{\mathbb{P}} \sigma_{h^{(s)},(u_1,u_2)}^2$ .

Similarly, define  $\widehat{\sigma}_{g^{(C_0^*)},(u_1,u_2)}^2$

$$\widehat{\sigma}_{g^{(C_0^*)},(u_1,u_2)}^2 = \frac{\widehat{\sigma}_{C_0^*,u_2}^2}{\widehat{C}_{0,T_1^{(N)}}(X,u_1)^2} - 2 \frac{\widehat{C}_{0,T_1^{(N)}}(X,u_2)}{\widehat{C}_{0,T_1^{(N)}}(X,u_1)^3} \widehat{\Sigma}_{C_0^*,(u_1,u_2)}^2 + \frac{\widehat{C}_{0,T_1^{(N)}}(X,u_2)^2}{\widehat{C}_{0,T_1^{(N)}}(X,u_1)^4} \widehat{\sigma}_{C_0^*,u_1}^2,$$

with  $\widehat{C}_{0,T_1^{(N)}}(X,u_i)$  as in (10) and  $\widehat{\Sigma}_{C_0^*,(u_1,u_2)}^2$  as in (Conjecture).

Take a confidence level  $\alpha \in (0, 1)$  and set  $q_{1-\frac{\alpha}{2}}$  such that  $\mathbb{P}(|N(0, 1)| \leq q_{1-\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}$ . We define the symmetric test  $\phi_{T^{(N)}}$  with asymptotic level  $\alpha$  as

$$\phi_{T_1^{(N)}, T_2^{(N)}} = \mathbf{1} \left\{ \sqrt{\frac{|T_1^{(N)}|}{\widehat{\sigma}_{R_{u_1, u_2}}^2}} \left| \widehat{R}_{T_1^{(N)}}(u_1, u_2) - \widehat{R}_{T_2^{(N)}}^{H_0}(u_1, u_2) \right| \geq q_{1-\frac{\alpha}{2}} \right\}, \quad (22)$$

where  $\widehat{\sigma}_{R_{u_1, u_2}}^2$  is the consistent variance estimator built before.

**Comments** The test statistic is well defined for all levels  $u \neq \mu$ , similarly to [Biermé et al. \(2019\)](#) where we could not use the level  $u = 0$ . Here  $\mu$  is unknown and choosing  $u_1$  close to  $\mu$  will make the ratio  $\widehat{R}_T(u_1, u_2)$  unstable. A solution is to modify the test statistics and consider the difference  $\widetilde{R}_T(u_1, u_2) = \widehat{C}_{0,T}(X, u_2) - \widehat{C}_{0,T}(X, u_1)$  instead. The test could be modified accordingly at the expense of more tedious calculations, in particular due to the fact that the unknown quantity  $w := \frac{\lambda}{\sigma^2}$  will appear in the test statistics  $\widetilde{R}_T(u_1, u_2)$  under  $H_0$ .

Furthermore, under  $H_0$  we do not impose any constraint on the shape of the covariance function nor on the spectral moment other than Assumptions (A1) and (A2). However, contrary to the case where the field is centered with unit variance (see [Biermé et al. \(2019\)](#)), the test statistic under  $H_0$  depends on unknown quantities  $(\mu, \sigma^2)$ .

Establishing the consistency of the test, even for particular alternatives  $H_1$ , is a delicate problem as it requires the knowledge of a limit theorem for  $\widehat{C}_{0,T}(X, u)$  as well as a good understanding of the value  $R(u_1, u_2)$  under the alternative. One possible hypothesis  $H_1$  is

$$H_1 : \exists k \geq 3, X \text{ is Student random field with } k \text{ unknown degrees of freedom.}$$

Under  $H_1$ , we can derive the ratio

$$R(u_1, u_2) := \frac{C_0^*(X, u_2)}{C_0^*(X, u_1)} \stackrel{\text{under } H_1}{=} \frac{u_2}{u_1} \left( \frac{k + u_2^2}{k + u_1^2} \right)^{\frac{1-k}{2}} := R^{H_1}(u_1, u_2).$$

As explained in Section 3.1 in [Biermé et al. \(2019\)](#), if a joint central limit theorem for Lipschitz-Killing curvatures for Student fields was known this would entail the consistency of the test in (22), *i.e.*,  $\mathbb{P}_{H_1}(\phi_{T_1^{(N)}, T_2^{(N)}} = 1) \xrightarrow{N \rightarrow \infty} 1$ .

### 3.2.2 Interval containing the unknown location $\mu$

We propose the following construction for an interval containing the unknown location  $\mu$  of the field based on the observations  $T \cap E_X(u_1)$ , for  $J$  fixed levels  $u_1 < \dots < u_J$ .

*Procedure to build interval for the unknown location of the field*

*Input :*

Let  $J \in \mathbb{N}$  be fixed, the field  $X$  is observed at levels  $u_1 < \dots < u_J$ .

*Estimation :*

Let  $\hat{\sigma}_{s_{u_j}}^2 := \hat{\Sigma}_{s, (u_i, u_i)}^2$ , with  $j \in \{1, \dots, J\}$  and  $\hat{\Sigma}_{s, (u_i, u_i)}^2$  as in (18).

Define  $\hat{j} = \underset{j \in \{1, \dots, J\}}{\operatorname{argmin}} \hat{\sigma}_{s_{u_j}}^2$  and

$$\hat{j}^{\pm 1} = \begin{cases} \hat{j} + 1 & \text{if } \hat{\sigma}_{s_{u_{\hat{j}+1}}}^2 \leq \hat{\sigma}_{s_{u_{\hat{j}-1}}}^2, \\ \hat{j} - 1 & \text{if } \hat{\sigma}_{s_{u_{\hat{j}-1}}}^2 \leq \hat{\sigma}_{s_{u_{\hat{j}+1}}}^2 \end{cases},$$

*Final output :*

The following interval contains  $\mu$  with large probability:  $I_\mu := [\min\{u_{\hat{j}}, u_{\hat{j}^{\pm 1}}\}, \max\{u_{\hat{j}}, u_{\hat{j}^{\pm 1}}\}]$

with convention that if  $\hat{j} = 0$ ,  $I_\mu := [-u_1, u_0]$  and if  $\hat{j} = J$ ,  $I_\mu := [u_J, -u_{J-1}]$ .

**Proposition 3.3.** *It holds that  $\mathbb{P}(\mu \in [\min\{u_{\hat{j}}, u_{\hat{j}^{\pm 1}}\}, \max\{u_{\hat{j}}, u_{\hat{j}^{\pm 1}}\}]) \xrightarrow{T \nearrow \mathbb{R}^2} 1$ .*

Proof of Proposition 3.3 is postponed to Section A.2.

## 4 Test to compare two images of excursion sets

Let  $Y$  and  $Z$  be two Gaussian fields satisfying Assumptions (A0) and (A1) with possibly different mean, variance, spectral moment or correlation function. Suppose one has two images of the excursion sets of these fields observed at the respective levels  $u_Y$  and  $u_Z$  and wants to know if it is possible to compare the Lipschitz-Killing curvatures of these two images, *i.e.* one wants to test whether the *effective level* of  $Y$  (denoted by  $s_{u_Y}(Y)$ ) is equal to the *effective level* of  $Z$  (denoted by  $s_{u_Z}(Z)$ ):

$$H_0 : s_{u_Y}(Y) = s_{u_Z}(Z) \quad H_1 : s_{u_Y}(Y) \neq s_{u_Z}(Z).$$

Introduce the quantity  $S(Y, Z) := s_{u_Y} - s_{u_Z}$  and  $\widehat{S}_{T^{(N)}}(Y, Z) := \widehat{s}_{u_Y, T^{(N)}} - \widehat{s}_{u_Z, T^{(N)}}$  its empirical counterpart. Notice that  $S(Y, Z) = 0$  under the null hypothesis. We can now state the auto-normalised central limit theorem for the test statistics  $\widehat{S}_{T^{(N)}}(Y, Z)$ .

**Corollary 4.1.** *Let  $X$  be a Gaussian random field satisfying Assumptions (A0) and (A1). It holds that*

$$\sqrt{\frac{1}{\widehat{\Sigma}_{Y,Z}}} \widehat{S}_{T^{(N)}}(Y, Z) \xrightarrow[N \rightarrow \infty]{d, H_0} \mathcal{N}(0, 1), \quad \text{where} \quad \widehat{\Sigma}_{Y,Z} = \widehat{\sigma}_{s, u_Y}^2 + \widehat{\sigma}_{s, u_Z}^2,$$

with  $\widehat{\sigma}_{s, u_Y}^2 := \widehat{\Sigma}_{s, (u_Y, u_Y)}^2$  (resp.  $\widehat{\sigma}_{s, u_Z}^2 := \widehat{\Sigma}_{s, (u_Z, u_Z)}^2$ ), the consistent estimator of the variance in (18).

Corollary 4.1 is proved by using Proposition 2.1, Corollary 3.1 and Theorem 3.1. Notice that due to the independence of the two considered images, we get a simplified expression for the asymptotic variance.

Let  $q_{1-\frac{\alpha}{2}}$  be such that  $\mathbb{P}(|N(0, 1)| \leq q_{1-\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}$ . Finally, we introduce the symmetric test  $\phi_{T^{(N)}}$  with asymptotic level  $\alpha$ :

$$\phi_{T^{(N)}} = \mathbf{1} \left\{ \sqrt{\frac{1}{\widehat{\Sigma}_{Y,Z}}} \left| \widehat{S}_{T^{(N)}}(Y, Z) \right| \geq q_{1-\frac{\alpha}{2}} \right\}. \quad (23)$$

Clearly the test proposed in Equation (23) is consistent: under  $H_1 : s_{u_Y}(Y) \neq s_{u_Z}(Z)$ , the test statistics  $\widehat{S}_{T^{(N)}}(Y, Z)$  diverges.

In the following, we evaluate the performances of the proposed test  $\phi_{T^{(N)}}$  in (23) on simulated mammograms (from a digital texture model, see Section 4.1) and on real mammograms (from mini-MIAS database, see Section 4.2). In particular, we aim to test whether the *effective level* of an image is equal to the *effective level* of another one.

## 4.1 Comparing images of excursion sets: a synthetic mammograms study

**The data-set** In this section we consider images from a recent solid breast texture model inspired by the morphology of medium and small scale fibro-glandular and adipose tissue observed in clinical breast computed tomography (bCT) images (UC Davis database). Each adipose compartment is modeled as a union of overlapping ellipsoids and the underlying dynamic is dictated by a spatial marked point process. The contour of each ellipsoid is blurred to render the model more realistic (for details see [Li et al. \(2016\)](#), Section 2.2 and Figure 1). Finally, the synthetic mammograms images were simulated by x-ray projection. Evaluation provided in [Li et al. \(2016\)](#) has shown that simulated mammograms and digital breast tomosynthesis images are visually similar, according to medical experts.

We consider 15 simulated 2D images generated by this texture model. The images were kindly provided by GE Healthcare France, department *Mammography*. From a clinical point of view, radiologists use the Breast Imaging Reporting and Data System (or BI-RADS) to classify breast density into four categories. They go from almost all fatty tissue to extremely dense tissue with very little fat. In this latter category, it can be hard to see small tumors in or around the dense tissue. The images we studied belong to three morphologic situation groups :

- (F) Almost entirely adipose breasts;
- (FG) Scattered fibro-glandular dense breasts;
- (D) Heterogeneously dense breasts.

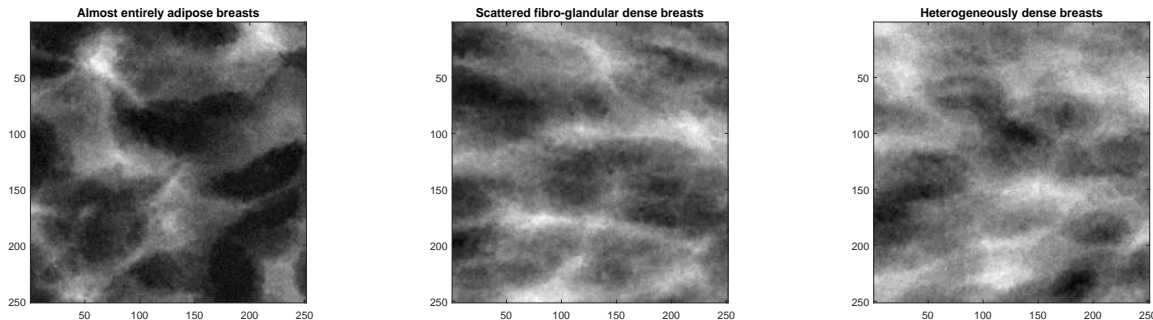


Figure 4: Synthetic digital mammograms study. Image 1.F from group (F) (left), 1.FG from group (FG) (center) and 1.D from group (D) (right). Image size:  $251 \times 251$ .

The first image from each group is reported in Figure 4. As observed in Section 3 in [Li et al. \(2016\)](#),

these simulated digital mammograms from groups (F), (FG) and (D) show a high visual realism compared to real images in these 3 different clinical situations (see also Figure 8).

**Importance of the *effective level*** In Figure 5 (left) we estimate the *effective levels*  $\widehat{s}_u$  for different values of  $u$ . One can appreciate that the estimated *effective levels* are different for each group of mammograms. Moreover, it seems there exists an order between the three groups of mammograms in terms of *effective levels*. This means that if one want to compare images from different groups, one cannot use the same value of  $u$  to perform the comparison. Indeed, in Figure 5 (right-up) we display the excursion sets of the first image of each groups (denoted image 1.F, 1.FG and 1.D in Figure 4) for the same fixed level  $u = 2200$ . The resulting excursion sets look completely different from one group to the other. However this difference is not necessarily the result of a intrinsic difference between the images, but a problem in the calibration of the level  $u$  used for the comparison. On this example, as we observed in Figure 5 (left) a difference between the groups of *effective levels*, we calibrate a level  $u$  for each group such that they all have an *effective level* close to 0, *i.e.*, adaptive levels  $\tilde{u}$ , such that  $|\widehat{s}_{\tilde{u}}| < \epsilon$ , for  $\epsilon = 10^{-2}$ . Notice that this choice guarantees a minimal variance for the estimated *effectively level* (see Corollary 2.1). Figure 5 (right-down) is obtained by considering  $\tilde{u}_{1.F} = 2133$ ,  $\tilde{u}_{1.FG} = 2291$ ,  $\tilde{u}_{1.D} = 2518$ .

We observe now that the excursion sets in Figure 5 (right-down) look visually similar. Moreover, at these levels  $\tilde{u}$ , we theoretically expect to get estimated Euler characteristic values close to zero (see Equation (7)). This behavior is illustrated in the boxplots gathered in Figure 6.



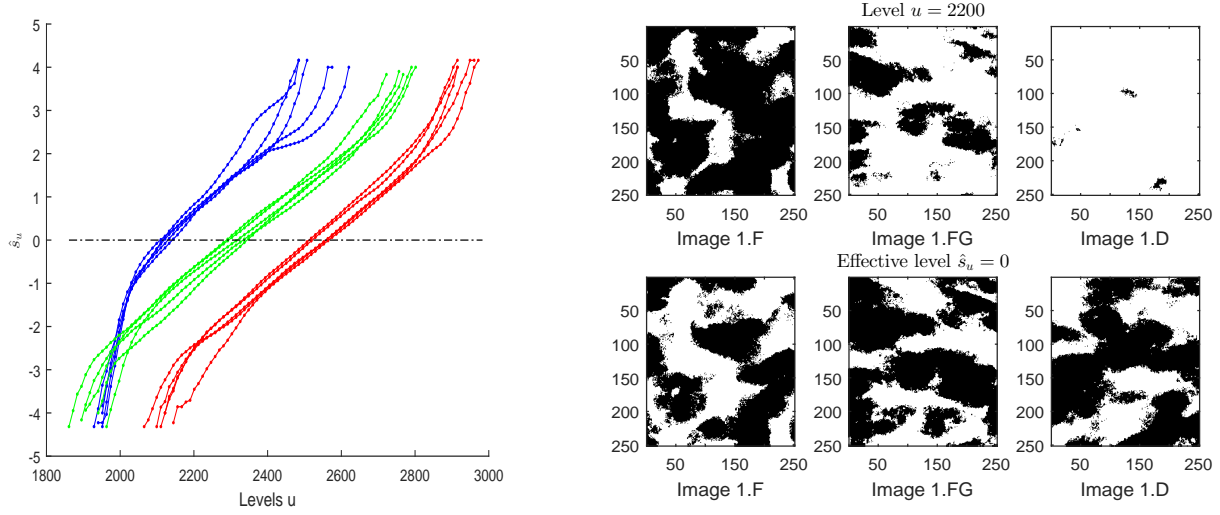


Figure 5: Synthetic digital mammograms study. **Left:** Estimated  $\hat{s}_u$  for several values of  $u$  for each image: group (F) in blue curves, (FG) in green curves and (D) in red ones. **Right:** Excursion sets for a fixed level  $u = 2200$  (first row) and for the three adaptive levels  $\tilde{u}$ , such that for each  $\tilde{u}$  it holds that  $|\hat{s}_{\tilde{u}}| < \epsilon$ , for  $\epsilon = 10^{-2}$  (second row).

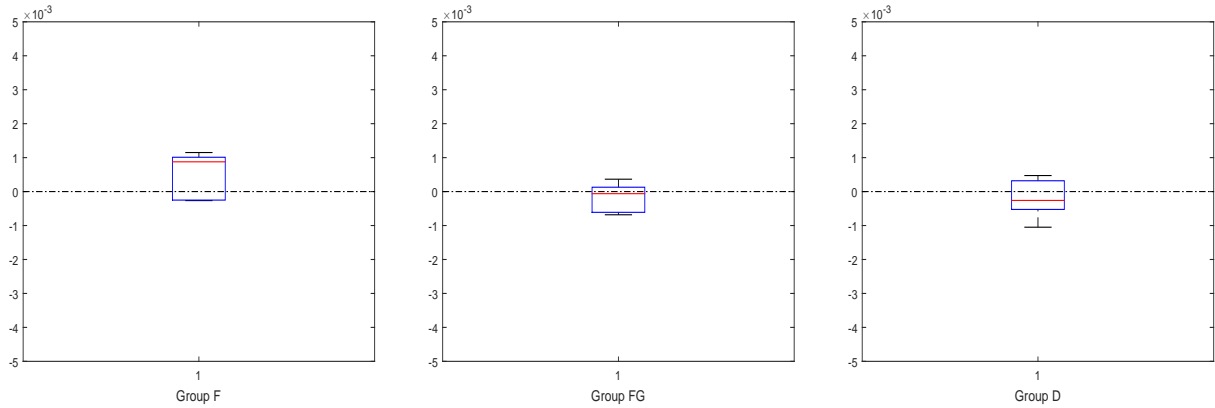


Figure 6: Synthetic digital mammograms study. Boxplots of the estimates  $\hat{C}_{0,T}(\tilde{u})$  in (10), for each group of images (group (F) in left panel, (FG) in center and (D) in right one) with adaptive levels  $\tilde{u}$ , such that for each  $\tilde{u}$  it holds that  $|\hat{s}_{\tilde{u}}| < \epsilon$ , for  $\epsilon = 10^{-2}$ .

Then, any geometric estimated quantity (as for instance the Lipschitz-Killing curvatures) on these excursion sets cannot be compared at the same levels but at the same *effective levels*.

**Testing two images of excursion sets** We now test for the 1.F, 1.FG, 1.D images in Figure 4:

$$H_0 : s_{\tilde{u}_Y}(Y) = s_{\tilde{u}_Z}(Z) \quad \text{versus} \quad H_1 : s_{\tilde{u}_Y}(Y) \neq s_{\tilde{u}_Z}(Z), \quad (24)$$

for  $Y, Z \in \{1.F, 1.FG, 1.D\}$ , where  $\tilde{u}_Y$  and  $\tilde{u}_Z$  are the adaptive levels previously defined and such that  $|\widehat{s}_{\tilde{u}}| < 10^{-2}$ , *i.e.*, the associated  $\widehat{C}_{0,T}(\tilde{u})$  is zero (see right panel of Figure 5 and boxplots in Figure 6).

Indeed, for these levels  $\tilde{u}_{1.F}$ ,  $\tilde{u}_{1.FG}$ ,  $\tilde{u}_{1.D}$ , the  $p$ -values of the test (24), gathered in Table 1 below, lead to accept the  $H_0$  hypothesis. This preliminary adjustment of the image level is necessary to properly compare the excursion sets. In case only two excursion sets are available and not the whole images, this test (24) can be a preliminary prerequisite to determine whether or not it is legitimate to perform a comparison between these images.

1.F versus 1.FG	1.F versus 1.D	1.FG versus 1.D
0.9858	0.9511	0.9642

Table 1: Synthetic digital mammograms study.  $p$ -values associated to the test  $H_0 : s_{\tilde{u}_Y}(Y) = s_{\tilde{u}_Z}(Z)$  for excursion sets in Figure 5 (right, second row) with  $\tilde{u}_{1.F} = 2133$ ,  $\tilde{u}_{1.FG} = 2291$ ,  $\tilde{u}_{1.D} = 2518$ .

We now consider 1000 different values of  $u$  and for each  $u$  we perform the following test

$$H_0 : s_u(Y) = s_u(Z) \quad \text{versus} \quad H_1 : s_u(Y) \neq s_u(Z), \quad (25)$$

for  $Y, Z$  images of this synthetic mammograms data-set. Notice that, contrary to the test (24), we did not previously choose adaptive levels. Then the test (25) is performed for possible different *effective levels*  $s_u$ . In Table 2 we display the number of  $p$ -values associated to the 1000 values of  $u$  that are smaller than the significant level  $\alpha = 0.2$ . In particular we perform an intra-class analysis (left panel in Table 2) and an inter-classes analysis (right panel in Table 2).

Groups	$\alpha = 0.2$	Intra-class Analysis			
F	Image	2.F	3.F	4.F	5.F
	1.F	118	153	153	<b>204</b>
	2.F		89	126	<b>226</b>
	3.F			117	<b>211</b>
	4.F				<b>250</b>
FG	Image	2.FG	3.FG	4.FG	5.FG
	1.FG	14	105	2	4
	2.FG		74	58	21
	3.FG			129	123
	4.FG				5
D	Image	2.D	3.D	4.D	5.D
	1.D	3	8	168	110
	2.D		82	190	188
	3.D			70	5
	4.D				48

$\alpha = 0.2$	Inter-classes Analysis				
Image	1.FG	2.FG	3.FG	4.FG	5.FG
1.F	<b>1000</b>	<b>1000</b>	<b>1000</b>	<b>1000</b>	<b>1000</b>
2.F	<b>694</b>	<b>1000</b>	<b>1000</b>	<b>712</b>	<b>1000</b>
3.F	<b>686</b>	<b>1000</b>	<b>1000</b>	<b>715</b>	<b>1000</b>
4.F	<b>641</b>	<b>1000</b>	<b>1000</b>	<b>633</b>	<b>818</b>
5.F	<b>1000</b>	<b>1000</b>	<b>1000</b>	<b>1000</b>	<b>1000</b>
Image	1.FG	2.FG	3.FG	4.FG	5.FG
1.D	<b>1000</b>	<b>1000</b>	<b>1000</b>	<b>1000</b>	<b>1000</b>
2.D	<b>1000</b>	<b>774</b>	<b>1000</b>	<b>1000</b>	<b>890</b>
3.D	<b>1000</b>	<b>1000</b>	<b>1000</b>	<b>1000</b>	<b>1000</b>
4.D	<b>1000</b>	<b>1000</b>	<b>1000</b>	<b>1000</b>	<b>1000</b>
5.D	<b>1000</b>	<b>1000</b>	<b>1000</b>	<b>1000</b>	<b>1000</b>
Image	1.D	2.D	3.D	4.D	5.D
1.F	<b>1000</b>	<b>1000</b>	<b>1000</b>	<b>1000</b>	<b>1000</b>
2.F	<b>1000</b>	<b>1000</b>	<b>1000</b>	<b>1000</b>	<b>1000</b>
3.F	<b>1000</b>	<b>1000</b>	<b>1000</b>	<b>1000</b>	<b>1000</b>
4.F	<b>1000</b>	<b>1000</b>	<b>1000</b>	<b>1000</b>	<b>1000</b>
5.F	<b>1000</b>	<b>1000</b>	<b>1000</b>	<b>1000</b>	<b>1000</b>

Table 2: Synthetic digital mammograms study. Number of  $p$ -values associated to the 1000 different values of  $u$  that are smaller than the significant level  $\alpha = 0.2$ . The numbers larger than  $\alpha \times 1000 = 200$  for which  $H_0$  is rejected.

In the intra-class analysis only tests involving the image 5.F lead to a number of  $p$ -values slightly larger than  $\alpha \times 1000 = 200$ , for which  $H_0$  is rejected<sup>1</sup>. In all the other cases, the test accepts the  $H_0$  hypothesis (see left-side Table 2). Conversely in the inter-classes analysis all the obtained numbers of  $p$ -values are much higher than 200 (right-side table). This means that, to properly compare excursion set of a given level  $u$  for images belonging to different classes, a previous *effective level*

<sup>1</sup>The same atypical behavior for figure 5.F has been previously observed in [Biermé et al. \(2019\)](#), where the same data-set is studied.

scaling is necessary.

Finally, we provide a graphical illustration of the test in (25) for three couple of images (2.F and 3.F, first panel; 1.F versus 5.D, second panel; 1.F and 3.FG, third panel). In bold marked points we represent the cases when the test (25) rejects  $H_0$  for at level  $\alpha = 0.2$ . These points are drawn on the estimates  $\hat{s}_u$ , for 1000 values of  $u$ . Coherently, the test accepts  $H_0$  (see Table 2) in the first panel of Figure 7 (intra-classes analysis) except for some extreme values of level  $u$ . Conversely, the test (25) rejects the  $H_0$  hypothesis for the values of  $u$  in the considered grid in the last two panels of Figure 7 (inter-classes analysis).

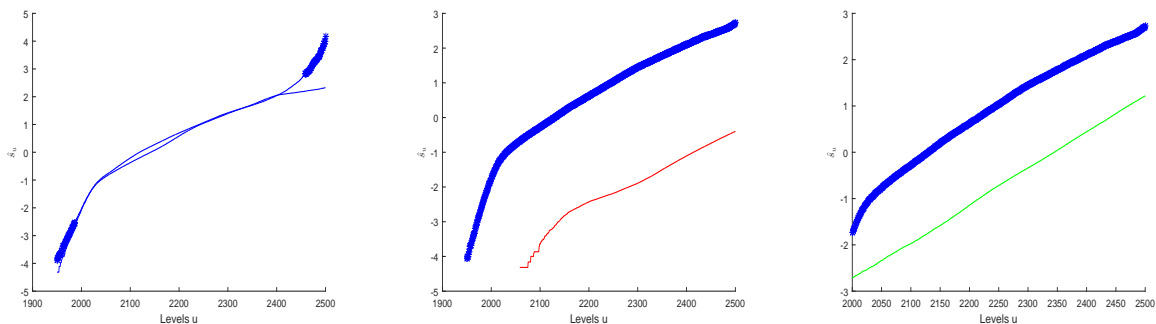


Figure 7: Synthetic digital mammograms study. Estimation of  $\hat{s}_u$  for 1000 different values of  $u$  and couples of images: 2.F and 3.F (first panel); 1.F versus 5.D (second panel); 1.F and 3.FG (third panel). In bold marked points we represent the cases where the test (25) rejects  $H_0$  for a significant level  $\alpha = 0.2$ . Group (F) is displayed using blue curves, (FG) green curves and (D) red ones.

## 4.2 Comparing images of excursion sets: a real digital mammograms study

**The data-set** The Mammographic Image Analysis Society (MIAS) is an organisation of UK research groups interested in the understanding of mammograms and has produced a database of real digital mammograms. The range of intensity in all images is represented from 0 to 255 and we consider images size of  $250 \times 250$  pixels. We study 211 mammograms classified in terms of the character of background tissue: fatty tissue group (F) (66 images), fatty-glandular (FG) (67 images), and dense (D) (77 images). Mammographic images are available online: <http://peipa.essex.ac.uk/info/mias.html> (see also Suckling et al. (1994)). One image of each group is displayed in Figure 8.

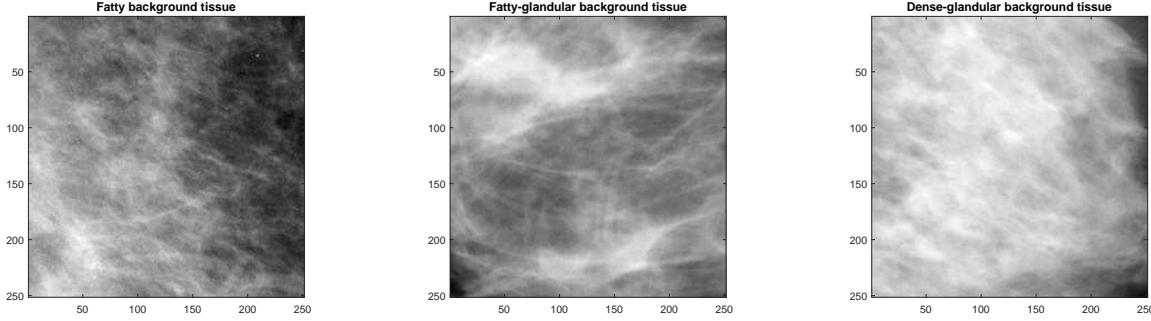


Figure 8: Real digital mammograms study. Image from group fatty background tissue (F) (left), from group fatty-glandular (FG) (center) and from group dense (D) (right). Image size:  $251 \times 251$ .

**Inference and testing for *effective level*** For this real mammograms data-set, similarly to Figure 7, we perform the test in (25) and we display in bold marked points the cases when the test rejects  $H_0$  for a significant level  $\alpha = 0.2$  (see Figure 9 below). These points are drawn on the estimates  $\hat{s}_u$ , for the considered values of  $u$ . For the sake of brevity we only display here the inter-classes analysis. Notice that in the first panel of Figure 9 we choose the difficult comparison between the two closest images between groups F and FG.

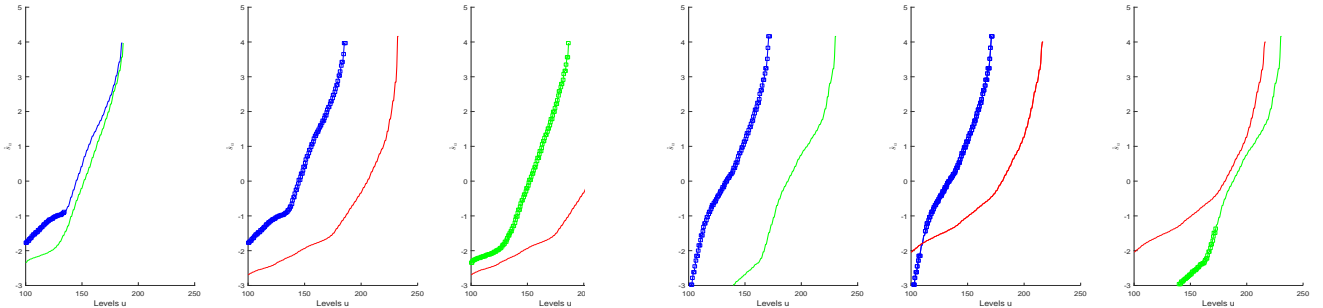


Figure 9: Real digital mammograms study. Estimation of  $\hat{s}_u$  for several levels  $u$  and a couple of images. From left to right: 1.F versus 1.FG, 1.F versus 1.D, 1.FG versus 1.D, 27.F versus 19.FG, 27.F versus 20.D, 19.FG versus 20.D. In bold marked points we represent the relative  $u$  values such that the test (25) rejects  $H_0$  for a significant level  $\alpha = 0.2$ . Group (F) is displayed using blue curves, (FG) green curves and (D) red ones.

Remark that the  $H_0$  hypothesis is accepted for almost all the levels  $u$  in the considered grid in the first and the last panels of Figure 9, due to the visible proximity of the *effective levels*  $\hat{s}_u$ . In the second, third and fourth panel, the  $H_0$  hypothesis is rejected for all  $u$  and the associated excursion

sets can not be compared without a preliminary image processing. The fifth panel in Figure 9 represents an interesting hybrid situation. Almost everywhere  $H_0$  is rejected except for a small interval of  $u$ 's values where  $H_0$  is accepted. In this small range of  $u$ , the relative excursion sets can be considered visually similar.

Analogously to Table 2, we now perform the test (25) to compare all possible combinations of excursion sets of images in this data-set in a grid of 200 values of level  $u \in [100, 240]$ . In boxplots of Figure 10 we display the number of  $p$ -values associated to intra and inter-classes analyses for 200 different values of  $u \in [100, 240]$  that are smaller than the significant level  $\alpha = 0.2$ . The reference value is represented by the horizontal line at level  $\alpha \times 200 = 40$ , above which  $H_0$  is rejected.

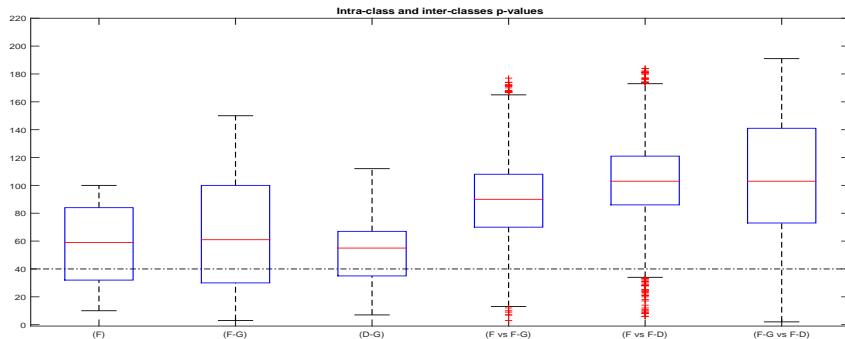


Figure 10: Real digital mammograms study. **Intra and inter-classes analysis** for the test to comparing images from the considered 3 groups. We display the boxplots of number of  $p$ -values associated to the 200 different values of  $u \in [100, 240]$  that are smaller than the significant level  $\alpha = 0.2$ . In horizontal line we display the reference threshold  $\alpha \times 200 = 40$ , above which  $H_0$  is rejected.

As one can expect, in Figure 10 we observe differences in terms of the medians and of the variances: the intra-classes boxplots (first three boxplots) have smaller median and variance values with respect to the inter-classes ones (last three boxplots). As in Section 4.1, this study suggests that the *effective level* scaling procedure seems to be necessary especially in the comparison of two images of excursion sets belonging to two different background tissue groups.

## 5 Conclusions and discussion

In this paper, we have presented new statistics based on the average LK curvatures of the excursion set of a stationary non-standard isotropic Gaussian field  $X$  on  $\mathbb{R}^2$ , in particular on the Euler characteristic, the half perimeter and the area. These tools allow to build consistent inference procedures based only on a sparse observation of the Gaussian 2D random field with unknown location and scale parameters. A byproduct is the construction of a test to determine if two images of excursion sets can be compared. Here we discuss some potential improvements of the results proposed in this work.

Firstly, notice that in Sections 3.2 and 4, our testing procedure results relied mostly on the area and not on the Euler characteristic devise. The reason being that for the Euler characteristic we did not fully establish the consistency of its asymptotic variance estimator, therefore limiting the possibility to properly calibrate statistical procedures. If Conjecture 3.1 was proved, this would enlarge the options for testing and inference. Besides, if a joint auto-normalized central limit theorem for  $(C_0^{/T}(X, u), C_2^{/T}(X, u))$  was available, this would imply results on the joint behavior of  $(\widehat{S}_{u,T}, \widehat{w}_{u,T})$  in Definition 2.1 and the possibility to build consolidated tests using both quantities.

Secondly, inspired by the analysis of Sections 4.1 and 4.2, the *effective level* could be useful to build a “classification criterion” between the three groups (F), (FG) and (D). Indeed, the estimated  $\widehat{s}_u$  seems globally able to distinguish images coming from different groups relying exclusively on the sparse information of the excursion set with a same level  $u$ .

To explore this idea, in Figure 11 we represent for the real digital data-set the adaptive level  $u$  ( $y$  axis) for all considered 211 images such that the estimated *effective level*  $\widehat{s}_u$  is approximated equal to  $\delta$ , for several values of  $\delta$  and where the different groups are distinguished in notation.

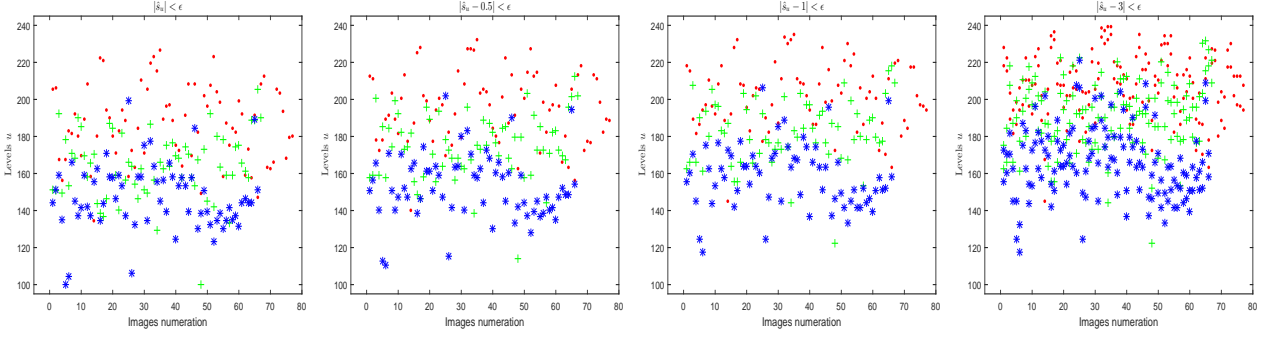


Figure 11: Real digital mammograms study. Adaptive level  $u$  (*i.e.*, level  $u$  such that  $s_u$  is close to a given value  $\delta$ ,  $y$  axis) for all images ( $x$  axis), such that  $|\widehat{s}_u - \delta| < \epsilon$ , for  $\epsilon = 10^{-2}$  and  $\delta = 0$  (first panel)  $\delta = 0.5$  (second panel),  $\delta = 1$  (third panel) and  $\delta = 3$  (fourth panel). Images from group fatty (F) are displayed by blue stars, fatty-glandular group (FG) by green crosses and dense group (D) by red points.

Visually, it appears that for small values of  $\delta$  (*i.e.*, for intermediate values of  $u$ ) the groups might be classified from their *effective level*. Obviously, for large values of  $\delta$  (*i.e.*, for extreme values of  $u$ ) the quality of the separation of groups is less good, as one can expect (see last panel of Figure 11). Moreover, one can expect that for large values of the level  $u$ , the obtained classes will be more distant also due to the large variance of the *effective level* estimator (see Corollary 2.1). However globally, this tool seems adequate to recognize the underlying background breast tissue. However an investigation on a rigorous classification procedure based on *effective levels* remains an open point and is left for future works.

Finally, in the present paper, the empirical estimator of the variance is obtained from the cutting of the domain  $T^{(N)}$ , as described in Section 3.1 and illustrated in Figure 3. While the theoretical constraints of the cutting procedure are clear, its practical and numerical implementation has to be treated carefully. Indeed we need simultaneously that  $M_N \rightarrow \infty$  (*i.e.*, a large number of subwindow domains) and that  $\text{dist}(V^{(N,(i,j))}, V^{(N,(i',j'))}) \rightarrow \infty$  (*i.e.*, the size of each subwindow domain is large). As previously described in Bulinski et al. (2012), this is the crucial and well known compromise between variance and bias. Indeed, Bulinski et al. (2012) provide a numerical study in the very specific case of a spherical covariance model to find the best subwindow size in order to minimize the mean error for the variance estimator for three intermediate levels  $u$ . The appropriate size and the form (*e.g.*, square, rectangular, ...) of the subwindow is a crucial issue in applications (see



Section 4). Furthermore, the cutting of  $T^{(N)}$  has to guarantee that in each subwindow one can find observations of the excursion set at the chosen level  $u \in \mathbb{R}$ . Then, the subwindow procedure is also implicitly related to the choice of the (intermediate, large or extremes) value of observation level  $u$ .

**Acknowledgments:** The authors express their gratitude to two anonymous Referees for their valuable comments on this article. The authors are grateful to Anne Estrade for fruitful discussions. The authors acknowledge Z. Li, (GE Healthcare France, department *Mammography*) for simulated 2D digital mammograms data-sets in Section 4.1 and the related discussion. Mammographic Image Analysis Society database can be found here <http://peipa.essex.ac.uk/info/mias.html>. This work has been supported by the French government, through the 3IA Côte d’Azur Investments in the Future project managed by the National Research Agency (ANR) with the reference number ANR-19-P3IA-0002. This work has been partially supported by the project ANR MISTIC (ANR-19-CE40-0005).

**Supporting information.** Additional information for this article is available online.

## References

- Adler, R. J., Subag, E., and Taylor, J. E. (2012). Rotation and scale space random fields and the Gaussian kinematic formula. *The Annals of Statistics*, 40(6):2910–2942.
- Adler, R. J. and Taylor, J. E. (2007). *Random fields and geometry*. Springer Monographs in Mathematics. Springer, New York.
- Adler, R. J. and Taylor, J. E. (2011). *Topological complexity of smooth random functions*, volume 2019 of *Lecture Notes in Mathematics*. Springer, Heidelberg. Lectures from the 39th Probability Summer School held in Saint-Flour, 2009, École d’Été de Probabilités de Saint-Flour. [Saint-Flour Probability Summer School].
- Azaïs, J. M. and Chassan, M. (2020). Discretization error for the maximum of a Gaussian field. *Stochastic Processes and their Applications*, 130(2):545–559.
- Azaïs, J.-M. and Pham, V.-H. (2016). Asymptotic formula for the tail of the maximum of smooth stationary Gaussian fields on non locally convex sets. *Stochastic Processes and their Applications*, 126(5):1385 – 1411.

- Azaïs, J. M. and Wschebor, M. (2009). *Level sets and extrema of random processes and fields*. John Wiley & Sons.
- Berzin, C. (2018). Estimation of Local Anisotropy Based on Level Sets. *ArXiv e-prints:1801.03760*.
- Biermé, H. and Desolneux, A. (2016). On the perimeter of excursion sets of shot noise random fields. *The Annals of Probability*, 44(1):521–543.
- Biermé, H., Di Bernardino, E., Duval, C., and Estrade, A. (2019). Lipschitz-Killing curvatures of excursion sets for two-dimensional random fields. *Electronic Journal of Statistics*, 13(1):536–581.
- Breuer, P. and Major, P. (1983). Central limit theorems for non-linear functionals of Gaussian fields. *Journal of Multivariate Analysis*, 13(3):425–441.
- Bulinski, A., Spodarev, E., and Timmermann, F. (2012). Central limit theorems for the excursion set volumes of weakly dependent random fields. *Bernoulli*, 18(1):100–118.
- Cammarota, V. and Marinucci, D. (2020). A reduction principle for the critical values of random spherical harmonics. *Stochastic Processes and their Applications*, 130(4):2433 – 2470.
- Casaponsa, B., Crill, B., Colombo, L., Danese, L., Bock, J., Catalano, A., Bonaldi, A., Basak, S., Bonavera, L., Coulais, A., et al. (2016). Planck 2015 results: Xvi. Isotropy and statistics of the CMB.
- Cheng, D. (2016). Excursion probability of certain non-centered smooth gaussian random fields. *Stochastic Processes and their Applications*, 126(3):883 – 905.
- David, O. S. and Worsley, K. (1995). Testing for a Signal with Unknown Location and Scale in a Stationary Gaussian Random Field. *The Annals of Statistics*, 23(2):608–639.
- Di Bernardino, E., Estrade, A., and León, J. R. (2017). A test of Gaussianity based on the Euler characteristic of excursion sets. *Electronic Journal of Statistics*, 11(1):843–890.
- Ebner, B., Henze, N., Klatt, M. A., and Mecke, K. (2018). Goodness-of-fit tests for complete spatial randomness based on Minkowski functionals of binary images. *Electronic Journal of Statistics*, 12(2):2873–2904.

- Estrade, A. and León, J. (2016). A Central Limit Theorem for the Euler characteristic of a Gaussian excursion set. *Annals of Probability*, 44(6):3849–3878.
- Gott, J. R., Colley, W. N., Park, C.-G., Park, C., and Mugnolo, C. (2007). Genus topology of the cosmic microwave background from the WMAP 3-year data. *Monthly Notices of the Royal Astronomical Society*, 377(4):1668–1678.
- Gott, J. R., Hambrick, D. C., Vogeley, M. S., Kim, J., Park, C., Choi, Y.-Y., Cen, R., Ostriker, J. P., and Nagamine, K. (2008). Genus topology of structure in the sloan digital sky survey: Model testing. *The Astrophysical Journal*, 675(1):16.
- Hug, D., Last, G., and Schulte, M. (2016). Second-order properties and central limit theorems for geometric functionals of Boolean models. *The Annals of Applied Probability*, 26(1):73–135.
- Kratz, M. and Vadlamani, S. (2018). Central limit theorem for Lipschitz–Killing curvatures of excursion sets of Gaussian random fields. *Journal of Theoretical Probability*, 31(3):1729–1758.
- Lachièze-Rey, R. (2019). Bicovariograms and Euler characteristic of random fields excursions. *Stochastic Processes and their Applications*, 129(11):4687 – 4703.
- Li, Z., Desolneux, A., Muller, S., and Carton, A. K. (2016). A novel 3D stochastic solid breast texture model for X-Ray breast imaging. In Tingberg, A., Lång, K., and Timberg, P., editors, *Breast Imaging*, pages 660–667, Cham. Springer International Publishing.
- Lindgren, G. (2000). Wave analysis by Slepian models. *Probabilistic engineering mechanics*, 15(1):49–57.
- Longuet-Higgins, M. S. (1957). The statistical analysis of a random, moving surface. *Philosophical Transactions of the Royal Society A*, 249(966):321–387.
- Marinucci, D. and Rossi, M. (2015). Stein-Malliavin approximations for nonlinear functionals of random eigenfunctions on  $S^d$ . *Journal of Functional Analysis*, 268(8):2379 – 2420.
- Mattfeldt, T., Meschenmoser, D., Pantle, U., and Schmidt, V. (2011). Characterization of mammary gland tissue using joint estimators of Minkowski functionals. *Image Analysis & Stereology*, 26(1):13–22.

- Müller, D. (2017). A central limit theorem for Lipschitz–Killing curvatures of Gaussian excursions. *Journal of Mathematical Analysis and Applications*, 452(2):1040–1081.
- Nieto-Reyes, A., Cuesta-Albertos, J. A., and Gamboa, F. (2014). A random-projection based test of Gaussianity for stationary processes. *Computational Statistics & Data Analysis*, 75:124–141.
- Nourdin, I. and Peccati, G. (2012). Normal approximation with Malliavin calculus. *Cambridge Tracts in Mathematics*, 192.
- Panigrahi, S., Taylor, J., and Vadlamani, S. (2019). Kinematic formula for heterogeneous gaussian related fields. *Stochastic Processes and their Applications*, 129(7):2437 – 2465.
- Pantle, U., Schmidt, V., and Spodarev, E. (2010). On the estimation of integrated covariance functions of stationary random fields. *Scandinavian Journal of Statistics*, 37(1):47–66.
- Pham, V.-H. (2013). On the rate of convergence for central limit theorems of sojourn times of Gaussian fields. *Stochastic Processes and their Applications*, 123(6):2158 – 2174.
- Reddy, T. R., Vadlamani, S., and Yogeshwaran, D. (2018). Central limit theorem for exponentially quasi-local statistics of spin models on cayley graphs. *Journal of Statistical Physics*.
- Schmalzing, J. and Górski, K. M. (1998). Minkowski functionals used in the morphological analysis of cosmic microwave background anisotropy maps. *Monthly Notices of the Royal Astronomical Society*, 297(2):355–365.
- Shashkin, A. P. (2002). Quasi-associatedness of a Gaussian system of random vectors. *Russian Mathematical Surveys*, 57(6):1243–1244.
- Suckling, J., Parker, J., and Dance, D. (1994). The Mammographic Image Analysis Society Digital Mammogram Database. *Excerpta Medica, International Congress Series 1069*.
- Szegő, G. (1959). *Orthogonal Polynomials*. Number vol. 23,ptie. 1 in American Mathematical Society. American Mathematical Society.
- Taqqu, M. S. (1977). Law of the iterated logarithm for sums of non-linear functions of Gaussian variables that exhibit a long range dependence. *Probability Theory and Related Fields*, 40(3):203–238.

Wschebor, M. (2006). *Surfaces aléatoires: mesure géométrique des ensembles de niveau*, volume 1147. Springer.

**Corresponding author's address:** Elena Di Bernardino, Laboratoire J.A. Dieudonné, UMR CNRS 7351, Université Côte d'Azur, Parc Valrose, 06108 Nice, Cedex 2, France; elenadb@unice.fr

## A Proofs

### A.1 Preliminary results

In the following we prove two auxiliary consistency results for  $\widehat{C}_{0,T^{(N)}}(X, u)$  and  $\widehat{C}_{2,T^{(N)}}(X, u)$ .

**Proposition A.1.** *Let  $X$  be a Gaussian random field satisfying Assumptions (A0) and (A2).*

(i) *Let  $T_1, \dots, T_m$  be  $m$  cubes in  $\mathbb{R}^d$  such that  $|T_1| = \dots = |T_m|$  and  $\text{dist}(T_i, T_j) > 0$  for all  $i \neq j \in \{1, \dots, m\}$ . Let  $u_1, \dots, u_m$  be  $m$  levels in  $\mathbb{R}$ , for any integer  $N > 0$  and  $\widehat{C}_{0,T_i^{(N)}}(X, u_i)$  as in (10), let*

$$Q_i^{(N)} := \sqrt{|T_i^{(N)}|} (\widehat{C}_{0,T_i^{(N)}}(X, u_i) - C_0^*(X, u_i)), \quad \text{for } i \in \{1, \dots, m\}.$$

*As  $N \rightarrow +\infty$ ,  $(Q_1^{(N)}, \dots, Q_m^{(N)})$  converges in distribution to a centered Gaussian vector with covariance matrix  $\text{diag}(\sigma_{C_0^*, u_1}^2, \dots, \sigma_{C_0^*, u_m}^2)$  where  $\sigma_{C_0^*, u_i}^2 < +\infty, \forall i \in \{1, \dots, m\}$ .*

(ii) *Let  $T$  be a cube in  $\mathbb{R}^d$ ,  $u_1, \dots, u_m$  be  $m$  levels in  $\mathbb{R}$ . For any integer  $N > 0$  and  $\widehat{C}_{0,T^{(N)}}(X, u_i)$  as in (10), let*

$$S_i^{(N)} := \sqrt{|T^{(N)}|} (\widehat{C}_{0,T^{(N)}}(X, u_i) - C_0^*(X, u_i)), \quad \text{for } i \in \{1, \dots, m\}.$$

*As  $N \rightarrow +\infty$ ,  $(S_1^{(N)}, \dots, S_m^{(N)})$  converges in distribution to a centered Gaussian vector with covariance matrix  $(\Sigma_{C_0^*, (u_i, u_j)}^2)_{1 \leq i, j \leq m}$  with  $\Sigma_{C_0^*, (u_i, u_j)}^2 < +\infty$ , for  $i, j \in \{1, \dots, m\}$ .*

*Proof.* Let  $i \in \{1, \dots, m\}$ . The  $i$ th coordinate of the considered multivariate central limit theorem

can be written using the following decomposition

$$\begin{aligned}
Q_i^{(N)} &= \sqrt{|T_i^{(N)}|}(\widehat{C}_{0,T_i^{(N)}}(X, u_i) - C_0^*(X, u_i)) = \sqrt{|T_i^{(N)}|} \left( C_0^{/T_i^{(N)}}(X, u_i) - \mathbb{E}[C_0^{/T_i^{(N)}}(X, u_i)] \right) \\
&\quad - \frac{1}{\pi} \sqrt{|T_i^{(N)}|} \left( C_1^{/T_i^{(N)}}(X, u_i) - \mathbb{E}[C_1^{/T_i^{(N)}}(X, u_i)] \right) \frac{|\partial T_i^{(N)}|_1}{|T_i^{(N)}|} \\
&\quad + \sqrt{|T_i^{(N)}|} \left( C_2^{/T_i^{(N)}}(X, u_i) - \mathbb{E}[C_2^{/T_i^{(N)}}(X, u_i)] \right) \left( \frac{1}{2\pi} \frac{|\partial T_i^{(N)}|_1^2}{|T_i^{(N)}|^2} - \frac{1}{|T_i^{(N)}|} \right) \\
&:= I_0(T_i^{(N)}) + I_1(T_i^{(N)}) + I_2(T_i^{(N)}).
\end{aligned}$$

For a Gaussian random field satisfying Assumptions (A0) and (A2), Theorem 1.1 in [Kratz and Vadlamani \(2018\)](#) and Theorem 2.1 in [Müller \(2017\)](#) give  $\sqrt{|T_i^{(N)}|} (C_1^{/T_i^{(N)}}(X, u_i) - \mathbb{E}[C_1^{/T_i^{(N)}}(X, u_i)])$  admits a Gaussian centered limit distribution and therefore  $I_1(T_i^{(N)}) \xrightarrow[T \nearrow \mathbb{R}^2]{\mathbb{P}} 0$ . The same discussion holds for  $I_2(T_i^{(N)})$ . Let

$$Z_i^{(N)} = \sqrt{|T_i^{(N)}|} (C_0^{/T_i^{(N)}}(X, u_i) - \mathbb{E}[C_0^{/T_i^{(N)}}(X, u_i)]) \text{ for } i \in \{1, \dots, m\}.$$

One can adapt the proof of the CLT in Proposition 5 (a) in [Di Bernardino et al. \(2017\)](#) to prove that for  $X$  satisfying Assumptions (A0) and (A2),  $X(0) \sim \mathcal{N}(\mu, \sigma^2)$  and as  $N \rightarrow +\infty$ ,  $(Z_1^{(N)}, \dots, Z_m^{(N)})$  converges in distribution to a centered Gaussian vector with diagonal covariance matrix and finite elements. Finally, applying the multivariate Slutsky's theorem we get the item (i). The proof of (ii) comes down in a similar way using item (b) of Theorem 2.5 in [Estrade and León \(2016\)](#).  $\square$

**Proposition A.2** (Asymptotic normality of  $(\widehat{C}_{2,T^{(N)}}(X, u_1), \dots, \widehat{C}_{2,T^{(N)}}(X, u_m))$ ). *Let  $X$  be a Gaussian random field satisfying Assumptions (A0) and (A1). For a positive integer  $N$ , consider  $T^{(N)} = \{Nt : t \in T\}$  and  $\widehat{C}_{2,T^{(N)}}(X, u)$  the estimator defined in (8) built on the observation  $T^{(N)} \cap E_X(u_i)$ , where  $u_1, \dots, u_m$  are fixed. Then,*

$$\sqrt{|T^{(N)}|} (\widehat{C}_{2,T^{(N)}}(X, u_1) - C_2^*(X, u_1), \dots, \widehat{C}_{2,T^{(N)}}(X, u_m) - C_2^*(X, u_m))$$

converges in distribution to a centered Gaussian vector with  $m \times m$  covariance matrix  $(\Sigma_{C_2^*, (u_i, u_j)}^2)_{1 \leq i, j \leq m}$  given by  $\Sigma_{C_2^*, (u_i, u_j)}^2 = \frac{1}{2\pi} \int_{\mathbb{R}^2} \int_0^{\rho(t)} g_{(u_i, u_j)}(r) dr dt \in (0, +\infty)$  where

$$g_{(u_i, u_j)}(r) = \frac{1}{\sqrt{1-r^2}} \exp \left\{ - \frac{(u_i - \mu)^2 - 2r(u_i - \mu)(u_j - \mu) + (u_j - \mu)^2}{2\sigma^2(1-r^2)} \right\}.$$

The proof of Proposition A.2 comes down from Theorem 4 in [Bulinski et al. \(2012\)](#), together with [Shashkin \(2002\)](#) ensuring Gaussian fields are quasi-associated.

## A.2 Proofs of the obtained results

### Proof of Proposition 1.1

The Gaussian kinematic formula provides the mean LK curvatures of excursion sets of  $X$  within a rectangle  $T$  (see, *e.g.*, Theorem 15.9.5 in Adler and Taylor (2007) or Theorem 4.8.1, 4.3.1 in Adler and Taylor (2011)), for  $u \in \mathbb{R}$  and  $i = 0, 1, 2$ ,

$$\mathbb{E}[\mathcal{L}_i(X, T, u)] = \sum_{l=0}^{2-i} \begin{bmatrix} i+l \\ l \end{bmatrix} (2\pi)^{-l/2} \left(\frac{\lambda}{\sigma^2}\right)^{l/2} \mathcal{M}_l(X, u) \mathcal{L}_{i+l}(T)$$

where  $\mathcal{L}_j(T)$ ,  $j = 0, 1, 2$  are defined in (5),  $\begin{bmatrix} i+l \\ l \end{bmatrix} = \binom{i+l}{l} \frac{\omega_{l+i}}{\omega_l \omega_i}$  with  $\omega_k$  the Lebesgue measure of the  $k$ -dimensional unit ball ( $w_0 = 1$ ,  $w_1 = 2$  and  $w_2 = \pi$ ), and, following Formula (3.5.2) in Adler and Taylor (2011), the coefficients  $\mathcal{M}_l(X, u)$ ,  $l = 0, 1, 2$  are obtained having an expansion in  $\theta$  at order 2 of

$$\begin{aligned} \mathbb{P}\left(G(0) \geq \frac{u-\mu}{\sigma} - \theta\right) &= \psi\left(\frac{u-\mu}{\sigma}\right) - \theta\psi'\left(\frac{u-\mu}{\sigma}\right) + \frac{1}{2}\theta^2\psi''\left(\frac{u-\mu}{\sigma}\right) + O(\theta^3) \\ &= \mathcal{M}_0(X, u) + \theta\mathcal{M}_1(X, u) + \frac{1}{2}\theta^2\mathcal{M}_2(X, u) + O(\theta^3), \end{aligned}$$

where  $G(0)$  is a Gaussian random variable with zero mean and unit variance and  $\psi$  its tail distribution. This concludes the proof.  $\square$

### Proof of Proposition 2.2

Set  $h : s \mapsto s \exp\{-s^2/2\}/(2\pi)^{3/2}$  and note that  $|T_1^{(N)}| = |T_2^{(N)}|$ . The  $i$ th coordinate of the vector  $|T_1^{(N)}|^{1/2}(\widehat{w}_{u_i, T_1^{(N)}, T_2^{(N)}} - w)$  can be decomposed as follows

$$\sqrt{|T_1^{(N)}|} \left[ \frac{h(s_{u_i})}{h(\widehat{s}_{u,2})} \left( \frac{\widehat{C}_{0, T_1^{(N)}}(X, u_i)}{h(s_{u_i})} - \frac{C_0^*(X, u_i)}{h(s_{u_i})} \right) + C_0^*(X, u_i) \left( \frac{1}{h(\widehat{s}_{u,2})} - \frac{1}{h(s_{u_i})} \right) \right],$$

where  $\widehat{s}_{u,2} := \widehat{s}_{u, T_2^{(N)}}$ . Since  $h$  is continuous, using Proposition 2.1 we get  $h(s_{u_i})/h(\widehat{s}_{u_i, T_2^{(N)}}) \xrightarrow[N \rightarrow \infty]{\mathbb{P}} 1$ , for all  $1 \leq i \leq m$ . Then, by using Propositions 2.1 and A.1, the multivariate delta method and that  $\text{dist}(T_1^{(N)}, T_2^{(N)}) \rightarrow \infty$  as  $N \rightarrow \infty$ , we obtain that  $\sqrt{|T_1^{(N)}|}(\widehat{w}_{u_1, T_1^{(N)}, T_2^{(N)}} - w, \dots, \widehat{w}_{u_m, T_1^{(N)}, T_2^{(N)}} - a)$  converges in distribution to a centered Gaussian vector with covariance matrix given by  $A \Sigma_{C_0^*}^2 A^t + B \Sigma_{\frac{1}{h}}^2 B^t$ , where  $\cdot^t$  denotes the matrix transposition,  $A$  and  $B$  are the diagonal matrices

$$A = \text{diag}\left(\frac{1}{h(s_{u_1})}, \dots, \frac{1}{h(s_{u_m})}\right), \quad B = \text{diag}\left(C_0^*(X, u_1), \dots, C_0^*(X, u_m)\right),$$

$\Sigma_{C_0^*}^2 = (\Sigma_{C_0^*,(u_i,u_j)}^2)_{1 \leq i,j \leq m}$  is defined in Proposition A.1,  $\Sigma_{\frac{1}{h}}^2 = J_{\frac{1}{h}} \Sigma_s^2 J_{\frac{1}{h}}^t$  with  $\Sigma_s^2 = (\Sigma_{s,(u_i,u_j)}^2)_{1 \leq i,j \leq m}$  defined in Proposition 2.1 and  $J_{\frac{1}{h}}$  the  $m \times m$  Jacobian matrix

$$J_{\frac{1}{h}} = (2\pi)^{\frac{3}{2}} \text{diag} \left( \frac{(s_{u_1}^2 - 1)}{s_{u_1}^2} \exp \left\{ \frac{s_{u_1}^2}{2} \right\}, \dots, \frac{(s_{u_m}^2 - 1)}{s_{u_m}^2} \exp \left\{ \frac{s_{u_m}^2}{2} \right\} \right).$$

After computations we get  $A \Sigma_{C_0^*}^2 A^t = \left( \frac{\Sigma_{C_0^*,(u_i,u_j)}^2}{h(s_{u_i})h(s_{u_j})} \right)_{1 \leq i,j \leq m}$  and

$$B \Sigma_{\frac{1}{h}}^2 B^t = \left( (2\pi)^3 C_0^*(X, u_i) C_0^*(X, u_j) \frac{(s_{u_i}^2 - 1)(s_{u_j}^2 - 1)}{s_{u_i}^2 s_{u_j}^2} \exp \left\{ \frac{1}{2}(s_{u_i}^2 + s_{u_j}^2) \right\} \Sigma_{s,(u_i,u_j)}^2 \right)_{1 \leq i,j \leq m}.$$

Adding these expressions provides the desired formula.  $\square$

### Proof of Proposition 3.2

First, write

$$\begin{aligned} \sqrt{|T_1^{(N)}|} (\widehat{R}_{T_1^{(N)}}(u_1, u_2) - \widehat{R}_{T_2^{(N)}}^{H_0}(u_1, u_2)) &= \sqrt{|T_1^{(N)}|} (\widehat{R}_{T_1^{(N)}}(u_1, u_2) - R^{H_0}(u_1, u_2)) \\ &\quad + \sqrt{|T_2^{(N)}|} (R^{H_0}(u_1, u_2) - \widehat{R}_{T_2^{(N)}}^{H_0}(u_1, u_2)). \end{aligned} \quad (26)$$

For the first term in (26) we use Proposition A.1 (ii) and the delta method with the function  $g : (x, y) \mapsto \frac{x}{y}$ . We get after computations, using that  $\widehat{R}_{T_1}(u_1, u_2) = g(\widehat{C}_{0,T_1}(X, u_2), \widehat{C}_{0,T_1}(X, u_1))$ ,

$$\sqrt{|T_1^{(N)}|} \left( g(\widehat{C}_{0,T_1^{(N)}}(X, u_2), \widehat{C}_{0,T_1^{(N)}}(X, u_1)) - g(C_0^*(X, u_2), C_0^*(X, u_1)) \right) \xrightarrow[N \rightarrow \infty]{d, H_0} \mathcal{N} \left( 0, \sigma_{g(C_0^*), (u_1, u_2)}^2 \right)$$

where

$$\sigma_{g(C_0^*), (u_1, u_2)}^2 = \frac{\sigma_{C_0^*, u_2}^2}{C_0^*(X, u_1)^2} - 2 \frac{C_0^*(X, u_2)}{C_0^*(X, u_1)^3} \Sigma_{C_0^*, (u_1, u_2)}^2 + \frac{C_0^*(X, u_2)^2}{C_0^*(X, u_1)^4} \sigma_{C_0^*, u_1}^2.$$

For the second term in (26) we use Proposition 2.1 and apply the delta method with the function  $h : (x, y) \mapsto \frac{x}{y} \exp\{-\frac{1}{2}(x^2 - y^2)\}$ . We get after computations, using that  $\widehat{R}_{T_2^{(N)}}^{H_0}(u_1, u_2) = h(\widehat{s}_{u_2, T_2^{(N)}}, \widehat{s}_{u_1, T_2^{(N)}})$ ,

$$\sqrt{|T_2^{(N)}|} \left( h(s_{u_2}, s_{u_1}) - h(\widehat{s}_{u_2, T_2^{(N)}}, \widehat{s}_{u_1, T_2^{(N)}}) \right) \xrightarrow[N \rightarrow \infty]{d, H_0} \mathcal{N} \left( 0, \sigma_{h(s), (u_1, u_2)}^2 \right)$$

where  $\sigma_{h(s), (u_1, u_2)}^2 = \nabla h(s_{u_2}, s_{u_1})' \Sigma_{s, (u_1, u_2)}^2 \nabla h(s_{u_2}, s_{u_1})$ ,  $\Sigma_{s, (u_1, u_2)}^2$  is defined as in Proposition 2.1,

$$\sigma_{h(s), (u_1, u_2)}^2 = e^{-(s_{u_2}^2 - s_{u_1}^2)} \left[ \sigma_{s_{u_2}}^2 \left( \frac{1 - s_{u_2}^2}{s_{u_1}} \right)^2 + 2 \Sigma_{s, (u_1, u_2)}^2 \left( \frac{1 - s_{u_2}^2}{s_{u_1}} \right) \left( -\frac{s_{u_2}}{s_{u_1}^2} + s_{u_2} \right) + \sigma_{s_{u_1}}^2 \left( -\frac{s_{u_2}}{s_{u_1}^2} + s_{u_2} \right)^2 \right].$$

Using  $\text{dist}(T_1^{(N)}, T_2^{(N)}) \rightarrow \infty$  as  $N \rightarrow \infty$  and  $\sigma_{R_{u_1, u_2}}^2 = \sigma_{g(C_0^*), (u_1, u_2)}^2 + \sigma_{h(s), (u_1, u_2)}^2$  we get the result.

$\square$



### Proof of Proposition 3.3

Since  $J$  is a finite set and using Corollary 3.1, it immediately holds that

$$\left(\widehat{\sigma}_{s_{u_1}}^2, \dots, \widehat{\sigma}_{s_{u_J}}^2\right) \xrightarrow[T \nearrow \mathbb{R}^2]{\mathbb{P}} \left(\sigma_{s_{u_1}}^2, \dots, \sigma_{s_{u_J}}^2\right).$$

Denote by  $j^* = \operatorname{argmin}_{j \in \{1, \dots, J\}} \sigma_{s_{u_j}}^2$ , it follows

$$\begin{aligned} \mathbb{P}(\widehat{j} \neq j^*) &= \mathbb{P}\left(\bigcup_{j \neq j^*} \{\widehat{\sigma}_{s_{u_j}}^2 \leq \widehat{\sigma}_{s_{u_{j^*}}}^2\}\right) \leq \sum_{j \neq j^*} \mathbb{P}(\widehat{\sigma}_{s_{u_j}}^2 - \sigma_{s_{u_j}}^2 + (\sigma_{s_{u_j}}^2 - \sigma_{s_{u_{j^*}}}^2) \leq \widehat{\sigma}_{s_{u_{j^*}}}^2 - \sigma_{s_{u_{j^*}}}^2) \\ &\leq \sum_{j \neq j^*} \mathbb{P}(|\sigma_{s_{u_j}}^2 - \sigma_{s_{u_{j^*}}}^2| \leq |\widehat{\sigma}_{s_{u_{j^*}}}^2 - \sigma_{s_{u_{j^*}}}^2| + |\widehat{\sigma}_{s_{u_j}}^2 - \sigma_{s_{u_j}}^2|) \\ &\leq \sum_{j \neq j^*} \left(\mathbb{P}(\tfrac{1}{2}|\sigma_{s_{u_j}}^2 - \sigma_{s_{u_{j^*}}}^2| \leq |\widehat{\sigma}_{s_{u_{j^*}}}^2 - \sigma_{s_{u_{j^*}}}^2|) + \mathbb{P}(\tfrac{1}{2}|\sigma_{s_{u_j}}^2 - \sigma_{s_{u_{j^*}}}^2| \leq |\widehat{\sigma}_{s_{u_j}}^2 - \sigma_{s_{u_j}}^2|)\right) \xrightarrow[T \nearrow \mathbb{R}^2]{} 0, \end{aligned}$$

where we used that  $J$  is finite,  $\forall j \neq j^*$ ,  $|\sigma_{s_{u_j}}^2 - \sigma_{s_{u_{j^*}}}^2| > 0$  and Corollary 3.1. Finally, we derive that  $\widehat{j} \xrightarrow{\mathbb{P}} j^*$ , for  $T \nearrow \mathbb{R}^2$ , together with the definition of  $j^*$  and the fact that  $u \mapsto \sigma_{s_u}^2$  is decreasing on  $(-\infty, \mu)$  and increasing on  $(\mu, \infty)$ , the result follows.  $\square$

### A.3 Proof of Proposition 3.1

**Preliminaries on the Itô-Wiener chaos decomposition for  $\mathcal{L}_2$**  Let  $G := (X - \mu)/\sigma$  be the centered and unit variance Gaussian random field associated to  $X$ , for all fixed levels  $u$ . Recall that  $s_u = (u - \mu)/\sigma$ , then  $\mathcal{L}_{2,T}(X, u) = \mathcal{L}_{2,T}(G, s_u)$ , which is a square-integrable functional of the Gaussian field  $G$ . It admits an orthogonal decomposition into Itô-Wiener chaos in the  $L^2$  sense (see, *e.g.*, Marinucci and Rossi (2015), Nourdin and Peccati (2012)):

$$\mathcal{L}_2(G, s_u, T) = \sum_{q=0}^{+\infty} \frac{\beta_q(s_u)}{q!} \int_T H_q(G(t)) dt,$$

where  $H_q$  is the  $q$ -th Hermite polynomial, *i.e.*, for  $z \in \mathbb{R}$ ,  $H_0(z) = 1$  and

$$H_q(z) = (-1)^q \exp\{z^2/2\} \frac{d^q}{dz^q} \exp\{-z^2/2\}, \quad \text{if } q \geq 1.$$

The above series converges in  $L^2(\mathbb{P})$  and for  $Z \sim \mathcal{N}(0, 1)$ ,  $\beta_q(s_u) := \mathbb{E}[\mathbf{1}_{\{Z \geq s_u\}} H_q(Z)]$ . The chaotic coefficients  $(\beta_q(s_u))_{q \geq 0}$  for  $\mathcal{L}_2(G, s_u, T)$  are given by:  $\beta_0(s_u) = \psi(s_u)$  and

$$\beta_q(s_u) = \int_{s_u}^{+\infty} \varphi(z) \frac{(-1)^q}{\varphi(z)} \frac{d^q}{dz^q} \varphi(z) dz = (-1)^{q-1} \frac{d^{q-1}}{dz^{q-1}} \varphi(s_u) = \varphi(s_u) H_{q-1}(s_u),$$

for  $q \geq 1$ , where  $\varphi$  is the probability density function of  $Z$ . Denoting  $H_q^\phi$  the ‘‘physicist Hermite polynomials’’ we have  $H_q^\phi(x) = 2^{q/2}H_q(\sqrt{2}x)$  and it holds  $\forall x, |(2^q q! \sqrt{\pi})^{-1/2} H_q^\phi(x) e^{-x^2/2}| \leq C_\infty / (q+1)^{1/12}$  (see [Szegő \(1959\)](#) for the value of the constant  $C_\infty$ ). As, it holds  $\forall x \in \mathbb{R}$

$$H_q^\phi(x) e^{-x^2/2} = 2^{q/2} H_q(\sqrt{2}x) e^{-x^2/2} = e^{x^2/2} 2^{q/2} H_q(\sqrt{2}x) e^{-(\sqrt{2}x)^2/2},$$

it follows  $\forall x \in \mathbb{R}$

$$|\beta_{q+1}(\sqrt{2}x)| = |H_q(\sqrt{2}x) e^{-(\sqrt{2}x)^2/2}| \leq e^{-x^2/2} \sqrt{q! \sqrt{\pi}} C_\infty / (q+1)^{1/12} \leq C_\infty \pi^{1/4} \frac{\sqrt{q!}}{(q+1)^{1/12}}.$$

We derive the following inequality, frequently used in the sequel

$$\|\beta_0\|_\infty \leq 1 \quad \text{and} \quad \|\beta_q\|_\infty \leq c_\beta \frac{\sqrt{(q-1)!}}{q^{1/12}}, \quad q \geq 1. \quad (27)$$

The mean of  $\mathcal{L}_2(G, s_u, T)$  is its projection onto the 0-th Itô-Wiener chaos  $\mathbb{E}[\mathcal{L}_2(G, s_u, T)] = \psi(s_u)|T|$ . The remaining of the proof consists in controlling the order of the fourth moment of  $\mathcal{L}_2(G, s, T)$ , computed using the chaos decomposition as follows

$$\begin{aligned} & \mathbb{E}[\mathcal{L}_2(G, s_{u_1}, T) \mathcal{L}_2(G, s_{u_2}, T) \mathcal{L}_2(G, s_{u_3}, T') \mathcal{L}_2(G, s_{u_4}, T')] \\ &= \sum_{k_1, k_2, k_3, k_4=0}^{+\infty} \frac{\beta_{k_1}(s_{u_1}) \beta_{k_2}(s_{u_2}) \beta_{k_3}(s_{u_3}) \beta_{k_4}(s_{u_4})}{k_1! k_2! k_3! k_4!} \\ & \quad \times \int_T \int_T \int_{T'} \int_{T'} \mathbb{E}[H_{k_1}(G(t_1)) H_{k_2}(G(t_2)) H_{k_3}(G(t_3)) H_{k_4}(G(t_4))] dt_1 dt_2 dt_3 dt_4. \end{aligned}$$

**Control of a fourth moment of  $\mathcal{L}_2(G, s, T)$**  We sometimes denote  $T_1 = T_2 := T$  and  $T_3 = T_4 := T'$  and use that  $|T| = |T'|$ . We write,

$$\mathbb{E}[\mathcal{L}_2(G, s_{u_1}, T) \mathcal{L}_2(G, s_{u_2}, T) \mathcal{L}_2(G, s_{u_3}, T') \mathcal{L}_2(G, s_{u_4}, T')] = V_0 + V_1 + V_2 + V_3 + V_4, \quad (28)$$

where

$$\begin{aligned} V_0 &:= |T|^4 \beta_0(s_{u_1}) \beta_0(s_{u_2}) \beta_0(s_{u_3}) \beta_0(s_{u_4}) \\ V_1 &:= |T|^3 \sum_{j=1}^4 \frac{\beta_0(s_{u_1}) \beta_0(s_{u_2}) \beta_0(s_{u_3}) \beta_0(s_{u_4})}{\beta_0(s_{u_j})} \sum_{k_j=1}^{+\infty} \frac{\beta_{k_j}(s_{u_j})}{k_j!} \int_{T_j} \mathbb{E}[H_{k_j}(G(t_j))] dt_j \end{aligned}$$

$$\begin{aligned}
V_2 &:= |T|^2 \sum_{j_1 \neq j_2 \in \{1, \dots, 4\}} \sum_{\substack{k_{i_1}, k_{i_2}=1 \\ i_1, i_2 \neq j_1, j_2}}^{+\infty} \frac{\beta_{k_{i_1}}(s_{u_{i_1}}) \beta_{k_{i_2}}(s_{u_{i_2}}) \beta_0(s_{u_{j_1}}) \beta_0(s_{u_{j_2}})}{k_{i_1}! k_{i_2}!} \int_{T_{i_1}} \int_{T_{i_2}} \mathbb{E}[H_{k_{i_1}}(G(t_{i_1})) H_{k_{i_2}}(G(t_{i_2}))] dt_{i_1} dt_{i_2} \\
V_3 &:= |T| \sum_{j=1}^4 \beta_0(s_{u_j}) \sum_{\substack{k_{i_1}, k_{i_2}, k_{i_3}=1 \\ i_1 \neq j, i_2 \neq j, i_3 \neq j}}^{+\infty} \frac{\beta_{k_{i_1}}(s_{u_{i_1}}) \beta_{k_{i_2}}(s_{u_{i_2}}) \beta_{k_{i_3}}(s_{u_{i_3}})}{k_{i_1}! k_{i_2}! k_{i_3}!} \\
&\quad \times \int_{T_{i_1}} \int_{T_{i_2}} \int_{T_{i_3}} \mathbb{E}[H_{k_{i_1}}(G(t_{i_1})) H_{k_{i_2}}(G(t_{i_2})) H_{k_{i_3}}(G(t_{i_3}))] dt_{i_1} dt_{i_2} dt_{i_3} \\
V_4 &:= \sum_{k_1, k_2, k_3, k_4=1}^{+\infty} \frac{\beta_{k_1}(s_{u_1}) \beta_{k_2}(s_{u_2}) \beta_{k_3}(s_{u_3}) \beta_{k_4}(s_{u_4})}{k_1! k_2! k_3! k_4!} \\
&\quad \times \int_T \int_T \int_{T'} \int_{T'} \mathbb{E}[H_{k_1}(G(t_1)) H_{k_2}(G(t_2)) H_{k_3}(G(t_3)) H_{k_4}(G(t_4))] dt_1 dt_2 dt_3 dt_4.
\end{aligned}$$

It holds  $V_0 = \psi(s_{u_1})\psi(s_{u_2})\psi(s_{u_3})\psi(s_{u_4})|T|^4$  and by orthogonality of  $H_{k_1}$  with  $H_0 = 1$ ,  $\forall k_1 \geq 1$ , we get  $V_1 = 0$ . It remains to control the terms  $V_2$ ,  $V_3$  and  $V_4$ .

*Control of  $V_2$ .* Using the fact that  $\mathbb{E}[H_k(G(t))H_l(G(s))] = \delta_{k,l}k!\rho(t-s)^k$  (see *e.g.* Equation (2.1) in Breuer and Major (1983)), we get

$$\begin{aligned}
V_2 &= |T|^2 \sum_{j_1 \neq j_2 \in \{1, \dots, 4\}} \beta_0(s_{u_{j_1}}) \beta_0(s_{u_{j_2}}) \sum_{\substack{k=1 \\ i_1, i_2 \neq j_1, j_2}}^{+\infty} \frac{\beta_k(s_{u_{i_1}}) \beta_k(s_{u_{i_2}})}{k!} \int_{T_{i_1}} \int_{T_{i_2}} \rho(t_{i_1} - t_{i_2})^k dt_{i_1} dt_{i_2} \\
&= |T|^2 \left( \sum_{j_1 \neq j_2 \in \mathcal{I}_{2,1}} + \sum_{j_1 \neq j_2 \in \mathcal{I}_{2,2}} \right) \beta_0(s_{u_{j_1}}) \beta_0(s_{u_{j_2}}) \sum_{\substack{k=1 \\ i_1, i_2 \neq j_1, j_2}}^{+\infty} \frac{\beta_k(s_{u_{i_1}}) \beta_k(s_{u_{i_2}})}{k!} \int_{T_{i_1}} \int_{T_{i_2}} \rho(t_{i_1} - t_{i_2})^k dt_{i_1} dt_{i_2} \\
&=: V_{2,1} + V_{2,2},
\end{aligned}$$

where  $\mathcal{I}_{2,1} := \{(1, 3), (1, 4), (2, 3), (2, 4)\}$  and  $\mathcal{I}_{2,2} := \{(1, 2), (3, 4)\}$ . Using (27), (A1) together with the fact that on  $\mathcal{I}_{2,1}$  the integration is made on distinct rectangles  $T$  and  $T'$ , we get

$$\begin{aligned}
|V_{2,1}| &\leq 4c_\beta^2 |T|^2 \sum_{k=1}^{+\infty} \frac{(k-1)!}{k! k^{1/6}} \int_T \int_{T'} |\rho(t-t')|^k dt dt' \leq 4c_\beta^2 |T|^2 \sum_{k=1}^{+\infty} \frac{1}{k^{1+1/6}} \int_T \int_{T'} \frac{1}{(1+\|t-t'\|)^\gamma} dt dt' \\
&\leq 4c_\beta^2 |T|^2 \sum_{k=1}^{+\infty} \frac{1}{k^{1+1/6}} \frac{1}{\text{dist}(T, T')^{\gamma k-2}} \int_T \int_{T'} \frac{1}{(1+\|t-t'\|)^2} dt dt' \\
&\leq 4c_\beta^2 |T|^2 \sum_{k=1}^{+\infty} \frac{1}{\text{dist}(T, T')^{\gamma k-2}} \int_{\mathbb{R}^2} \frac{1}{(1+\|t-t'\|)^2} dt dt' \leq C_{2,1} \frac{|T|^2}{\text{dist}(T, T')^{\gamma-2}},
\end{aligned}$$

where we used Assumption (A1) and  $\text{dist}(T, T') > 2$ , then  $C_{2,1}$  is given by  $8c_\beta^2 \int_{\mathbb{R}^2} (1+\|t-t'\|)^{-2} dt dt'$ .

For  $V_{2,2}$ , using (27) and Assumption (A1), we have

$$\begin{aligned}
|V_{2,2}| &\leq |T|^2 \left( |\beta_0(s_{u_1})\beta_0(s_{u_2})| \sum_{k=1}^{\infty} \frac{|\beta_k(s_{u_3})\beta_k(s_{u_4})|}{k!} \int_{T'} \int_{T'} |\rho(t-s)|^k dt ds \right. \\
&\quad \left. + |\beta_0(s_{u_3})\beta_0(s_{u_4})| \sum_{k=1}^{\infty} \frac{|\beta_k(s_{u_1})\beta_k(s_{u_2})|}{k!} \int_T \int_T |\rho(t-s)|^k dt ds \right) \\
&\leq 2c_\beta^2 |T|^2 \sum_{k=1}^{\infty} \frac{1}{k^{1+1/6}} \int_T \int_T \frac{dt ds}{(1+\|t-s\|)^\gamma} \leq C_{2,2} |T|^2,
\end{aligned}$$

where  $C_{2,2} = 2c_\beta^2 \int_{\mathbb{R}^2} (1+\|t-t'\|)^{-2} dt dt' \sum_{k \geq 1} k^{-1-\frac{1}{6}}$  is a constant independent of  $T$ . It follows that  $V_2 = O(|T|^2)$ .

*Control of  $V_3$ .* To control this term we rely on the diagram formula and the following decomposition

$$\begin{aligned}
V_3 &= |T| \sum_{j=1}^4 \beta_0(s_{u_j}) \sum_{\substack{k_{i_1}, k_{i_2}, k_{i_3}=1 \\ i_1 \neq j, i_2 \neq j, i_3 \neq j}}^{+\infty} \frac{\beta_{k_{i_1}}(s_{u_{i_1}}) \beta_{k_{i_2}}(s_{u_{i_2}}) \beta_{k_{i_3}}(s_{u_{i_3}})}{k_{i_1}! k_{i_2}! k_{i_3}!} \\
&\quad \times \int_{T_{i_1}} \int_{T_{i_2}} \int_{T_{i_3}} \mathbb{E}[H_{k_{i_1}}(G(t_{i_1})) H_{k_{i_2}}(G(t_{i_2})) H_{k_{i_3}}(G(t_{i_3}))] dt_{i_1} dt_{i_2} dt_{i_3} \\
&= |T| \sum_{j=1}^4 \beta_0(s_{u_j}) \left( \sum_{\substack{k_{i_1}, k_{i_2}, k_{i_3}=1 \\ i_1 \neq j, i_2 \neq j, i_3 \neq j \\ k_{i_1} + k_{i_2} + k_{i_3} \leq 2N_{T, T'}}}^{+\infty} + \sum_{\substack{k_{i_1}, k_{i_2}, k_{i_3}=1 \\ i_1 \neq j, i_2 \neq j, i_3 \neq j \\ k_{i_1} + k_{i_2} + k_{i_3} > 2N_{T, T'}}}^{+\infty} \right) \frac{\beta_{k_{i_1}}(s_{u_{i_1}}) \beta_{k_{i_2}}(s_{u_{i_2}}) \beta_{k_{i_3}}(s_{u_{i_3}})}{k_{i_1}! k_{i_2}! k_{i_3}!} \\
&\quad \times \int_{T_{i_1}} \int_{T_{i_2}} \int_{T_{i_3}} \mathbb{E}[H_{k_{i_1}}(G(t_{i_1})) H_{k_{i_2}}(G(t_{i_2})) H_{k_{i_3}}(G(t_{i_3}))] dt_{i_1} dt_{i_2} dt_{i_3} \\
&= V_{3, \leq N_{T, T'}} + V_{3, > N_{T, T'}},
\end{aligned}$$

for some positive integer  $N_{T, T'}$  depending on  $T$  and  $T'$  and such that  $N_{T, T'} \rightarrow \infty$  as  $\text{dist}(T, T') \rightarrow \infty$ .

First, note that  $V_{3, > N_{T, T'}} = o(|T|^2)$  as  $\text{dist}(T, T') \rightarrow \infty$ : using that  $\mathcal{L}_2(G, s_u, T) \leq |T|$  *a.s.* and that the Itô-Wiener chaos decomposition holds in the  $L^2$  sense, we write

$$V_{3, > N_{T, T'}} \leq C |T|^2 \max_{u \in \{u_1, u_2, u_3, u_4\}} \mathbb{E} \left[ \left( \sum_{q \geq N_{T, T'}} \frac{\beta_q(s_u)}{q!} \int_T H_q(G(t)) dt \right)^2 \right],$$

which tends to 0 as the remainder of a convergent series. To control this term, we take advantage of the cutting described in Figure 3.

We now focus on  $V_{3, \leq N_{T, T'}}$ , by symmetry we set  $k_4 = 0$  and compute  $T_{k_1, k_2, k_3} := \mathbb{E}[H_{k_1}(G_1) H_{k_2}(G_2) H_{k_3}(G_3)]$  for  $k_j \geq 1, j \in \{1, 2, 3\}$  and  $(G_1, G_2, G_3)$  a standard Gaussian (see Taqqu (1977) Definition 3.1), *i.e.*

a centered Gaussian vector with  $\mathbb{E}[G_i^2] = 1$  and  $\rho_{i,j} := \mathbb{E}[G_i G_j]$  such that  $|\rho_{i,j}| \leq 1$ ,  $1 \leq i, j \leq 3$ .

The diagram formula (see [Taqqu \(1977\)](#), Lemma 3.2) gives

$$T_{k_1, k_2, k_3} = \begin{cases} \frac{k_1! k_2! k_3!}{2^q q!} \sum_{\mathcal{I}(k_1, k_2, k_3)} \rho_{i_1, j_1} \dots \rho_{i_q, j_q} & \text{if } k_1 + k_2 + k_3 = 2q, 1 \leq k_1, k_2, k_3 \leq q \\ 0 & \text{otherwise} \end{cases} \quad (29)$$

where the set of indices  $\mathcal{I}(k_1, k_2, k_3)$  is the set of all indices  $(i_1, j_1, \dots, i_q, j_q)$  such that

- i)  $(i_1, j_1, \dots, i_q, j_q) \in \{1, 2, 3\}^{2q}$ ,
- ii)  $i_1 \neq j_1, \dots, i_q \neq j_q$ ,
- iii) there are  $k_1$  indices 1,  $k_2$  indices 2 and  $k_3$  indices 3.

As  $\rho_{i,j} = \rho_{j,i}$ , after multiplying  $T_{k_1, k_2, k_3}$  by  $2^q$ ,  $\mathcal{I}(k_1, k_2, k_3)$  simplifies in  $\tilde{\mathcal{I}}(k_1, k_2, k_3)$  the set of all indices  $(i_1, j_1, \dots, i_q, j_q)$  such that

i+ii)  $((i_1, j_1), \dots, (i_q, j_q)) \in \{(1, 2), (1, 3), (2, 3)\}^q$ ,

- iii) there are  $q_1$  times (1, 2),  $q_2$  times (1, 3) and  $q_3$  times (2, 3) with  $q_1 = k_1 + k_2 - q = q - k_3$ ,  
 $q_2 = k_1 + k_3 - q = q - k_2$  and  $q_3 = k_2 + k_3 - q = q - k_1$ .

The cardinality of  $\tilde{\mathcal{I}}(k_1, k_2, k_3)$  is  $\binom{q}{q_1, q_2, q_3} = \frac{q!}{q_1! q_2! q_3!}$ . Set  $\mathcal{J}_3(q) = \{(k_1, k_2, k_3) \in \{1, \dots, q\}^3, k_1 + k_2 + k_3 = 2q\}$ , the set of indices where  $T_{k_1, k_2, k_3}$  is nonzero. Using (27), Assumption (A1) and the cardinality of  $\tilde{\mathcal{I}}(k_1, k_2, k_3)$ , it follows from (29) that,

$$\begin{aligned} |V_{3, \leq N_{T, T'}}| &\leq 4 |T| \max_{u^* \in (u_1, u_2, u_3, u_4)} |\beta_0(s_{u^*})|^4 \sum_{q=4}^{N_{T, T'}} \sum_{(k_1, k_2, k_3) \in \mathcal{J}_3(q)} \frac{|\beta_{k_1}(s_{u^*}) \beta_{k_2}(s_{u^*}) \beta_{k_3}(s_{u^*})|}{q_1! q_2! q_3!} \\ &\quad \times \int_T \int_T \int_{T'} |\rho(t_1 - t_2)|^{q-k_3} |\rho(t_1 - t_3)|^{q-k_2} |\rho(t_2 - t_3)|^{q-k_1} dt_1 dt_2 dt_3 \\ &\leq 4c_\beta^3 |T| \sum_{q=4}^{N_{T, T'}} \sum_{(k_1, k_2, k_3) \in \mathcal{J}_3(q)} \frac{\sqrt{(k_1 - 1)!(k_2 - 1)!(k_3 - 1)!}}{(k_1 k_2 k_3)^{\frac{1}{2}} (q - k_3)! (q - k_2)! (q - k_1)!} \\ &\quad \times \int_T \int_T \int_{T'} \frac{dt_1 dt_2 dt_3}{(1 + \|t_1 - t_2\|)^{\gamma(q-k_3)} (1 + \|t_1 - t_3\|)^{\gamma(q-k_2)} (1 + \|t_2 - t_3\|)^{\gamma(q-k_1)}}. \end{aligned}$$

Set

$$A_q(k_1, k_2, k_3) := \frac{\sqrt{(k_1 - 1)!(k_2 - 1)!(k_3 - 1)!}}{(k_1 k_2 k_3)^{\frac{1}{2}} (q - k_1)! (q - k_2)! (q - k_3)!}, \quad k_1 + k_2 + k_3 = 2q.$$

Consider the decomposition,

$$\begin{aligned}
|V_{3, \leq N_{T, T'}}| &\leq 4c_\beta^3 |T| \sum_{q=4}^{N_{T, T'}} \left( \sum_{k_3=1}^{\lfloor q/4 \rfloor} + \sum_{k_3=\lfloor q/4 \rfloor+1}^q \right) \sum_{\substack{k_1, k_2 \in \{1, \dots, q\} \\ k_1+k_2+k_3=2q}} A_q(k_1, k_2, k_3) \\
&\quad \times \int_T \int_T \int_{T'} \frac{dt_1 dt_2 dt_3}{(1 + \|t_1 - t_2\|)^{\gamma(q-k_3)} (1 + \|t_1 - t_3\|)^{\gamma(q-k_2)} (1 + \|t_2 - t_3\|)^{\gamma(q-k_1)}} \\
&=: 4c_\beta^3 |T| (V_{3,1} + V_{3,2}). \tag{30}
\end{aligned}$$

First, we consider  $V_{3,1}$ , *i.e.* the set  $\{1 \leq k_3 \leq \lfloor q/4 \rfloor\}$ . Straightforward computations and the fact that  $k_3 = 2q - k_1 - k_2$ , lead to

$$V_{3,1} \leq |T| \sum_{q=4}^{N_{T, T'}} \sum_{k_3=1}^{\lfloor q/4 \rfloor} \frac{1}{(\text{dist}(T, T'))^{\gamma k_3}} \sum_{\substack{k_1, k_2 \in \{1, \dots, q\} \\ k_1+k_2=2q-k_3}} A_q(k_1, k_2, k_3) \int_T \int_T \frac{dt_1 dt_2}{(1 + \|t_1 - t_2\|)^{\gamma(q-k_3)}}. \tag{31}$$

For the integral in (31), after two successive changes of variables we get, for all  $Q > 1$  that

$$\begin{aligned}
\int_T \int_T \frac{dt_1 dt_2}{(1 + \|t_1 - t_2\|)^Q} &= \int_{\mathbb{R}} \mathbf{1}_{t_1 \in T} \int_{\mathbb{R}} \frac{\mathbf{1}_{u \in T-t_1}}{(1 + \|u\|)^Q} du dt_1 = \int_T \left( \int_0^{2\pi} \int_{\mathbb{R}_+} \frac{\mathbf{1}_{re^{i\theta} \in T-t_1}}{(1+r)^Q} r dr d\theta \right) dt_1 \\
&\leq 2\pi |T| \int_0^\infty \frac{r dr}{(1+r)^Q} \leq C |T| \frac{1}{Q^2}, \quad \forall Q > 1, \tag{32}
\end{aligned}$$

where the last inequality follows from an integration by part and  $C$  is a positive constant. Next, for all  $k_3 \leq \lfloor q/4 \rfloor$ , it holds

$$\begin{aligned}
\sum_{\substack{k_1, k_2 \in \{1, \dots, q\} \\ k_1+k_2=2q-k_3}} A_q(k_1, k_2, k_3) &= \frac{\sqrt{(k_3-1)!}}{k_3^{\frac{1}{12}} (q-k_3)!} \sum_{\substack{k_1, k_2 \in \{1, \dots, q\} \\ k_1+k_2=2q-k_3}} \frac{\sqrt{(k_1-1)!(k_2-1)!}}{(k_1 k_2)^{\frac{1}{12}} (q-k_1)!(q-k_2)!} \\
&\leq 2 \frac{\sqrt{(k_3-1)!}}{k_3^{\frac{1}{12}} (q-k_3)!} \sum_{k=q-k_3}^q \frac{\sqrt{(k-1)!(2q-k-k_3-1)!}}{(k(2q-k-k_3))^{\frac{1}{12}} (q-k)!(k+k_3-q)!} \\
&= 2 \frac{\sqrt{(k_3-1)!}}{k_3^{\frac{1}{12}} (q-k_3)!} \sum_{\ell=0}^{k_3} \frac{\sqrt{(\ell+q-k_3-1)!(q-\ell-1)!}}{((\ell+q-k_3)(q-\ell))^{\frac{1}{12}} (k_3-\ell)! \ell!} \\
&\leq 2 \left(\frac{4}{3}\right)^{\frac{1}{12}} \frac{\sqrt{(k_3-1)!(q-k_3-1)!(q-1)!}}{q^{\frac{1}{6}} k_3^{\frac{1}{12}} (q-k_3)! k_3!} \sum_{\ell=0}^{k_3} \binom{k_3}{\ell} \leq 2 \left(\frac{4}{3}\right)^{\frac{7}{12}} \frac{2^{k_3}}{q^{1+\frac{1}{6}} k_3^{\frac{7}{12}}} \sqrt{\binom{q}{k_3}}, \tag{33}
\end{aligned}$$

where we used that  $k_3 \leq \lfloor q/4 \rfloor$  and that  $\ell \mapsto \sqrt{(\ell+q-k_3-1)!(q-\ell-1)!}$  is symmetric and maximal for  $\ell = 0$  (or  $\ell = k_3$ ). Injecting (32) and (33) in (31) and using that  $1 \leq k_3 \leq \lfloor q/4 \rfloor$ , lead –for a positive constant  $C$  whose value may change from line to line– to

$$V_{3,1} \leq C |T|^2 \sum_{q=4}^{N_{T, T'}} \frac{1}{q^{3+\frac{1}{6}}} \sum_{k_3=1}^{\lfloor q/4 \rfloor} \frac{2^{k_3}}{(\text{dist}(T, T'))^{\gamma k_3}} \sqrt{\binom{q}{k_3}}.$$

Set  $\eta_{T,T'} := (2/(\text{dist}(T, T'))^\gamma)$  and note that  $u_k : k \mapsto \eta_{T,T'}^k \sqrt{\binom{q}{k}}$  is decreasing and bounded by  $\sqrt{q}$  iff

$$\frac{u_{k+1}}{u_k} = \eta_{T,T'} \sqrt{\frac{q+1}{k+1} - 1} \leq 1 \iff k \geq \frac{q+1}{\eta_{T,T'}^2 + 1} \eta_{T,T'}^2 - 1$$

which always holds if  $q \leq (\eta_{T,T'})^{-2}$ . Finally, fix  $N_{T,T'} := \lfloor (\eta_{T,T'})^{-2} \rfloor$ , note that we have  $N_{T,T'} \rightarrow \infty$  as  $\text{dist}(T, T') \rightarrow \infty$ . It follows that

$$V_{3,1} \leq C \frac{|T|^2}{(\text{dist}(T, T'))^\gamma} \sum_{q=4}^{N_{T,T'}} \frac{q\sqrt{q}}{q^{3+\frac{1}{6}}} \leq C \frac{|T|^2}{(\text{dist}(T, T'))^\gamma} \sum_{q=4}^{\infty} q^{-\frac{5}{3}} = o(|T|^2). \quad (34)$$

Second, we study the set  $\{\lfloor q/4 \rfloor + 1 \leq k_3 \leq q\}$ , similarly we get

$$V_{3,2} \leq |T| \sum_{q=4}^{N_{T,T'}} \sum_{k_3=\lfloor q/4 \rfloor + 1}^q \sum_{\substack{k_1, k_2 \in \{1, \dots, q\} \\ k_1 + k_2 + k_3 = 2q}} \frac{A_q(k_1, k_2, k_3)}{(\text{dist}(T, T'))^{\gamma k_3}} \int_T \int_T \frac{dt_1 dt_2}{(1 + \|t_1 - t_2\|)^{\gamma(q-k_3)}}.$$

Note that  $(k_1, k_2, k_3, \text{ subject to } k_1 + k_2 + k_3 = 2q) \mapsto A_q(k_1, k_2, k_3)$  is maximal for  $k_1 \asymp k_2 \asymp k_3 \asymp \lfloor 2q/3 \rfloor$ , and the Stirling formula gives  $A_q(\lfloor 2q/3 \rfloor, \lfloor 2q/3 \rfloor, \lfloor 2q/3 \rfloor) = O(q^{-1}2^q)$ . Moreover,  $\text{Card}((k_1, k_2, k_3), 1 \leq k_1, k_2 \leq q, k_3 \geq \lfloor q/4 \rfloor, k_1 + k_2 + k_3 = 2q) \leq q^2$ . It follows that for some positive constant  $C$ ,

$$V_{3,2} \leq C |T| \int_{\mathbb{R}^2} \frac{dt_1 dt_2}{(1 + \|t_1 - t_2\|)^\gamma} \sum_{q=4}^{N_{T,T'}} \frac{2^q}{q} q^2 \frac{1}{(\text{dist}(T, T'))^{\gamma q/4}}. \quad (35)$$

It follows that for  $\text{dist}(T, T') > 2^{4/\gamma}$ ,  $V_{3,2} = o(|T|)$ . Observe that the condition for convergence of this term depends on  $\gamma$ : the larger  $\gamma$ , the weaker is the condition on  $\text{dist}(T, T')$  is. Gathering (34) and (35) in (30), we derive that  $V_{3, \leq N_{T,T'}} = o(|T|^3)$ , this together with  $V_{3, > N_{T,T'}} = o(|T|)$  implies that  $V_3 = o(|T|^3)$ .

*Control of  $V_4$ .* Similarly to  $V_3$ , we rely on the diagram formula and the following decomposition

$$\begin{aligned} V_4 &= \left( \sum_{q \geq 4}^{N_{T,T'}} + \sum_{q=N_{T,T'}+1}^{\infty} \right) \sum_{(k_1, k_2, k_3, k_4) \in \mathbb{N} \setminus \{0\}} \frac{\beta_{k_1}(s_{u_1}) \beta_{k_2}(s_{u_2}) \beta_{k_3}(s_{u_3}) \beta_{k_4}(s_{u_4})}{k_1! k_2! k_3! k_4!} \\ &\quad \times \int_T \int_T \int_{T'} \int_{T'} \mathbb{E}[H_{k_1}(G(t_1)) H_{k_2}(G(t_2)) H_{k_3}(G(t_3)) H_{k_4}(G(t_4))] dt_1 dt_2 dt_3 dt_4 \\ &= V_{4, \leq N_{T,T'}} + V_{4, > N_{T,T'}}, \end{aligned}$$

for some positive integer  $N_{T,T'}$  depending on  $T$  and  $T'$  and such that  $N_{T,T'} \rightarrow \infty$  as  $\text{dist}(T, T') \rightarrow \infty$ .

Using similar arguments as for  $V_{3, > N_{T,T'}}$ , we derive that  $V_{4, > N_{T,T'}} = O(|T|^2)$  as  $\text{dist}(T, T') \rightarrow \infty$ .

We focus on  $V_{4, \leq N_{T,T'}}$  and compute  $T_{k_1, k_2, k_3, k_4} := \mathbb{E}[H_{k_1}(G_1)H_{k_2}(G_2)H_{k_3}(G_3)H_{k_4}(G_4)]$  for  $k_j \geq 1$ ,  $j \in \{1, 2, 3, 4\}$  and  $(G_1, G_2, G_3, G_4)$  a standard Gaussian (see [Taqqu \(1977\)](#) Definition 3.1) with  $\rho_{i,j} := \mathbb{E}[G_i G_j] = \rho_{j,i}$  such that  $|\rho_{i,j}| \leq 1$ , for  $1 \leq i, j \leq 4$ . The diagram formula –taking into account that  $\rho_{i,j} = \rho_{j,i}$ , which comes down to multiplying the one given in [Taqqu \(1977\)](#), Lemma 3.2, by  $2^q$  (see also proof of Lemma 10.7 in [Azaïs and Wschebor \(2009\)](#)– gives

$$T_{k_1, k_2, k_3, k_4} = \begin{cases} \frac{k_1!k_2!k_3!k_4!}{q!} \sum_{\mathcal{I}(k_1, k_2, k_3, k_4)} \rho_{i_1, j_1} \cdots \rho_{i_q, j_q} & \text{if } k_1 + \dots + k_4 = 2q, 1 \leq k_1, k_2, k_3, k_4 \leq q, \\ 0 & \text{otherwise,} \end{cases} \quad (36)$$

where the set of indices  $\mathcal{I}(k_1, k_2, k_3, k_4)$  is the set of all indices  $(i_1, j_1, \dots, i_q, j_q)$  such that

i+ii)  $((i_1, j_1), \dots, (i_q, j_q)) \in \{(1, 2), (1, 3), (1, 4), (2, 3), (2, 4), (3, 4)\}^q =: \mathcal{I}_{pair}^q$ ,

iii) there are  $k_1$  indices 1,  $k_2$  indices 2,  $k_3$  indices 3 and  $k_4$  indices 4.

For  $q \in \mathbb{N}$ ,  $q \geq 4$  define the set of indices  $\mathcal{J}_4(q) = \{(k_1, k_2, k_3, k_4) \in \{1, \dots, q\}^4, k_1 + k_2 + k_3 + k_4 = 2q\}$ .

Formulae (36) and (27) lead to, for a positive constant  $C$  whose value may change from line to line,

$$\begin{aligned} V_{4, \leq N_{T,T'}} &= \sum_{q=4}^{N_{T,T'}} \sum_{(k_1, k_2, k_3, k_4) \in \mathcal{J}_4(q)} \frac{\beta_{k_1}(s_{u_1})\beta_{k_2}(s_{u_2})\beta_{k_3}(s_{u_3})\beta_{k_4}(s_{u_4})}{k_1!k_2!k_3!k_4!} T_{k_1, k_2, k_3, k_4} \\ &\leq C \sum_{q=4}^{N_{T,T'}} \frac{1}{q!} \sum_{(k_1, k_2, k_3, k_4) \in \mathcal{J}_4(q)} \frac{\sqrt{(k_1-1)!(k_2-1)!(k_3-1)!(k_4-1)!}}{(k_1 k_2 k_3 k_4)^{\frac{1}{12}}} \\ &\quad \times \sum_{\mathcal{I}(k_1, k_2, k_3, k_4)} \int_T \int_T \int_{T'} \int_{T'} \prod_{a \in \mathcal{I}_{pair}} (1 + \|t_i - t_j\|)^{-\gamma_{q_a}} dt_1 dt_2 dt_3 dt_4, \end{aligned}$$

where  $q_a$  for is the number of occurrences of the pair  $w$  and  $\mathcal{I}(k_1, k_2, k_3, k_4)$  reduces to the set of indices such that

$$\left\{ \begin{array}{l} q_{(1,2)} + q_{(1,3)} + q_{(1,4)} + q_{(2,3)} + q_{(2,4)} + q_{(3,4)} = q, \\ q_{(1,2)} + q_{(1,3)} + q_{(1,4)} = k_1, \\ q_{(1,2)} + q_{(2,3)} + q_{(2,4)} = k_2, \\ q_{(1,3)} + q_{(2,3)} + q_{(3,4)} = k_3, \\ q_{(1,4)} + q_{(2,4)} + q_{(3,4)} = k_4. \end{array} \right. \quad (37)$$



Denote  $\mathcal{I}_{pair}(q) = \{q_a, a \in \mathcal{I}_{pair}, \sum_{a \in \mathcal{I}_{pair}} q_a = q, \sum_{(i,j)} q_{(i,j)} \mathbf{1}_{\{i=l\} \cup \{j=l\}} = k_l\}$ , the latter can be rewritten as

$$\begin{aligned} V_{4, \leq N_{T, T'}} &\leq C \sum_{q=4}^{N_{T, T'}} \frac{1}{q!} \sum_{\mathcal{I}_{pair}(q)} \frac{\sqrt{(k_1-1)!(k_2-1)!(k_3-1)!(k_4-1)!}}{(k_1 k_2 k_3 k_4)^{\frac{1}{12}}} \\ &\quad \left( \begin{matrix} q \\ q_{(1,2)}, q_{(1,3)}, q_{(1,4)}, q_{(2,3)}, q_{(2,4)}, q_{(3,4)} \end{matrix} \right) \int_T \int_T \int_{T'} \int_{T'} \prod_{a \in \mathcal{I}_{pair}} (1 + \|t_i - t_j\|)^{-\gamma q_a} dt_1 dt_2 dt_3 dt_4 \\ &= C \sum_{q=4}^{N_{T, T'}} \sum_{\mathcal{I}_{pair}(q)} \frac{\sqrt{(k_1-1)!(k_2-1)!(k_3-1)!(k_4-1)!}}{(k_1 k_2 k_3 k_4)^{\frac{1}{12}} q_{(1,2)}! q_{(1,3)}! q_{(1,4)}! q_{(2,3)}! q_{(2,4)}! q_{(3,4)}!} \\ &\quad \int_T \int_T \int_{T'} \int_{T'} \prod_{a \in \mathcal{I}_{pair}} (1 + \|t_i - t_j\|)^{-\gamma q_a} dt_1 dt_2 dt_3 dt_4. \end{aligned}$$

Similarly to  $V_{3, \leq N_{T, T'}}$ , we decompose this majorant of  $V_{4, \leq N_{T, T'}}$  in  $V_{4,1} + V_{4,2}$  according to  $\{q_{(1,2)} + q_{(3,4)} \leq \lfloor 3q/4 \rfloor\}$  (which plays the same role as  $q - k_3$  in  $V_{3, \leq N_{T, T'}}$ ) and its complementary set.

First, we study the set  $\{q_{(1,2)} + q_{(3,4)} \leq \lfloor 3q/4 \rfloor\}$ , where  $q_{(1,3)} + q_{(1,4)} + q_{(2,3)} + q_{(2,4)} \geq \frac{q}{4}$ , it follows

$$\begin{aligned} V_{4,1} &\leq C \left( \int_T \int_T \frac{dt_1 dt_2}{(1 + \|t_1 - t_2\|)^\gamma} \right)^2 \sum_{q=4}^{N_{T, T'}} \frac{1}{(\text{dist}(T, T'))^{\gamma q/4}} \\ &\quad \times \sum_{\substack{\mathcal{I}_{pair}(q) \\ q_{(1,2)} + q_{(3,4)} \leq \lfloor 3q/4 \rfloor}} \frac{\sqrt{(k_1-1)!(k_2-1)!(k_3-1)!(k_4-1)!}}{(k_1 k_2 k_3 k_4)^{\frac{1}{12}} q_{(1,2)}! q_{(1,3)}! q_{(1,4)}! q_{(2,3)}! q_{(2,4)}! q_{(3,4)}!} \quad (38) \\ &\leq C \left( \int_T \int_T \frac{dt_1 dt_2}{(1 + \|t_1 - t_2\|)^\gamma} \right)^2 \sum_{q=4}^{N_{T, T'}} \frac{1}{(\text{dist}(T, T'))^{\gamma q/4}} q^5 \frac{\left(\frac{3q}{6} - 1\right)!^2}{\left(\frac{q}{6}\right)!^6}, \end{aligned}$$

where we used that  $k_i \geq 1$ , the cardinality of  $\mathcal{I}_{pair}(q)$  is bounded by  $q^5$  and that the ratio in (38) is maximal for  $q_a = \lfloor q/6 \rfloor, \forall a \in \mathcal{I}_{pair}$ . Using that  $\sqrt{2\pi m} \left(\frac{m}{e}\right)^m \leq m! \leq 2\sqrt{2\pi m} \left(\frac{m}{e}\right)^m$  provides the bound  $\left(\frac{3q}{6}\right)!^2 / \left(\frac{q}{6}\right)!^6 \leq C 3^q / q^2$  and we get for  $\text{dist}(T, T') > 3^{4/\gamma}$  that

$$V_{4,1} \leq C \left( \int_{\mathbb{R}^2} \frac{dt_1 dt_2}{(1 + \|t_1 - t_2\|)^\gamma} \right)^2 \sum_{q=4}^{+\infty} q \frac{3^q}{(\text{dist}(T, T'))^{\gamma q/4}} = O(1). \quad (39)$$

Second, consider the set  $\{q_{(1,2)} + q_{(3,4)} > \lfloor 3q/4 \rfloor\}$ ,

$$\begin{aligned} V_{4,2} &\leq C \sum_{q=4}^{N_{T, T'}} \sum_{\substack{\mathcal{I}_{pair}(q) \\ q_{(1,2)} + q_{(3,4)} > \lfloor 3q/4 \rfloor}} \frac{\sqrt{(k_1-1)!(k_2-1)!(k_3-1)!(k_4-1)!}}{(k_1 k_2 k_3 k_4)^{\frac{1}{12}} q_{(1,2)}! q_{(1,3)}! q_{(1,4)}! q_{(2,3)}! q_{(2,4)}! q_{(3,4)}!} \\ &\quad \int_T \int_T \int_{T'} \int_{T'} \prod_{a \in \mathcal{I}_{pair}} (1 + \|t_i - t_j\|)^{-\gamma q_a} dt_1 dt_2 dt_3 dt_4. \end{aligned}$$

Note that, if  $q_{(1,2)} + q_{(3,4)} > \lfloor 3q/4 \rfloor$ , it follows that  $\max\{q_{(1,2)}, q_{(3,4)}\} > \lfloor 3q/8 \rfloor$ , then using (37) we derive that on this set  $(k_1 k_2 k_3 k_4)^{-1/12} \leq Cq^{-1/6}$ , for some positive constant  $C$ . Next, we decompose  $V_{4,2}$  according to the values of  $k = q_{(1,3)} + q_{(1,4)} + q_{(2,3)} + q_{(2,4)} \leq \lfloor q/4 \rfloor$  and use Assumption (A1), (32) and (37) to get

$$\begin{aligned} V_{4,2} &\leq C|T|^2 \sum_{q=4}^{N_{T,T'}} q^{-1/6} \sum_{k=0}^{\lfloor q/4 \rfloor} \frac{1}{(\text{dist}(T, T'))^{\gamma k}} \sum_{q_{(1,3)}+q_{(1,4)}+q_{(2,3)}+q_{(2,4)}=k} \frac{1}{q_{(1,3)}!q_{(1,4)}!q_{(2,3)}!q_{(2,4)}!} \\ &\quad \times \sum_{q_{(1,2)}=1}^{q-k-1} \frac{\sqrt{(q_{(1,2)} + q_{(1,3)} + q_{(1,4)} - 1)!(q_{(1,2)} + q_{(2,3)} + q_{(2,4)} - 1)!}}{q_{(1,2)}!q_{(1,2)}^2} \\ &\quad \times \frac{\sqrt{(q_{(1,3)} + q_{(2,3)} + q - k - q_{(1,2)} - 1)!(q_{(1,4)} + q_{(2,4)} + q - k - q_{(1,2)} - 1)!}}{(q - k - q_{(1,2)})!(q - k - q_{(1,2)})^2}, \end{aligned}$$

where the constant  $C$  contains  $\int_{\mathbb{R}^2} (1 + \|t - s\|)^{-2} dt ds$ . Observe that the function in  $q_{(1,2)}$  in the last summand is symmetric with respect to  $(q - k)/2$  and is maximal for  $q_{(1,2)} \in \{1, q - k - 1\}$ . It follows that

$$\begin{aligned} V_{4,2} &\leq C|T|^2 \sum_{q=4}^{N_{T,T'}} q^{-1/6} \sum_{k=0}^{\lfloor q/4 \rfloor} \frac{1}{(\text{dist}(T, T'))^{\gamma k}} \sum_{q_{(1,3)}+q_{(1,4)}+q_{(2,3)}+q_{(2,4)}=k} \frac{1}{q_{(1,3)}!q_{(1,4)}!q_{(2,3)}!q_{(2,4)}!} \\ &\quad \times \frac{\sqrt{(q_{(1,3)} + q_{(1,4)})!(q_{(2,3)} + q_{(2,4)})!(q_{(1,3)} + q_{(2,3)} + q - k - 2)!(q_{(1,4)} + q_{(2,4)} + q - k - 2)!}}{(q - k - 1)!(q - k - 1)}. \end{aligned}$$

Considering the function in  $(q_{(1,3)}, q_{(1,4)}, q_{(2,3)}, q_{(2,4)})$  under the square root in the last display, it is easy —considering all possible cases— to derive that it is maximal at  $(k, 0, 0, 0)$  (the maximum is not unique). We obtain

$$\begin{aligned} V_{4,2} &\leq C|T|^2 \sum_{q=4}^{N_{T,T'}} q^{-\frac{1}{6}} \sum_{k=0}^{\lfloor q/4 \rfloor} \frac{1}{(\text{dist}(T, T'))^{\gamma k} k!} \sum_{\substack{q_{(1,3)}+q_{(1,4)}+ \\ q_{(2,3)}+q_{(2,4)}=k}} \frac{k!}{q_{(1,3)}!q_{(1,4)}!q_{(2,3)}!q_{(2,4)}!} \frac{\sqrt{k!(q-2)!(q-k-2)!}}{(q-k-1)!(q-k-1)} \\ &= C|T|^2 \sum_{q=4}^{N_{T,T'}} q^{-\frac{1}{6}} \sum_{k=0}^{\lfloor q/4 \rfloor} \frac{4^k}{(\text{dist}(T, T'))^{\gamma k}} \frac{\sqrt{(q-2)!}}{(q-k-1)^2 \sqrt{k!(q-k-2)!}}, \end{aligned}$$

where we used the multinomial theorem. Therefore, we derive that

$$V_{4,2} \leq C|T|^2 \sum_{q=4}^{N_{T,T'}} q^{-2-\frac{1}{6}} \sum_{k=0}^{\lfloor q/4 \rfloor} \frac{4^k}{(\text{dist}(T, T'))^{\gamma k}} \sqrt{\binom{q}{k}}$$

Finally, as for  $V_{3,2}$  if we set  $\eta_{T,T'} := (4/(\text{dist}(T, T'))^\gamma)$  the function  $k \mapsto \eta_{T,T'}^k \sqrt{\binom{q}{k}}$  is decreasing if  $q \leq (\eta_{T,T'})^{-2}$ . Fix  $N_{T,T'} := \lfloor (\eta_{T,T'})^{-2} \rfloor$ , note that we have  $N_{T,T'} \rightarrow \infty$  as  $\text{dist}(T, T') \rightarrow \infty$ . It

follows that

$$V_{4,2} \leq C|T|^2 \sum_{q=4}^{N_{T,T'}} \frac{q}{q^{2+\frac{1}{6}}} \leq C|T|^2. \quad (40)$$

Gathering (39) and (40) shows that  $V_{4,\leq N_{T,T^4}} = O(|T|^2)$  and finally that  $V_4 = O(|T|^2)$ .

Gathering the control of four terms,  $V_0$  to  $V_4$ , leads to the announced result in Proposition 3.1.  $\square$