HAL open science

A test of Gaussianity based on the Euler characteristic of excursion sets

Elena Di Bernardino, Anne Estrade, José Rafael León

▶ To cite this version:

Elena Di Bernardino, Anne Estrade, José Rafael León. A test of Gaussianity based on the Euler characteristic of excursion sets. 2016. hal-01345225v1

HAL Id: hal-01345225

https://hal.science/hal-01345225v1

Preprint submitted on 13 Jul 2016 (v1), last revised 21 Dec 2016 (v2)

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.

A test of Gaussianity based on the Euler characteristic of excursion sets

Elena Di Bernardino

CNAM, Paris, EA4629, Département IMATH
292 rue Saint-Martin, Paris Cedex 03, France
e-mail: elena.di_bernardino@cnam.fr

Anne Estrade

MAP5 UMR CNRS 8145, Université Paris Descartes
45 rue des Saints-Peres, 75270 Paris Cedex 06, France
e-mail: anne.estrade@parisdescartes.fr

José R. León*

Escuela de Matemática, Facultad de Ciencias
Universidad Central de Venezuela, Av. Paseo de los Ilustres
Los Chaguaramos, Caracas 1040, Venezuela
e-mail: jose.leon@ciens.ucv.ve

Abstract: In the present paper, we deal with a stationary random field $X : \mathbb{R}^d \to \mathbb{R}$ and we assume it is partially observed through some level functionals. We aim at providing a methodology for a test of Gaussianity based on this information. More precisely, the level functionals are given by the Euler characteristic of the excursion sets above some levels. On the one hand, we study the properties of these level functionals under the hypothesis that the random field X is Gaussian. In particular, we focus on the mapping that associates to any u the expected Euler characteristic of the excursion set above level u. On the other hand, we study the same level functionals under alternative distributions of X, such as chi-square, harmonic oscillator and shot-noise. In order to validate our methodology, a part of the work consists in numerical experimentations. We generate Monte-Carlo samples of Gaussian and non-Gaussian random fields and compare, from a statistical point of view, their level functionals. Simulations are performed both in one dimensional case ($d = 1$) and in two dimensional case ($d = 2$), using R.

MSC 2010 subject classifications: Primary 62G10 ; secondary 60G10, 60G15, 60G60.
Keywords and phrases: Test of Gaussianity, Gaussian fields, Excursion sets, Euler characteristic, Crossings.

Introduction

The problem of determining whether an i.i.d random sample comes from a Gaussian distribution has been studied extensively. In the case where the mean and the variance of the random variable are known, one can use a classical goodness of fit test. However, if these parameters need to be estimated the test of Lillifors and a variant of the Cramer-Von Mises test, with estimated parameters, are well adapted. These tests are no more distribution-free and depend on the true value of the parameters. The p-values must be obtained by simulations.
The situation becomes more complicated when the sample comes from a stationary stochastic process satisfying some mixing conditions. For this type of problem, some tests have been designed to determine whether the one dimensional marginals are Gaussian. We can cite, by way of illustration: the Eps test

*Partially supported by the INRIA International Chairs program

[9] based on the empirical characteristic function; the test built by Lobato and Velasco [14] that uses a test of symmetry and kurtosis; the test of Subba and Gabr [20] where the bi-spectral density is the basis for the test. A remarkable exception is the test built in Cuesta-Albertos et al.[7]. There, the method for constructing the test uses a one dimensional random projection, after which the projected sample is subjected to a test that infers whether the one dimensional marginal is Gaussian. Actually, the random projection procedure allows one to test whether the whole distribution of the process is Gaussian and is not limited to marginal distributions.

In the present article, we deal with a real valued stationary random field -or a multivariate process- and we use the information given by level functionals of the process to build a test of Gaussianity. The level functionals are the Euler characteristic (EC) of the excursion sets above some levels. Our first motivation comes from the article [1] Section 7 (Model identification). In the aforementioned paper, Adler suggests to use the EC of the excursion sets as a way to determine what is the actual distribution of the observed process. His words can describe better than ours the main goal *"Suppose that we are given a real valued random field f on $\mathbb{R}^3$, which we observe on the unit cube $[0,1]^3$. Our problem is that we are not certain what the distribution of the field is. For simplicity, we shall assume that the choices are Gaussian and χ^2, where both are assumed to be stationary and isotropic. Then one way to choose between the two is to calculate, from the data, an empirical Euler characteristic curve, ..."* c.f. [1]. If the data is Gaussian, then the theoretical curve is a very precise one, depending on the second spectral moment of the process and some other invariant quantities. Otherwise, if it is a χ^2-process, or whatever any other model, a completely different curve appears. In what follows, we offer a methodology that tries to implement these ideas.

The idea of observing level functionals of a random field in order to infer some information on the distribution of the field is not new. In [13], Lindgren offers estimators for the second spectral moment of an univariate Gaussian process that are based on the number of upcrossings at various levels. In [6], Cabaña builds a test of isotropy, for a two dimensional random field, that is based on the perimeter and the area of excursion sets. In [21], the covariance function of a bivariate Gaussian field is inferred from the excursion sets Euler characteristic. The same idea has inspired many precursors working in materials science, see for instance [18], [15]. In those papers, the modelling of images or slices of a two-phases materials (even more complicated) is achieved by using a two dimensional stationary Gaussian field that has been thresholded at a certain level. The observed data are the lengths of the respective two phases along any line extracted from the image. The aim of these studies is to identify the Gaussian covariance function. Let us also mention [2] where the authors start from the observation of a neurological space-time signal at some moderate levels and deduce some parameters that help in estimating the probability of exceeding very high values. Not far from this thema, one can find the question of exceedances or the study of extreme values, when considering high levels. We will not go further in that direction and, at the opposite, stay with the observation of moderate levels. In all the mentioned papers, the field that is under study is assumed to be Gaussian. On the contrary, in the present paper, Gaussianity is not assumed but has to be inferred without knowing the spatial correlation.

To be more precise, we aim at proving that the function that associates u to the mean excursions EC at level u provides a signature of the distribution of the random field under study. Although a so complex information as the knowledge of the whole distribution cannot be summarize in a single function, our guess is that the shape of its graph could be enough to discriminate between Gaussianity and non Gaussianity. Our main tool will be a Central Limit Theorem for the EC of an excursion set of an isotropic Gaussian random field. This asymptotic normality takes place when the domain grows to $\mathbb{R}^d$. The result is proved in [10] with the help of a Breuer-Major theorem [16]. We will also need some generalisations, extensions and explicit computations of this result. In order to keep the methodological spirit of the present paper, we will give precise statements but we will refer to the companion paper

[8] for the proofs.

Outline of the paper. Section 1 contains the general setting and the definition of the observation tool, namely the Euler characteristic of excursion sets. In Section 2, we focus on the Gaussian hypothesis. Assuming that the field under study X is a stationnary Gaussian field, we give explicit formulas for the first two moments of the excursions EC (the first moment formula is well known, the second moment formula is new, see Proposition 1 and Proposition 2). We also recall the Central Limit Theorem satisfied by the excursions EC when the domain tends to $\mathbb{R}^d$ (see [10]) and give an extension for the joint excursions EC concerned with different levels and disjoint domains, see Proposition 3. It allows us to describe the situation in term of a statistical model with observations whose distribution is known under the null hypothesis **H0**. Section 3 is concerned with the study of three alternative distributions of the observed random field: χ^2, Kramer oscillator and shot noise. In all cases, we give an explicit formula for the mean EC of excursion sets. Section 4 is devoted to numerical illustrations for univariate processes. We generate trajectory samples of stationary processes, Gaussian and non Gaussian, and compare the theoretical mean function of the excursions EC and the empirical one. We also build some chi-square statistics in order to quantify the deviation between Gaussian and non Gaussian case and to illustrate the Central Limit Theorem presented above. At last, in Section 5, we go further in the numerical study by considering two dimensional random fields. We generate Gaussian and χ^2 samples and compare with the theoretical situation. All the random generations and numerical computations are performed with `R`.

1. Setting

All over the paper, we consider a real valued random field X defined on $\mathbb{R}^d$ that satisfies the following assumption.

Assumption (A): The random field X is stationary, isotropic, $\mathbb{E}(X(0)) = 0$, $\mathrm{Var}(X(0)) = 1$ and almost all realisations belong to $C^3(\mathbb{R}^d)$. Let $\nabla^2 X(t)$ stand for the $\frac{1}{2}d(d+1)$ random vector that contains the upper coefficients of the symmetric Hessian matrix $X''(t)$ and $\mathbf{X}(t)$ for the $d + \frac{1}{2}d(d+1) + 1$ random vector $(X'(t), \nabla^2 X(t), X(t))$. For any fixed t in $\mathbb{R}^d$,

$$\text{the covariance matrix of the random vector } \ (\mathbf{X}(0), \mathbf{X}(t)) \ \text{ has full rank.}$$

We denote by r the covariance function of X, $r(t) = \mathrm{Cov}(X(0), X(t))$, $t \in \mathbb{R}^d$, which belongs to $C^6(\mathbb{R}^d)$. The isotropy assumption implies that the Hessian matrix $r''(0)$ is equal to $-\lambda I_d$ for some $\lambda > 0$, usually named as *second spectral moment*. At last, r is such that,

$$\psi(t) \to 0 \text{ when } \|t\| \to +\infty \ \text{ and } \ \psi \in L^1(\mathbb{R}^d),$$

where $\psi(t) = \max\left(\left|\frac{\partial^{\mathbf{k}} r}{\partial t^{\mathbf{k}}}(t)\right|; \mathbf{k} = (k_1, \ldots, k_d) \in \mathbb{N}^d,\ k_1 + \ldots + k_d \leq 4\right)$.

Notations.
• for any $u \in \mathbb{R}$ and any compact $T \subset \mathbb{R}^d$, we call "excursion set of X above the level u within the domain T" the following set $\{t \in T : X(t) \geq u\}$,
• $p_Z(.)$ denotes the probability density function of any random vector Z (assuming it exists),
• $|.|$ denotes without any ambiguity, either the absolute value, or the d-dimensional Lebesgue measure.

Euler characteristic.
The Euler characteristic of a compact domain K in $\mathbb{R}^d$ can be heuristically defined in the case $d = 1$ as

the number of connected components of K, or in the case $d = 2$ as the number of connected components minus the number of holes in K. In the case where K is an excursion set $\{t \in T : X(t) \geq u\}$, with T a rectangle in $\mathbb{R}^d$ and u a real number, there exists a rather tractable formula that uses the theory of Morse functions, see [3] Chapter 9 for instance. It states that the Euler characteristic of $\{t \in T : X(t) \geq u\}$ is equal to a sum of two terms. The first one only depends on the restriction of X to the interior of T, it is given by the quantity $\varphi(X, T, u)$ defined in Equation (1) below. The second one exclusively depends on the behaviour of X on the l-dimensional faces of T, with $0 \leq l < d$. From now on, we focus on the term $\varphi(X, T, u)$, named as "modified Euler characteristic" in [10], and we still call it Euler characteristic (EC). It is defined by the following,

$$\begin{aligned} \varphi(X, T, u) &= \sum_{k=0}^{d} (-1)^k \mu_k(T, u), \quad \text{where} \\ \mu_k(T, u) &= \#\{t \in \mathring{T} : X(t) \geq u,\, X'(t) = 0, \text{ index } (X''(t)) = d - k\}, \end{aligned} \tag{1}$$

and the "index" stands for the number of negative eigenvalues.

Special case 1: dimension one. When $d = 1$, writing $[0, T]$ instead of T for a while, Equation (1) becomes

$$\begin{aligned} \varphi(X, [0, T], u) &= \#\{\text{local maxima of } X \text{ above } u \text{ in } (0, T)\} \\ &\quad -\#\{\text{local minima of } X \text{ above } u \text{ in } (0, T)\}. \end{aligned} \tag{2}$$

Morse's theorem says that this quantity is linked with the number of up-crossings,

$$U(X, [0, T], u) = \#\{t \in [0, T] : X(t) = u,\, X'(t) \geq 0\},$$

by the relation

$$\varphi(X, [0, T], u) + \mathbf{1}_{\{X(0) > u, X'(0) < 0\}} + \mathbf{1}_{\{X(T) > u, X'(T) > 0\}} = U(X, [0, T], u) + \mathbf{1}_{\{X(0) > u\}}.$$

Taking expectation in both expressions and using stationarity yield the next formula that we will use in Sections 3 and 4,

$$\mathbb{E}[\varphi(X, [0, T], u)] = \mathbb{E}[U(X, [0, T], u)]. \tag{3}$$

Special case 2: dimension two. When $d = 2$, Equation (1) can be rewritten in the following way. With the notations introduced within this equation, $\mu_0(T, u)$ denotes the number of local maxima above u, $\mu_2(T, u)$ denotes the number of local minima above u and $\mu_1(T, u)$ the number of local saddle points above u. Hence,

$$\begin{aligned} \varphi(X, T, u) &= \#\{\text{local extrema of } X \text{ above } u \text{ in } \mathring{T}\} \\ &\quad - \#\{\text{local saddle points of } X \text{ above } u \text{ in } \mathring{T}\}. \end{aligned} \tag{4}$$

2. Under Gaussian hypothesis

In this section, we assume that X is Gaussian and satisfies all the assumptions described in Section 1.

2.1. First two moments of the Euler characteristic of an excursion set

Let T be a cube in $\mathbb{R}^d$. This section is devoted to explicit formulas for the first two moments of $\varphi(X,T,u)$. They are based on the decomposition (1) and on Rice formulas for the factorial moments of $\mu_k(T,u)$ (see for instance [3] Chapter 11 or [4] Chapter 6).

In particular, using the stationarity of X, the expectation can be computed as follows

$$\begin{aligned}\mathbb{E}[\varphi(X,T,u)] &= (-1)^d \int_T \mathbb{E}[\mathbf{1}_{[u,\infty)}(X(t))\,\det(X''(t))\,|\,X'(t)=0]\,p_{X'(t)}(0)\,dt \\ &= (-1)^d\,|T|\,(2\pi\lambda)^{-d/2}\,\mathbb{E}[\mathbf{1}_{[u,\infty)}(X(0))\,\det(X''(0))]\,.\end{aligned}$$

Moreover, it is proved in [3] Lemma 11.7.1, through a regression and thanks to Wick's formula, that

$$\mathbb{E}[\mathbf{1}_{[u,\infty)}(X(0))\,\det(X''(0))] = (-1)^d\,(2\pi)^{-1/2}\,\lambda^d\,H_{d-1}(u)\,e^{-u^2/2}\,,$$

where H_k stands for the Hermite polynomial of order k. Hence, the next formula holds

$$\mathbb{E}[\varphi(X,T,u)] = |T|\,(2\pi)^{-(d+1)/2}\,\lambda^{d/2}\,H_{d-1}(u)\,e^{-u^2/2}\,. \tag{5}$$

In what follows, we will be particularly interested in the next function

$$C(u) = (2\pi)^{-(d+1)/2}\,\lambda^{d/2}\,H_{d-1}(u)\,e^{-u^2/2}\,, \tag{6}$$

that yields $\mathbb{E}[\varphi(X,T,u)] = |T|\,C(u)$. Equation (6) shows that $C(u)$ implicitly depends on X through its dimension parameter d and its second spectral moment λ. Whenever necessary in the next sections, we will emphasize this dependence by writing $C(u) = C(u,\lambda)$.

For the second moment, a so nice formula as (5) seems to be out of reach. Nevertheless, in the next proposition, we provide the second moment as an integral that can be numerically evaluated (see also [22] for another formula, which is valid under restrictive assumptions on X).
We will use the following functions, defined for $u \in \mathbb{R}$ and $t \in \mathbb{R}^d$,

$$\left.\begin{aligned}&g(u) = \mathbb{E}[\,\mathbf{1}_{[u,\infty)}(X(0))\,|\det(X''(0))|\,] \\ &D(t) = (2\pi)^{2d}\,\det(\lambda^2 I_d - r''(t)^2) \\ &G(u,t) = \mathbb{E}[\mathbf{1}_{[u,\infty)}(X(0))\,\mathbf{1}_{[u,\infty)}(X(t))\,\det(X''(0))\,\det(X''(t))\,|\,X'(0)=X'(t)=0]\end{aligned}\right\} \tag{7}$$

Proposition 1. *Assume that X is Gaussian and satisfies Assumption* **(A)**.
Then, for any $u \in \mathbb{R}$, the map $t \mapsto G(u,t)\,D(t)^{-1/2}$ is integrable on any compact set in $\mathbb{R}^d$ and

$$\mathbb{E}[\varphi(X,T,u)^2] = \int_{\mathbb{R}^d} |T \cap (T-t)|\,G(u,t)\,D(t)^{-1/2}\,dt + |T|\,(2\pi\lambda)^{-d/2}\,g(u)\,.$$

Proof. We give here a sketch of the proof. For more details, see [8].
Integrability comes from [10] Proposition 1.1 since $X \in C^3$. Equation (1) yields

$$\begin{aligned}\varphi(X,T,u)^2 &= \sum_{0\le k\le d} \mu_k(T,u) + \sum_{0\le k\le d} \mu_k(T,u)(\mu_k(T,u)-1) \\ &\quad + \sum_{0\le k\ne l\le d} (-1)^{k+l}\mu_k(T,u)\mu_l(T,u).\end{aligned}$$

The expectation of the first term is evaluated thanks to Rice formula,

$$\mathbb{E}[\sum_{0\le k\le d}\mu_k(T,u)] \quad = \quad |T|\, g(u)\, p_{X'(0)}(0)\,.$$

For the second and third terms, we use again a Rice formula, adapted from the second factorial moment formula, to get

$$\mathbb{E}[\varphi(X,T,u)^2] = \int_{\mathbb{R}^d} |T\cap(T-t)|\, G(u,t)\, p_{X'(0),X'(t)}(0,0)\, dt + |T|\, g(u)\, p_{X'(0)}(0).$$

It remains to compute the probability density function of $(X'(0), X'(t))$. The covariance matrix of this vector is equal to $\begin{pmatrix} \lambda I_d & -r''(t) \\ -r''(t) & \lambda I_d \end{pmatrix}$ and so $p_{X'(0),X'(t)}(0,0) = D(t)^{-1/2}$. □

2.2. Asymptotic variance

In the next proposition, we let the cube T grow to $\mathbb{R}^d$ and we give a closed formula for the asymptotic variance of $\varphi(X,T,u)$. Actually, we consider

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

the image of a fixed cube T by the dilation $t\mapsto Nt$ and we let N grow to $+\infty$.

Proposition 2. *Assume that X is Gaussian and satisfies Assumption* **(A)** *and let T be a cube in $\mathbb{R}^d$. Then for any u in $\mathbb{R}$,*

$$\lim_{N\to+\infty} \mathrm{Var}[|T^{(N)}|^{-1/2}\,\varphi(X,T^{(N)},u)] = V(u) \;<+\infty$$

with

$$V(u) = \int_{\mathbb{R}^d} (G(u,t)\, D(t)^{-1/2} - C(u)^2)\, dt + (2\pi\lambda)^{-d/2}\, g(u)\,, \tag{8}$$

where $C(u)$, $g(u)$, $G(u,t)$, $D(t)$ have been defined in (6) *and* (7).

Note that a more explicit formula for (8) can be performed by using Gaussian regressions for the computation of the expectation functions $G(u,t)$ and $g(u)$ (see [8] in the one dimensional case).

Proof. From Proposition 1, for a fixed cube T,

$$\begin{aligned} & \mathrm{Var}[|T|^{-1/2}\varphi(X,T,u)] \\ = & \int_{\mathbb{R}^d} \frac{|T\cap(T-t)|}{|T|}\, G(u,t)\, D(t)^{-1/2}\, dt + (2\pi\lambda)^{-d/2}\, g(u) - |T|^{-1}\,(\mathbb{E}\varphi(X,T,u))^2 \\ = & \int_{\mathbb{R}^d} \frac{|T\cap(T-t)|}{|T|}\, (G(u,t)\, D(t)^{-1/2} - C(u)^2)\, dt + (2\pi\lambda)^{-d/2}\, g(u)\,, \end{aligned}$$

where we have used the relation $|T|^2 = \int_{\mathbb{R}^d} |T\cap(T-t)|\, dt$ to get the last line. Hence, the asymptotic formula can easily be derived using Lebesgue dominated convergence Theorem conditionaly to the fact that $t\mapsto G(u,t)\; D(t)^{-1/2} - C(u)^2$ belongs to $L^1(\mathbb{R}^d)$. This point is proved in [8] Lemma 3. □

Beyond the existence of a finite asymptotic variance as stated in the previous proposition, $\varphi(X,T^{(N)},u)$

satifies a central limit theorem. Actually, Theorem 2.6 of [10] states that the next convergence holds in distribution

$$|T^{(N)}|^{-1/2}\left(\varphi(X,T^{(N)},u)-\mathbb{E}[\varphi(X,T^{(N)},u)]\right) \underset{N\to\infty}{\overset{distr}{\longrightarrow}} \mathcal{N}(0,V(u))\,, \tag{9}$$

where $\mathcal{N}(0,V(u))$ stands for the centered Gaussian distribution with variance $V(u)$.

2.3. Disjoint domains and various levels

We now consider two domains T_1 and T_2 that are disjoint and two levels u_1 and u_2 that can be equal or not.

Proposition 3. *Assume that X is Gaussian and satisfies Assumption* **(A)**.
(a) *Let T_1 and T_2 be two cubes in $\mathbb{R}^d$ such that $|T_1|=|T_2|$ and $dist(T_1,T_2)>0$ and let u_1 and u_2 belong to $\mathbb{R}$. For any integer $N>0$, we introduce*

$$Z_i^{(N)}=|T_i^{(N)}|^{-1/2}\left(\varphi(X,T_i^{(N)},u_i)-\mathbb{E}[\varphi(X,T_i^{(N)},u_i)]\right) \quad for \quad i=1,2.$$

As $N\to+\infty$, $\left(Z_1^{(N)},Z_2^{(N)}\right)$ converges in distribution to a centered Gaussian vector with diagonal covariance matrix $\begin{pmatrix} V(u_1) & 0\\ 0 & V(u_2)\end{pmatrix}$ where $V(u_i)$ is prescribed by (8).
(b) *Let T be a cube in $\mathbb{R}^d$ and let u_1 and u_2 belong to $\mathbb{R}$. For any integer $N>0$, we introduce*

$$\zeta_i^{(N)}=|T^{(N)}|^{-1/2}\left(\varphi(X,T^{(N)},u_i)-\mathbb{E}[\varphi(X,T^{(N)},u_i)]\right) \quad for \quad i=1,2.$$

As $N\to+\infty$, $\left(\zeta_1^{(N)},\zeta_2^{(N)}\right)$ converges in distribution to a centered Gaussian vector with covariance matrix $\begin{pmatrix} V(u_1) & V(u_1,u_2)\\ V(u_1,u_2) & V(u_2)\end{pmatrix}$.

Proof. We give here a sketch of proof of point **(a)**. For more details, see Proposition 4 in [8].
As the distance between $T_1^{(N)}$ and $T_2^{(N)}$ goes to infinity, the random variables $Z_1^{(N)}$ and $Z_2^{(N)}$ are asymptotically independent in such a way that $\mathrm{Cov}(Z_1^{(N)},Z_2^{(N)})$ goes to 0. Together with the convergence result in (9), it implies that the covariance matrix of the random vector $(Z_1^{(N)},Z_2^{(N)})$ tends to $\begin{pmatrix} V(u_1) & 0\\ 0 & V(u_2)\end{pmatrix}$. Hence, using the same arguments, it is not difficult to establish that any linear combination $xZ_1^{(N)}+yZ_2^{(N)}$ has a Gaussian limit in distribution.
Point **(b)** is proved in [10] Theorem 2.5. There, the covariance $V(u_k,u_l)$ is prescribed by a series. Let us mention that, although convergent, the series expansion is so awkward that it cannot be used in practise to evaluate $V(u_k,u_l)$. □

2.4. Statistical model

We are now able to built a test for the hypothesis **H0**: "the random field X is Gaussian".

Actually, we assume that X is observed through the family $(Y_k^i)_{1\le i\le m;1\le k\le p}$ where

$$Y_k^i=\frac{\varphi(X,T_i,u_k)}{|T_i|}\,,\ i=1,\dots,m\,;\ k=1,\dots,p\,,$$

for disjoint domains $T_1, \ldots, T_m$ that have the same large volume and are at large distance one from the others, and various levels $u_1 \leq \ldots \leq u_p$.

Thanks to Proposition 3 and Equation (5), one can write the following approximation valid under **H0**,

$$Y_k^i = C(u_k, \lambda) \ + \ \epsilon_k^i \, , \, i = 1, \ldots, m \, ; \, k = 1, \ldots, p \, , \tag{10}$$

where $(\epsilon_k^i)_{i,k}$ is a mp-dimensional centered Gaussian vector with covariance

$$\mathrm{Cov}(\epsilon_k^i, \epsilon_l^j) = \mathrm{Cov}(Y_k^i, Y_l^j) = \delta_{ij} \, V(u_k, u_l)/|T_i| \, . \tag{11}$$

The deterministic value $C(u_k, \lambda)$ is given by Equation (6). It depends on the level u_k and on the field X only through its second spectral moment λ. For $k = l$, $V(u_k, u_l) = V(u_k)$ is given by (8), and for $k \neq l$, since no explicit formula is available for $V(u_k, u_l)$ in practise, it has to be estimated.

Using the statistical model (10)-(11), we will focus on two particular cases.

(a) **Diagonal case**. We take m disjoint domains $T_1, \ldots, T_m$ and m levels $u_1, \ldots, u_m$, such that *each level u_k is associated to a single domain T_k*. In this setting we have m observations $Y_1^1, \ldots, Y_m^m$ and a diagonal covariance matrix equal to $diag(V(u_1)/|T_1|, \ldots, V(u_m)/|T_m|)$.

(b) **Crossed case**. We take m disjoint domains $T_1, \ldots, T_m$ and p levels $u_1, \ldots, u_p$, such that *different levels u_k are associated to the same domain T_i*. In this setting we have mp observations $(Y_k^i)_{i,k}$ and their covariance matrix is given by (11).

3. Under alternative hypothesis

3.1. χ^2 hypothesis

In this section, we deal with χ^2 distributions instead of Gaussian ones. Let us fix s, a non negative integer, as the degrees of freedom.

We start with $\{X_i(.)\}_{i=1}^s$, an independent sample of centered stationary Gaussian fields on $\mathbb{R}^d$ that satisfy Assumption **(A)** of Section 1. We denote by r_X their covariance function and recall that $r_X(0) = 1$. Consider now the following stationary fields

$$\chi_s^2(\cdot) = \sum_{i=1}^s (X_i(\cdot))^2 \ \text{ and } \ Z^{(s)}(.) = \frac{1}{\sqrt{2s}} (\sum_{i=1}^s (X_i(\cdot))^2 - s) \, .$$

Note that for any $t \in \mathbb{R}^d$, $\chi_s^2(t)$ is a Chi-square random variable with s degrees of freedom. One get readily that

$$\mathbb{E}[Z^{(s)}(0)] = 0, \, \mathrm{Var}[Z^{(s)}(0)] = 1, \, \mathbb{E}[Z^{(s)}(t) Z^{(s)}(0)] = r_X(t)^2.$$

Therefore, $Z^{(s)}$ also satisfies Assumption **(A)**. Moreover, its second spectral moment is equal to λ with $\lambda = -((r_X)^2)''(0) = -2r_X''(0)$.

We are now interested in the expectation of the Euler characteristic of

$$\{t \in T \, : \, Z^{(s)}(t) \geq u\} = \{t \in T \, : \, \chi_s^2(t) \geq s + u\sqrt{2s}\} \, ,$$

for a fixed cube $T \subset \mathbb{R}^d$ and a fixed level u in $\mathbb{R}$. A formula for the mean EC of excursion sets of χ^2 fields is given in [23], Theorem 3.5 (see also Theorem 15.10.1 in [3]). We can then establish that

$$\begin{aligned} &\mathbb{E}[\varphi(Z^{(s)}, T, u)] \\ &= |T| \left(\frac{\lambda}{\pi}\right)^{d/2} \frac{e^{-(s+u\sqrt{2s})/2} \, (s + u\sqrt{2s})^{(s-d)/2}}{2^{(s-2+2d)/2} \Gamma(s/2)} \, P_{d,s}(s + u\sqrt{2s}) \, \mathbf{1}_{[0,\infty)}(s + u\sqrt{2s}) \, , \end{aligned} \tag{12}$$

where $P_{d,s}(.)$ is a polynomial of degree $d-1$ with integer coefficients. In particular, $P_{1,s}(u)=1$ and $P_{2,s}(u)=u-s+1$. Let us quote that we have to handle carefully with the second spectral moment of the X_i's , which is equal to $\lambda/2$ in our context.

Let us recall that, in dimension one, the mean Euler characteristic of the excursion above the level u is equal to the mean number of upcrossings at level u (see Equation (3)). With this point of view, Formula (12) can also be found in [19] for instance. Keeping with dimension one, it is proved in [8] (see Proposition 6) that there exists a finite asymptotic variance $v_s(u)$ such that

$$\lim_{N\to+\infty} \mathrm{Var}\left(|T^{(N)}|^{-1/2}\,\varphi(Z^{(s)},T^{(N)},u)\right) = v_s(u)\,. \tag{13}$$

3.2. Kramer oscillator hypothesis

We work in dimension $d=1$. Let us consider the following system of stochastic differential equations, well known as Kramer oscillator system,

$$\left.\begin{aligned} dQ(t) &= P(t)dt,\\ dP(t) &= \sigma dW(t) - (cP(t)+V'(Q(t))dt,\end{aligned}\right\} \tag{14}$$

where $V(q)=a_0q^4-a_1q^2$ for positive constants a_0 and a_1, σ and c are also positive constants, and W is a Brownian motion. The asymptotic properties of such a system have been studied for instance in [25]. Besides, it is well known that the Markov process (Q,P) has an invariant measure μ that can be written (up to a numerical constant factor) as

$$d\mu(p,q) = \exp\left(-\frac{2c}{\sigma^2}(\frac{p^2}{2}+V(q))\right)\,dpdq.$$

From now on, we assume that $(Q(0),P(0))$'s distribution is proportional to μ, so that (Q,P) is stationary. Here, we are interested in the process Q. It is stationary, centered, but certainly not Gaussian since its distribution is proportional to $\exp\big(-\frac{2c}{\sigma^2}V(q)\big)\,dq$. Nevertheless, the distribution of its derivative process, $Q'=P$, is actually Gaussian with zero mean and variance equal to $\frac{\sigma^2}{2c}$. An application of Rice formula gives the mean number of upcrossings of Q in $[0,T]$ at any level u and hence, using (3), we get

$$\begin{aligned} \mathbb{E}[\varphi(Q,[0,T],u)] &= \mathbb{E}[U(Q,[0,T],u)]\\ &= T\left(\frac{1}{\sqrt{2\pi(\sigma^2/2c)}}\int_0^\infty p\exp-(\frac{c}{\sigma^2}p^2)dp\right)p_Q(u)\\ &= \frac{T\,\sigma}{2\sqrt{\pi c}}\,p_Q(u) \quad\text{with}\quad p_Q(u)=\frac{\exp(-\frac{2c}{\sigma^2}V(u))}{\int_{\mathbb{R}}\exp(-\frac{2c}{\sigma^2}V(q))dq} \end{aligned} \tag{15}$$

Moreover, let us remark that a suitable choice of the parameters σ,c,a_0,a_1 allows us to get $\mathrm{Var}(Q(0))=1$ and $\mathrm{Var}(Q'(0))=\lambda$, so that Q satisfies the same moments constraints as the generic process X in Section 1. Actually, it is sufficient to prescribe $\frac{\sigma^2}{2c}=\lambda$ and to solve the following non linear equation

$$\int_{\mathbb{R}}\exp\big(-\frac{2c}{\sigma^2}(a_0q^4-a_1q^2)\big)\,dq = \int_{\mathbb{R}}q^2\exp\big(-(\frac{2c}{\sigma^2}(a_0q^4-a_1q^2)\big)\,dq. \tag{16}$$

3.3. Shot noise hypothesis

Let us consider a last alternative hypothesis. We work in dimension $d = 1$ and we introduce the following shot noise process S,

$$S(t) = \left(\sum_{\xi \in \Phi} \mathbf{1}_{[0,a]}(t - \xi) \right) - \lambda a \ , \ t \in \mathbb{R} \, , \tag{17}$$

where $a > 0$ and Φ is a homogeneous Poisson point process on $\mathbb{R}$ with intensity $\lambda > 0$. The process S is clearly stationary with zero mean and variance equal to λa.

Moreover, its values almost surely belong to the discrete set $\{k - \lambda a, \, k \in \mathbb{N}\}$. If a whole trajectory of S could be observed, its discrete shape would clearly indicate the right choice between Gaussian and Poissonian hypothesis. If only excursion sets of S above a few levels are observed, the choice remains as clear when considering the mean number of "upcrossings". Note that the notion of upcrossings has to be properly defined in this context, since S is not continuous. Actually, it is proved in [5] (see also [11] for a different approach) that for any level $u \in \mathbb{R} \setminus \{k - \lambda a, \, k \in \mathbb{N}\}$ and any interval T,

$$\mathbb{E}[U(S, T, u)] = |T| \, \lambda e^{-\lambda a} \sum_{k \geq 0} \frac{(\lambda a)^k}{k!} \mathbf{1}_{\{k < u + \lambda a < k+1\}} \, . \tag{18}$$

This formula, which obviously differs from (5), is an indicator that enables to discrimate between a Gaussian or a Poissonian model. Once again, let us emphasize that the EC of excursion sets really contains information about the process under study. See Section 4.6 for a numerical illustration.

4. Univariate numerical illustrations

In this section, we focus on the one dimensional case and hence only deal with univariate processes.

4.1. First and second moments: $\mathbb{E}[\varphi(X, T, u)]$ and $V(u)$

We rewrite formulas (5) and (8) of Section 2 in the case $d = 1$ under **H0** hypothesis, i.e. when X is a stationary centered Gaussian process with variance 1, covariance function r and second spectral moment $\lambda = -r''(0)$:

$$\mathbb{E}[\varphi(X, T, u)] = |T| \, C(u, \lambda) \ \text{ with } \ C(u, \lambda) = (2\pi)^{-1} \, \lambda^{1/2} \, e^{-u^2/2} \, , \tag{19}$$

and $\mathrm{Var}[\varphi(X, T, u)] \sim |T| \, V(u)$ as $|T| \to +\infty$, with

$$V(u) = \int_{\mathbb{R}} (G(u, t) \, D(t)^{-1/2} - C(u, \lambda)^2) \, dt + (2\pi\lambda)^{-1/2} \, g(u) \, , \tag{20}$$

with $D(t) = (2\pi)^2 \, (\lambda^2 - r''(t)^2)$ and $g(u), G(u, t)$ are given by (7). In [8] Proposition 8, we give explicit formulas for these functions in the case $d = 1$. They are used in the following to give a numerical evaluation of $V(u)$ for various values of u.

In order to illustrate (19) and (20), we generate a 300 Monte-Carlo sample of a stationary centered Gaussian process with covariance function $r(t) = e^{-t^2}$. Note that it implies the second spectral moment $\lambda = 2$. In order to evaluate $\varphi(X, T, u)$ on each realization of X, we use Equation (2) for various values of u. Comparison between theoretical formulas and empirical counterparts are shown in Figure 1.

Furthermore, as shown in Figure 1, Equation (19) can be used to understand the behaviour of the rate of exceedances above high thresholds u. This aspect can be very useful in many applications as described in the seminal work of Rice [17] or in [24].

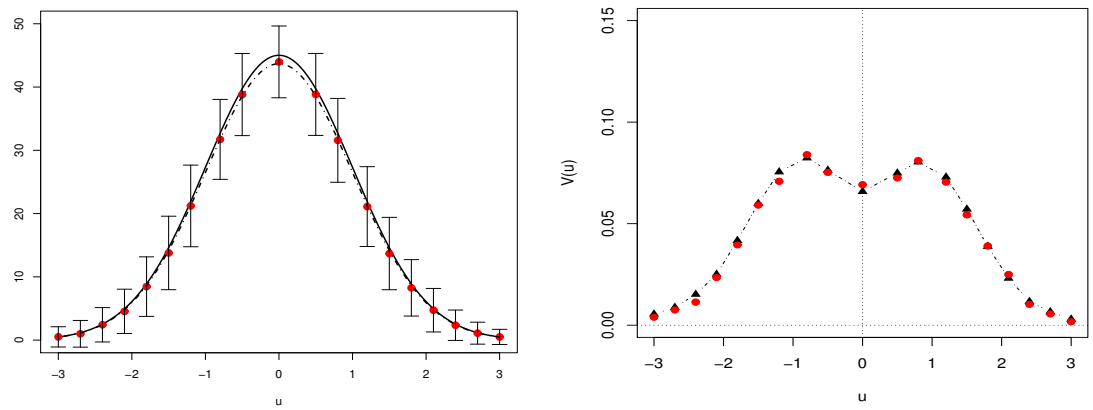

FIG 1. ***Left:*** *Theoretical* $u \mapsto \mathbb{E}[\varphi(X,T,u)]$ *from Equation* (19) *for* $|T| = 200$ *(full line). We also display, for different levels* u*, the empirical counterpart* $\widehat{\mathbb{E}}[\varphi(X,T,u)]$ *(red dots) based on* 300 *Monte Carlo simulations. The map* $u \mapsto |T|\, C(u,\widehat{\lambda})$ *from Equation* (19)*, by using the estimated spectral moment* $\widehat{\lambda}$ *as in Section 4.2, is represented in dashed line.* ***Right:*** *Theoretical* $u \mapsto V(u)$ *from Equation* (20) *for various values of* u *(black triangles) and empirical variance of* $|T|^{-1/2}\,\varphi(X,T,u)$ *(red dots) based on 300 Monte Carlo simulations with* $|T| = 200$ *and the same values of* u*.* $X(\cdot)$ *is a Gaussian univariate process* ($d = 1$) *with* $\mathbb{E}(X(0)) = 0$, $\mathrm{Var}(X(0)) = 1$ *and covariance function* $r(t) = e^{-t^2}$*. In this case* $\lambda = 2$*.*

4.2. Estimation of the second spectral moment

It has been already quoted that the second spectral moment of X, denoted by λ, plays an important role in Equation (19). Actually, all the influence of X in $\mathbb{E}[\varphi(X,T,u)]$ is summarised through this parameter. From our statistical point of view, since we aim at inferring the distribution of X from observations of the excursions of X, parameter λ is unknown *a priori*. It has to be estimated from the observed level functionals X.

We estimate $\gamma = \lambda^{1/2}$ by using an unbiased estimator introduced by Lingren [13] for stationary, zero-mean, Gaussian processes. It is based on p different levels $u_1 < u_2 < \ldots < u_p$ by the following prescription,

$$\widehat{\gamma} = \sum_{k=1}^{p} c_k \widehat{\gamma}_{u_k} \;\text{ with }\; \widehat{\gamma}_{u_k} = 2\,\pi\, T^{-1}\, e^{u_k^2/2}\, \varphi(X,[0,T],u_k)\,. \tag{21}$$

Actually, in the paper of Lingren, the number of up-crossings of level u_k in the interval $[0,T]$, namely $U(X,[0,T],u_k)$, is used instead of $\varphi(X,[0,T],u_k)$ (recall the analogy described in Section 1). As a general rule, considering

$$p = 3\,,\; (u_1,u_2,u_3) = (-u,0,u) \;\text{ with }\; u = \frac{2}{3}\sqrt{Var(X(t))}\,,\; c_1 = c_2 = c_3 = \frac{1}{3}$$

seems to be an acceptable choice (see discussion in [13]). In this case $\widehat{\lambda}^{1/2} = \widehat{\gamma} = \frac{1}{3}\,(\widehat{\gamma}_{-u} + \widehat{\gamma}_0 + \widehat{\gamma}_{-u})$. An illustration is given in Figure 2. This estimation of $\lambda^{1/2}$ is also used in Figure 1 (left), dashed line.

4.3. Chi-square statistics

In the following we will consider particular sub-cases of the two cases presented in Section 2.4, diagonal case (a) and crossed case (b). In particular, we focus on the four next models.

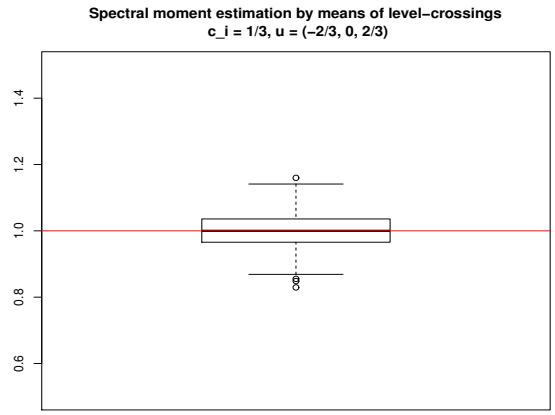

FIG 2. *Boxplot, based on 300 Monte Carlo simulations, of the ratio between theoretical value* $\lambda^{1/2}$ *and estimated value* $\widehat{\gamma}$ *given by* (21). $X(\cdot)$ *is a Gaussian univariate process* ($d = 1$) *with* $\mathbb{E}(X(0)) = 0$, $\mathrm{Var}(X(0)) = 1$ *and covariance function* $r(t) = e^{-t^2}$. *In this case* $\lambda = 2$. *We choose* $p = 3$, $c_i = 1/3$ *and* $u = (-2/3, 0, 2/3)$.

(a) **Diagonal case**. Each level u_k is associated to a single domain T_k. We have m observations $Y_1^1, \ldots, Y_m^m$ in the statistical model (10).
   - **(a.1)** Case with same level u, i.e. $u_1 = \ldots = u_m = u$, and m disjoint domains $T_1, \ldots, T_m$.
   - **(a.2)** Case with different levels $u_1 < \ldots < u_m$ and m disjoint domains $T_1, \ldots, T_m$.

(b) **Crossed case**. Different levels u_k can be associated to the same domain T_i. We have mp observations $(Y_k^i)_{1 \le i \le m, 1 \le k \le p}$ in (10).
   - **(b.1)** Case with one single domain T ($m = 1$) and p different levels $u_1 < \ldots < u_p$.
   - **(b.2)** Case with m disjoint domains $T_1, \ldots, T_m$ and p different levels $u_1 < \ldots < u_p$.

We will now associate Chi-square statistics with each of the four aforementioned models. For the **diagonal case (a.1)**, we consider

$$F_{a1} := \sum_{i=1}^{m} \left(\frac{\varphi(X, T_i, u) - \mathbb{E}[\varphi(X, T_i, u)]}{\sqrt{|T_i|\, V(u)}} \right)^2 \quad \widetilde{F}_{a1} := \sum_{i=1}^{m} \left(\frac{\varphi(X, T_i, u) - \widehat{\mathbb{E}}[\varphi(X, T_i, u)]}{\sqrt{\widehat{Var}(\varphi(X, T_i, u))}} \right)^2$$

where $\mathbb{E}[\varphi(X, T_i, u)] = |T_i|\, C(u, \lambda)$ is given by (19) and $V(u)$ is given by (20). Furthermore, $\widehat{\mathbb{E}}$ and $\widehat{Var}$ in $\widetilde{F}_{a1}$ are respectively the empirical mean and the empirical variance on considered Monte-Carlo sample generations.

Under **H0** hypothesis, a consequence of (10)-(11) is that both random variables F_{a1} and $\widetilde{F}_{a1}$ are approximately χ^2_m distributed, i.e. central Chi-square with m degrees of freedom. We evaluate F_{a1} and $\widetilde{F}_{a1}$ on 300 Monte Carlo simulations. We choose $m = 3$ and $u = 1.2$. The QQplot comparison between the obtained empirical quantiles with the theoretical quantiles of a centered χ^2_m distribution is gathered in Figure 3 (first and second panels).

Consider now the **diagonal case (a.2)**. Let

$$F_{a2} := \sum_{i=1}^{m} \left(\frac{\varphi(X, T_i, u_i) - \mathbb{E}[\varphi(X, T_i, u_i)]}{\sqrt{|T_i| V(u_i)}} \right)^2 \quad \widetilde{F}_{a2} := \sum_{i=1}^{m} \left(\frac{\varphi(X, T_i, u_i) - \widehat{\mathbb{E}}[\varphi(X, T_i, u_i)]}{\sqrt{\widehat{Var}(\varphi(X, T_i, u_i))}} \right)^2$$

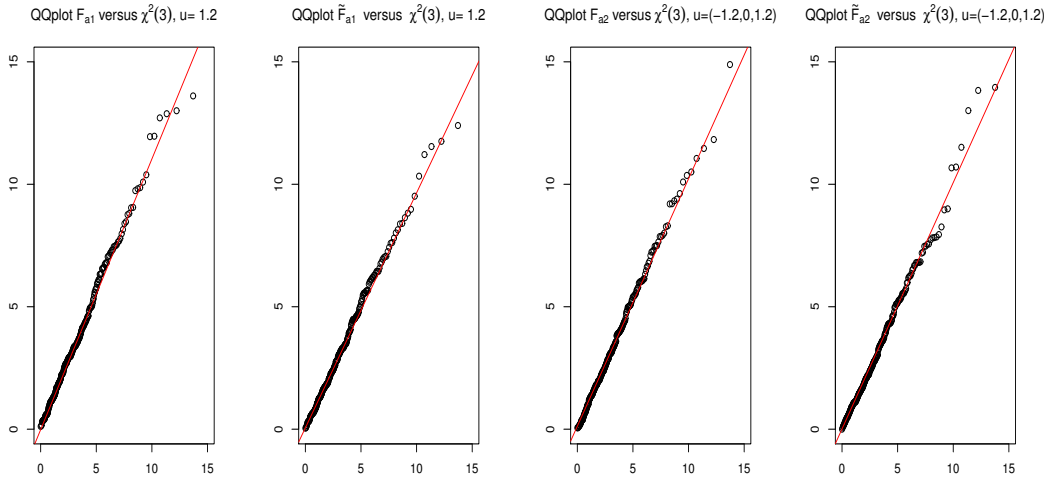

FIG 3. ***First and second panels***. *QQplot based on 300 Monte Carlo simulations: quantiles of F_{a1}, $\widetilde{F}_{a1}$ versus quantiles of the χ^2_m distribution for $m = 3$ disjoint domains T_1, T_2, T_3 with $|T_i| = 200$ and a single level $u = 1.2$.* ***Third and fourth panels***. *QQplot based on 300 Monte Carlo simulations: quantiles of F_{a2}, $\widetilde{F}_{a2}$ versus quantiles of the χ^2_m distribution for $m = 3$ disjoint domains T_1, T_2, T_3 with $|T_i| = 200$ and different levels $u_1 = -1.2, u_2 = 0, u_3 = 1.2$. $X(\cdot)$ is a Gaussian univariate process ($d = 1$) with $\mathbb{E}(X(0)) = 0$, $\mathrm{Var}(X(0)) = 1$ and covariance function $r(t) = e^{-t^2}$ with $\lambda = 2$.*

where $\widehat{\mathbb{E}}$ and $\widehat{Var}$ are respectively the empirical mean and the empirical variance on considered Monte Carlo sample generations. By using again the statistical model in (10)-(11), both F_{a2} and $\widetilde{F}_{a2}$ are approximately χ^2_m distributed. An illustration is presented in Figure 3 (third and fourth panels), by choosing 300 Monte Carlo simulations, $m = 3$, $|T_i| = 200$ and different levels $u_1 = -1.2, u_2 = 0, u_3 = 1.2$.

For the **crossed case (b.1)**, let us now consider different levels $u_1 < \ldots < u_p$ and one single domain T, i.e. $m = 1$. Let us define

$$F_{b1} := ||\Lambda^{-1}(\mathbf{Z} - \mathbb{E}[\mathbf{Z}])||^2, \quad \widetilde{F}_{b1} := ||\widehat{\Lambda}^{-1}(\mathbf{Z} - \widehat{\mathbb{E}}[\mathbf{Z}])||^2,$$

where $\mathbf{Z}$ is the p-dimensional Gaussian vector given by

$$\mathbf{Z} = \left(\frac{\varphi(X, T, u_k)}{|T|^{1/2}} \right)_{1 \le k \le p},$$

$\widehat{\mathbb{E}}$ is the empirical mean on the Monte-Carlo simulations and the matrices Λ and $\widehat{\Lambda}$ are defined below. Let $\Gamma = (V(u_k, u_l))_{1 \le kl, \le p}$ be the (theoretical) covariance matrix of $\mathbf{Z}$ and let Λ stand for any square root of Γ. Similarly, let $\widehat{\Gamma}$ be the empirical covariance matrix of $\mathbf{Z}$ evaluated on the Monte-Carlo sample generations, and let $\widehat{\Lambda}$ be any of its square root matrix. Hence, F_{b1} and $\widetilde{F}_{b1}$ are both approximately χ^2_p distributed. An illustration of the behaviour of $\widetilde{F}_{b1}$ is presented in Figure 4 (first panel), by choosing 300 Monte-Carlo simulations, $m = 1$, $|T| = 200$ and $p = 3$ different levels $u_1 = -1.5, u_2 = 0, u_3 = 1.5$.

Consider now the **crossed case (b.2)**. In this setting we have m disjoint domains $T_1, \ldots, T_m$ and p different levels $u_1 < \ldots < u_p$. Let

$$F_{b2} := \sum_{i=1}^{m} ||\Lambda^{-1}(\mathbf{Z}^i - \mathbb{E}[\mathbf{Z}^i])||^2, \quad \widetilde{F}_{b2} := \sum_{i=1}^{m} ||\widehat{\Lambda}_{(i)}{}^{-1}(\mathbf{Z}^i - \widehat{\mathbb{E}}[\mathbf{Z}^i])||^2,$$

where for any $i \in \{1, \dots, m\}$, $\mathbf{Z}^i$ is the p-dimensional Gaussian vector given by

$$\mathbf{Z}^i = \left(\frac{\varphi(X, T_i, u_k)}{|T_i|^{1/2}} \right)_{1 \le k \le p}$$

and $\widehat{\mathbb{E}}$ is the empirical mean on the Monte-Carlo simulations. Let $\Gamma = (V(u_k, u_l))_{1 \le kl, \le p}$ be the (theoretical) covariance matrix of $\mathbf{Z}^i$. Let Λ stand for any square root of Γ. Similarly, let $\widehat{\Gamma}_{(i)}$ be the empirical covariance matrix of $\mathbf{Z}^i$ evaluated on the Monte-Carlo sample generations, and let $\widehat{\Lambda}_{(i)}$ be any of its square root matrix. Hence, $||\Lambda^{-1}(\mathbf{Z}^i - \mathbb{E}[\mathbf{Z}^i])||^2$ and $||\widehat{\Lambda}_{(i)}{}^{-1}(\mathbf{Z}^i - \widehat{\mathbb{E}}[\mathbf{Z}^i])||^2$ are both approximately χ^2_p distributed.
Moreover, since for $1 \le i \ne j \le m$, the Gaussian vectors $\mathbf{Z}^i$ and $\mathbf{Z}^j$ are independent, F_{b2} and $\widetilde{F}_{b2}$ are still centered χ^2 distributed with now mp degrees of freedom.

An illustration of the behaviour of $\widetilde{F}_{b2}$ is presented in Figure 4 (second panel), by choosing 300 Monte Carlo simulations, $m = 3$ disjoint domains T_1, T_2, T_3 with $|T_i| = 200$ and $p = 3$ different levels $u_1 = -1.5, u_2 = 0, u_3 = 1.5$.

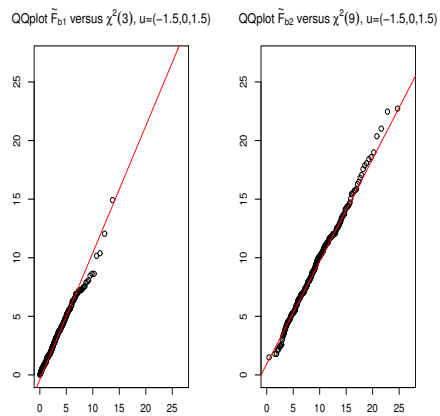

FIG 4. ***First panel***. *QQplot based on 300 Monte Carlo simulations: quantiles of $\widetilde{F}_{b1}$ versus quantiles of the χ^2_{mp} distribution for $m = 1$, with $|T| = 200$ and $p = 3$ different levels $u_1 = -1.5, u_2 = 0, u_3 = 1.5$.* ***Second panel***. *QQplot based on 300 Monte Carlo simulations: quantiles of $\widetilde{F}_{b2}$ versus quantiles of the χ^2_{mp} distribution for $m = 3$ disjoint domains T_1, T_2, T_3 with $|T_i| = 200$ and $p = 3$ different levels $u_1 = -1.5, u_2 = 0, u_3 = 1.5$. $X(\cdot)$ is a Gaussian univariate process ($d = 1$) with $\mathbb{E}(X(0)) = 0$, $\mathrm{Var}(X(0)) = 1$ and covariance function $r(t) = e^{-t^2}$ with $\lambda = 2$.*

4.4. First alternative: χ^2 process

In conformity with Section 3.1, we consider

$$Z^{(s)}(\cdot) = \frac{1}{\sqrt{2s}}(\chi^2_s(\cdot)) - s) \quad \text{with} \quad \chi^2_s(\cdot) = \sum_{1 \le i \le s} X_i(\cdot)^2 \,,$$

where the X_i's are independent copies of a centered stationary Gaussian process with covariance function $r(t) = e^{-t^2/2}$. Hence, we get $\lambda = -2r''(0) = 2$ and then the obtained process $Z^{(s)}$ has the same variance and the same second spectral moment λ as the previous Gaussian one presented in Section 4.1.

In the first and third panels of Figure 5, we display the boxplot for the ratio between the empirical mean of 300 Monte Carlo values of $\varphi(Z^{(s)}, T, u)$ and the theoretical mean given by (12) in the case $d = 1$,

$$\mathbb{E}[\varphi(Z^{(s)}, T, u)] = |T| \left(\frac{\lambda}{\pi}\right)^{1/2} \frac{(s + u\sqrt{2s})^{(s-1)/2}}{2^{s/2}\Gamma(s/2)} e^{-(s+u\sqrt{2s})/2} \mathbf{1}_{[0,\infty)}(s + u\sqrt{2s}). \tag{22}$$

This expectation has to be compared with the expectation given by (19) when X is a stationary centered Gaussian process with variance equal to 1 and second spectral moment $\lambda = 2$. Therefore, in the second and fourth panels of Figure 5, we display the boxplot for the ratio between the empirical 300 Monte Carlo mean value of $\varphi(Z^{(s)}, T, u)$ and the Gaussian theoretical expectation given in (19).

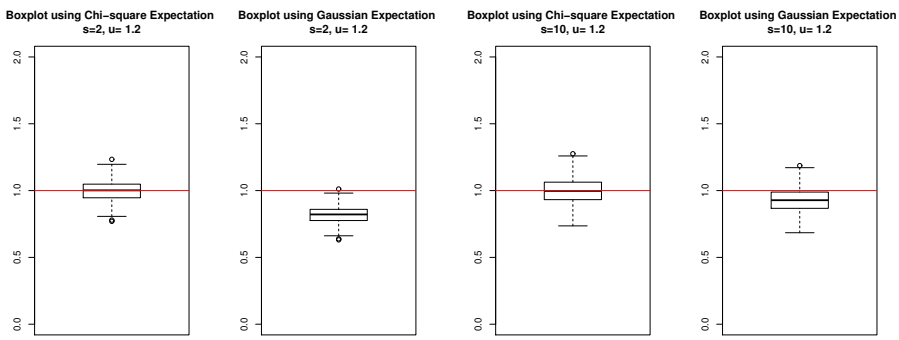

FIG 5. **First and third panels**: *Boxplot for the ratio between the empirical 300 Monte Carlo values of* $\varphi(Z^{(s)}, T, u)$ *and the theoretical mean given by Equation* (22). **Second and fourth panels**: *Boxplot for the ratio between the same empirical values and the theoretical mean in the Gaussian case given by Equation* (19). *In both cases* $\lambda = 2$. *The degrees of freedom* s *is chosen equal to 2 and 10 respectively.*

Remark. When the degrees of freedom s tend to infinity, the CLT implies that $Z^{(s)}$ tends in distribution to a stationary centered Gaussian process with covariance function equal to $t \mapsto r(t)^2$, which implies a variance equal to 1 and a second spectral moment equal to 2 (exactly as the Gaussian process X in Section 4.1). On the other hand, using Stirling approximation for the Γ function, one can prove that the right-hand side of (22) tends to the right-hand side of (19), i.e.,

$$\mathbb{E}[\varphi(Z^{(s)}, T, u)] \to |T|\, \lambda^{1/2}\, (2\pi)^{-1}\, e^{-u^2/2} \text{ when } s \to \infty.$$

See also formula (3.4) in [19] for the same remark. The comparison between the second and fourth panels in Figure 5, as well as Figure 6, illustrate this convergence.

Similarly to Section 4.3, we now consider chi-square statistics. We focus on the diagonal case **(a.1)** with a unique level u and m disjoint domains $T_1, \ldots, T_m$, and we introduce

$$F := \sum_{i=1}^{m} \left(\frac{\varphi(Z^{(s)}, T_i, u) - \mathbb{E}[\varphi(Z^{(s)}, T_i, u)]}{|T_i|^{1/2}} \right)^2 \quad \widetilde{F} := \sum_{i=1}^{m} \left(\frac{\varphi(Z^{(s)}, T_i, u) - \widehat{\mathbb{E}}[\varphi(Z^{(s)}, T_i, u)]}{|T_i|^{1/2}} \right)^2 \tag{23}$$

where $\mathbb{E}[\varphi(Z^{(s)}, T_i, u)]$ is given by (22) and $\widehat{\mathbb{E}}[\varphi(Z^{(s)}, T_i, u)]$ is the empirical mean based on the Monte-Carlo sample generations. Unlike the situation described in Section 4.3, hypothesis **H0** is not assumed

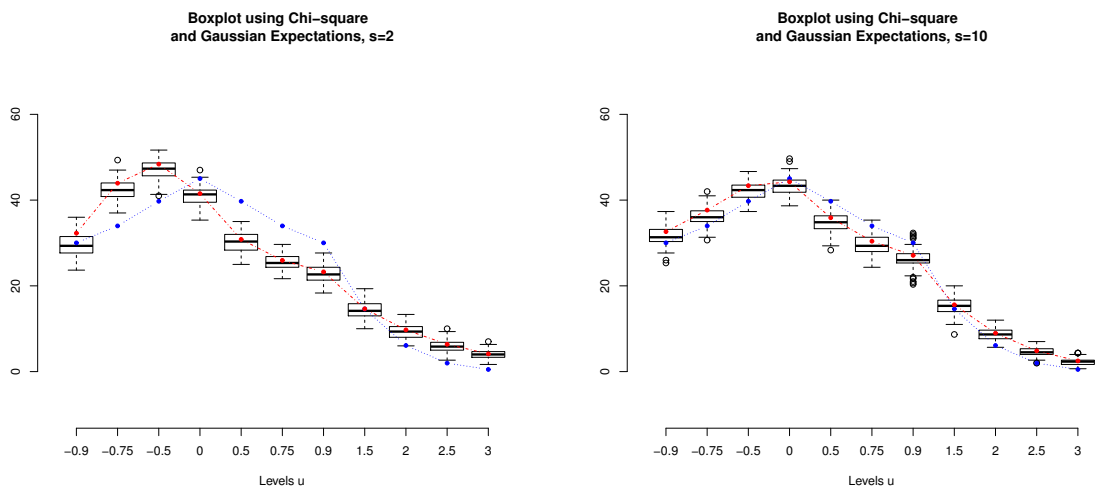

FIG 6. *Boxplots of the empirical 300 Monte Carlo values of* $\varphi(Z^{(s)}, T, u)$ *for the considered chi-square process* $Z^{(s)}$ *and different values of* u*. Red points represent the theoretical means given by Equation* (22) *for the same values of* u*; blue ones are the Gaussian means given by Equation* (19)*. In this case* $\lambda = 2$*,* $|T| = 200$*,* $s = 2$ *(left panel) and* $s = 10$ *(right panel).*

to be valid and the distributions of F and $\widetilde{F}$ are not known to be χ^2_m. However, in Figure 7, we display the quantiles of F (for $s = 2$ in the first panel, and for $s = 10$ in the fourth panel) and of $\widetilde{F}$ (Panel 3 for $s = 2$ and Panel 6 for $s = 10$) versus the quantiles of χ^2_m. The pretty good alignment of the QQplot suggests that a Central Limit Theorem should be valid as in the Gaussian case under **H0**. Note that the finiteness of the asymptotic variance in (13) is a first step in that direction. Remark also that the QQplot lines do not coincide with the bisectors of the orthant, conformly to (13).

In Figure 7 -second and fifth panels-, we also show the QQplot of the statistics F, where the theoretical mean of $\varphi(Z^{(s)}, T_i, u)$ has been replaced by the theoretical mean of $\varphi(X, T_i, u)$ with X a Gaussian process as in Section 4.1, versus the quantiles of χ^2_m. Since the considered χ^2 process $Z^{(s)}$ is not Gaussian, a deviation can be observed.

4.5. Second alternative: Kramer oscillator process

In this section we generate a 300 Monte-Carlo sample of a Kramer oscillator process as defined in Section 3.2. In order to obtain a process Q with zero mean, unit variance and second spectral moment equal to 2, we solve Equation (16) and choose

$$\sigma = 2 \ , \ c = 1 \ , \ a_0 = 1 \ , \ a_1 = 2.3373 \ , \ C = 4.886.$$

The generation procedure is the following. We define a discretized schema to simulate the solution $(Q(t), P(t))$ of the stochastic differential system in (14). Actually, we use the Metropolization of the Euler-Verlet schema with a sufficient small discretization step. The interested reader is referred to Algorithm 2.11 (Generalized Hybrid Monte-Carlo) in [12].

The comparison between the (theoretical) expectation given by (15) and the empirical one is shown in Figure 8 below.

As we did in previous Sections 4.3 and 4.4, we now consider some chi-square statistics. Precisely, we introduce F and $\widetilde{F}$, defined in a similar way as in (23), but for the harmonic oscillator process Q

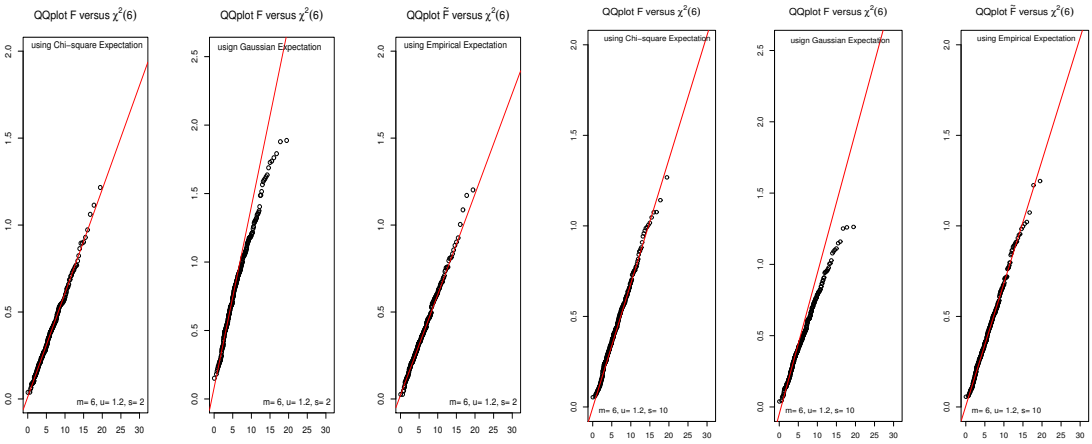

FIG 7. *QQplot: quantiles of F and $\widetilde{F}$ versus quantiles of the χ^2_m distribution for fixed level $u = 1.2$ and $m = 6$ disjoint domains. Here $|T_i| = 200$ for $i \in \{1, \dots, m\}$. We take 300 Monte Carlo simulations. The considered process $Z^{(s)}(\cdot)$ is a chi-square univariate process ($d = 1$) with $s = 2$ (first, second and third panels), $s = 10$ (fourth, fifth and sixth panels), $\mathbb{E}(Z^{(s)}(0)) = 0$, $\mathrm{Var}(Z^{(s)}(0)) = 1$ and $\lambda = 2$.*

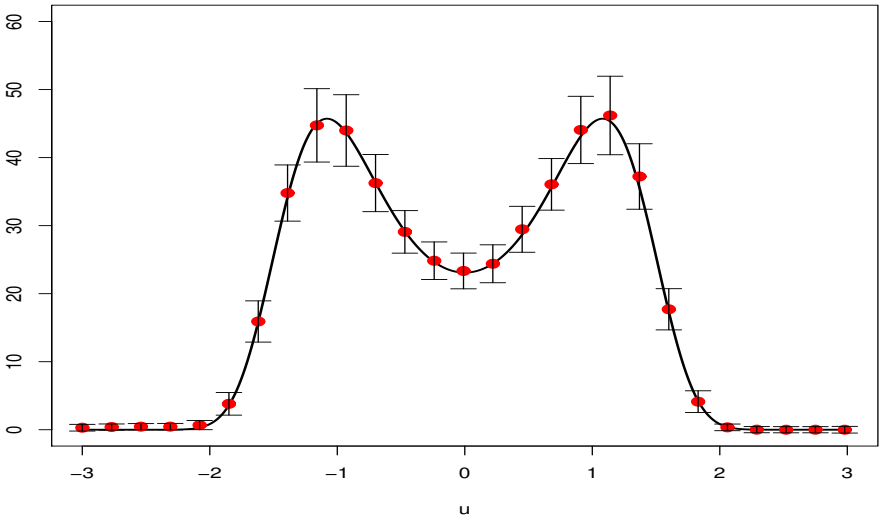

FIG 8. ***Left:*** *Theoretical $u \mapsto \mathbb{E}[\varphi(Q, T, u)]$ from Equation (15) for $|T| = 200$ (full line). We also display, for different levels of u, the empirical counterpart $\widehat{\mathbb{E}}[\varphi(Q, T, u)]$ (red dots), with associated empirical intervals, based on 300 Monte Carlo simulations. $Q(\cdot)$ is a Kramer oscillator process as defined in Section 3.2 with $\mathbb{E}(Q(0)) = 0$, $\mathrm{Var}(Q(0)) = 1$ and second spectral moment $\lambda = 2$.*

instead of the process $Z^{(s)}$. In this case, the expectation $\mathbb{E}[\varphi(Q, T, u)]$ is given by Equation (15) and $\widehat{\mathbb{E}}[\varphi(Q, T, u)]$ is the empirical mean evaluated on the Monte Carlo sample. The QQplot of F and $\widetilde{F}$ versus χ^2_m distribution are shown in the first and third panels of Figure 9 respectively. The second panel shows the QQplot of F versus χ^2_m distribution, when the expectation is comptuted following (19) as if hypothesis **H0** was true (i.e., Gaussian case). Since the considered process Q is not Gaussian,

a deviation is observed. Besides, let us remark that the almost linear graphs observed in Panels 1 and 3 allow us to believe that $\varphi(Q, T, u)$ satisfies a Central Limit Theorem when T grows to $\mathbb{R}$.

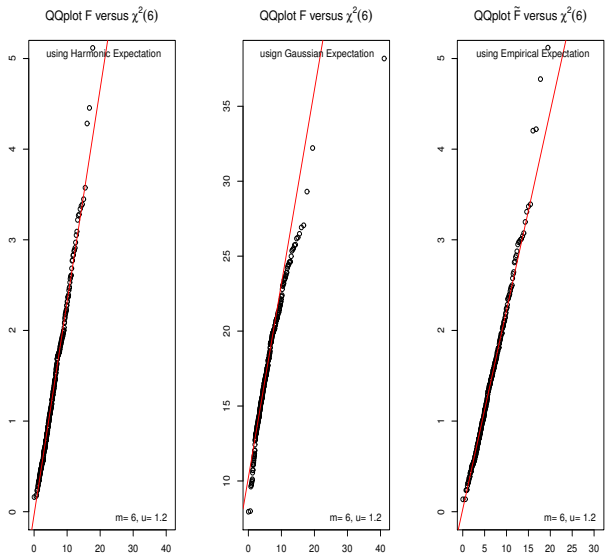

FIG 9. *QQplot: quantiles of F and $\widetilde{F}$ versus quantiles of the χ^2_m distribution for fixed level $u = 1.2$ and $m = 6$ disjoint domains. Here $|T_i| = 200$ for $i \in \{1, \ldots, m\}$. We take 300 Monte Carlo simulations. The considered process $Q(\cdot)$ is a Kramer oscillator process as defined in Section 3.2 with $\mathbb{E}(Q(0)) = 0$, $\mathrm{Var}(Q(0)) = 1$ and second spectral moment $\lambda = 2$.*

4.6. Third alternative: shot noise process

We consider a shot noise process as introduced in Equation (17) with $a = \lambda = 1$, i.e.,

$$S(t) = \left(\sum_{\xi \in \Phi} \mathbf{1}_{[0,1]}(t - \xi) \right) - 1 \ , \ t \in \mathbb{R} \, ,$$

and generate 300 trajectories of such a process on a fixed interval T. Since $a\lambda = 1$, then $\mathrm{Var}(S(0)) = 1$. Moreover, Equation (18) becomes

$$\mathbb{E}[U(S, T, u)] = |T| \, e^{-1} \sum_{k \geq 0} \frac{1}{k!} \mathbf{1}_{\{k-1 < u < k\}} \, . \tag{24}$$

A comparison between this theoretical expectation and the Monte-Carlo empirical mean is presented in Figure 10.

5. Bivariate numerical illustration

5.1. Under H0 *hypothesis*

We consider a stationary centered Gaussian random field $X = \{X(t) : t \in \mathbb{R}^2\}$. Its restriction to a finite regular grid included can be seen as a model for a grey level image. The modified Euler characteristic

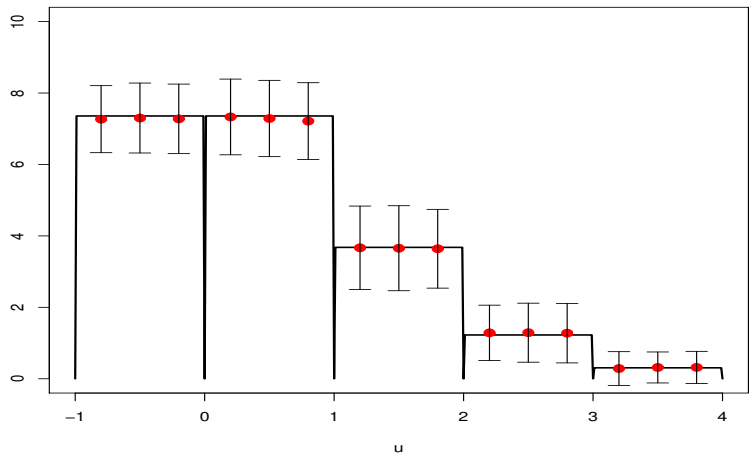

FIG 10. *Theoretical* $u \mapsto \mathbb{E}[U(S,T,u)]$ *from Equation* (24) *for* $|T| = 200$ *(full line). We also display, for different levels* u*, the empirical counterpart* $\widehat{\mathbb{E}}[U(S,T,u)]$ *(red dots) with associated empirical intervals based on* 300 *Monte Carlo simulations.* $S(\cdot)$ *is a shot noise univariate process (*$d = 1$*) as defined in Section 3.3 with* $\mathbb{E}(S(0)) = 0$, $\mathrm{Var}(S(0)) = 1$.

of an excursion set in the rectangle domain T above level u is given by Equation (4). On the other hand, Equation (5) in dimension $d = 2$ gives its expectation,

$$\mathbb{E}[\varphi(X,T,u)] = |T|\, C_2(u,\lambda) \quad \text{with} \quad C_2(u,\lambda) = (2\pi)^{-3/2}\, \lambda\, u\, e^{-u^2/2}. \tag{25}$$

In what follows, we generate a 300 Monte-Carlo sample of a bivariate stationary centered Gaussian random field X with covariance function $r(t) = e^{-||t||^2}$, $t \in \mathbb{R}^2$. In that case, $\mathbb{E}(X(0)) = 0$, $\mathrm{Var}(X(0)) = 1$ and the second spectral moment λ is equal to 2. We use (4) in order to compute $\varphi(X,T,u)$ for a fixed cube T and various values of u. The local extremum points of X are given by the `R` function `extrema2dC` in the `EMD` package. To identify the saddle points, we find all the stationary points of X in the considered domain T and we exclude the points previously identified as local extremum points. In Figure 11, the comparison between the theoretical mean given by (25) and the empirical mean based on the simulations is illustrated.

In Figure 12, we consider the chi-square statistics F and $\widetilde{F}$ introduced in (23) for a unique level u and m disjoint domains $T_1, \ldots, T_m$. The considered process is now the Gaussian bivariate process X. For F (see left panel), the expectation is given by (25) and for $\widetilde{F}$ (see right panel), $\widehat{\mathbb{E}}$ is the empirical mean on the Monte Carlo sample generations.

5.2. Alternative: χ^2 *field*

As an alternative to hypothesis **H0**, let us consider

$$Z^{(s)}(\cdot) = \frac{1}{\sqrt{2s}}(\chi_s^2(\cdot)) - s) \quad \text{with} \quad \chi_s^2(\cdot) = \sum_{1 \le i \le s} X_i(\cdot)^2\,,$$

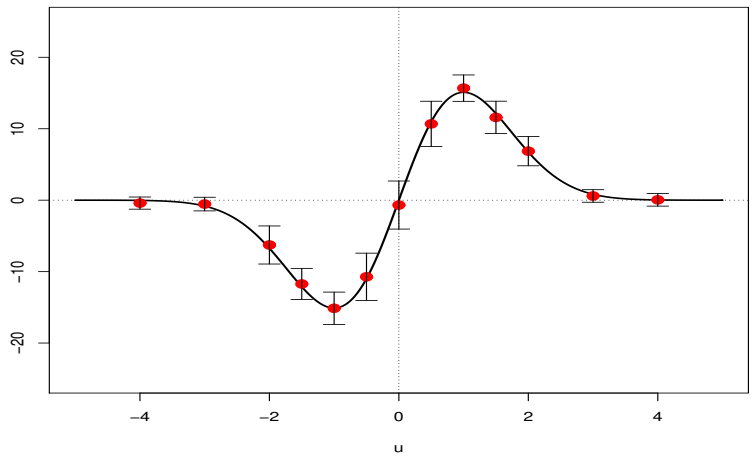

FIG 11. *Theoretical* $u \mapsto \mathbb{E}[\varphi(X,T,u)]$ *from Equation* (25) *for* $|T| = 196$ *(full line). We also display, for different levels* u*, the empirical counterpart* $\widehat{\mathbb{E}}[\varphi(X,T,u)]$ *(red dots), with associated empirical intervals, based on* 300 *Monte Carlo simulations.* $X(\cdot)$ *is a Gaussian bivariate process* ($d = 2$) *with covariance function* $r(t) = e^{-||t||^2}$, $t \in \mathbb{R}^2$*. In this case* $\mathbb{E}(X(0)) = 1$, $\mathrm{Var}(X(0)) = 1$ *and* $\lambda = 2$.

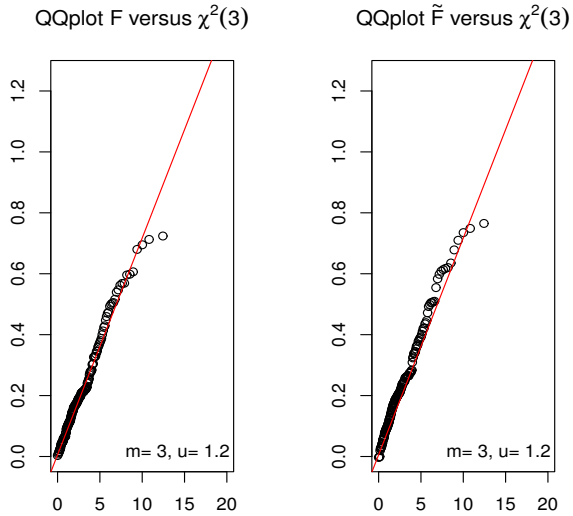

FIG 12. *QQplot: quantiles of* F *and* $\widetilde{F}$ *versus quantiles of* χ^2_m *distribution for fixed level* $u = 1.2$ *and* $m = 3$ *disjoint domains. Here* $|T_i|^2 = 196$ *for* $i \in \{1, \ldots, m\}$*. We take 300 Monte Carlo simulations.* $X(\cdot)$ *is a Gaussian bivariate process* ($d = 2$) *with covariance function* $r(t) = \exp(-||t||^2)$*. In this case* $\lambda = 2$.

where the X_i's are independent copies of a centered stationary Gaussian two dimensional field with covariance function $r(t) = e^{-||t||^2/2}$. As described in Section 3.1, for any fixed $t \in \mathbb{R}^2$, $\chi^2_s(t)$ has a central χ^2 distribution with s degrees of freedom. Furthermore, the field $Z^{(s)}$ is centered, stationary and its covariance function is given by $r(t) = e^{-||t||^2}$. Hence its variance is equal to 1 and its second

spectral moment is equal to $\lambda = -2r''(0) = 2$. The expectation of $\varphi(Z^{(s)}, T, u)$ is given by (12) with $d = 2$.

We generate a sample of 300 realizations of the bivariate process $Z^{(s)}$ for $s = 2$ on a fixed cube T. We use Equation (4) to compute $\varphi(Z^{(2)}, T, u)$ for various values of u. On the other hand, Equation (12) yields the following in the case $d = s = 2$,

$$\mathbb{E}[\varphi(Z^{(2)}, T, u)] = |T| \left(\frac{\lambda}{4\pi}\right) (2u+1)\, e^{-(u+1)}\, \mathbf{1}_{[0,\infty)}(u+1). \tag{26}$$

In Figure 13, we compare Equation (26) and its empirical counterpart based on the Monte Carlo simulations.

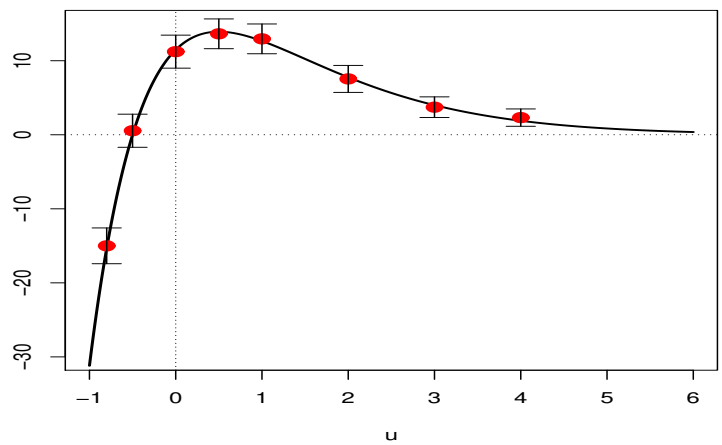

FIG 13. *Theoretical* $u \mapsto \mathbb{E}[\varphi(Z^{(2)}, T, u)]$ *from Equation* (26) *for* $|T| = 196$ *(full line). We also display, for different levels* u*, the empirical counterpart* $\widehat{\mathbb{E}}[\varphi(Z^{(2)}, T, u)]$ *(red dots), with associated empirical intervals, based on* 300 *Monte Carlo simulations.* $Z^{(2)}(\cdot)$ *is a* χ^2 *bivariate process* ($d = 2$) *with covariance function* $r(t) = e^{-||t||^2}$*. In this case* $\mathbb{E}(Z^{(2)})(0) = 1$, $\mathrm{Var}(Z^{(2)}(0)) = 1$ *and* $\lambda = 2$.

References

[1] Adler R. (2008). Some New Random Field Tools for Spatial Analysis. *Stochastic Environmental Research and Risk Assessment*, **22**, 809-822.

[2] Adler R.J., Bartz K., Kou S., Monod A. (2014). Estimating thresholding levels for random fields via Euler characteristics. *Preprint*.

[3] Adler R.J., Taylor J.E. (2007). *Random Fields and Geometry.* Springer Monographs in Mathematics. Springer.

[4] Azaïs J-M., Wschebor M. *Level Sets and Extrema of Random Processes and Fields.* Wiley (2009).

[5] Biermé H., Desolneux A. (2012). A Fourier approach for the level crossings of Shot Noise processes with jumps. *J. Appl. Probab.*, **49**(1), 100–113.

[6] Cabaña E.M. (1987). Affine processes: a test of isotropy based on level sets. *SIAM Journal on Applied Mathematics*, **47**(4), 886-891.

[7] Cuesta-Albertos J.A., Gamboa, F. and Nieto-Reyes, A. (2014). A random-projection based procedure to test if a stationary process is Gaussian. *Computational Statistics and Data Analysis*, **75**, 124-141.

[8] Di Bernardino E., Estrade A., León J.R. (2016). Supplementary material to "A test of Gaussianity based on the Euler characteristic of excursion sets". *HAL-archives ouvertes*.

[9] Epps, T. W. (1987). Testing that a stationary time series is Gaussian. *Ann. Stat.* **15**(4), 1683-1698.

[10] Estrade A., León J.R. (2015). A central limit theorem for the Euler characteristic of a Gaussian excursion set. Preprint, hal-00943054 (2015). *Ann. of Probab.* (to appear).

[11] Hsing T. (1987). On the intensity of crossings by a shot noise process. *Adv. Appl. Probab.*, **19**, 743-745.

[12] Lelievre T., Rousset M. and Stolz G. (2010). *Free Energy Computations.* Imperial College Press.

[13] Lindgren G. (1974). Spectral Moment Estimation by Means of Level Crossings. *Biometrika*, **61**(3), 401-418.

[14] Lobato, I.N. and Velasco, C. (2004). A simple test of normality for time series. *Econometric Theory*, **20**(4), 671-689.

[15] Nott, D.J., Wilson, R.J. (2000). Multi-phase image modelling with excursion sets. *Signal Processing*, **80**, 125-139.

[16] Nourdin I., Peccati G. and Podolskij M. (2011). Quantitative Breuer-Major theorems. *Stochastic Processes and their Applications*, **121**, 793-812.

[17] Rice S.O. (1945). Mathematical Analysis of Random Noise. *Bell System Technical Journal*, **23** and **24**.

[18] Roberts, A.P., Torquato, S. (1999). Chord-distribution functions of three-dimensional random media: approximate first-passage times of Gaussian processes. *Phys. Rev. E* **59**(5), 4953-4963.

[19] Sharpe K. (1978). Some properties of the crossings process generated by a stationary χ^2 process. *Adv. App. Prob.* **10**, 373-391.

[20] Subba Rao, T. and Gabr, M.M. (1980). A test for linearity of stationary time series. *J. Time Ser. Anal.* **1**, 145-158.

[21] Taheriyoun A.R. (2012). Testing the covariance function of stationary Gaussian random fields. *Statist. Probab. Lett.* **82**, 606-613.

[22] Taheriyoun A.R., Shafie K., Jafari Jozani M. (2009). A note on the higher moments of the Euler characteristic of the excursion sets of random fields. *Statist. Probab. Lett.* **79**, 1074-1082.

[23] Worsley K.J. (1994). Local maxima and the expected Euler characteristic of excursion sets of χ^2, F and t fields. *Adv. Appl. Probab.* **26**, 13-42.

[24] Worsley K.J. (1995). Estimating the number of peaks in a random field using the Hadwiger characteristic of excursion sets, with applications to medical images. *Ann. Stat.* **23**, 640-669.

[25] L. Wu (2001). Large and moderate deviations and exponential convergence for stochastic damping Hamiltonian systems. *Stoch. Processes and Appl.* **91**, 205-238.