

An approximation to peak detection power using Gaussian random field theory

Yu Zhao ^{a,*}, Dan Cheng ^c, Armin Schwartzman ^{a,b}

^a Division of Biostatistics, Herbert Wertheim School of Public Health and Human Longevity Science, University of California San Diego, 9500 Gilman Dr., La Jolla, CA 92093, USA

^b Halicioğlu Data Science Institute, University of California San Diego, 9500 Gilman Dr., La Jolla, CA 92093, USA

^c School of Mathematical and Statistical Sciences, Arizona State University, 900 S. Palm Walk, Tempe, AZ 85281, USA

ARTICLE INFO

MSC:

primary 62H12

secondary 62F12

Keywords:

Gaussian random field

Image analysis

Peak detection

Power calculations

ABSTRACT

We study power approximation formulas for peak detection using Gaussian random field theory. The approximation, based on the expected number of local maxima above the threshold u , $\mathbb{E}[M_u]$, is proved to work well under three asymptotic scenarios: small domain, large threshold, and sharp signal. An adjusted version of $\mathbb{E}[M_u]$ is also proposed to improve accuracy when the expected number of local maxima $\mathbb{E}[M_{-\infty}]$ exceeds 1.

Cheng and Schwartzman (2018) developed explicit formulas for $\mathbb{E}[M_u]$ of smooth isotropic Gaussian random fields with zero mean. In this paper, these formulas are extended to allow for rotational symmetric mean functions, making them applicable not only for power calculations but also for other areas of application that involve non-centered Gaussian random fields. We also apply our formulas to 2D and 3D simulated datasets, and the 3D data is induced by a group analysis of fMRI data from the Human Connectome Project to measure performance in a realistic setting.

1. Introduction

Detection of peaks (local maxima) is an important topic in image analysis. For example, a fundamental goal in fMRI analysis is to identify the local hotspots of brain activity [1,2], which are typically captured by peaks in the fMRI signal. The detection of such peaks can be posed as a statistical testing problem intended to test whether the underlying signal has a peak at a given location. This is challenging because such tests are conducted only at locations of observed peaks, which depend on the data. Therefore, the height distribution of the observed peak is conditional on a peak being observed at that location. This is a nonstandard problem. Solutions exist using random field theory (RFT) [3]. RFT is a statistical framework that can be used to perform topological inference and modeling. RFT-based peak detection has been studied in Cheng and Schwartzman [4] and Schwartzman and Telschow [5], which provide the peak height distribution for isotropic noise under the null hypothesis of no signal anywhere, often referred to in the multiple testing literature as the “complete null hypothesis”.

In general, for any statistical testing problem, accurate power calculations help researchers decide the minimum sample size required for an informative test, and thus reduce cost. For this reason, power calculation is crucial in image analysis [6–8], yet it is often overlooked or lacks serious consideration [9,10]. Power calculation formulas exist for common univariate tests, such as z-tests and t-tests. However, particular challenges arise when we perform power calculations in peak detection settings. Due to the nature of imaging data, the number and location of the signal peaks are unknown. Besides, the power is affected by other spatial aspects

* Corresponding author.

E-mail address: yuz867@ucsd.edu (Y. Zhao).

of the problem, such as the shape of the peak function and the spatial autocorrelation of the noise. Considering these difficulties, it requires some extra effort to derive a power formula for peak detection.

A formal definition of power in peak detection is necessary to perform power calculations. In Cheng and Schwartzman [4] and Durnez et al. [11], the authors explored approaches to control the false discovery rate (FDR). For the entire domain, average peakwise power, i.e., power averaged over all non-null voxels, is a natural choice for these approaches. For a local domain (a subset of the entire domain) where a single peak exists, the power can be defined as the probability of successfully detecting that peak. Following this idea, we describe the null and alternative hypothesis and the definition of detection power. For the entire domain, we can approximate the average peakwise power by taking the arithmetic mean of the approximation proposed in this paper over non-null voxels. We do so informally here for didactic purposes and present formal definitions in Section 2.

Consider a local domain where a single peak may exist, and consider the hypotheses:

H_0 : the signal is equal to 0 in the local domain,

H_1 : the signal has at least one peak with height greater than 0 in the local domain.

Suppose we observe a random field to be used as test statistic at every location, typically as the result of statistical modeling of the data. For a fixed threshold u , the existence of observed peaks with height greater than u would lead to rejecting the null hypothesis. Therefore, we define the type I error and power as the probability of existing at least one local maximum above u under H_0 and H_1 respectively:

Type I error: $\mathbb{P}\{\text{a peak is observed in the local domain with height greater than } u \text{ when } H_0 \text{ is true}\},$

Power: $\mathbb{P}\{\text{a peak is observed in the local domain with height greater than } u \text{ when } H_1 \text{ is true}\}.$ (1)

Formulas for type I error have been developed for stationary fields in 1D and isotropic fields in 2D and 3D [4,12,13]. However, there is no formula to calculate power. In order to get an appropriate estimate of power, we need to know the peak height distribution for non-centered (the mean function is not 0) random fields. Generally speaking, it is very difficult to calculate the peak height distribution especially when the random field has non-zero mean. Durnez et al. [11] suggests using Gaussian distribution to describe the non-null peaks and truncated Gaussian distribution to approximate the overshoot distribution. This approach is easy to implement but not very accurate because the peak height distribution is in reality always skewed and not close to any Gaussian distribution.

In this article, we propose to approximate the probability (1) by calculating the expected number of peaks above the threshold u . We show that the approximation, which is also an upper bound, works well under certain scenarios. For the entire domain, we can approximate the average peakwise power by taking the arithmetic mean of the approximation proposed in this paper over non-null voxels.

The proposed approximation makes the problem more tractable, but in general, it does not have an explicit form. In order to make it applicable in practice, we further simplify the formula under the isotropy assumption and show its explicit form in 1D, 2D, and 3D. The explicit results are validated through 2D and 3D computer simulations carried out in MATLAB. The simulation also covers multiple scenarios by modifying the parameters used to generate the data. The performance of power approximation and its conservative adjustment under these scenarios are discussed.

Finally, to assess the real-data performance of our power approximation method, we apply it to a 3D simulation induced by a real brain imaging dataset, where the parameters are estimated from the Human Connectome Project [14] fMRI data. By testing the method in a realistic setting, we also demonstrate how effect size and other parameters affect the power.

The paper is organized as follows. We first show in Section 2 the problem setup and theoretical results in certain scenarios. In Section 3, we derive the explicit formulas under isotropy. Simulation in 2D is conducted in Section 4. Details regarding how to apply our formula in application settings are discussed in Section 5. The methodology is applied to a 3D real dataset in Section 6.

2. Power approximation

2.1. Setup

Let $Y(s) = \sigma(s)Z(s) + \mu(s)$, where $Z = \{Z(s), s \in D\}$ representing the noise is a centered (zero-mean), smooth (as defined by Assumption 1 below), unit-variance Gaussian random field on an N -dimensional, non-empty, bounded, open domain $D \subset \mathbb{R}^N$, $\sigma(s)$ is the standard deviation of the noise and $\mu(s)$ is the mean function. Let $X(s) = Y(s)/\sigma(s) = Z(s) + \theta(s)$ where the ratio $\theta(s) = \mu(s)/\sigma(s)$ is the standardized mean function, which we assume to be C^2 . Here a.s. C^3 is a sufficient smoothness condition for Z , and this will be clarified in Assumption 1 below.

Let

$$X_i(s) = \frac{\partial X(s)}{\partial s_i}, \nabla X(s) = (X_1(s), \dots, X_N(s)), X_{ij}(s) = \frac{\partial^2 X(s)}{\partial s_i \partial s_j}, \nabla^2 X(s) = (X_{ij}(s))_{1 \leq i, j \leq N},$$

$$Z_i(s) = \frac{\partial Z(s)}{\partial s_i}, \nabla Z(s) = (Z_1(s), \dots, Z_N(s)), Z_{ij}(s) = \frac{\partial^2 Z(s)}{\partial s_i \partial s_j}, \nabla^2 Z(s) = (Z_{ij}(s))_{1 \leq i, j \leq N}.$$

We will make use of the following assumptions:

Assumption 1. $Z \in C^2(D)$ almost surely and its second derivatives satisfy the mean-square Hölder condition: for any $s_0 \in D$, there exist positive constants L, γ and ϵ such that

$$\mathbb{E}[Z_{ij}(s) - Z_{ij}(t)]^2 \leq L^2 \|s - t\|^{2\gamma}, \quad \forall t, s \in U(s_0, \epsilon) \cap D, i, j \in \{1, \dots, N\},$$

where $U(s_0, \epsilon) = s_0 \oplus (-\epsilon/2, \epsilon/2)^N$ is the N dimensional open cube of side length ϵ centered at s_0 .

Assumption 1 is satisfied, for example, if Z is $C^3(D)$.

Assumption 2. For every pair $(t, s) \in D \times D$ with $s \neq t$, the Gaussian random vector

$$(Z(s), \nabla Z(s), \text{vech}(\nabla^2 Z(s)), Z(t), \nabla Z(t), \text{vech}(\nabla^2 Z(t)))$$

is non-degenerate, i.e., its covariance matrix has full rank. Here $\text{vech}(\cdot)$ is the half vectorization operator [15].

Assumption 2 is satisfied, for example, for random fields obtained as a convolution of white noise with a kernel (26), as used in the data example section below.

2.2. Peak detection

Following the notation in the problem setup, the null and alternative hypotheses can be written as:

$$\begin{aligned} H_0 : \theta(s) &= 0 \text{ for all } s \in D, \\ H_1 : \theta(s) &> 0, \nabla \theta(s) = 0, \nabla^2 \theta(s) < 0 \text{ for some } s \in D, \end{aligned}$$

where $\nabla^2 \theta(s) < 0$ denotes that the Hessian matrix is negative definite. The mean function $\theta(s)$ is not directly observed, so the hypothesis is tested based on the peak height of $X(s)$. For a peak detection procedure that aims to test this hypothesis, a threshold u for the peak height of $X(s)$ needs to be set in advance. If a local maximum with height greater than u is observed, we would choose to reject the null hypothesis due to the strong evidence against it. The probability that existing a peak of X exceeds u

$$\mathbb{P}(\exists s \in D \text{ s.t. } X(s) > u, \nabla X(s) = 0, \nabla^2 X(s) < 0) \quad (2)$$

is the type I error under H_0 and power under H_1 . The threshold u can be obtained based on the peak height distribution under H_0 . A formula for peak height distribution of smooth isotropic Gaussian random fields has been derived in Cheng and Schwartzman [13] and it can also be derived directly from a special case of the formulas presented in this paper. Usually, u is set to be some quantile of the null distribution of peak height to maintain the nominal α type I error. More details about selecting the threshold will be discussed in the real data example. Selecting u is not the main focus of this paper and our method can be applied to any choice of u .

2.3. Power approximation

Let $M_u(D)$ be the number of local maxima of the random field X above u over the domain D . The power defined in (2) can be represented as a function $F(u, D) = \mathbb{P}[M_u(D) \geq 1]$. We call this the power function, seen as a function of the threshold u . In what follows, to simplify the notation, we write $M_u(D)$ and $F(u, D)$ as M_u and $F(u)$, respectively, unless a domain other than D is specified. Note that

$$F(u) = \mathbb{P}[M_u \geq 1] = \sum_{k=1}^{\infty} \mathbb{P}[M_u = k] \leq \sum_{k=1}^{\infty} k \mathbb{P}[M_u = k] = \mathbb{E}[M_u]. \quad (3)$$

On the other hand,

$$\mathbb{E}[M_u] - F(u) = \sum_{k=2}^{\infty} (k-1) \mathbb{P}[M_u = k] \leq \frac{1}{2} \sum_{k=2}^{\infty} k(k-1) \mathbb{P}[M_u = k] = \frac{1}{2} \mathbb{E}[M_u(M_u - 1)]. \quad (4)$$

Thus, we have

$$\mathbb{E}[M_u] - \frac{1}{2} \mathbb{E}[M_u(M_u - 1)] \leq F(u) \leq \mathbb{E}[M_u]. \quad (5)$$

This inequality tells us that for any fixed u , the power is bounded within an interval of length $\mathbb{E}[M_u(M_u - 1)]/2$. Thus, $\mathbb{E}[M_u]$ is a good approximation of power if one of the two conditions below is satisfied:

- (i) The factorial moment $\mathbb{E}[M_u(M_u - 1)]$ converges to 0 and $\mathbb{E}[M_u]$ does not.
- (ii) They both converge to 0 and $\mathbb{E}[M_u(M_u - 1)]$ converges faster than $\mathbb{E}[M_u]$.

The convergence above refers to conditions on the signal and noise parameters. Later in this section, we introduce four related results and demonstrate how they are useful in application. The first result can be useful for simplifying the power function and the other three results give different scenarios where one of the conditions above holds.

We have provided evidence supporting the use of $\mathbb{E}[M_u]$ to approximate power, as demonstrated by (5). However, $\mathbb{E}[M_u]$ alone might not be sufficient for power approximation since it only gives an upper bound. Also, unlike power, $\mathbb{E}[M_u]$ sometimes exceeds one. To correct for this, we define the adjusted $\mathbb{E}[M_u]$ as

$$\mathbb{E}[M_u]_{\text{adj}} = \mathbb{E}[M_u] / \max(1, \mathbb{E}[M_{-\infty}]). \quad (6)$$

The adjusted $\mathbb{E}[M_u]$ is the same as $\mathbb{E}[M_u]$ when the expected number of local maxima $\mathbb{E}[M_{-\infty}]$ is less or equal to 1. When $\mathbb{E}[M_{-\infty}]$ is greater than 1, we divide $\mathbb{E}[M_u]$ by $\mathbb{E}[M_{-\infty}]$ to make sure it never exceeds one. The adjusted $\mathbb{E}[M_u]$ is more conservative, and we conjecture that it is a lower bound of power when there exists at least one local maximum in the domain D . In applications, people are interested in a conservative estimator so that the test is guaranteed to have enough power. Combining $\mathbb{E}[M_u]$ and $\mathbb{E}[M_u]/\mathbb{E}[M_{-\infty}]$, we can get an approximate range of the true power. Later in this paper, we will compare $\mathbb{E}[M_u]$ and adjusted $\mathbb{E}[M_u]$ through simulations.

2.4. Main results

Our first result does not concern the approximation (5) yet, but it offers a simplification of the power function and $\mathbb{E}[M_u]$ that will be used later. The proposition below states that the power function and $\mathbb{E}[M_u]$ for peak detection are translation equivariant with respect to peak height.

Proposition 1 (Height Equivariance). Let \tilde{M}_u and $\tilde{F}(u)$ be the number of local maxima above u and the power function for $\tilde{X}(s) = X(s) + \theta_0$, respectively. Then M_u , $\mathbb{E}[M_u]$ and $F(u)$ are translation equivariant with respect to the peak height θ_0 , that is, $\tilde{M}_u = M_{u-\theta_0}$, $\mathbb{E}[\tilde{M}_u] = \mathbb{E}[M_{u-\theta_0}]$ and $\tilde{F}(u) = F(u - \theta_0)$.

Proposition 1 indicates the shape of the power and $\mathbb{E}[M_u]$ curve remains unaffected by peak height. Next, we give three scenarios where the two equalities in (5) can be achieved asymptotically: small domain size, large threshold, and sharp signal.

For the first asymptotic scenario, we demonstrate that as the size of the local domain where a single peak exists converges to 0, the error of using $\mathbb{E}[M_u]$ to approximate power (i.e., the factorial moment $\mathbb{E}[M_u(M_u - 1)]$) goes to 0 faster than $\mathbb{E}[M_u]$, satisfying condition i in 2.3. This is helpful, for example, if we want to detect smaller, localized effects.

Theorem 1 (Small Domain). Consider a local domain $D_\epsilon = U(s_0, \epsilon) \cap D$ for any fixed $s_0 \in D$ where $U(s_0, \epsilon) = s_0 \oplus (-\epsilon/2, \epsilon/2)^N$ is the N -dimensional open cube of side ϵ centered at s_0 . As $\epsilon \rightarrow 0$ and with a fixed threshold u , the power function $F(u, D_\epsilon)$, $\mathbb{E}[M_u(D_\epsilon)]$, and $\mathbb{E}[M_u(D_\epsilon)]_{\text{adj}}$ satisfy the following equation

$$F(u, D_\epsilon) = \mathbb{E}[M_u(D_\epsilon)](1 - o(1)) = \mathbb{E}[M_u(D_\epsilon)]_{\text{adj}}(1 - o(1)). \quad (7)$$

The second asymptotic scenario is also closely relevant to applications, as the threshold u is usually chosen to be large to control the type I error. In the following theorem, we show that as $u \rightarrow \infty$, condition i in 2.3 holds true, so that power can be precisely approximated by $\mathbb{E}[M_u]$.

Theorem 2 (Large Threshold). For any fixed domain D , as $u \rightarrow \infty$

$$F(u) = \mathbb{E}[M_u](1 - o(e^{-\omega u^2})), \quad (8)$$

where the error term $o(e^{-\omega u^2})$ is non-negative and $\omega > 0$ is some constant.

Notice that the threshold u does not affect the value of $\mathbb{E}[M_{-\infty}]$ which is part of the adjusted $\mathbb{E}[M_u]$. By (8)

$$F(u) = \mathbb{E}[M_u]_{\text{adj}}(1 - o(e^{-\omega u^2})) \max(1, \mathbb{E}[M_{-\infty}]).$$

Therefore, the adjusted $\mathbb{E}[M_u]$ might be overly conservative as $u \rightarrow \infty$ when $\mathbb{E}[M_{-\infty}] > 1$.

The third asymptotic scenario involves power approximation when the signal peak is sharp. Here condition ii in 2.3 is satisfied. Interestingly, while the power function is generally non-Gaussian, it becomes closer to Gaussian as the signal peak becomes sharper.

Theorem 3 (Sharp Signal). Let $\theta(s) = ah(s) + \theta_0$ where $h(s)$ is a unimodal (i.e., its Hessian matrix is negative definite when gradient equals 0) mean function with maximum equal to 0 at s_0 , $a > 0$, and θ_0 represents the height. For any fixed threshold u , as $a \rightarrow \infty$

$$F(u) = \mathbb{E}[M_u] + o(1) = \mathbb{E}[M_u]_{\text{adj}} + o(1) = \Phi(\theta_0 - u)(1 + o(1)), \quad (9)$$

where $\Phi(x)$ is CDF of the standard Gaussian distribution.

3. Explicit formulas

We have shown that the power for peak detection can be approximated by the expected number of local maxima above u , $\mathbb{E}[M_u]$, under certain scenarios. Although the Kac–Rice formula can be applied to calculate $\mathbb{E}[M_u]$, it remains difficult to evaluate it explicitly for $N > 1$ without making any further assumptions. In this section, we focus on the computation of $\mathbb{E}[M_u]$ and provide a general formula when the noise field is isotropic. Furthermore, explicit formulas when $N \in \{1, 2, 3\}$ are derived for application purposes.

3.1. Isotropic Gaussian fields

Suppose Z is a zero-mean unit-variance isotropic random field. We can write the covariance function of Z as $\mathbb{E}\{Z(s)Z(t)\} = \rho(\|t - s\|^2)$ for an appropriate function $\rho(\cdot) : [0, \infty) \rightarrow \mathbb{R}$. Denote

$$\rho' = \rho'(0), \quad \rho'' = \rho''(0), \quad \kappa = -\rho'/\sqrt{\rho''}, \quad (10)$$

where ρ' and ρ'' are first and second derivative of function ρ respectively.

The following lemma comes from Cheng and Schwartzman [13].

Lemma 1. For each $s \in \mathbb{R}^N$ and $i, j, k, \ell \in \{1, \dots, N\}$,

$$\mathbb{E}\{Z_i(s)Z(s)\} = \mathbb{E}\{Z_i(s)Z_{jk}(s)\} = 0, \quad \mathbb{E}\{Z_i(s)Z_j(s)\} = -\mathbb{E}\{Z_{ij}(s)Z(s)\} = -2\rho'\delta_{ij}, \quad \mathbb{E}\{Z_{ij}(s)Z_{k\ell}(s)\} = 4\rho''(\delta_{ij}\delta_{k\ell} + \delta_{ik}\delta_{j\ell} + \delta_{i\ell}\delta_{jk}),$$

where ρ' and ρ'' are defined in (10) and δ_{ij} is the Kronecker delta function.

In particular, it follows from Lemma 1 that $\text{Var}(Z_i(s)) = -2\rho'$ and $\text{Var}(Z_{ii}(s)) = 12\rho''$ for any $i \in \{1, \dots, N\}$, implying $\rho' < 0$ and $\rho'' > 0$ and hence $\kappa > 0$.

We can use theoretical results from Gaussian Orthogonally Invariant (GOI) matrices to make the calculation of $\mathbb{E}[M_u]$ easier. GOI matrices were first introduced in Schwartzman et al. [16], and used for the first time in the context of random fields in Cheng and Schwartzman [13]. It is a class of Gaussian random matrices that are invariant under orthogonal transformations, and can be useful for computing the expected number of critical points of isotropic Gaussian fields. We call an $N \times N$ random matrix $G = (G_{ij})_{1 \leq i, j \leq N}$ GOI with covariance parameter c , denoted by $\text{GOI}(c)$, if it is symmetric and all entries are centered Gaussian variables such that

$$\mathbb{E}[G_{ij}G_{k\ell}] = \frac{1}{2}(\delta_{ik}\delta_{j\ell} + \delta_{i\ell}\delta_{jk}) + c\delta_{ij}\delta_{k\ell}. \quad (11)$$

The following lemma is Lemma 3.4 from Cheng and Schwartzman [13].

Lemma 2. Let the assumptions in Lemma 1 hold. Let \tilde{G} and G be $\text{GOI}(1/2)$ and $\text{GOI}((1 - \kappa^2)/2)$ matrices respectively. I_N denotes $N \times N$ identity matrix.

- (i) The distribution of $\nabla^2 Z(s)$ is the same as that of $\sqrt{8\rho''}\tilde{G}$.
- (ii) The distribution of $(\nabla^2 Z(s)|Z(s) = z)$ is the same as that of $\sqrt{8\rho''}[G - (\kappa z/\sqrt{2})I_N]$.

Lemma 2 shows the distribution and conditional distribution of the Hessian matrix of a centered random field $Z(s)$. Next, we establish the corresponding result for non-centered random field $X(s) = Z(s) + \theta(s)$.

Lemma 3. Let \tilde{G} and G be $\text{GOI}(1/2)$ and $\text{GOI}((1 - \kappa^2)/2)$ matrices respectively.

- (i) The distribution of $\nabla^2 X(s)$ is the same as that of $\sqrt{8\rho''}\tilde{G} + \nabla^2 \theta(s)$.
- (ii) The distribution of $(\nabla^2 X(s)|X(s) = x)$ is the same as that of

$$\sqrt{8\rho''} \left[G - \frac{\kappa(x - \theta(s))}{\sqrt{2}} I_N \right] + \nabla^2 \theta(s).$$

3.2. General formula under isotropy

Theorem 4. Let $X(s) = Z(s) + \theta(s)$, where $Z(s)$ is a smooth zero-mean unit-variance isotropic Gaussian random field satisfying Assumptions 1, 2. Let $\theta(s)$ a C^3 mean function such that $\nabla^2 \theta(s)$ is a non-singular matrix with ordered eigenvalues $\theta''_1(s) \dots \theta''_N(s)$ at all critical points s . Then for any domain D

$$\mathbb{E}[M_u] = \left(\frac{2\rho''}{-\pi\rho'} \right)^{N/2} \int_D \exp\left(\frac{\|\nabla \theta(s)\|^2}{4\rho'} \right) \int_u^\infty \phi(x - \theta(s)) \mathbb{E} [|\det(\text{Matrix}(s))| \mathbb{1}_{\{\text{Matrix}(s) < 0\}}] dx ds, \quad (12)$$

where $\phi(x)$ is the PDF of the standard Gaussian distribution, $\text{Matrix}(s) = G - \kappa(x - \theta(s))I_N/\sqrt{2} + \text{diag}\{\theta''_1(s), \dots, \theta''_N(s)\}/\sqrt{8\rho''}$, G as in Lemma 3 represents $\text{GOI}((1 - \kappa^2)/2)$, and $\mathbb{1}_{\{\cdot\}}$ denotes the indicator function.

The expression (12) can be simplified further if we further assume the mean function $\theta(s)$ to be a rotationally symmetric paraboloid centered at s_0 . In this case, the Hessian of $\theta(s)$ is the identity matrix multiplied by a constant, i.e.,

$$\theta'' = \theta''_1(s) = \theta''_2(s) = \dots = \theta''_N(s).$$

Then we can write the mean function as $\theta(s) = \theta_0 + \theta''\|s - s_0\|^2/2$. Define

$$\eta = \frac{\theta''}{2\kappa\sqrt{\rho''}} = \frac{\theta''}{-2\rho'} = \frac{\theta''}{\text{Var}(Z_1(s))}. \quad (13)$$

and

$$H(\tilde{x}) = \mathbb{E}_{\text{GOI}((1-\kappa^2)/2)}^N \left[\prod_{j=1}^N \lambda_j - \frac{\kappa \tilde{x}}{\sqrt{2}} \right] \mathbb{1}_{\{\lambda_N < \frac{\kappa \tilde{x}}{\sqrt{2}}\}}. \quad (14)$$

$\mathbb{E}[M_u]$ can be simplified as

$$\mathbb{E}[M_u] = \left(\frac{2\rho''}{-\pi\rho'} \right)^{N/2} \int_D \exp\left(\frac{\theta'^2 \|s - s_0\|^2}{4\rho'} \right) \int_{\tilde{u}(s)}^\infty \phi(\tilde{x} + \eta) H(\tilde{x}) d\tilde{x} ds, \quad (15)$$

where we make a change of variable $\tilde{x} = x - \theta(s) - \eta$ and $\tilde{u}(s) = u - \theta(s) - \eta$. Note that the parameter κ depends on the correlation structure of $Z(s)$.

3.3. Explicit formulas in 1D, 2D and 3D

In (15), a general formula for $\mathbb{E}[M_u]$ under isotropy was derived. To make the formula easier to apply in practice, we have the following results for computing it in 1D, 2D, and 3D. When $N = 1$, the derivation is simple enough that we do not need additional assumptions on the mean function $\theta(s)$ except those in Theorem 4, and it follows directly from Kac–Rice formula. When $N = 2$ and 3, we assume the mean function $\theta(s)$ is a rotationally symmetric paraboloid centered at s_0 . $\mathbb{E}[M_u]$ is calculated by first obtaining explicit formulas for $H(\tilde{x})$, and plugging H into (15).

Proposition 2. Let $N = 1$, $X(s) = Z(s) + \theta(s)$, where $Z(s)$ is a smooth zero-mean unit-variance Gaussian process and $\theta(s)$ is a C^3 mean function. Assume additionally that $Z(s)$ is stationary, then

$$\mathbb{E}[M_u] = \int_D \frac{\sqrt{-2\rho'(3-\kappa^2)}}{\kappa} \phi\left(\frac{\theta'(s)}{\sqrt{-2\rho'}} \right) \int_u^\infty \phi(x - \theta(s)) \psi\left(\frac{\kappa[x - \theta(s) - \eta(s)]}{\sqrt{3-\kappa^2}} \right) dx ds, \quad (16)$$

where the function ψ is defined as

$$\psi(x) = \int_{-\infty}^x \Phi(y) dy = \phi(x) + x\Phi(x), \quad x \in \mathbb{R}.$$

Note that when $N = 1$,

$$H(\tilde{x}) = \phi\left(\frac{\kappa \tilde{x}}{\sqrt{3-\kappa^2}} \right) + \frac{\kappa \tilde{x}}{\sqrt{3-\kappa^2}} \Phi\left(\frac{\kappa \tilde{x}}{\sqrt{3-\kappa^2}} \right) = \psi\left(\frac{\kappa \tilde{x}}{\sqrt{3-\kappa^2}} \right). \quad (17)$$

We need the following lemmas to calculate $H(\tilde{x})$ explicitly when $N = 2$ and $N = 3$. They result from a direct application of integral by parts.

Lemma 4. Let $N = 2$. For constant $a > -\frac{1}{2}$ and $b \in \mathbb{R}$, we have

$$\begin{aligned} & \int_{\mathbb{R}^2} \exp\left(-\frac{1}{2} \sum_{i=1}^2 \lambda_i^2 - \frac{a}{2} \left(\sum_{i=1}^2 \lambda_i \right)^2 \right) \left(\prod_{i=1}^2 |\lambda_i - b| \right) |\lambda_1 - \lambda_2| \mathbb{1}_{\{\lambda_1 < \lambda_2 < b\}} d\lambda_1 d\lambda_2 \\ &= \frac{\sqrt{2\pi}}{\sqrt{1+a}} \exp\left(-\frac{1+2a}{2(1+a)} b^2 \right) \Phi\left(\frac{1+2a}{\sqrt{1+a}} b \right) + \left(2b^2 - \frac{1+4a}{1+2a} \right) \frac{\sqrt{\pi}}{\sqrt{1+2a}} \Phi(\sqrt{2(1+2a)}b) + \frac{b}{1+2a} \exp(-(1+2a)b^2). \end{aligned} \quad (18)$$

Lemma 5. Let $N = 3$. For constant $a > 0$ and $b \in \mathbb{R}$, we have

$$\begin{aligned} & \int_{\mathbb{R}^3} \exp\left(-\frac{1}{2} \sum_{i=1}^3 \lambda_i^2 + \frac{a}{2(2+3a)} \left(\sum_{i=1}^3 \lambda_i \right)^2 \right) \left(\prod_{i=1}^3 |\lambda_i - b| \right) \prod_{1 \leq i < j \leq 3} |\lambda_i - \lambda_j| \mathbb{1}_{\{\lambda_1 < \lambda_2 < \lambda_3 < b\}} d\lambda_1 d\lambda_2 d\lambda_3 \\ &= \left[\frac{a^3 + 6a^2 + 12a + 24}{2(a+2)^2} b^2 + \frac{2a^3 + 3a^2 + 6a}{4(a+2)} + \frac{3}{2} \right] \frac{1}{\sqrt{\pi(a+2)}} \exp\left(-\frac{b^2}{a+2} \right) \Phi\left(\frac{2\sqrt{2}b}{\sqrt{(a+2)(3a+2)}} \right) \\ &+ \left[\frac{a+1}{2} b^2 + \frac{a^2 - a}{2} - 1 \right] \frac{1}{\sqrt{\pi(a+1)}} \exp\left(-\frac{b^2}{a+1} \right) \Phi\left(\frac{\sqrt{2}b}{\sqrt{(a+1)(3a+2)}} \right) \\ &+ \left(a+6 + \frac{3a^3 + 12a^2 + 28a}{2(a+2)} \right) \frac{b}{2\pi(a+2)\sqrt{3a+2}} \exp\left(-\frac{3b^2}{3a+2} \right) + b \left[b^2 + \frac{3(a-1)}{2} \right] [\Phi_{\Sigma_1}(0, b) + \Phi_{\Sigma_2}(0, b)], \end{aligned}$$

where

$$\Sigma_1 = \begin{pmatrix} \frac{3}{2} & -1 \\ -1 & \frac{a+2}{2} \end{pmatrix}, \quad \Sigma_2 = \begin{pmatrix} \frac{3}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{a+1}{2} \end{pmatrix}.$$

Proposition 3. Let $N = 2$ and the assumptions in Theorem 4 hold. Then the function H defined in (14) can be written explicitly as

$$H(\tilde{x}) = \frac{\sqrt{2\pi}}{\sqrt{3-\kappa^2}} \phi\left(\frac{\kappa\tilde{x}}{\sqrt{3-\kappa^2}}\right) \Phi\left(\frac{\kappa\tilde{x}}{\sqrt{(2-\kappa^2)(3-\kappa^2)}}\right) + \frac{\kappa^2}{2}(\tilde{x}^2 - 1)\Phi\left(\frac{\kappa\tilde{x}}{\sqrt{2-\kappa^2}}\right) + \frac{\kappa\sqrt{2-\kappa^2}\tilde{x}}{2}\phi\left(\frac{\kappa\tilde{x}}{\sqrt{2-\kappa^2}}\right). \quad (19)$$

Proposition 4. Let $N = 3$ and the assumptions in Theorem 4 hold. Then the function H defined in (14) can be written explicitly as

$$\begin{aligned} H(\tilde{x}) = & \left[\frac{\kappa^2 [(1-\kappa^2)^3 + 6(1-\kappa^2)^2 + 12(1-\kappa^2) + 24]}{4(3-\kappa^2)^2} \tilde{x}^2 \right. \\ & + \left. \frac{2(1-\kappa^2)^3 + 3(1-\kappa^2)^2 + 6(1-\kappa^2)}{4(3-\kappa^2)} + \frac{3}{2} \right] \frac{\phi(\frac{\kappa\tilde{x}}{\sqrt{3-\kappa^2}})}{\sqrt{\pi(3-\kappa^2)}} \Phi\left(\frac{2\kappa\tilde{x}}{\sqrt{(3-\kappa^2)(5-3\kappa^2)}}\right) \\ & + \left[\frac{\kappa^2(2-\kappa^2)}{4} \tilde{x}^2 - \frac{\kappa^2(1-\kappa^2)}{2} - 1 \right] \frac{\phi(\frac{\kappa\tilde{x}}{\sqrt{2-\kappa^2}})}{\sqrt{\pi(2-\kappa^2)}} \Phi\left(\frac{\kappa\tilde{x}}{\sqrt{(2-\kappa^2)(5-3\kappa^2)}}\right) \\ & + \left[7 - \kappa^2 + \frac{(1-\kappa^2)[3(1-\kappa^2)^2 + 12(1-\kappa^2) + 28]}{2(3-\kappa^2)} \right] \frac{\kappa\tilde{x}\phi(\sqrt{\frac{3}{5-3\kappa^2}}\kappa\tilde{x})}{2\sqrt{2\pi}(3-\kappa^2)\sqrt{5-3\kappa^2}} \\ & + \frac{\kappa^3}{2\sqrt{2}} \tilde{x}(\tilde{x}^2 - 3) \left[\Phi_{\Sigma_1}(0, \kappa\tilde{x}/\sqrt{2}) + \Phi_{\Sigma_2}(0, \kappa\tilde{x}/\sqrt{2}) \right], \end{aligned}$$

where

$$\Sigma_1 = \begin{pmatrix} \frac{3}{2} & -1 \\ -1 & \frac{3-\kappa^2}{2} \end{pmatrix}, \quad \Sigma_2 = \begin{pmatrix} \frac{3}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{2-\kappa^2}{2} \end{pmatrix}.$$

Note that for $N = 2$ and $N = 3$, we need to solve an integral over the domain (see (15)) to get $\mathbb{E}[M_u]$. Although we cannot derive the explicit form for the entire formula, this can be evaluated in applications with the help of numerical algorithms.

3.4. Isotropic unimodal mean function

We have calculated the explicit formulas assuming the mean function is a concave paraboloid. This is a very strong assumption. However, in a general setting, where the unimodal mean function is rotationally symmetric of any shape, we can apply a multivariate Taylor expansion at the peak and use the second-order approximation to estimate power. For example, suppose the shape of the mean function is proportional to a rotationally symmetric Gaussian density

$$\theta(s) = \theta_0 \exp\left(-\frac{\|s - s_0\|^2}{2\xi^2}\right), \quad (20)$$

where s_0 is the center of the mean function and ξ is the signal bandwidth. The Taylor expansion at the center is

$$\theta(s) = \theta_0 - \frac{\theta_0}{2\xi^2} \|s - s_0\|^2 + o(\|s - s_0\|^2). \quad (21)$$

When the domain size gets small, we neglect the remainder term, and use its quadratic approximation as the mean function. With quadratic mean function, it becomes convenient to use compute $\mathbb{E}[M_u]$. We will evaluate the performance of this approach for different domain sizes in the simulation study.

4. Simulations

In Section 2 above, we discussed power approximation under different scenarios. We showed the factorial moment $\mathbb{E}[M_u(M_u - 1)]$ decays faster than $\mathbb{E}[M_u]$ under some circumstances so that we can use $\mathbb{E}[M_u]$ or adjusted $\mathbb{E}[M_u]$ to approximate power. In this section, a simulation study is conducted to validate each scenario as well as visualize the power function, $\mathbb{E}[M_u]$, and adjusted $\mathbb{E}[M_u]$. Through simulation, we could also get a better sense of applying them to real data.

4.1. Paraboloidal mean function

We generate $B = 100,000$ centered, unit-variance, smooth isotropic 2D Gaussian random fields over a grid of size 50×50 pixels as $Z(s)$, each field obtained as the convolution of white Gaussian noise with a Gaussian kernel of standard deviation 5, and normalized to have unit variance. For the mean function $\mu(s)$, we use a concave paraboloid centered at $s_0 = (25, 25)$. The equation of the paraboloid is

$$\theta(s) = \theta_0 - \frac{\|s - s_0\|^2}{2\xi^2}, \quad (22)$$

Table 1
2D simulation: default value of each parameter.

Parameter	Default value
$\text{Rad}(D)$	10
ξ	7
θ_0	3

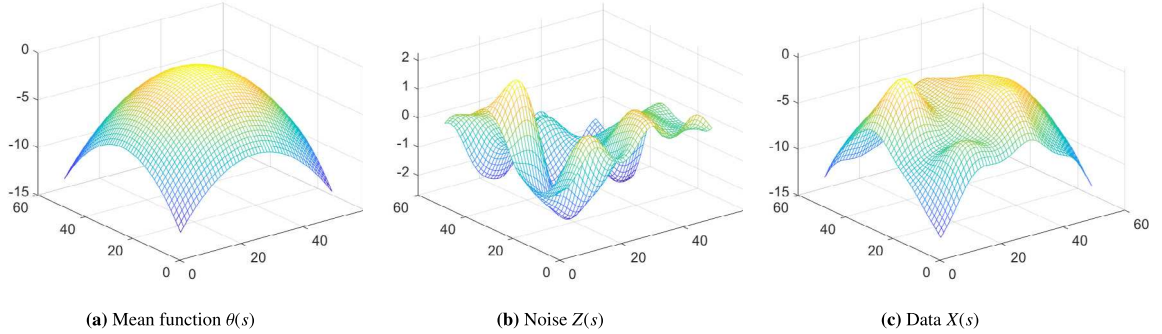


Fig. 1. 2D simulation: A single instance of the mean function $\theta(s)$, the isotropic noise field $Z(s)$ and their resulting sum $X(s) = \theta(s) + Z(s)$.

where ξ controls the sharpness of the mean function. The smaller ξ is, the sharper the paraboloid will be. θ_0 controls the height of the signal. To maintain the rotationally symmetric property of $\theta(s)$, we take closed disks centered at s_0 as domain D . The size of D is measured by the radius $\text{Rad}(D)$. The default value of each parameter is listed in Table 1 unless otherwise specified.

The first two panels of Fig. 1 display two instances of $\theta(s)$ and $Z(s)$ respectively. The third panel displays the resulting sum $X(s)$ which is calculated by the signal-plus-noise model.

In the simulation, we validate and visualize the scenarios presented above, and check the effect of different choices of parameters on the power function, $\mathbb{E}[M_u]$ and adjusted $\mathbb{E}[M_u]$. Four different scenarios are considered as we discussed in Section 2:

- (i) Height equivariance
- (ii) Small domain size $\text{Rad}(D)$
- (iii) Large threshold u
- (iv) Sharp signal (small ξ)

For each simulated random field, we record the height of its highest peak if there exists at least one, and then for any threshold u , we calculate the empirical estimate of detection power (1) and $\mathbb{E}[M_u]$:

$$\hat{F}(u) = \frac{1}{B} \sum_{i=1}^B \mathbb{1}(\exists \text{ a peak in } D \text{ with height } > u \text{ for } i\text{th simulated sample}), \quad (23)$$

$$\hat{\mathbb{E}}[M_u] = \frac{1}{B} \sum_{i=1}^B \# \{ \text{peaks in } D \text{ with height } > u \text{ for } i\text{th simulated sample} \}. \quad (24)$$

Fig. 2 displays simulated power, $\mathbb{E}[M_u]$ and adjusted $\mathbb{E}[M_u]$ curves under the four scenarios. The first panel is to validate Scenario 1 (height equivariance). As stipulated by Proposition 1, the power, $\mathbb{E}[M_u]$ and adjusted $\mathbb{E}[M_u]$ curves are parallel for different signal height h having other parameters remain the same. In the second panel, both the $\mathbb{E}[M_u]$ and adjusted $\mathbb{E}[M_u]$ curve are close to the power curve under Scenario 2 (small domain) which indicates that for a smaller domain (quantified by $\text{Rad}(D)$), using $\mathbb{E}[M_u]$ and adjusted $\mathbb{E}[M_u]$ to approximate power becomes more accurate as stipulated by Theorem 1. We can also find in all three panels that when u becomes larger, the $\mathbb{E}[M_u]$ curve converges to the power curve as u increases, as stipulated by Theorem 2. The adjusted $\mathbb{E}[M_u]$ curve also converges to the power curve but with a slower rate compared to $\mathbb{E}[M_u]$. The third panel shows the power, $\mathbb{E}[M_u]$ and adjusted $\mathbb{E}[M_u]$ curve all converge to the Gaussian CDF for sharp signal (small ξ), as stipulated by Theorem 3.

4.2. Constant mean function

When the mean function $\theta(s)$ is constant, i.e., it does not depend on location s , $X(s)$ reduces to a centered isotropic Gaussian random field. Within the context of this paper, $\theta(s) = 0$ can be seen as the null hypothesis and the power function becomes the probability of Type I error. We use the peak height distribution when $\theta(s) = 0$ [13] to decide the cutoff point such that the test meets the nominal type I error. The simulation result when $\theta(s) = 0$ is displayed in Fig. 3.

The performance of Type I error approximation when the mean function is 0 is similar to what we find when the mean function is quadratic (Scenario 4 is ignored since the shape parameter does not exist when the mean function is constant). The conclusion is that under large u (which is guaranteed in order to control the Type I error) or small domain, we have good Type I error approximation.

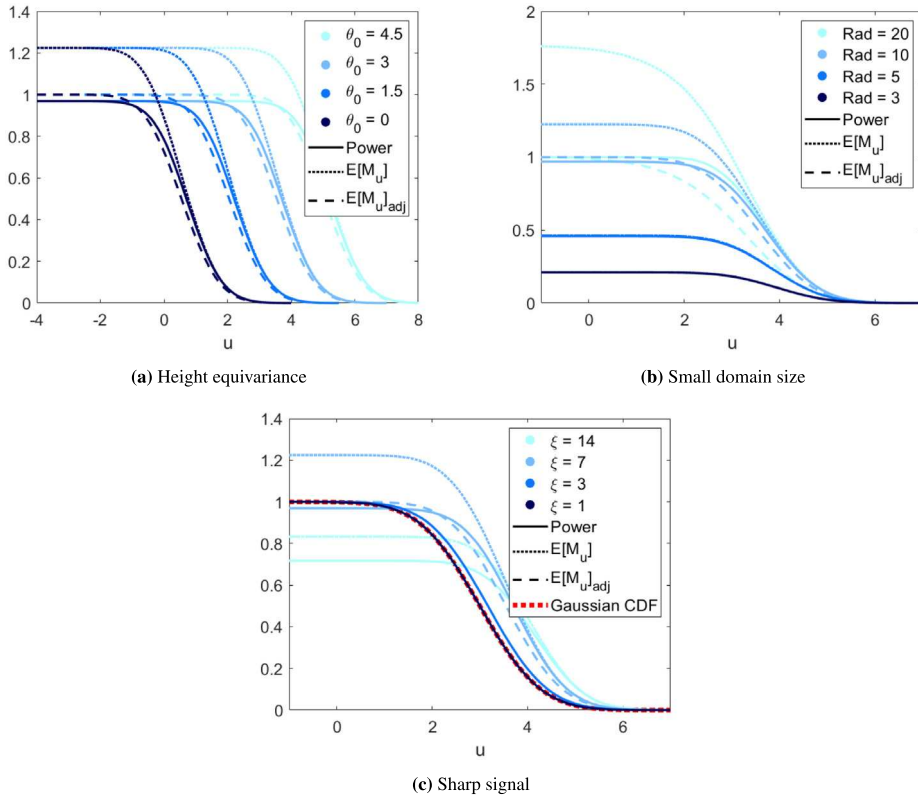


Fig. 2. 2D simulation: Simulated power (solid), $\mathbb{E}[M_u]$ (dotted) and adjusted $\mathbb{E}[M_u]$ (dashed) against the threshold u under four different scenarios (the large threshold scenario is displayed in all three panels) when the mean function is quadratic. The ordered colors represent ordered values of parameters.

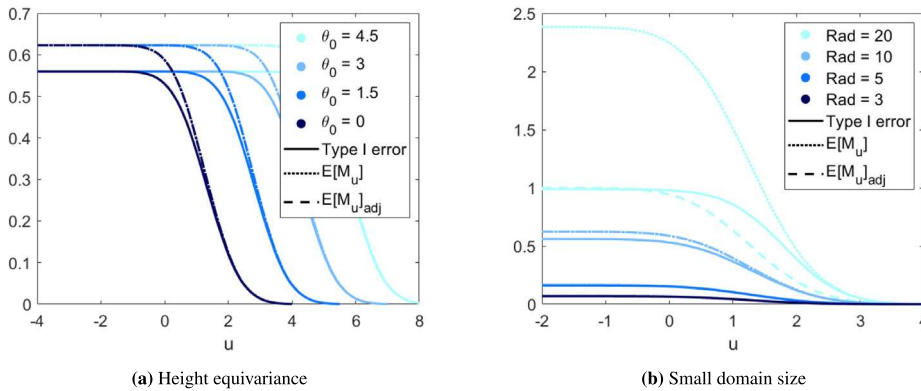


Fig. 3. 2D simulation: Simulated power (solid), $\mathbb{E}[M_u]$ (dotted) and adjusted $\mathbb{E}[M_u]$ (dashed) against the threshold u under three different scenarios (the large threshold scenario is displayed in both panels, and the sharp signal scenario is ignored since the shape parameter does not exist here) when the mean function is constant. The ordered colors represent ordered values of parameters.

4.3. Gaussian mean function

The simulation results under Gaussian mean are displayed in Fig. 4. For Scenarios 2 and 3, the results are consistent with those under quadratic mean. For Scenario 1, since θ_0 controls both the signal height and sharpness, the power, $\mathbb{E}[M_u]$ and adjusted $\mathbb{E}[M_u]$ are no longer equivariant in terms of θ_0 . For Scenario 4, if we look at Fig. 4(c) with Fig. 4(d), it can be seen that as the signal becomes sharper, the power, $\mathbb{E}[M_u]$ and adjusted $\mathbb{E}[M_u]$ curve converges to the Gaussian CDF only when the domain size (quantified by $\text{Rad}(D)$) is small. In this case, the asymptotic curve is a mixture of Gaussian CDF and $\mathbb{E}[M_u]$ under constant mean. This is due to the shape of Gaussian density as it converges to 0 if we expand the domain.

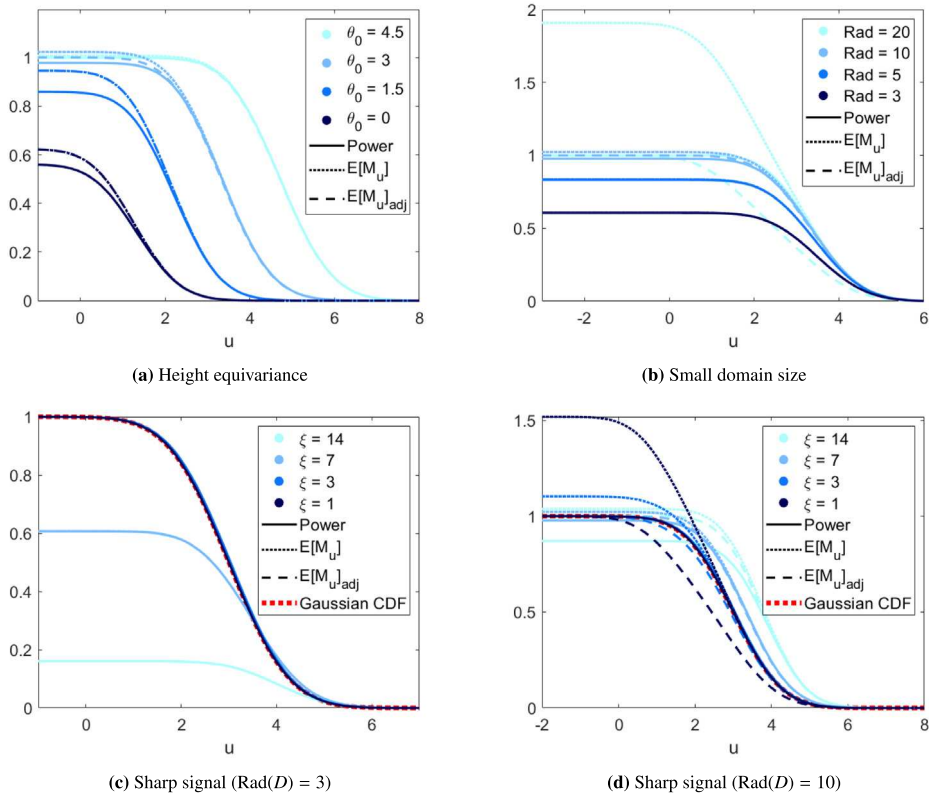


Fig. 4. 2D simulation: Simulated power (solid), $\mathbb{E}[M_u]$ (dotted) and adjusted $\mathbb{E}[M_u]$ (dashed) against the threshold u under four different scenarios (the large threshold scenario is displayed in all four panels, and the sharp signal scenario is displayed for small and large domain sizes) when the mean function is Gaussian. The ordered colors represent ordered values of parameters.

In conclusion, for Gaussian mean function, we recommend applying our method to approximate power only when the domain size is small.

5. Estimation from data

To use our power approximation formula in real peak detection problems, we need to estimate the spatial covariance function of the noise as well as the mean function from the data. In this section, we demonstrate the 3D application setting and how to estimate the noise spatial covariance function and the mean function. Consider an imaging dataset with n subjects, and let $Y_i(s)$ represent the signal plus noise for subject i ($i \in \{1, \dots, n\}$)

$$Y_i(s) = \mu(s) + \sigma(s)\varepsilon_i(s),$$

where $s = (s_1, s_2, s_3)^T \in \mathbb{R}^3$, the signal $\mu(s)$, standard deviation $\sigma(s)$ and noise $\varepsilon(s)$ are assumed to be smooth C^3 functions. If we compute the standardized mean of all n subjects, we will get a standardized random field

$$X(s) = \bar{Y}(s)/\text{SE}(\bar{Y}(s)) = \sqrt{n}\mu(s)/\sigma(s) + \sqrt{n}\bar{\varepsilon}(s), \quad (25)$$

where $\bar{Y}(s)$ and $\bar{\varepsilon}(s)$ are the mean of $Y_i(s)$ and $\varepsilon_i(s)$ over all subjects. This standardized random field $X(s)$ has constant standard deviation of 1. We can treat $\sqrt{n}\mu(s)/\sigma(s)$ as the new signal and $\sqrt{n}\bar{\varepsilon}(s)$ as the new noise of the standardized field. Let $\theta(s) = \sqrt{n}\mu(s)/\sigma(s)$ and $Z(s) = \sqrt{n}\bar{\varepsilon}(s)$. We propose using the following method to estimate the new signal and noise.

5.1. Estimation of the noise spatial covariance function

We consider the noise $Z(s)$ to be constructed by convolving Gaussian white noise with a kernel:

$$Z(s) = \int_{\mathbb{R}^N} K(t-s)dB(t), \quad (26)$$

where $K(\cdot)$ is a N dimensional kernel function, and $dB(s)$ is Gaussian white noise. Assume that the kernel is rotationally symmetric so that the noise $Z(s)$ is isotropic. Under model (26), we would be able to simulate the noise if we were able to estimate the kernel function from the data.

It can be shown that the autocorrelation of $Z(s)$ is the convolution of the kernel with itself:

$$\text{Cor}(Z(s), Z(s')) = \int_{\mathbb{R}^N} K(t-s)K(t-s')dt = \int_{\mathbb{R}^N} K(t-(s-s'))K(t)dt.$$

By the convolution theorem, convolution in the original domain equals point-wise multiplication in the Fourier-transformed domain. Thus the kernel function can be estimated empirically using the following method:

- (i) Determine a location s_0 of interest (for example, center of the peak), and calculate the empirical correlation vectors between $Y(s_0)$ and $Y(s)$ where s lies on the three orthogonal axes centered at s_0 , and belongs to a subdomain of interest.
- (ii) Take the average of the three estimated correlation vectors (forcing the noise to be isotropic) and perform Fourier transform.
- (iii) Take the square root of the Fourier coefficients, then the estimated kernel function can be obtained by performing the inverse Fourier transform.

5.2. Estimation of the mean function

Our explicit formulas are derived assuming the Hessian of the mean function is a constant times the identity matrix. Therefore, we aim to find a rotationally symmetric paraboloid $\hat{\theta}(s)$ that mimics the shape of the real mean function:

$$\theta(s) = \beta_0 + \beta_1 \|s\|^2 + (\beta_2, \beta_3, \beta_4) \cdot s, \quad (27)$$

where the dot represents the vector inner product in \mathbb{R}^3 . Note that all the quadratic terms share the same coefficient which is due to the rotational symmetry. To estimate (27), we fit a linear regression using all $X(s)$ within the subdomain as outcome. We do not consider other models because looking for the best model to estimate the mean is not the main focus of this paper and our method works no matter what type of model is used as long as the shape of the fitted curve meets our requirements.

6. A 3D real data example

As an application, we illustrate the methods in a group analysis of fMRI data from the Human Connectome Project (HCP) [14]. The data consists of the 2-back vs 0-back working memory contrast [17] from 80 unrelated subjects and is used here to get realistic 3D signal and noise parameters from which to do 3D simulations as well as evaluate the performance of our formulas in power approximation. For each subject, the size of the fMRI image is $91 \times 109 \times 91$ voxels. The mean and standard deviation of the data are displayed in Figs. 5(a) and 5(b). Regions like the dorsal and ventral prefrontal cortex are activated in the contrast [17] as demonstrated by the standardized mean in Fig. 5(c).

6.1. Data preprocessing

Gaussian kernel smoothing is applied to the dataset to make the mean function unimodal around the peak and increase the signal-to-noise ratio. The standard deviation of the smoothing kernel we use in this example is 1 voxel which translates to full width at half maximum (FWHM) being around 2.235. It is obvious from Fig. 5(b) that the standard deviation of the noise is not a constant for different locations, thus we use the transformation described in Section 5 to standardize the smoothed data before the analysis. The standardized mean of the smoothed data $X(s)$ is displayed in Fig. 5(c).

6.2. Estimation of the autocorrelation and mean functions

After standardizing the data, our next step is to estimate the mean and kernel functions using the methods described in Section 5. Here a $15 \times 15 \times 15$ subdomain (the red box in Fig. 5(c)) around the peak is taken to estimate the kernel. Since we assume the noise to be isotropic, the correlation around the peak is supposed to be the same along any dimension. However, this is not always true in real data. To tackle this, we construct an isotropic kernel that closely matches the observed kernel function derived from the data. This is achieved through the following steps. For each of the three dimensions, we first save the correlation $\text{Cor}(X(s), X(s_0))$ around the peak s_0 as a vector and create a new vector by flipping the saved correlation vector. Then we take the mean of the two vectors so that it is guaranteed to be symmetric. The empirical correlation after such symmetrization and the corresponding isotropic kernel function are displayed in Fig. 6.

We consider two approaches to estimate the mean function, nonparametric and parametric. The nonparametric mean estimation is obtained as a voxelwise average over subjects.

$$\hat{\theta}(s) = \frac{1}{n} \sum_{i=1}^n X_i(s) = \bar{X}(s). \quad (28)$$

The parametric mean estimation is obtained by fitting a linear regression model (27) using all observed data $X(s)$ within the subdomain of size $6 \times 6 \times 6$ (the blue box in Fig. 5(c)) as outcome and their corresponding location variables $\|s\|^2$, s as covariates. The least square estimate of the mean is

$$\hat{\theta}(s) = 13.03 - 0.26\|s\|^2 + (0.20, 0.11, 0.39) \cdot s. \quad (29)$$

We will compare the difference in simulated power and $\mathbb{E}[M_u]$ when the mean function is estimated by the nonparametric approach (28) vs the parametric approach (29).

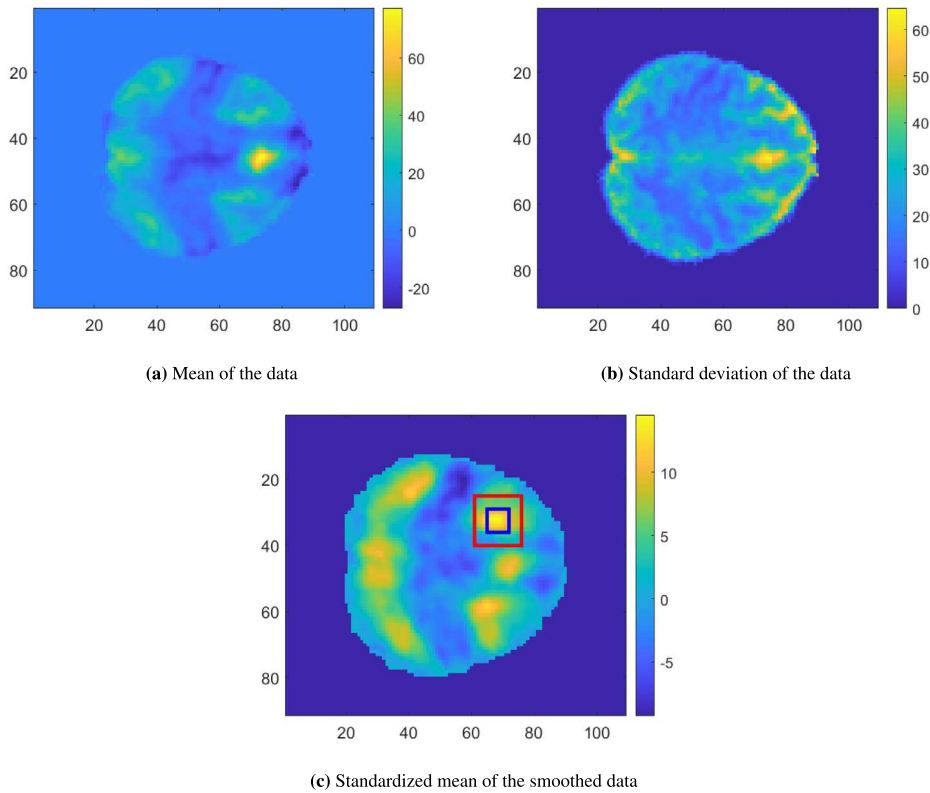


Fig. 5. HCP data: Mean, standard deviation of the data, and standardized mean of the smoothed data (transverse sliced at the peak of the image along the third dimension). The blue box represents the subdomain of the peak and the red box represents the subdomain we use to estimate the noise spatial covariance function.

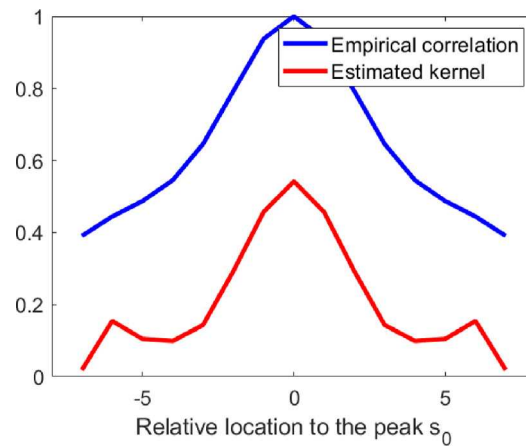


Fig. 6. Cross-section of the isotropic empirical correlation after symmetrization, and the estimated kernel from a subdomain of HCP data.

6.3. 3D simulation induced by data

We have done several simulation studies under a well-designed 2D setting where the formulas are supposed to work well, but eventually, we want to apply the formulas to real-life data which is more complicated. Besides, in terms of fMRI data analysis, the image is always 3D by nature. Considering all these factors, a 3D simulation study induced by real data is necessary to validate the performance of the formulas under a more realistic setting.

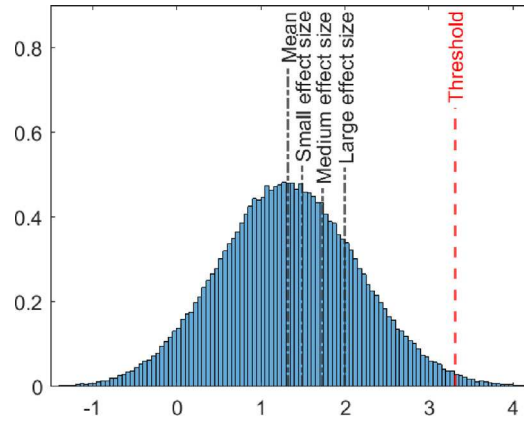


Fig. 7. 3D simulation induced by data: Simulated Peak height distribution under the null (zero mean) with different levels of effect sizes and threshold u .

In the previous two subsections, we have studied the signal and noise of the HCP data. For the simulation, we would like to generate 3D images using the estimated mean and kernel function. The noise field is generated by convolving the estimated kernel (displayed in Fig. 6) and Gaussian white noise. For each simulation setting, 10,000 such noise fields are generated.

The signal from the standardized data is very strong (see Fig. 5(c)). For illustrative purposes, we choose to weaken the signal by scaling down the estimated mean function (29) while maintaining the same shape. Here signal strength is measured by effect size, and the amount of scaling is determined by different levels of effect size. In traditional t-test or z-test, Cohen's d values of 0.2, 0.5, and 0.8 (corresponding to 0.58th, 0.69th and 0.79th quantiles of the standard Gaussian distribution) are considered as small, medium, and large effect sizes [18]. The peak height distribution under the null hypothesis (zero mean function) is displayed in Fig. 7, and does not follow a Gaussian distribution. Therefore, we take 0.58th, 0.69th, and 0.79th quantile of the null distribution minus the mean as small (0.16), medium (0.40), and large (0.65) effect size (see the black dash-dot lines in Fig. 7). For simplicity, we see the peak height of the mean function as effect size in this simulation. However, this is not the most accurate way of defining effect size in the peak detection setting. More details will be discussed in 7.2. The threshold u for peak detection is chosen as the 0.99 th quantile of the peak height distribution under the null (≈ 3.42) according to Cheng and Schwartzman [4] (see the red dashed line in Fig. 7).

Similar to the 2D simulation, the search domain D is assumed to be a sphere centered at the true peak, and we use radius of D to control the domain size. Signal sharpness is fixed since it is estimated from the data. The empirical power and $\mathbb{E}[M_u]$ are computed using (23) and (24).

To derive the explicit formulas for $\mathbb{E}[M_u]$, we assume the mean function to be a rotationally symmetric paraboloid, and this assumption might cause some bias in applications. In Fig. 8, we demonstrate the differences in simulated power and $\mathbb{E}[M_u]$ when the mean function is estimated by the nonparametric approach (28) vs the parametric approach (29). It can be observed that the quadratic approximation only has a small impact on power and $\mathbb{E}[M_u]$ in this example. In Fig. 9, we validate the theoretical formula for $\mathbb{E}[M_u]$ (15) and adjusted $\mathbb{E}[M_u]$. As we can see, the theoretical curve for $\mathbb{E}[M_u]$ and adjusted $\mathbb{E}[M_u]$ closely mirrors the empirical curve. The figure also shows that the power approximation using $\mathbb{E}[M_u]$ is accurate for large u as stipulated by Theorem 2. Power curves using three different effect sizes, and comparisons between large and small domain sizes are displayed in Fig. 10. We can see from the figure that $\mathbb{E}[M_u]$ performs quite well with small sample sizes and $\mathbb{E}[M_u]_{\text{adj}}$ performs better than $\mathbb{E}[M_u]$ with large sample sizes. When the domain size becomes smaller, we observe an improved power approximation performance from $\mathbb{E}[M_u]$, as stipulated by Theorem 1. We also use 3D simulation to compare the performance of our power approximation method versus the method proposed in Durnez et al. [11]. Fig. 10 demonstrates that our method outperforms the power approximation method proposed in Durnez et al. [11] which overly underestimates power except when the sample size is large.

7. Discussion

7.1. Explicit formulas and approximations

Calculating power for peak detection (1) has been a difficult problem in random field theory due to the lack of formula that can compute it directly. In this paper, we have discussed the rationale of using $\mathbb{E}[M_u]$ and $\mathbb{E}[M_u]_{\text{adj}}$ to approximate peak detection power under different scenarios (Section 2.4) and derived formulas to compute $\mathbb{E}[M_u]$ assuming isotropy. Isotropy is assumed so that we are able to use the GOI matrix [13] as a tool to calculate $\mathbb{E}[M_u]$ via the Kac–Rice formula.

Assuming the mean function to be a rotationally symmetric paraboloid, we further simplified the formula and defined the expectation with respect to the GOI matrix as $H(\bar{x})$ (14). We showed explicit formulas for $H(\bar{x})$ when $N \in \{1, 2, 3\}$ by applying the probability density function for the eigenvalues of GOI matrices. Then $\mathbb{E}[M_u]$ can be calculated by plugging $H(\bar{x})$ to (15). The integration in (15), however, cannot be evaluated explicitly. In practice, we may evaluate it with numerical integration tools. For higher dimensions ($N > 3$), it remains difficult to get an explicit form of $H(\bar{x})$ due to the fact inferred by Propositions 2, 3 and 4 that the integration becomes extremely complicated as N increases.

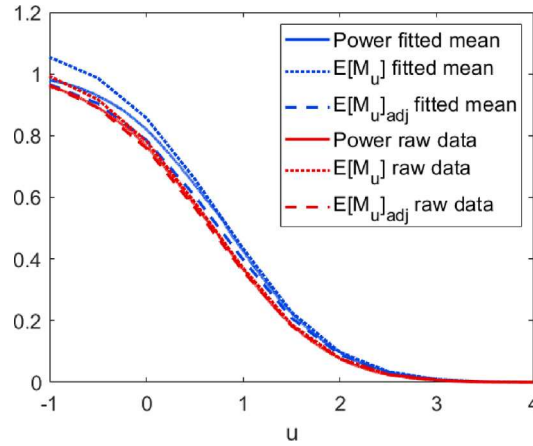


Fig. 8. 3D simulation induced by data ($\text{Rad}(D) = 3$, medium effect size): Simulated power (solid), $\mathbb{E}[M_u]$ (dotted) and adjusted $\mathbb{E}[M_u]$ (dashed) against the threshold u when the mean function is obtained from raw data vs quadratic estimation (29). The red curves represent the raw data, and the blue curves represent the quadratic estimation.

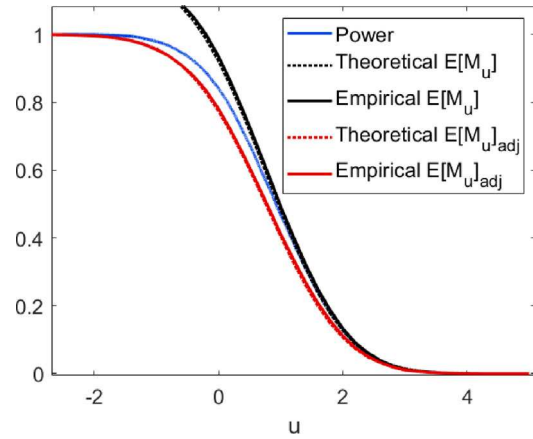


Fig. 9. 3D Simulation induced by data ($\text{Rad}(D) = 6$, medium effect size): Simulated (solid) vs theoretical (dotted) $\mathbb{E}[M_u]$ and adjusted $\mathbb{E}[M_u]$ against the threshold u . The power curve is provided for reference.

7.2. Effect size

We want to emphasize that the power depends on both the signal strength parameter θ_0 and shape parameter η . In a traditional z-test or t-test which tests a single null hypothesis that the mean value is equal to 0, the detection power depends only on a single parameter we call effect size. For our use case, the test is conditional on the point being a local maximum. Applying a simple z-test or t-test, one could reject the null hypothesis if the peak height θ_0 exceeds the pre-specified threshold. This approach is not accurate since the peak height does not follow a Gaussian or t distribution. To address this, the threshold can be determined by the null distribution of peak height [13] to control the type I error at a nominal level. However, power calculation based on the test over peak height is still biased since the true effect size depends both on the signal height and curvature. The height of the peak affects the likelihood of exceeding the threshold and the curvature affects the likelihood of existing such peak in the domain. It follows that a sharp and high peak is easier to detect compared to a flat and low peak, leading to a larger detection power.

For an interpretation of the parameter $\eta = \theta''/(-2\rho')$, we consider two types of mean function: paraboloid and Gaussian. Suppose the noise is the result of the convolution of white noise with a Gaussian kernel with spatial standard deviation v resulting in the covariance function with $\rho(r) = \exp(-r/(2v^2))$ as specified in Section 3.1. This is the same noise as we simulated in Section 5. When the mean function is paraboloid, i.e., $\theta(s) = -\|s\|^2/(2\xi^2) + \theta_0$ as in (22), we obtain $\theta'' = -1/\xi^2$ and $\rho' = -1/(2v^2)$, yielding $\eta = \theta''/(-2\rho') = -v^2/\xi^2$. Thus, η is a shape parameter representing the relative sharpness of the mean function with respect to the curvature of the noise. When the mean function is Gaussian, consider $\theta(s) = a \exp(-\|s\|^2/(2\tau^2))$. This expression is obtained, for example, if the signal is the result of the convolution of a delta function with a Gaussian kernel with spatial standard deviation τ . We obtain $\theta'' = -a/\tau^2$ and $\rho' = -1/(2v^2)$, yielding $\eta = \theta''/(-2\rho') = -av^2/\tau^2$. Thus, η is the height of the signal a , scaled by the ratio

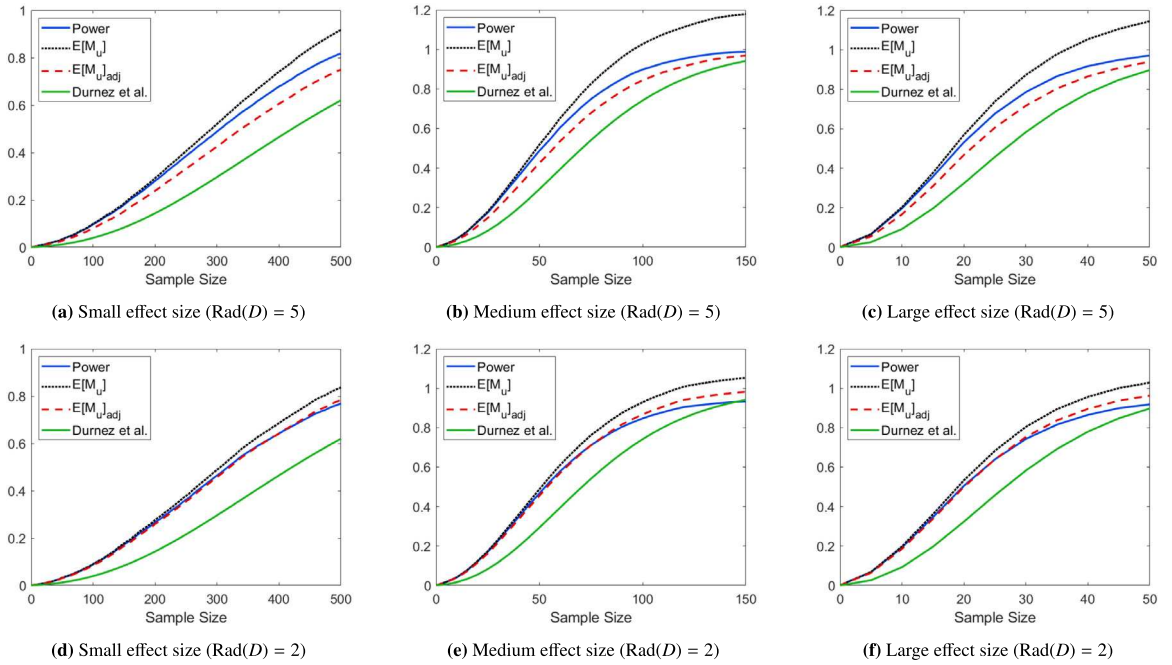


Fig. 10. 3D Simulation induced by data: Simulated power (solid), $\mathbb{E}[M_u]$ (dotted) and adjusted $\mathbb{E}[M_u]$ (dashed) against the threshold u when the signal has small, medium, and large effect sizes, and comparisons between large and small domain sizes. The green curve represents the simulated power from the method proposed in Durnez et al. [11] for performance comparison.

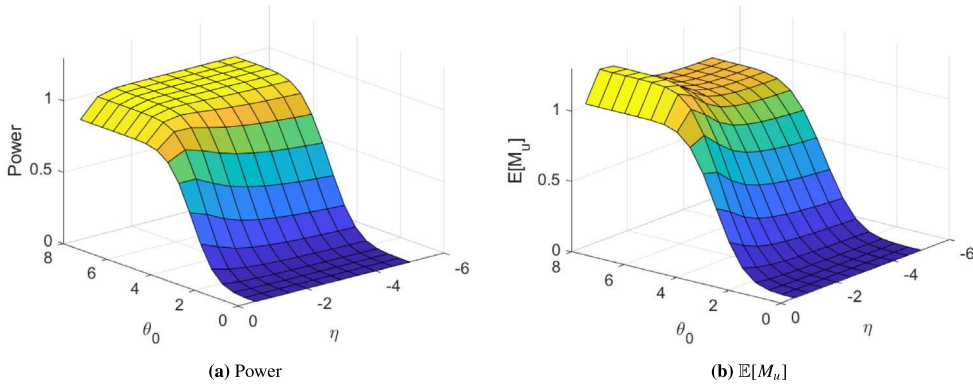


Fig. 11. 2D simulation ($u = 3.92$ and $\text{Rad}(D) = 10$): Power and $\mathbb{E}[M_u]$ for different θ_0 and η . Since θ_0 and η together determine the effect size, these can be interpreted as the plot of power and $\mathbb{E}[M_u]$ against the effect size.

of the spatial extent of the noise and signal filters. In both cases, the parameter η , and thus the power, are invariant under isotropic scaling of the domain, in a similar fashion to the peak height distribution under the null hypothesis [19].

Fig. 11 illustrates how θ_0 and η affect power and $\mathbb{E}[M_u]$ in the 2D simulation described in Section 4. As we have explained, θ_0 and η together determine the effect size. Although deriving an explicit form of effect size as a function of θ_0 and η is difficult, we are able to roughly show how the two parameters relate to power. θ_0 which can be seen as signal-to-noise ratio (SNR) plays a major role. Having η stay the same, the power monotonically increases with respect to θ_0 . On the other hand, power monotonically decreases with respect to η having θ_0 stays the same. In this simulation example, the impact of θ_0 on power is about 10 times stronger than η if we fit a linear model of power using θ_0 and η . We can also observe from the figure that the effect of η on power is stronger for large θ_0 compared to that for small θ_0 .

7.3. Application to data

Compared to simpler power approximation approaches which typically ignore the spatial correlation, our RFT-based method fully accounts for the spatial structure of the data. We demonstrate through simulation that our power approximation method outperforms

the one proposed in Durnez et al. [11] that uses the truncated normal distribution to approximate the distribution of peak height in settings that closely mirror real-world conditions.

To use our formula in practice, one needs to assume the peak to be a certain type such as paraboloid or Gaussian. While specific, this assumption is not too restrictive for the data we have analyzed. Fig. 8 shows that the estimated power using the formula under a paraboloidal assumption is not far from the power using the true peak shape.

Regarding the conjecture of $\mathbb{E}[M_u]_{\text{adj}}$ being a lower bound when there exists at least one local maximum in the domain D , it remains difficult to prove in general, but as we showed in the real data example, it seems to be correct in practice. When it comes to a real-life problem, we can take both $\mathbb{E}[M_u]$ and the $\mathbb{E}[M_u]_{\text{adj}}$ into consideration to get a better understanding of the true sample size. As demonstrated in simulation, we recommend $\mathbb{E}[M_u]$ when the sample size is small, and considering $\mathbb{E}[M_u]_{\text{adj}}$ when the sample size is large. $\mathbb{E}[M_u]_{\text{adj}}$ also gives a more conservative estimate of power compared to $\mathbb{E}[M_u]$ which is useful to guarantee that the test is powerful enough when we design future studies. Because of its difficulty, we leave further study of $\mathbb{E}[M_u]_{\text{adj}}$ for future work.

CRediT authorship contribution statement

Yu Zhao: Formal analysis, Investigation, Methodology, Project administration, Software, Validation, Visualization, Writing – original draft. **Dan Cheng:** Conceptualization, Methodology, Writing – review & editing. **Armin Schwartzman:** Conceptualization, Data curation, Funding acquisition, Investigation, Resources, Supervision, Writing – review & editing.

Acknowledgments

Y.Z., D.C. and A.S. were partially supported by NIH grant R01EB026859 and NSF grant 1811659. D.C. acknowledges also support from NSF Grants DMS-1902432, DMS-2220523 and the Simons Foundation Collaboration Grant 854127. Data were provided in part by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

Appendix. Proofs

Proof of Proposition 1. By the definitions, it is easy to show $M_{u-\theta_0} = \tilde{M}_u$, and $\mathbb{E}[M_{u-\theta_0}] = \mathbb{E}[\tilde{M}_u]$. Considering the definition of power, we have

$$\tilde{F}(u) = P[\tilde{M}_u \geq 1] = P[M_{u-\theta_0} \geq 1] = F(u - \theta_0). \quad \square$$

Proof of Theorem 1. The proof is based on the proof of Lemma 3 in Piterbarg [20] and Lemma 4.1 in Cheng and Schwartzman [12].

$$\begin{aligned} \mathbb{E}[M_u(D_\epsilon)(M_u(D_\epsilon) - 1)] &= \int_{D_\epsilon} \int_{D_\epsilon} \int_u^\infty \int_u^\infty \mathbb{E} \left[|\det \nabla^2 X(s)| |\det \nabla^2 X(t)| \frac{X(s) = x_1, X(t) = x_2}{\nabla X(s) = \nabla X(t) = 0} \right] \\ &\quad P_{X(s), X(t), \nabla X(s), \nabla X(t)}(x_1, x_2, 0, 0) dx_1 dx_2 ds dt. \end{aligned} \quad (30)$$

Let

$$E_1(s, t) = \mathbb{E} \left[|\det \nabla^2 X(s)| |\det \nabla^2 X(t)| \frac{X(s) = x}{\nabla X(s) = \nabla X(t) = 0} \right],$$

and replace one of the integration limits in (30) by $-\infty$, we have

$$\mathbb{E}[M_u(D_\epsilon)(M_u(D_\epsilon) - 1)] \leq \int_{D_\epsilon} \int_{D_\epsilon} P_{\nabla X(s), \nabla X(t)}(0, 0) ds dt \int_u^\infty E_1(s, t) P_{X(s)}(x \mid \nabla X(s) = \nabla X(t) = 0) dx.$$

Then we can take the Taylor expansion

$$\nabla X(t) = \nabla X(s) + \nabla^2 X(s)(t - s) + \|t - s\|^{1+\gamma} Y_{s,t},$$

where $Y_{s,t} = (Y_{s,t}^1, \dots, Y_{s,t}^N)^\top$ is a Gaussian vector field and γ is a positive constant. Note that the determinant of $\nabla^2 X(s)$ is equal to the determinant of

$$\begin{pmatrix} 1 & -(t_1 - s_1) & \dots & -(t_N - s_N) \\ 0 & & & \\ \vdots & & \nabla^2 X(s) & \\ 0 & & & \end{pmatrix}. \quad (31)$$

For $i \in \{2, \dots, N+1\}$, multiply the i th column of this matrix by $(t_i - s_i)/\|t - s\|^2$, take the sum of all such columns and add the result to the first column. Since $\nabla X(s) = \nabla X(t) = 0$, we can derive $\nabla^2 X(s)(t - s) = -\|t - s\|^{1+\gamma} Y_{s,t}$, and obtain the matrix below with the same determinant as (31),

$$\begin{pmatrix} 0 & -(t_1 - s_1) & \dots & -(t_N - s_N) \\ -\|t - s\|^{-1+\gamma} Y_{s,t}^1 & & & \\ \vdots & & \nabla^2 X(s) & \\ -\|t - s\|^{-1+\gamma} Y_{s,t}^N & & & \end{pmatrix}.$$

Let

$$A_{s,t} = \begin{pmatrix} 0 & -(t_1 - s_1)/\|t - s\| & \dots & -(t_N - s_N)/\|t - s\| \\ Y_{s,t}^1 & & & \\ \vdots & & \nabla^2 X(s) & \\ Y_{s,t}^N & & & \end{pmatrix}.$$

We have $|\det \nabla^2 X(s)| = \|t - s\|^\gamma |\det A_{s,t}|$, implying $E_1(s, t) = \|t - s\|^\gamma E_2(s, t)$, where

$$E_2(s, t) = \mathbb{E} \left[|\det A_{s,t}| \det \nabla^2 X(t) \middle| \begin{matrix} X(s) = x, \nabla X(s) = 0 \\ \nabla^2 X(s)(t - s) = -\|t - s\|^{1+\gamma} Y_{s,t} \end{matrix} \right].$$

Using the inequality of arithmetic and geometric means [20], we can bound the determinant

$$|\det \nabla^2 X(t)| \leq N^{2N-2} \sum_{i,j} |X_{ij}(t)|^N, \quad |\det A_{s,t}| \leq (N+1)^{2N} \sum_{i,j} |a_{ij}|^{N+1},$$

where a_{ij} is the i, j entry of $A_{s,t}$. Apply the inequality again

$$|\det \nabla^2 X(t)| |\det A_{s,t}| \leq \frac{1}{2} N^{2N-2} (N+1)^{2N+1} \left(\sum_{i,j} |X_{ij}(t)|^{2N} + \sum_{i,j} |a_{ij}|^{2N+2} \right). \quad (32)$$

For any Gaussian variable X and integer $N \geq 0$, the following inequality holds

$$\mathbb{E}[X^{2N}] \leq 2^{2N} (\mathbb{E}[X]^{2N} + C_N \text{Var}(X)^{2N}), \quad (33)$$

where C_N is a constant depending on N . Next, we can focus on the conditional expectation and conditional variance of $X_{ij}(t)$ and $Y_{s,t}$.

Using the fact that the conditional variance of a Gaussian variable is less or equal to the unconditional variance, we can use (33) to bound the conditional expectation of $|X_{ij}(t)|^{2N}$ and $|a_{ij}|^{2N+2}$. By Assumptions 1 and 2, $X_{ij}(t)$ and $Y_{s,t}$ have finite first and second moments. Therefore, the conditional expectation of the right-hand side of (32) is bounded above by some constant.

Summarizing the results above,

$$\sup_{s,t \in D_\epsilon, s \neq t} |E_2(s, t)| \leq C_1$$

for some constant $C_1 > 0$ and

$$E_1(s, t) \leq \|t - s\|^\gamma \mathbb{E}_2(s, t) \leq C_1 \|t - s\|^\gamma.$$

Combine the results above and with a fixed threshold u

$$\begin{aligned} \int_u^\infty E_1(s, t) P_{X(s)}(x \mid \nabla X(s) = \nabla X(t) = 0) dx &\leq C_1 \|t - s\|^\gamma \int_u^\infty P_{X(s)}(x \mid \nabla X(s) = \nabla X(t) = 0) dx \\ &= C_1 \|t - s\|^\gamma \int_u^\infty \exp(-(Ax - B)^2) dx = C_2 \|t - s\|^\gamma, \end{aligned}$$

for some constant A, B , and $C_2 > 0$.

Next, by the proof of Lemma 4.1 in Cheng and Schwartzman [12]

$$p_{\nabla X(s), \nabla X(t)}(0, 0) \leq C_3 \|t - s\|^{-N},$$

for some constant $C_3 > 0$.

Therefore, there exists C_4 such that

$$\mathbb{E}[M_u(D_\epsilon)(M_u(D_\epsilon) - 1)] \leq C_4 \int_{D_\epsilon} \int_{D_\epsilon} \frac{1}{\|t - s\|^{N-\gamma}} dt ds = o(\epsilon^N).$$

For $\mathbb{E}[M_u(D_\epsilon)]$, by Kac-Rice formula in Adler and Taylor [21]

$$\mathbb{E}[M_u(D_\epsilon)] = \int_{D_\epsilon} p_{\nabla X(s)}(0) \mathbb{E} \left[|\det \nabla^2 X(s)| \mathbb{1}_{\{\nabla^2 X(s) < 0\}} \mathbb{1}_{\{X(s) > u\}} \mid \nabla X(s) = 0 \right] ds.$$

Denote the integrand by $g(s)$. The function $g(s)$ is continuous and positive over the non-empty domain D_ϵ . Thus $\inf_{s \in D_\epsilon} g(s) \geq g_0 > 0$, implying

$$\mathbb{E}[M_u(D_\epsilon)] \geq g_0 \epsilon^N.$$

Then (7) is an immediate consequence of (5).

For $\mathbb{E}[M_{-\infty}(D_\epsilon)]$, by Kac–Rice formula

$$\mathbb{E}[M_{-\infty}(D_\epsilon)] = \int_{D_\epsilon} p_{\nabla X(s)}(0) \mathbb{E} \left[|\det \nabla^2 X(s)| \mathbb{1}_{\{\nabla^2 X(s) < 0\}} | \nabla X(s) = 0 \right] ds.$$

The integrand is also continuous and positive over the domain D_ϵ indicating $\mathbb{E}[M_{-\infty}(D_\epsilon)] = o(1)$ as $\epsilon \rightarrow 0$. Thus, as $\epsilon \rightarrow 0$, we have

$$\mathbb{E}[M_u(D_\epsilon)]_{\text{adj}} = \mathbb{E}[M_u(D_\epsilon)] / \max(1, \mathbb{E}[M_{-\infty}(D_\epsilon)]) = \mathbb{E}[M_u(D_\epsilon)] / \max(1, o(1)) = \mathbb{E}[M_u(D_\epsilon)]. \quad \square$$

Proof of Theorem 2. By lemma 3 of Piterbarg [20], as $u \rightarrow \infty$, the factorial moment is super-exponentially small. That means $\exists \omega > 0$ such that

$$\mathbb{E}[M_u(M_u - 1)] = o \left(e^{-\frac{u^2}{2} - \omega u^2} \right).$$

Also, by (14.0.13) of Adler and Taylor [21], as $u \rightarrow \infty$, $\exists C > 0$ such that

$$\mathbb{E}[M_u] \geq \mathbb{P}[M_u \geq 1] = F(u) \geq \mathbb{P}[\sup X(s) \geq u] \geq Cu^{N-1} e^{-\frac{u^2}{2}}.$$

Thus, the factorial moment decays exponentially faster than $\mathbb{E}[M_u]$. Combining the above two results,

$$\frac{\mathbb{E}[M_u(M_u - 1)]}{\mathbb{E}[M_u]} = o \left(\frac{e^{-\frac{u^2}{2} - \omega u^2}}{Cu^{N-1} e^{-\frac{u^2}{2}}} \right) = o(e^{-\omega u^2}).$$

Using (5) and the results above, we have

$$\mathbb{E}[M_u] - F(u) \leq \mathbb{E}[M_u(M_u - 1)] = \mathbb{E}[M_u] o(e^{-\omega u^2}). \quad \square$$

Proof of Theorem 3. By lemma A.1 of Cheng and Schwartzman [4], as $a \rightarrow \infty$

$$\mathbb{P}(M_{-\infty} = 1) \geq 1 - O(e^{-Ca^2}),$$

where $C > 0$ is some constant. Therefore $M_{-\infty} \xrightarrow{p} 1$.

Since $M_u \leq M_{-\infty}$ and both of them only take non-negative integer values, $|M_u(M_u - 1)|$ and $|M_{-\infty}(M_{-\infty} - 1)|$ are bounded above by $|M(M - 1)|$ where M is the number of critical points of the random field X . Apply Kac–Rice formula

$$\mathbb{E}[M(M - 1)] = \int_D \int_D \mathbb{E} \left[|\det \nabla^2 X(s)| |\det \nabla^2 X(t)| | \nabla X(s) = \nabla X(t) = 0 \right] P_{\nabla X(s), \nabla X(t)}(0, 0) ds dt.$$

Denote the integrand by $g(s, t, a)$. The function $g(s, t, a)$ is continuous and positive over the domain D and $M(M - 1) \xrightarrow{p} 0$ as $a \rightarrow \infty$. Thus there exists $g_0 > 0$ such that $\mathbb{E}[M(M - 1)] \leq g_0$. Then by dominated convergence theorem

$$\mathbb{E}[M_u(M_u - 1)] \rightarrow 0,$$

as $a \rightarrow \infty$. Since $M_{-\infty} \xrightarrow{p} 1$, the adjusted $\mathbb{E}[M_u]$

$$\mathbb{E}[M_u]_{\text{adj}} = \mathbb{E}[M_u] \max(1, \mathbb{E}[M_{-\infty}]) = \mathbb{E}[M_u](1 + o(1)) = \mathbb{E}[M_u] + o(1).$$

To calculate $\mathbb{E}[M_u]$, apply Kac–Rice formula

$$\begin{aligned} \mathbb{E}[M_u] &= \int_D p_{\nabla X(s)}(0) \mathbb{E} \left[|\det \nabla^2 X(s)| \mathbb{1}_{\{\nabla^2 X(s) < 0\}} \mathbb{1}_{\{X(s) > u\}} | \nabla X(s) = 0 \right] ds \\ &= \int_D p_{\nabla X(s)}(0) \mathbb{E} \left[|\det \nabla^2 X(s)| \mathbb{1}_{\{\nabla^2 X(s) < 0\}} \mathbb{1}_{\{X(s) > u\}} | \nabla X(s) = 0 \right] ds \\ &= \int_D \frac{1}{(2\pi)^{N/2} \sqrt{\det(\Lambda)}} \exp(-a^2 (\nabla h(s))^T \Lambda^{-1} \nabla h(s) / 2) \mathbb{E} \left[|\det(\nabla^2 Z(s) + a \nabla^2 h(s))| \mathbb{1}_{\{\nabla^2 X(s) < 0\}} \mathbb{1}_{\{X(s) > u\}} | \nabla X(s) = 0 \right] ds, \end{aligned} \quad (34)$$

where Λ is the covariance matrix of $\nabla h(s)$. Let $f(s) = (\nabla h(s))^T \Lambda^{-1} \nabla h(s) / 2$ which attains its minimum 0 only at s_0 . Similar to the proof of A.4 in Cheng and Schwartzman [4], as $a \rightarrow \infty$, (34) can be approximated by applying Laplace's method

$$\mathbb{E}[M_u] = \frac{\det(a \nabla^2 h(s_0))}{(2\pi)^{N/2} \sqrt{\det(\Lambda)}} \left(\frac{(2\pi)^N \det(\Lambda)}{a^{2N} \det(\nabla^2 h(s_0))} \right)^{1/2} \Phi(\theta_0 - u) + O(a^{-2}) = \Phi(\theta_0 - u) + O(a^{-2}).$$

This finishes the proof. \square

Proof of Lemma 3. Part (i) is a direct consequence of Lemma 2. For part (ii), note that $(\nabla^2 X(s) | X(s) = x)$ is equivalent to $(\nabla^2 Z(s) | Z(s) = x - \theta(s)) + \nabla^2 \theta(s)$, and the result follows immediately from Lemma 2. \square

Proof of Theorem 4. By the Kac–Rice formula and Lemma 1

$$\begin{aligned}
 \mathbb{E}[M_u] &= \int_D p_{\nabla X(s)}(0) \mathbb{E} \left[|\det \nabla^2 X(s)| \mathbb{1}_{\{\nabla^2 X(s) < 0\}} \mathbb{1}_{\{X(s) > u\}} | \nabla X(s) = 0 \right] ds \\
 &= \int_D p_{\nabla Z(s) + \nabla \theta(s)}(0) \mathbb{E} \left[|\det \nabla^2 X(s)| \mathbb{1}_{\{\nabla^2 X(s) < 0\}} \mathbb{1}_{\{X(s) > u\}} | \nabla X(s) = 0 \right] ds \\
 &= \int_D \frac{1}{(2\pi)^{N/2} (-2\rho')^{N/2}} \exp \left(-\frac{\|\nabla \theta(s)\|^2}{4\rho'} \right) \mathbb{E} \left[|\det \nabla^2 X(s)| \mathbb{1}_{\{\nabla^2 X(s) < 0\}} \mathbb{1}_{\{X(s) > u\}} | \nabla X(s) = 0 \right] ds \\
 &= \int_D \frac{(8\rho''^2)^{N/2}}{(2\pi)^{N/2} (-2\rho')^{N/2}} \exp \left(-\frac{\|\nabla \theta(s)\|^2}{4\rho'} \right) \int_u^\infty \phi(x - \theta(s)) \mathbb{E} \left[|\det(\text{Matrix}(s))| \mathbb{1}_{\{\text{Matrix}(s) < 0\}} \right] dx ds \\
 &= \int_D \left(\frac{2\rho''}{-\pi\rho'} \right)^{N/2} \exp \left(-\frac{\|\nabla \theta(s)\|^2}{4\rho'} \right) \int_u^\infty \phi(x - \theta(s)) \mathbb{E} \left[|\det(\text{Matrix}(s))| \mathbb{1}_{\{\text{Matrix}(s) < 0\}} \right] dx ds \\
 &= \left(\frac{2\rho''}{-\pi\rho'} \right)^{N/2} \int_D \exp \left(-\frac{\|\nabla \theta(s)\|^2}{4\rho'} \right) \int_u^\infty \phi(x - \theta(s)) \mathbb{E} \left[|\det(\text{Matrix}(s))| \mathbb{1}_{\{\text{Matrix}(s) < 0\}} \right] dx ds.
 \end{aligned}$$

Next, we show the derivation from the third to the fourth line in the equation above. Since we assume $\nabla^2 \theta(s)$ is a non-singular matrix at all critical points, then there exists an orthonormal matrix, denoted by $A(s)$, such that $A(s)^\top \nabla^2 \theta(s) A(s) = \text{diag}\{\theta_1''(s), \theta_2''(s), \dots, \theta_N''(s)\}$, where $\theta_1'' \leq \dots \leq \theta_N''(s)$ are ordered eigenvalues of $\nabla^2 \theta(s)$. On the other hand, GOI matrices are invariant under orthonormal transformations. By Lemma 3, the conditional expectation $\mathbb{E}[|\det(\nabla^2 X(s))| \mathbb{1}_{\{\nabla^2 X(s) < 0\}} | X(s) = x]$ is therefore

$$\begin{aligned}
 &\mathbb{E} \left[\left| \det \left(\sqrt{8\rho''} \left[G - \frac{\kappa(x - \theta(s))}{\sqrt{2}} I_N \right] + \nabla^2 \theta(s) \right) \right| \mathbb{1}_{\{\text{Matrix}(s) < 0\}} \right] \\
 &= \mathbb{E} \left[\left| \det \left(\sqrt{8\rho''} \left[G - \frac{\kappa(x - \theta(s))}{\sqrt{2}} I_N \right] + A(s)^\top \nabla^2 \theta(s) A(s) \right) \right| \mathbb{1}_{\{\text{Matrix}(s) < 0\}} \right] \\
 &= (\sqrt{8\rho''})^N \mathbb{E} \left[\left| \det \left(\left[G - \frac{\kappa(x - \theta(s))}{\sqrt{2}} I_N \right] + A(s)^\top \nabla^2 \theta(s) A(s) / \sqrt{8\rho''} \right) \right| \mathbb{1}_{\{\text{Matrix}(s) < 0\}} \right] \\
 &= (\sqrt{8\rho''})^N \mathbb{E} \left[\left| \det \left(G - \frac{\kappa(x - \theta(s))}{\sqrt{2}} I_N + \text{diag}\{\theta_1''(s), \dots, \theta_N''(s)\} / \sqrt{8\rho''} \right) \right| \mathbb{1}_{\{\text{Matrix}(s) < 0\}} \right]. \quad \square
 \end{aligned}$$

Proof of Proposition 2. Since we assume that $Z(s)$ is stationary, $Z'(s)$ is independent of $Z(s)$ and $Z''(s)$, and $\rho' = -\text{Var}(Z'(s))/2 = \mathbb{E}[Z(s)Z''(s)]/2$ and $\rho'' = \text{Var}(Z''(s))/12$ do not depend on s . Therefore,

$$\text{Var}(X(s)) = 1, \quad \text{Var}(X'(s)) = -\text{Cov}(X(s)X''(s)) = -2\rho', \quad \text{Var}(X''(s)) = 12\rho''.$$

Note that, by the formula of conditional Gaussian distributions,

$$X''(s) | X(s) = x \sim N(\theta''(s) + 2\rho'(x - \theta(s)), 12\rho'' - 4\rho'^2).$$

By the Kac–Rice formula

$$\begin{aligned}
 \mathbb{E}[M_u] &= \int_D p_{X'(s)}(0) \mathbb{E} \left[|X''(s)| \mathbb{1}_{\{X(s) > u\}} \mathbb{1}_{\{X''(s) < 0\}} | X'(s) = 0 \right] ds \\
 &= \int_D p_{X'(s)}(0) \int_u^\infty \phi(x - \theta(s)) \int_{-\infty}^0 (-x'') \frac{1}{\sqrt{12\rho'' - 4\rho'^2}} \phi \left(\frac{x'' - \theta''(s) - 2\rho'(x - \theta(s))}{\sqrt{12\rho'' - 4\rho'^2}} \right) dx'' dx ds \\
 &= \int_D p_{X'(s)}(0) \sqrt{12\rho'' - 4\rho'^2} \int_u^\infty \phi(x - \theta(s)) \psi \left(\frac{-2\rho'(x - \theta(s)) - \theta''(s)}{\sqrt{12\rho'' - 4\rho'^2}} \right) dx ds \\
 &= \int_D \frac{1}{\sqrt{-2\rho'}} \phi \left(\frac{\theta'(s)}{\sqrt{-2\rho'}} \right) \sqrt{12\rho'' - 4\rho'^2} \int_u^\infty \phi(x - \theta(s)) \psi \left(\frac{-2\rho'(x - \theta(s)) - \theta''(s)}{\sqrt{12\rho'' - 4\rho'^2}} \right) dx ds \\
 &= \int_D \frac{\sqrt{12\rho'' - 4\rho'^2}}{\sqrt{-2\rho'}} \phi \left(\frac{\theta'(s)}{\sqrt{-2\rho'}} \right) \int_u^\infty \phi(x - \theta(s)) \psi \left(\frac{-2\rho'(x - \theta(s)) - \theta''(s)}{\sqrt{12\rho'' - 4\rho'^2}} \right) dx ds.
 \end{aligned}$$

The second to third line is due to the fact that

$$\int_{-\infty}^0 (-x) \frac{1}{b} \phi \left(\frac{x+a}{b} \right) dx = \int_{-\infty}^0 \Phi \left(\frac{x+a}{b} \right) dx = b \int_{-\infty}^{a/b} \Phi(y) dy = b\psi \left(\frac{a}{b} \right).$$

Recall the κ (10) and η (13) parameters defined before. We can rewrite $\mathbb{E}[M_u]$ as

$$\int_D \frac{\sqrt{-2\rho'(3 - \kappa^2)}}{\kappa} \phi \left(\frac{\theta'(s)}{\sqrt{-2\rho'}} \right) \int_u^\infty \phi(x - \theta(s)) \psi \left(\frac{\kappa[x - \theta(s) - \eta(s)]}{\sqrt{3 - \kappa^2}} \right) dx ds.$$

This finishes the proof. \square

Proof of Proposition 3. By Lemma 4 above with $a = (\kappa^2 - 1)/(4 - 2\kappa^2)$ and $b = \kappa\tilde{x}/\sqrt{2}$, and Lemma 2.2 in Cheng and Schwartzman [13],

$$\begin{aligned} & \mathbb{E}_{\text{GOI}((1-\kappa^2)/2)}^N \left[\prod_{j=1}^N \left| \lambda_j - \frac{\kappa\tilde{x}}{\sqrt{2}} \right| I \left(\lambda_N < -\frac{\kappa\tilde{x}}{\sqrt{2}} \right) \right] \\ &= \frac{1}{2\sqrt{\pi(2-\kappa^2)}} \int_{\mathbb{R}^2} \exp \left(-\frac{1}{2}(\lambda_1^2 + \lambda_2^2) + \frac{1-\kappa^2}{4(2-\kappa^2)}(\lambda_1 + \lambda_2)^2 \right) (\lambda_2 - \lambda_1) \left| \lambda_1 - \frac{\kappa\tilde{x}}{\sqrt{2}} \right| \left| \lambda_2 - \frac{\kappa\tilde{x}}{\sqrt{2}} \right| \mathbb{1}_{\{\lambda_2 < \frac{\kappa\tilde{x}}{\sqrt{2}}\}} d\lambda_1 d\lambda_2 \\ &= \frac{1}{\sqrt{3-\kappa^2}} \exp \left(-\frac{\kappa^2\tilde{x}^2}{2(3-\kappa^2)} \right) \Phi \left(\frac{\kappa\tilde{x}}{\sqrt{(2-\kappa^2)(3-\kappa^2)}} \right) + \frac{\kappa^2}{2} (\tilde{x}^2 - 1) \Phi \left(\frac{\kappa\tilde{x}}{\sqrt{2-\kappa^2}} \right) + \frac{\kappa\sqrt{2-\kappa^2}\tilde{x}}{2\sqrt{2\pi}} \exp \left(-\frac{\kappa^2\tilde{x}^2}{2(2-\kappa^2)} \right). \end{aligned}$$

This simplifies to (19). \square

Proof of Proposition 4. This is a direct result of Lemma 5 above with $a = 1 - \kappa^2$ and $b = \kappa\tilde{x}/\sqrt{2}$. \square

References

- [1] C.R. Genovese, N.A. Lazar, T. Nichols, Thresholding of statistical maps in functional neuroimaging using the false discovery rate, *Neuroimage* 15 (4) (2002) 870–878.
- [2] R. Heller, D. Stanley, D. Yekutieli, N. Rubin, Y. Benjamini, Cluster-based analysis of fMRI data, *Neuroimage* 33 (2) (2006) 599–608.
- [3] S. Hayasaka, A.M. Peiffer, C.E. Hugenschmidt, P.J. Laurienti, Power and sample size calculation for neuroimaging studies by non-central random field theory, *Neuroimage* 37 (3) (2007) 721–730.
- [4] D. Cheng, A. Schwartzman, Multiple testing of local maxima for detection of peaks in random fields, *Ann. Statist.* 45 (2) (2017) 529–556.
- [5] A. Schwartzman, F. Telschow, Peak p-values and false discovery rate inference in neuroimaging, *NeuroImage* 197 (2019) 402–413.
- [6] J.E. Desmond, G.H. Glover, Estimating sample size in functional MRI (fMRI) neuroimaging studies: Statistical power analyses, *J. Neurosci. Methods* 118 (2) (2002) 115–128.
- [7] S. Geuter, G. Qi, R.C. Welsh, T.D. Wager, M.A. Lindquist, Effect size and power in fMRI group analysis, 2018, <http://dx.doi.org/10.1101/295048>, bioRxiv.
- [8] J. Suckling, J. Henty, C. Ecker, S.C. Deoni, M.V. Lombardo, S. Baron-Cohen, P. Jezzard, A. Barnes, B. Chakrabarti, C. Ooi, M.-C. Lai, S.C. Williams, D.G.M. Murphy, E. Bullmore, MRC AIMS Consortium, Are power calculations useful? A multicentre neuroimaging study, *Hum. Brain Mapp.* 35 (8) (2014) 3569–3577.
- [9] D. Szucs, J.P. Ioannidis, Sample size evolution in neuroimaging research: An evaluation of highly-cited studies (1990–2012) and of latest practices (2017–2018) in high-impact journals, *Neuroimage* 221 (2020) 117164.
- [10] B.O. Turner, E.J. Paul, M.B. Miller, A.K. Barbey, Small sample sizes reduce the replicability of task-based fMRI studies, *Commun. Biol.* 1 (1) (2018) 62.
- [11] J. Durnez, J. Degryse, B. Moerkerke, R. Seurinck, V. Sochat, R.A. Poldrack, T.E. Nichols, Power and sample size calculations for fMRI studies based on the prevalence of active peaks, 2016, <http://dx.doi.org/10.1101/049429>, bioRxiv.
- [12] D. Cheng, A. Schwartzman, Distribution of the height of local maxima of Gaussian random fields, *Extremes* 18 (2015) 213–240.
- [13] D. Cheng, A. Schwartzman, Expected number and height distribution of critical points of smooth isotropic Gaussian random fields, *Bernoulli* 24(4B) (2018) 3422–3446.
- [14] D.C. Van Essen, K. Ugurbil, E. Auerbach, D. Barch, T.E.J. Behrens, R. Bucholz, A. Chang, L. Chen, M. Corbetta, S.W. Curtiss, S. Della Penna, D. Feinberg, M.F. Glasser, N. Harel, A.C. Heath, L. Larson-Prior, D. Marcus, G. Michalareas, S. Moeller, R. Oostenveld, S.E. Petersen, F. Prior, B.L. Schlaggar, S.M. Smith, A.Z. Snyder, J. Xu, E. Yacoub, WU-Minn HCP Consortium, The Human Connectome Project: a data acquisition perspective, *Neuroimage* 62 (4) (2012) 2222–2231.
- [15] H.V. Henderson, S.R. Searle, Vec and vech operators for matrices, with some uses in Jacobians and multivariate statistics, *Canad. J. Statist.* 7 (1) (1979) 65–81.
- [16] A. Schwartzman, W. Mascarenhas, J. Taylor, Inference for eigenvalues and eigenvectors of Gaussian symmetric matrices, *Ann. Statist.* 36 (6) (2008) 2886–2919.
- [17] D.M. Barch, G.C. Burgess, M.P. Harms, S.E. Petersen, B.L. Schlaggar, M. Corbetta, M.F. Glasser, S. Curtiss, S. Dixit, C. Feldt, D. Nolan, E. Bryant, T. Hartley, O. Footer, J.M. Bjork, R. Poldrack, S. Smith, H. Johansen-Berg, A.Z. Snyder, D.C. Van Essen, WU-Minn HCP Consortium, Function in the human connectome: task-fMRI and individual differences in behavior, *Neuroimage* 80 (2013) 169–189.
- [18] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, L. Erlbaum Associates, 1988.
- [19] D. Cheng, A. Schwartzman, On critical points of Gaussian random fields under diffeomorphic transformations, *Statist. Probab. Lett.* 158 (2020) 108672.
- [20] V.I. Piterbarg, Rice's method for large excursions of Gaussian random fields, Technical Report NO 478, Center for Stochastic Processes, Univ. North Carolina, 1996.
- [21] R.J. Adler, J.E. Taylor, *Random Fields and Geometry*, Springer, New York, 2007.