MULTIPLE TESTING OF LOCAL MAXIMA FOR DETECTION OF PEAKS IN RANDOM FIELDS

Dan Cheng, **Armin Schwartzman**
Division of Biostatistics, University of California, San Diego

Abstract

A topological multiple testing scheme is presented for detecting peaks in images under stationary ergodic Gaussian noise, where tests are performed at local maxima of the smoothed observed signals. The procedure generalizes the one-dimensional scheme of [31] to Euclidean domains of arbitrary dimension. Two methods are developed according to two different ways of computing p-values: (i) using the exact distribution of the height of local maxima, available explicitly when the noise field is isotropic [9, 10]; (ii) using an approximation to the overshoot distribution of local maxima above a pre-threshold, applicable when the exact distribution is unknown, such as when the stationary noise field is non-isotropic [9]. The algorithms, combined with the Benjamini-Hochberg procedure for thresholding p-values, provide asymptotic strong control of the False Discovery Rate (FDR) and power consistency, with specific rates, as the search space and signal strength get large. The optimal smoothing bandwidth and optimal pre-threshold are obtained to achieve maximum power. Simulations show that FDR levels are maintained in non-asymptotic conditions. The methods are illustrated in the analysis of functional magnetic resonance images of the brain.

AMS 2000 subject classifications

Primary 62H35; secondary 62H15

Keywords and phrases

false discovery rate; Gaussian random field; kernel smoothing; image analysis; overshoot distribution; selective inference; topological inference

1. Introduction.

Detection of sparse localized signals embedded in smooth noise is a fundamental problem in image analysis, with applications in many scientific areas such as neuroimaging [40, 20, 38],

Dan Cheng, Division of Biostatistics University of California, San Diego 9500 Gilman Dr. La Jolla, CA 92093, dcheng2@ncsu.edu,
Armin Schwartzman, Division of Biostatistics University of California, San Diego 9500 Gilman Dr. La Jolla, CA 92093, armins@ucsd.edu.

microscopy [16, 19] and astronomy [7]. The key issue is to find a threshold to determine significant regions. This paper treats the thresholding problem as a multiple testing problem where tests are performed at local maxima of the observed image, allowing error rates and detection power to be topologically and geometrically defined in terms of detected spatial peaks, rather than pixels or voxels.

In neuroimaging, Keith Worsley pioneered the use of random field theory, especially the Euler characteristic heuristic, to approximate the null distribution of the global maximum of the observed image to control the family-wise error rate (FWER) of detected voxels [40, 41, 38]. On the other hand, initial attempts to control the false discovery rate (FDR), desirable for being less conservative, ignored the spatial structure in the data [20, 24]. Recognizing the need to make inferences about connected regions rather than voxels in imaging applications, multiple testing methods have since been developed for pre-defined regions [21, 5, 36] and for the harder problem of detecting unknown clusters [25, 26, 43]. It has been argued, however, that localized signal regions often present themselves as peaks in the image intensity profile, inviting a more powerful analysis based on local maxima of the observed data as the features of interest [28, 13, 14].

Schwartzman et al. [31] formalized peak detection by introducing a multiple testing paradigm where local maxima of the smoothed data are tested for significance. That work, was limited to one-dimensional spatial and temporal domains because the distribution of the height of local maxima, a key ingredient for calculation of p-values, has historically been known in closed-form only for one-dimensional stationary Gaussian processes [15]. Recently, however, Cheng and Schwartzman [9, 10] have obtained exact expressions for the height distribution of local maxima of isotropic Gaussian fields and an approximation to the overshoot distribution of local maxima of constant-variance Gaussian fields by applying techniques from random matrices theory [18]. These crucial developments allow us in the current paper to extend the multiple testing method of [31] to Euclidean domains of higher dimension.

Our "smoothing and testing of maxima" (STEM) algorithm consists of the following steps:

1. *Kernel smoothing*: to increase the signal-to-noise ratio (SNR) [39, 34].

2. *Candidate peaks*: find local maxima of the smoothed field above a pre-threshold.

3. *P-values*: computed at each local maximum under the null hypothesis of no signal in a local neighborhood.

4. *Multiple testing*: apply a multiple testing procedure and declare as detected peaks those local maxima whose p-values are significant.

The main conceptual difference with the algorithm of Schwartzman et al. [31], in addition to being multi-dimensional, is the introduction of a height pre-threshold in step 2. Pre-thresholding is often used in neuroimaging to reduce the number of candidate peaks or regions [43]. Considered formally here, it leads to two ways of applying the above algorithm. If the exact distribution of the height of local maxima for computing p-values in step 3 is known, such as for isotropic fields [9, 10], it is shown here that it is best not to apply pre-thresholding at all. However, if the height distribution is unknown, as is the case to

date for non-isotropic fields, then pre-thresholding is still valuable in that it enables the use of an approximation of the overshoot distribution of local maxima instead [9]. In step 4, for concreteness, we focus on the Benjamini-Hochberg (BH) procedure [6] for controlling FDR, although other procedures and error criteria could be used. The algorithm is illustrated by a toy example in Figure 1.

Following the reasoning of Schwartzman et al. [31], it is shown here that if the noise field is stationary and ergodic, then the proposed algorithm with the BH procedure provides asymptotic control of FDR and power consistency as both the search domain and the signal strength get large, the latter needing to grow only faster than the square root of the log of the former. The large domain assumption helps resolve an interesting aspect of inference for local maxima, namely the fact that the number of tests, equal to the number of observed local maxima, is random. The multiple testing literature usually assumes that the number of tests is fixed (an exception in a similar random processes setting is Siegmund et al. [32]). The large domain assumption implies that, by ergodicity, the number of tests behaves asymptotically as its expectation. On the other hand, the strong signal assumption asymptotically eliminates the false positives caused by the smoothed signal spreading into the null regions, causing each signal peak region to be represented by only one observed local maximum within the true domain with probability tending to one. Simulations show that FDR levels are maintained and high power is achieved at finite search domains and moderate signal strength. We also find that the optimal smoothing kernel is approximately that which is closest in shape and bandwidth to the signal peaks to be detected, akin to the matched filter theorem in signal processing [29, 33]. This bandwidth is much larger than the usual optimal bandwidth in nonparametric regression.

The results in this paper supercede those of Schwartzman et al. [31] in the sense that the latter can be seen as special cases when the domain is of one dimension and no pre-thresholding is applied. This paper provides specific rates for the asymptotic results, not available in Schwartzman et al. [31], as well as a more rigorous discussion of the optimal smoothing bandwidth. In addition, a multiscale procedure is proposed that allows searching over a finite set of bandwidths, eliminating the need to select a single optimal bandwidth. On the contrary, the procedure is adaptive to peaks of different spatial extents so that power is increased without compromising error control.

A similar idea on multiple testing of local maxima was also proposed by Chumbley et al. [14], which we shall refer to as CWFF following the authors' names. The statistical inference there, however, is unclear, as there is no formal definition of false positives (the signal is assumed to be nonzero over the entire domain) and no argument for error control. Moreover, p-values for local maxima are computed using an approximate formula for the overshoot distribution based on the expected Euler characteristic, requiring both a pre-threshold and a pre-selected domain (see Eqs. (1) and (2) in [14]). Our work is more rigorous not only in terms of error control, but also in using the exact height distribution of local maxima for computing p-values in isotropic Gaussian fields and a more accurate approximation of the overshoot distribution. Our overshoot distribution does not require a pre-selected domain because it only depends on local properties of the field [9]. Furthermore, the pre-threshold is optimized to maximize power.

In a broader statistical context, the STEM algorithm falls in the general category of *selective inference*, where inference is performed after selection [17, 22]. Similar to other high dimensional data problems, here too the number of tests is reduced by selecting particular features observed in the data, namely the observed local maxima. To be valid, the inference must take into account the selection process. This is done here by using the proper height distribution of local maxima [9, 10] mentioned above. A related p-value com-putation for the global maximum, also using the Kac-Rice formula, is given in [37].

Following the neuroimaging motivation, we show in this paper how inference for local maxima via the STEM algorithm can be used to find areas of brain activity in functional magnetic resonance imaging (fMRI). Specifically, we use a public fMRI data set from an experiment whose goal is to find brain regions that relate to inference of other people's mental states, in particular false beliefs about reality [23]. Following the usual fMRI analysis protocol, a linear regression is fitted at each voxel and then the STEM algorithm is used to make inferences about the contrast of interest as a random field.

The proofs for all the lemmas and theorems in this paper can be found in the Online Supplementary Material [12]. The data analysis and all simulations were implemented in Matlab.

2. The multiple testing scheme.

2.1. The model.

Consider the signal-plus-noise model

$$y(t) = \mu(t) + z(t), \quad t \in \mathbb{R}^N, \tag{1}$$

where the signal $\mu(t)$ is composed of unimodal positive peaks of the form

$$\mu(t) = \sum_{j=-\infty}^{\infty} a_j h_j(t), \quad a_j > 0,$$

and the peak shape $h_j(t) \geq 0$ has compact connected support $S_j = \{ t: h_j(t) > 0 \}$ and unit action $\int_{S_j} h_j(t)dt = 1$ for each j. Let $w(t) \geq 0$ be a unimodal kernel with compact connected support, mode at the origin and unit action, and let $w_\gamma(t) = w(t/\gamma)/\gamma^N$ with bandwidth parameter $\gamma > 0$. Convolving the process (1) with the kernel $w_\gamma(t)$ results in the smoothed random field

$$y_\gamma(t) = w_\gamma(t) * y(t) = \int_{\mathbb{R}^N} w_\gamma(t-s)y(s)ds = \mu_\gamma(t) + z_\gamma(t), \tag{2}$$

where the smoothed signal and smoothed noise are defined as

$$\mu_\gamma(t) = w_\gamma(t) * \mu(t) = \sum_{j=-\infty}^{\infty} a_j h_{j,\gamma}(t), \quad z_\gamma(t) = w_\gamma(t) * z(t). \tag{3}$$

The smoothed noise $z_\gamma(t)$ defined by (2) and (3) is assumed to be a zero-mean thrice differentiable stationary Gaussian field such that for any non-negative integers $k_1,\ldots,k_N$ with $\sum_{i=1}^{N} k_i = k \in \{0,1,2,3,4\}$,

$$
\int_{\mathbb{R}_+^N} \left| \frac{\partial^k r_\gamma(t)}{\partial t_1^{k_1} \cdots \partial t_N^{k_N}} \right| dt < \infty, \tag{4}
$$

where $\mathbb{R}_+^N = [0,\infty)^N$ and $r_\gamma(t) = \mathrm{E}\big[z_\gamma(t) z_\gamma(0)\big]$. The technical condition (4) is needed for obtaining the rates of FDR control and power consistency below, and by taking $k = 0$, it implies the ergodicity of $z_\gamma(t)$ [15]. It requires that the derivatives of the covariance function of the smoothed field $z_\gamma(t)$ should not decay too slowly. This can be easily obtained by using a Gaussian kernel $w_\gamma(t)$ in (2), regardless of the smoothness of the original noise.

For each j, the smoothed peak shape $h_{j,\gamma}(t) = w_\gamma(t) * h_j(t) \geq 0$ is uni-modal and has compact connected support $S_{j,\gamma}$ and unit action. For each j, we require that $h_{j,\gamma}(t)$ is twice differentiable in the interior of $S_{j,\gamma}$ and has no other critical points within its support. For simplicity, the theory requires that the supports $S_{j,\gamma}$ do not overlap although this is not crucial in practice.

Let $\tau_{j,\gamma} \in S_j$ be the unique point where the peak shape $h_{j,\gamma}(t)$ attains its maximum. We impose the following uniformity assumptions on the signal:

1. $\sup_j |S_{j,\gamma}| < \infty$ and $\inf_j M_{j,\gamma} > 0$, where $M_{j,\gamma} = h_{j,\gamma}(\tau_{j,\gamma})$.

2. There exists a universal $\delta > 0$ such that $I_{j,\gamma}^{\mathrm{mode}} := \big\{ t \in \mathbb{R}^N : \|t - \tau_{j,\gamma}\| \leq \delta \big\} \subset S_j$ for all j, $C_\gamma = \inf_j C_{j,\gamma} > 0$ and $D_\gamma = \inf_j D_{j,\gamma} > 0$, where

$$
C_{j,\gamma} = \inf_{t \in I_{j,\gamma}^{\mathrm{side}}} \| \nabla h_{j,\gamma}(t) \|, \quad I_{j,\gamma}^{\mathrm{side}} = S_{j,\gamma} \backslash I_{j,\gamma}^{\mathrm{mode}},
$$

$$
D_{j,\gamma} = - \sup_{t \in I_{j,\gamma}^{\mathrm{mode}}} \sup_{\|x\| = 1} x^T \nabla^2 h_{j,\gamma}(t) x,
$$

Here ∇f and $\nabla^2 f$ denote respectively the gradient and Hessian of a function f.

Assumption (1) indicates that the sizes of the supports $S_{j,\gamma}$ are bounded and that the heights of the peaks of $h_{j,\gamma}$ are uniformly positive. In assumption (2), the condition on $C_{j,\gamma}$ indicates that, uniformly for all j, $h_{j,\gamma}(t)$ has no critical points outside a ball $I_{j,\gamma}^{\mathrm{mode}}$ centered at its mode. The condition on $D_{j,\gamma}$ indicates that, uniformly for all j, $h_{j,\gamma}(t)$ is strictly concave within $I_{j,\gamma}^{\mathrm{mode}}$; the quantity $\sup_{\|x\|=1} x^T \nabla^2 h_{j,\gamma}(t) x$ is the largest (negative) eigenvalue of the matrix $\nabla^2 h_{j,\gamma}(t)$. These conditions allow for a large variety of unimodal functions which need not be concave within their support and whose level curves need not be convex. Figure 2 shows an example.

2.2. The STEM algorithm.

Suppose we observe $y(t)$ defined by (1) in the cube of length L centered at the origin, denoted by $U(L) = (-L/2, L/2)^N$, and suppose it contains J peaks. We call the following procedure STEM (Smoothing and TEsting of Maxima).

Algorithm 1 (STEM algorithm).

1. *Kernel smoothing*: Construct the field (2), ignoring the effects on the boundary of $U(L)$.

2. *Candidate peaks*: For a fixed pre-threshold $v \in [-\infty, \infty)$, find the set of local maxima exceeding level v for $y_\gamma(t)$ in $U(L)$

$$\tilde{T}_\gamma(v) = \left\{ t \in U(L) : y_\gamma(t) > v, \nabla y_\gamma(t) = 0, \nabla^2 y_\gamma(t) \prec 0 \right\}, \tag{5}$$

    where $\nabla^2 y_\gamma(t) \prec 0$ means that the Hessian matrix is negative definite.

3. *P-values*: For each local maximum $t \in t \in \tilde{T}_\gamma(v)$ with observed height $y_\gamma(t)$, compute the p-value $p_\gamma(t, v)$ for testing

$$\mathcal{H}_0(t) : \left\{ \exists \delta_0 > 0 \text{ such that } \mu(s) = 0 \text{ for all } s \in B(t, \delta_0) \right\} \quad \text{vs.}$$
$$\mathcal{H}_A(t) : \left\{ \forall \delta_0 > 0, \mu(s) > 0 \text{ for some } s \in B(t, \delta_0) \right\} \tag{6}$$

    where $B(t, \delta_0)$ is a ball of radius δ_0 centered at t.

4. *Multiple testing*: Let $\tilde{m}_\gamma(v) = \# \left\{ t \in \tilde{T}_\gamma(v) \right\}$ be the number of tested hypotheses. Apply a multiple testing procedure on the set of $\tilde{m}_\gamma(v)$ p-values $\left\{ p_\gamma(t, v), t \in \tilde{T}_\gamma(v) \right\}$, and declare significant all local maxima whose p-values are smaller than the significance threshold.

Steps 1 and 2 above are well defined under the model assumptions. In the context of selective inference, Step 2 is the selection step, such that only the random locations in (5) observed to attain local maxima are selected for testing. The computation of valid p-values in Step 3 accounts for this selection by using the proper conditional distribution, as detailed in Sections 3 and 4 below. For step 4, we use the BH procedure to control FDR (Section 3.2). Notice that, in contrast to the usual BH procedure, the number of tests $\tilde{m}_\gamma(v)$ is random.

When $v = -\infty$, we regard $\tilde{T}_\gamma = \tilde{T}_\gamma(-\infty)$ as the set of local maxima of $y_\gamma(t)$ in $U(L)$. In such case, Algorithm 1 becomes an N-dimensional version of the STEM algorithm proposed in [31] for one-dimensional domains. When $v > -\infty$, an option not available in [31], Algorithm 1 provides a different way of selecting candidate peaks and computing p-values by choosing a pre threshold v. In particular, this provides an effcient way to approximate the p-values for stationary and non-isotropic Gaussian noise (Section 4).

2.3. Error definitions.

As in [31], because the location of truly detected peaks may shift as a result of noise, a significant local maximum is called a true positive if it falls anywhere inside the support of a true peak; otherwise, it is called a false positive. This definition is consistent with (6).

Define the *signal region* $\mathbb{S}_1 = \cup_{j=1}^{J} S_j$ and *null region* $\mathbb{S}_0 = U(L)\backslash\mathbb{S}_1$. For a significance threshold u above the pre-threshold v, the total number of detected peaks and the number of falsely detected peaks are

$$R_\gamma(u) = \#\left\{t \in \tilde{T}_\gamma(u)\right\} \quad \text{and} \quad V_\gamma(u) = \#\left\{t \in \tilde{T}_\gamma(u) \cap \mathbb{S}_0\right\} \tag{7}$$

respectively. Both are defined as zero if $\tilde{T}_\gamma(u)$ is empty. The FDR is defined as the expected proportion of falsely detected peaks

$$\text{FDR}_\gamma(u) = \text{E}\left\{\frac{V_\gamma(u)}{R_\gamma(u) \vee 1}\right\}. \tag{8}$$

Kernel smoothing enlarges the signal support and increases the probability of obtaining false positives in the null regions neighboring the signal [26]. Define the *smoothed signal region* $\mathbb{S}_{1,\gamma} = \cup_{j=1}^{J} S_{j,\gamma} \supset \mathbb{S}_1$ and *smoothed null region* $\mathbb{S}_{0,\gamma} = U(L)\backslash\mathbb{S}_{1,\gamma} \subset \mathbb{S}_0$ We call the difference between the expanded signal support and the true signal support the *transition region* $\mathbb{T}_\gamma = \mathbb{S}_{1,\gamma}\backslash\mathbb{S}_1 = \mathbb{S}_0\backslash\mathbb{S}_{0,\gamma} = \cup_{j=1}^{J} T_{j,\gamma}$, where $T_{j,\gamma} = S_{j,\gamma} \mid S_j$ is the transition region corresponding to each peak j. While the transition region does not appear explicitly in the theorems below, it does play a crucial role in their proofs (online supplementary material).

In general, more than one significant local maximum may be obtained within the domain of a true peak, affecting the interpretation of definition (8). However, this has no effect asymptotically because, as shown in the proof of Theorem 3 below, each true peak is represented by exactly one local maximum of the smoothed observed field with probability tending to 1.

2.4. Power.

We define the power of Algorithm 1 as the expected fraction of true discovered peaks

$$\text{Power}_\gamma(u) = \text{E}\left(\frac{1}{J}\sum_{j=1}^{J} \mathbb{1}\left\{\tilde{T}_\gamma(u) \cap S_j \neq \varnothing\right\}\right) = \frac{1}{J}\sum_{j=1}^{J} \text{Power}_{j,\gamma}(u), \tag{9}$$

where $\text{Power}_{j,\gamma}(u)$ is the probability of detecting peak j

$$\text{Power}_{j,\gamma}(u) = \text{P}\left(\tilde{T}_\gamma(u) \cap S_j \neq \varnothing\right). \tag{10}$$

The indicator function in (9) ensures that only one significant local maximum is counted within the same peak support, so power is not inflated. Again, this has no effect asymptotically because each true peak is represented by exactly one local maximum of the smoothed observed process with probability tending to 1.

3.   Detection of peaks by the height distribution of local maxima.

3.1.   P-values.

For any local maximum $t \in \tilde{T}_\gamma(v)$ with observed height $y_\gamma(t)$, its p-value in step 3 of Algorithm 1 is the probability, under the null hypothesis (6), that such a height is observed given that t has been selected as an observed local maximum. As recognized in selective inference [17, 22, 37], the p-value must be computed as a conditional probability given the selection event. Specifically, the p-value is $p_\gamma(t, v) = F_\gamma(y_\gamma(t), v), t \in \tilde{T}_\gamma(v)$, where

$$F_\gamma(u, v) = P\big(z_\gamma(t) > u \,\big|\, t \in \tilde{T}_\gamma(v)\big) \tag{11}$$

denotes the right tail probability of the smoothed noise field $z_\gamma(t)$ at the local maximum $t \in \tilde{T}_\gamma(v)$. By convention, when $v = -\infty$, denote

$$F_\gamma(u) = F_\gamma(u, -\infty). \tag{12}$$

As shown in [9], the distribution (11) depends only on the local properties of the field z_γ at t. Therefore, the probability (11) is the same as that under the complete null model $\mu(s) = 0$ for all $s \in U(L)$, simplifying the calculations.

The conditional distribution (11) is a Palm distribution [3, Ch. 6] and requires careful evaluation because the conditioning event has probability zero. Unlike the marginal distribution of $z_\gamma(t)$, it is not Gaussian but stochastically greater. Generally, for a constant-variance Gaussian field, there is an implicit formula for $F_\gamma(\cdot, v)$ [9]. Theorem 2 below ([9], [3]) provides an implicit formula for $F_\gamma(\cdot, v)$ for stationary Gaussian fields.

Theorem 2. *Suppose the assumptions of* Section 2.1 *hold and that* $\mu(t) = 0, \; \forall t$. *Let* $\tilde{m}_{0,\gamma}(U(1), u)$ *denote the number of local maxima of* $z_\gamma(t)$ *exceeding level* u *in the unit cube* $U(1) = (-1/2, 1/2)^N$, *and let* $\tilde{m}_{0,\gamma}(U(1)) = \tilde{m}_{0,\gamma}(U(1), -\infty)$ *be the number of local maxima of* $z_\gamma(t)$ *in* $U(1)$. *Then the distributions* (11) *and* (12) *can be expressed as*

$$F_\gamma(u, v) = \frac{E\big[\tilde{m}_{0,\gamma}(U(1), u)\big]}{E\big[\tilde{m}_{0,\gamma}(U(1), v)\big]} \quad \text{and} \quad F_\gamma(u) = \frac{E\big[\tilde{m}_{0,\gamma}(U(1), u)\big]}{E\big[\tilde{m}_{0,\gamma}(U(1))\big]}. \tag{13}$$

Let $\sigma_\gamma^2 = \mathrm{Var}(z_\gamma(t))$ and $\Lambda_\gamma = \mathrm{Cov}(\nabla z_\gamma(t))$, both independent of t due to the stationarity of $z_\gamma(t)$. By the Kac-Rice formula [2], the expectations in (13) can be computed as

$$E\big[\tilde{m}_{0,\gamma}(U(1), u)\big] = E\Big[\big|\det \nabla^2 z_\gamma(t)\big| \mathbb{1}_{\{z_\gamma(t) > u\}} \mathbb{1}_{\{\nabla^2 z_\gamma(t) \prec 0\}} \,\Big|\, \nabla z_\gamma(t) = 0\Big] p_{\nabla z_\gamma(t)}(0), \tag{14}$$

where $p_{\nabla z_\gamma(t)}(0) = (2\pi)^{-N/2}\big(\det(\Lambda_\gamma)\big)^{-1/2}$ is the density function of $\nabla z_\gamma(t)$ evaluated at 0.

The expectations above involving the indicator function $\mathbb{1}_{\{\nabla^2 z_\gamma(t) \prec 0\}}$ are extremely hard to compute and thus the explicit formula for F_γ is usually unknown. The only exception, as far as we know, is the case when the field z_γ is isotropic. This is because, in such case, one may apply the Gaussian Orthogonal Ensemble (GOE) technique from random matrices theory to

compute these expectations [18]. The corresponding computable formula for F_γ for isotropic Gaussian fields was recently obtained in [9, Theorem 2.8] and [10]. In particular, when the dimension of the parameter space N is low, say $N \leq 3$, an explicit formula for F_γ is also available in [11]. This will be used to compute the p-values exactly, see Proposition 6 below. We emphasize, however, that the generic form of the distributions (13) is sufficient in order to show error control and power consistency of the STEM algorithm, as described next.

3.2. Error control.

For a fixed height threshold u, step 4 of Algorithm 1 amounts to selecting those local maxima with height greater than u. To control FDR at a fixed level α, the BH procedure is applied in step 4 of Algorithm 1, as follows. For a fixed $\alpha \in (0, 1)$, let k be the largest index for which the ith smallest p-value is less than $i\alpha/\widetilde{m}_\gamma(v)$. Then the null hypothesis $\mathcal{H}_0(t)$ at $t \in \widetilde{T}_\gamma(v)$ is rejected if

$$p_\gamma(t, v) < \frac{k\alpha}{\widetilde{m}_\gamma(v)} \quad \Leftrightarrow \quad y_\gamma(t) > \tilde{u}_{\mathrm{BH}, \gamma}(v) = F_\gamma(\,\cdot\,, v)^{-1}\!\left(\frac{k\alpha}{\widetilde{m}_\gamma(v)}\right), \tag{15}$$

where $k\alpha/\widetilde{m}_\gamma(v)$ is defined as 1 if $\widetilde{m}_\gamma(v) = 0$. Since $\tilde{u}_{\mathrm{BH}, \gamma}(v)$ is random, definition (8) is hereby modified to

$$\mathrm{FDR}_{\mathrm{BH}, \gamma}(v) = \mathrm{E}\left\{\frac{V_\gamma\big(\tilde{u}_{\mathrm{BH}, \gamma}(v)\big)}{R_\gamma\big(\tilde{u}_{\mathrm{BH}, \gamma}(v)\big) \vee 1}\right\}, \tag{16}$$

where $R_\gamma(\cdot)$ and $V_\gamma(\cdot)$ are defined in (7) and the expectation is taken over all possible realizations of the random threshold $\tilde{u}_{\mathrm{BH}, \gamma}(v)$.

Define the conditions:

(C1) The assumptions of Section 2.1 hold.

(C2) $L^N \to \infty$ and $a = \inf_j a_j \to \infty$, such that $\big(\log L^N\big)/a^2 \to 0$, $J/L^N = A_1 + O\big(a^{-2} + L^{-N/2}\big)$ and $|\mathbb{S}_{1,\gamma}|/L^N = A_{2,\gamma} + O\big(a^{-2} + L^{-N/2}\big)$ with $A_1 > 0$ and $A_{2,\gamma} \in [0,1)$.

The conditions on A_1 and $A_{2,\gamma}$ in (C2) above ensure that, as the search volume L^N grows, the signal does not asymptotically vanish nor it covers the entire search space.

Theorem 3. *Let conditions (C1) and (C2) hold.*

  i.   *Suppose that* Algorithm 1 *is applied with a fixed threshold $u > v$, then*

$$\mathrm{FDR}_\gamma(u) \leq \frac{\mathrm{E}\big[\widetilde{m}_{0,\gamma}(U(1), u)\big]\big(1 - A_{2,\gamma}\big)}{\mathrm{E}\big[\widetilde{m}_{0,\gamma}(U(1), u)\big]\big(1 - A_{2,\gamma}\big) + A_1} + O\big(a^{-2} + L^{-N/2}\big). \tag{17}$$

  ii.   *Suppose that* Algorithm 1 *is applied with the random threshold $\tilde{u}_{\mathrm{BH}, \gamma}(v)$ (15), then*

$$\text{FDR}_{\text{BH},\gamma}(v) \le \alpha \frac{\text{E}\big[\tilde{m}_{0,\gamma}(U(1),v)\big]\big(1 - A_{2,\gamma}\big)}{\text{E}\big[\tilde{m}_{0,\gamma}(U(1),v)\big]\big(1 - A_{2,\gamma}\big) + A_1} + O\big(a^{-1} + L^{-N/4}\big). \tag{18}$$

The proof of Theorem 3 is based on the fact that, as $L^N \to \infty$, the weak law of large numbers for ergodic fields guarantees that the number of observed local maxima above v is asymptotically proportional to the expectation $\text{E}\big[\tilde{m}_{0,\gamma}(U(1),v)\big]$ in (17) and (18) (as shown in the Supplementary Material, proving this with the right rates requires estimating the order of the variance of the number of local maxima of a stationary random field, a result previously known only for $N = 1$ [27]). Meanwhile, as $a \to \infty$, the number of local maxima within each true peak converges to 1 in probability, yielding the term A_1 in the denominators.

The proof of Theorem 3 is also based on the fact that, when the BH algorithm is used, the threshold (15) can be viewed as the smallest solution of the equation $\alpha \tilde{G}(u, v) \ge F_\gamma(u, v)$, where $\tilde{G}(u, v)$ is the empirical right cumulative distribution function of $y_\gamma(t), t \in \tilde{T}_\gamma(v)$ [20]. Thus, as $L \to \infty$, the random threshold $\tilde{u}_{\text{BH},\gamma}(v)$ converges asymptotically to the deterministic threshold

$$u^*_{\text{BH},\gamma}(v) = F_\gamma^{-1}\left(\frac{\alpha A_1 \text{E}\big[\tilde{m}_{0,\gamma}(U(1),v)\big]/\text{E}\big[\tilde{m}_{0,\gamma}(U(1))\big]}{A_1 + \text{E}\big[\tilde{m}_{0,\gamma}(U(1),v)\big]\big(1 - A_{2,\gamma}\big)(1 - \alpha)} \right). \tag{19}$$

Note that this is a strictly increasing function in v.

It can be seen from the proof of Theorem 3 that the inequalities in (17) and (18) become equalities asymptotically (without specific rates), so the bounds given in (17) and (18) are tight and can be regarded respectively as the asymptotic estimators of $\text{FDR}_\gamma(u)$ and $\text{FDR}_{\text{BH},\gamma}(v)$.

3.3. Power consistency.

Similar to the definition of $\text{FDR}_{\text{BH},\gamma}$ (16), since $\tilde{u}_{\text{BH},\gamma}(v)$ is random, define

$$\text{Power}_{\text{BH},\gamma}(v) = \text{E}\left(\frac{1}{J} \sum_{j=1}^{J} \mathbb{1}\big\{ \tilde{T}_\gamma\big(\tilde{u}_{\text{BII},\gamma}(v)\big) \cap S_j \ne \theta \big\} \right). \tag{20}$$

Since $\tilde{u}_{\text{BH},\gamma}(v)$ converges to the deterministic threshold $u^*_{\text{BH},\gamma}(v)$, which attains the minimum at $v = -\infty$, we see that the power is asymptotically maximized at $v = -\infty$ when γ is fixed. Intuitively, this occurs because pre-thresholding excludes certain true peaks below the threshold which can no longer be detected, yielding less power. This phenomenon will be reflected in the simulation studies below (Figure 4) and it implies that, if the exact height distribution of local maxima $F_\gamma(\cdot, v)$ or $F_\gamma(\cdot)$ is known, for example the smoothed noise z_γ is an isotropic Gaussian field, then we will choose to apply the original STEM algorithm without pre-thresholding (i.e., $v = -\infty$) to perform the test.

The following lemma provides an asymptotic approximation to the power at a fixed threshold.

Lemma 4. *Let conditions (C1) and (C2) hold. As $a_j \to \infty$, the power for peak j (10) can be approximated by*

$$\text{Power}_{j,\gamma}(u) = \Phi\left(\frac{a_j h_{j,\gamma}(\tau_{j,\gamma}) - u}{\sigma_\gamma}\right)\left(1 + O\left(a_j^{-2}\right)\right). \tag{21}$$

The next result indicates that the BH procedure is asymptotically consistent.

Theorem 5. *Let conditions (C1) and (C2) hold.*

i. *Suppose that* Algorithm 1 *is applied with a fixed threshold $u > v$, then*

$$\text{Power}_\gamma(u) = 1 - O\left(a^{-2}\right).$$

ii. *Suppose that* Algorithm 1 *is applied with the random threshold $\tilde{u}_{\text{BH},\gamma}(v)$ (15), then*

$$\text{Power}_{\text{BH},\gamma}(v) = 1 - O\left(a^{-2} + L^{-N/2}\right).$$

The proofs of Lemma 4 and Theorem 5 above are a consequence of the fact mentioned before that, as $a \to \infty$, the number of local maxima within each true peak converges to 1 in probability. Power consistency can be sustained if $L^N \to \infty$ as long as $(\log L^N)/a^2 \to 0$ (condition C2). However, if $L^N \to \infty$ faster, then it can be seen from the proofs that the power may be bounded above by a constant strictly less than one.

3.4. Optimal smoothing kernel.

The best smoothing kernel $w_\gamma(t)$ is that which maximizes the detection power under the true model. By Lemma 4, the power (21) is approximately maximized by maximizing the signal-to-noise ratio (SNR)

$$\text{SNR}_{j,\gamma} = \frac{a_j h_{j,\gamma}(\tau_{j,\gamma})}{\sigma_\gamma} = \frac{a_j \int_{\mathbb{R}^N} w_\gamma(s) h_j(s) ds}{\sigma \sqrt{\int_{\mathbb{R}^N} w_\gamma^2(s) ds}}, \tag{22}$$

where σ is the standard deviation of the observed process $y(t)$. The smoothing kernel $w_\gamma(t)$ that maximizes (22) is called a matched filter in signal processing [29, 33]. It is known in signal processing that if the peak locations are known, then the matched filter maximizes the detection power exactly. As shown in the simulations, the result only holds approximately in our case because the peak locations are unknown.

3.5. Isotropic Gaussian fields.

Explicit expressions for the distribution (11) are known in the special cases when $N \leq 3$ and $z_\gamma(t)$ is isotropic. They are obtained from [11] by standardizing the field in (11) as $F_\gamma(u,v) = P\left(z_\gamma(t)/\sigma_\gamma > u/\sigma_\gamma \mid t \in \tilde{T}_\gamma(v)\right)$ and are given by the following proposition.

Proposition 6. *Let the assumptions in Theorem 2 hold and let $z_\gamma(t)$ be an isotropic Gaussian field over $\mathbb{R}^N$ with correlation function*

$$\rho_\gamma\left(\|t-s\|^2\right) = \mathrm{E}\left[z_\gamma(t)z_\gamma(s)\right]/\sigma_\gamma^2.$$

Let $\kappa_\gamma = -\rho_\gamma'/\sqrt{\rho_\gamma''}$, where $\rho_\gamma' = \rho_\gamma'(0)$ and $\rho_\gamma'' = \rho_\gamma''(0)$. Then the distributions (12) *and* (11) *are given respectively by*

$$F_\gamma(u) = \int_{u/\sigma_\gamma}^\infty g_\gamma\left(x;\kappa_\gamma\right)dx \quad \text{and} \quad F_\gamma(u,v) = F_\gamma(u)/F_\gamma(v),$$

where $g_\gamma(x;\kappa_\gamma)$ is the density of the height distribution of local maxima for the standardized field $z_\gamma(t)/\sigma_\gamma$, explicitly formulated in [11] *for $N \le 3$.*

Also, explicit formulae for $\mathrm{E}\left[\tilde{m}_{0,\gamma}(U(1),u)\right]$ and $\mathrm{E}\left[\tilde{m}_{0,\gamma}(U(1))\right]$ and $\mathrm{E}\left[\tilde{m}_{0,\gamma}(U(1))\right]$ (14), which will be used in our simulations below for evaluating (19), the theoretical FDR and the theoretical power, can be found in [11] for $N \le 3$.

From a practical standpoint, if z_γ is isotropic, then as shown in [9],

$$\rho_\gamma' = -\frac{1}{2\sigma_\gamma^2}\mathrm{Var}\left(\frac{\partial z_\gamma(t)}{\partial t_i}\right), \quad \rho_\gamma'' = \frac{1}{12\sigma_\gamma^2}\mathrm{Var}\left(\frac{\partial^2 z_\gamma(t)}{\partial t_i^2}\right) \tag{23}$$

for any $i=1,\ldots,N$. Therefore, in oreder to estimate $\kappa_\gamma = -\rho_\gamma'/\sqrt{\rho_\gamma''}$, we only need to estimate the variances of derivatives of z_γ (or equivalently y_γ)

Example 7 (**Gaussian autocorrelation model**). Let the noise $z(t)$ in (1) be constructed as

$$z(t) = \sigma\int_{\mathbb{R}^N}\frac{1}{v^N}\phi_N\left(\frac{t-s}{v}\right)dB(s), \quad \sigma,v>0,$$

where $\phi_N(x) = (2\pi)^{-N/2}e^{-\|x\|^2/2}$ for all $x \in \mathbb{R}^N$ is the N-dimensional standard Gaussian density, $dB(s)$ is Gaussian white noise and $v>0$ ($z(t)$ is regarded by convention as Gaussian white noise when $v=0$). Convolving with a Gaussian kernel $w_\gamma(t) = (1/\gamma^N)\,\phi_N(t/\gamma)$ with $\gamma>0$ as in (3) produces a zero-mean infinitely differentiable stationary ergodic Gaussian field

$$z_\gamma(t) = w_\gamma(t)*z(t) = \sigma\int_{\mathbb{R}^N}\frac{1}{\xi^N}\phi_N\left(\frac{t-s}{\xi}\right)dB(s), \quad \xi = \sqrt{\gamma^2+v^2},$$

with $\sigma_\gamma^2 = \sigma^2/\left(2^N\pi^{N/2}\xi^N\right), \rho_\gamma' = -(2\xi)^{-2}, \rho_\gamma'' = (2\xi)^{-4}$ and $\kappa_\gamma = 1$. The above expressions may be used as approximations if the kernel, required to have finite support, is truncated at $[-\gamma d, \gamma d]^N$ for moderately large d, say $d=3$.

Suppose the signal peak j is a truncated Gaussian density

$$h_j(t) = \left(1/b_j^N\right)\phi_N\big[(t - \tau_j)/b_j\big]\mathbb{1}\left\{(t - \tau_j)/b_j \in \left[-c_j, c_j\right]^N\right\}, \quad b_j, c_j > 0$$

Ignoring the truncation, $h_{j,\gamma}(t) = w_\gamma(t) * h_j(t)$ in (22) is the convolution of two Gaussian densities with variances γ^2 and b_j^2, which is another Gaussian density with variance $\gamma^2 + b_j^2$. We have that

$$\text{SNR}_{j,\gamma} = \frac{a_j h_{j,\gamma}(\tau_j)}{\sigma_\gamma} = \frac{a_j}{\sigma\pi^{N/4}}\left[\frac{\gamma^2 + v^2}{\left(\gamma^2 + b_j^2\right)^2}\right]^{N/4}.$$

As a function of γ, the SNR is maximized at

$$\underset{\gamma}{\text{argmax}}\,\text{SNR}_{j,\gamma} = \begin{cases} \sqrt{b_j^2 - 2v^2}, & v < b_j/\sqrt{2} \\ 0, & v > b_j/\sqrt{2}. \end{cases} \tag{24}$$

In particular, when $v = 0$, we have that the optimal bandwidth for peak j is $\gamma = b_j$, the same as the signal bandwidth. We show in the simulations below that the optimal γ is indeed close to (24). It can be seen from (24) that as v gets larger, which means that $y(t)$ gets smoother, the optimal γ becomes smaller. If v is large enough, there is no need to smooth at all.

3.6. Multiscale testing procedure.

The optimal bandwidth γ may be hard to specify in practice. A data-driven procedure for selecting the optimal bandwidth was suggested in Schwartzman et al. [31]. However, because it depends on the shape and extent of the signal peak h_j, optimal detection of different signal peaks within the same image may require different values of γ. To avoid the problem of exact bandwidth selection and handle peaks of different widths, we consider a multiscale testing procedure as follows.

Let $\Gamma = \{\gamma_i\}_{1 \le i \le |\Gamma|}$ be a set of bandwidths, where $|\Gamma|$ denotes the length of Γ. Apply Algorithm 1 for each $\gamma_i \in \Gamma$ separately, and define the combined criteria

$$\text{FDR}_\Gamma(u) = \mathrm{E}\left\{\frac{\sum_{i=1}^{|\Gamma|} V_{\gamma_i}(u)}{\left(\sum_{i=1}^{|\Gamma|} R_{\gamma_i}(u)\right) \vee 1}\right\},$$

$$\text{FDR}_{\text{BH},\Gamma}(v) = \mathrm{E}\left\{\frac{\sum_{i=1}^{|\Gamma|} V_{\gamma_i}(\tilde{u}_{\text{BH},\gamma}(v))}{\left(\sum_{i=1}^{|\Gamma|} R_{\gamma_i}(\tilde{u}_{\text{BH},\gamma}(v))\right) \vee 1}\right\},$$

$$\mathrm{Power}_\Gamma(u) = \mathrm{E}\left(\frac{1}{J}\sum_{j=1}^{J} \mathbb{1}\left\{\cup_{i=1}^{|\Gamma|} \tilde{T}_{\gamma_i}(u) \cap S_j \neq \varnothing\right\}\right),$$

$$\mathrm{Power}_{\mathrm{BH},\Gamma}(v) = \mathrm{E}\left(\frac{1}{J}\sum_{j=1}^{J} \mathbb{1}\left\{\cup_{i=1}^{|\Gamma|} \tilde{T}_{\gamma_i}\left(\tilde{u}_{\mathrm{BH},\gamma}(v)\right) \cap S_j \neq \varnothing\right\}\right),$$

where v is the pre-threshold in Algorithm 1 and $u > v$ is some fixed number.

The combined FDR criterion above is the expected proportion of false discoveries across all bandwidths in Γ simultaneously. This definition accounts for multiplicity of false positives over bandwidths and can be asymptotically controlled, as shown below. It has the correct interpretation of the FDR denominator as the total number of observed significant local maxima because the analyst cannot know in a given analysis which peaks correspond to the same true peaks and which do not, even if the same true peak is detected with several bandwidths. The combined power above, on the other hand, counts each truly detected peak once regardless of the bandwidths used to detect it.

The above definitions are similar to those in [26] in the sense that we too consider the union of detections over scales. However, because significant local maxima are points rather than regions, we cannot take advantage of the union of overlapping regions to account for repeats, but must be more specific about how each significant local maximum is counted, as discussed above.

Theorem 8. *Let conditions (C1) and (C2) hold.*

    i.   *Suppose that* Algorithm 1 *is applied for each* $\gamma_i \in \Gamma$ *with a fixed thresh* $u > v$, *then*

$$\mathrm{FDR}_\Gamma(u) \leq \frac{\sum_{i=1}^{|\Gamma|} \mathrm{E}\left[\tilde{m}_{0,\gamma_i}(U(1),u)\right]\left(1 - A_{2,\gamma_i}\right)}{\sum_{i=1}^{|\Gamma|}\left(\mathrm{E}\left[\tilde{m}_{0,\gamma_i}(U(1),u)\right]\left(1 - A_{2,\gamma_i}\right) + A_1\right)} + O\left(a^{-2} + L^{-N/2}\right).$$

    ii.   *Suppose that* Algorithm 1 *is applied for each* $\gamma_i \in \Gamma$ *with the random threshold* $\tilde{u}_{\mathrm{BH},\gamma_i}(v)$ *(15), then*

$$\mathrm{FDR}_{\mathrm{BH},\Gamma}(v) \leq \alpha\frac{\sum_{i=1}^{|\Gamma|} \mathrm{E}\left[\tilde{m}_{0,\gamma_i}(U(1),v)\right]\left(1 - A_{2,\gamma_i}\right)}{\sum_{i=1}^{|\Gamma|}\left(\mathrm{E}\left[\tilde{m}_{0,\gamma_i}(U(1),v)\right]\left(1 - A_{2,\gamma_i}\right) + A_1\right)} + O\left(a^{-1} + L^{-N/4}\right). \quad (25)$$

The following result follows directly from Theorem 5 and the facts

$$\mathrm{Power}_\Gamma(u) \geq \max_{1 \leq i \leq |\Gamma|} \mathrm{Power}_{\gamma_i}(u),$$

$$\mathrm{Power}_{\mathrm{BH},\,\Gamma}(v) \geq \max_{1 \leq i \leq |\Gamma|} \mathrm{Power}_{\mathrm{BH},\,\gamma_i}(v)$$

Theorem 9. *Let conditions (C1) and (C2) hold.*

i. *Suppose that* Algorithm 1 *is applied for each* $\gamma_i \in \Gamma$ *with a fixed thresh old $u > v$, then*

$$\mathrm{Power}_{\Gamma}(u) \geq 1 - O\!\left(a^{-2}\right).$$

i. *Suppose that* Algorithm 1 *is applied for each* $\gamma_i \, \Gamma$ *with the random threshold* $\tilde{u}_{\mathrm{BH},\gamma i}(v)$ *(15), then*

$$\mathrm{Power}_{\mathrm{BH},\,\Gamma}(v) \geq 1 - O\!\left(a^{-2} + L^{-N/2}\right).$$

Because testing over a range of bandwidths exacerbates the multiple testing problem by increasing the number of tests and increasing the transition region, in practice the set Γ should not be too large and should reflect the range of signal peaks present in the data.

4. Detection of peaks by approximate overshoot distribution.

4.1. Approximating the overshoot distribution.

In the neuroimaging literature, it is common to pre-threshold the test statistic field and then perform inference on the supra-threshold statistics [43]. The rationale is that significance thresholds tend to be high, and so weak signals of no biological interest are not expected to survive a moderate pre-threshold. Moreover, it may seem that using a pre-threshold would increase power by reducing the multiplicity of the tests being performed. We showed theoretically in Section 3.3 and will confirm by simulations that, in the best case scenario where the exact distribution of the height of local maxima is known, pre-thresholding ($v = -\infty$) does not increase detection power. However, pre-thresholding is still very valuable if the exact distribution is unknown but an approximation is known. In that case, the question of which pre-threshold to use remains relevant.

As mentioned, if the Gaussian field is only stationary but not isotropic, then the explicit formula for $F_\gamma(u, v)$ (11) is unknown so far. However, by [9, Corollary 2.5], there exists $\varepsilon_0 > 0$ such that as $v \to \infty$ and $u > v$,

$$F_\gamma(u, v) = K_\gamma(u, v)\left(1 + o\!\left(e^{-\varepsilon_0 v^2}\right)\right),$$

where

$$K_\gamma(u, v) = \frac{H_{N-1}(u/\sigma_\gamma)e^{-u^2/(2\sigma_\gamma^2)}}{H_{N-1}(v/\sigma_\gamma)e^{-v^2/(2\sigma_\gamma^2)}}$$

and $H_{N-1}(x)$ is the Hermite polynomial of degree N - 1. A similar argument to the proof of [9, Corollary 2.5] yields that for a fixed v, as $u \to \infty$,

$$F_\gamma(u, v) = \beta_\gamma(v)K_\gamma(u, v)\left(1 + o\left(e^{-\varepsilon_0 u^2}\right)\right), \tag{26}$$

where

$$\beta_\gamma(v) = \frac{(2\pi)^{-(N+1)/2}\sigma_\gamma^{-N}(\det(\Lambda_\gamma))^{1/2}H_{N-1}(v/\sigma_\gamma)e^{-v^2/(2\sigma_\gamma^2)}}{E[\widetilde{m}_{0,\gamma}(U(1), v)]} \tag{27}$$

and $\Lambda_\gamma = \text{Cov}(\nabla z_\gamma(t))$. Note that $\beta_\gamma(v)$ is similar to the ratio of the expected Euler characteristic [2, Lemma 11.7.1] and the expected number of local maxima of $z_\gamma(t)$ over the unit cube $U(1)$. It is conjectured that $\beta_\gamma(v) < 1$ for all $v > 0$ (this is true for $N = 1$ and $N = 2$ [9]).

For a fixed threshold u, the control of FDR$_\gamma(u)$ and the consistency of power Power$_\gamma(u)$ in Theorem 10 are the same as those given in part (i) of Theorem 3 and part (i) in Theorem 5 respectively. When the BH algorithm is used, we have the following result.

Theorem 10. *Let conditions (C1) and (C2) hold. Suppose* Algorithm 1 *is applied with the random threshold $\tilde{u}_{\text{BH},\gamma}(v)$ by using $K_\gamma(u, v)$ instead of $F_\gamma(u, v)$ to compute p-values. Then as $v \to \infty$ such that $v^2/\log(L^N) \to 0$ and $v^2/\log(a) \to 0$,*

$$\text{FDR}_{\text{BH},\gamma}(v) \le \alpha\frac{E[\widetilde{m}_{0,\gamma}(U(1), v)](1 - A_{2,\gamma})\beta_\gamma(v)}{E[\widetilde{m}_{0,\gamma}(U(1), v)](1 - A_{2,\gamma}) + A_1}\left(1 + o\left(e^{-\varepsilon_0 v^2}\right)\right), \tag{28}$$

where $\varepsilon_0 > 0$ is some constant and $\beta_\gamma(v)$ is defined in (27), and moreover,

$$\text{Power}_{\text{BH},\gamma}(v) = 1 - O\left(a^{-2} + L^{-N/2}\right). \tag{29}$$

The proof of Theorem 10 requires careful attention to rates because both sides of (28) go to zero as v increases. Yet, (28) gives the right approximation to the FDR for finite v. The additional asymptotic conditions $v^2/\log(L^N) \to 0$ and $v^2/\log(a) \to 0$, in fact indicate that v should increase more slowly than $\log(L^N)$ and even more slowly than $\log(a)$, so in practice it need not be very large. In the simulations we will see that an appropriate value for v is about $2\sigma_\gamma$.

4.2. Optimal pre-threshold.

A question that arises when using the overshoot distribution is how to select the pre-threshold. From the proof of Theorem 10 in the Supplementary Material, we see that the random threshold $\tilde{u}_{\text{BH},\gamma}(v)$ converges asymptotically to the deterministic threshold

$$u_{\text{BH}, \gamma}^{*,*}(v) = F_\gamma^{-1}\left(\frac{\alpha A_1 \beta_\gamma(v) \text{E}\left[\widetilde{m}_{0,\gamma}(U(1), v)\right]/\text{E}\left[\widetilde{m}_{0,\gamma}(U(1))\right]}{A_1 + \text{E}\left[\widetilde{m}_{0,\gamma}(U(1), v)\right]\left(1 - A_{2,\gamma}\right)\left(1 - \alpha\beta_\gamma(v)\right)}\right)(1 + o(1)).$$ (30)

For fixed γ, the power (21) is maximized at the optimal pre-threshold minimizing $u_{\text{BH}, \gamma}^{*,*}(v)$, which is

$$v_{\text{opt}, \gamma} = \underset{v}{\text{argmax}} \frac{H_{N-1}(v/\sigma_\gamma)e^{-v^2/(2\sigma_\gamma^2)}}{A_1 + \text{E}\left[\widetilde{m}_{0,\gamma}(U(1), v)\right]\left(1 - A_{2,\gamma}\right)\left(1 - \alpha\beta_\gamma(v)\right)}.$$ (31)

When γ and α are fixed, we see that $v_{opt, \gamma}$ depends only on the covariance structure of $z_\gamma(t)$. A practical approximation to the optimal pre-threshold is found in the simulations section below.

4.3. Multiscale testing procedure.

The results about the multiscale procedure in Section 3.6 can be extended to the case where the overshoot distribution is used. Let $\Gamma = \{\gamma_i\}_{1 \le i \le |\Gamma|}$ be as in the defined Section 3.6. Applying overshoot distribution, we have the following result.

Theorem 11. *Let conditions (C1) and (C2) hold. Suppose that* Algorithm 1 *is applied for each* $\gamma_i \in \Gamma$, *with the random threshold* $\tilde{u}_{\text{BH}, \gamma i}(v)$ *by using* $K_{\gamma i}(u, v)$ *instead of* $F_{\gamma i}(u, v)$ *to compute p-values. Then, as* $v \to \infty$ *such that* $v^2/\log(L^N) \to 0$ *and* $v^2/\log(a) \to 0$,

$$\text{FDR}_{\text{BH}, \Gamma}(v) \le \alpha \frac{\sum_{i=1}^{|\Gamma|} \text{E}\left[\widetilde{m}_{0,\gamma_i}(U(1), v)\right]\left(1 - A_{2,\gamma_i}\right)\beta_{\gamma_i}(v)}{\sum_{i=1}^{|\Gamma|}\left(\text{E}\left[\widetilde{m}_{0,\gamma_i}(U(1), v)\right]\left(1 - A_{2,\gamma_i}\right) + A_1\right)}\left(1 + o\left(e^{-\varepsilon_0 v^2}\right)\right),$$

where $\varepsilon_0 > 0$ *is some constant and* $\beta_{\gamma i}(v)$ *is defined in (27), and moreover,*

$$\text{Power}_{\text{BH}, \Gamma}(v) \ge 1 - O\left(a^{-2} + L^{-N/2}\right).$$

5. Simulation studies.

5.1. Simulation setting.

Simulations were used to evaluate the performance and limitations of the STEM algorithm for finite range $L = 300$, finite number of peaks $J = 9$ and moderate signal strength a over $\mathbb{R}^2$ (i.e., $N = 2$). Adopting the notations in Example 7, the truncated Gaussian peaks $a_j h_j(t)$ are constructed with $a_j = a$, $b_j = 3$ and $c_j = 3$ for all $j = 1, \ldots, J$ and varying a, and $\{\tau_j\}_{1 \le j \le J} = \{(75i_1, 75i_2)\}_{i_1, i_2 = 1,2,3}$; the noise $z(t)$ is constructed with $\sigma = 1$ and varying $\nu \in \{0, 1, 2\}$ the smoothing kernel $w_\gamma(t)$ is constructed with $c = 3$ and varying γ. The noise parameters σ_γ, ρ_γ' and ρ_γ'' (note that $\kappa_\gamma = 1$) were estimated using the same smoothing kernel. The BH procedures were applied at level $a = 0.05$ and over 10,000 replications.

5.2. Detection of peaks by the height distribution of local maxima.

Figure 3 shows the realized FDR and power of the STEM algorithm with the BH procedure, evaluated according to (16) and (20) with $v = -\infty$. The range of values of the signal strength a was chosen between 35 and 55 to show interesting intermediate values of FDR and power, but notice that the signal is overall quite weak; the peaks in Figure 1 all have values of a of 75 and above, and all are virtually invisible in the presence of noise ($y(t)$ in Figure 1). As predicted by the theory, for every fixed bandwidth γ, the FDR is controlled below $a = 0.05$ for strong enough signal a, and the power increases to 1. The theoretical FDR curve (blue) is evaluated according to the upper bound in (18), while the theoretical power curve (blue) is derived by plugging the asymptotic threshold $u^*_{\mathrm{BH}, \gamma}$ (19) into the approximated power (21).

The discrepancy between the realized FDR and the theoretical FDR is caused by the boundary effects of kernel smoothing and the finite size of the search domain. For small v and small γ, say $v = 0$ and $\gamma = 1$, the discrepancy is larger because the representation of the smoothed field z_γ on the simulated discrete grid is not smooth enough. Note that the realized power curve follows the same patterns as the theoretical power curve. As the signal gets stronger, the bandwidth maximizing the realized power gets closer to its optimal value.

5.3. Multiscale testing procedure.

Tables 1 and 2 show the realized FDR and power for the multiscale procedure of Section 3.6. The theoretical FDR values in Table 1 were evaluated according to the upper bound in (25). Again, the FDR is controlled below $a = 0.05$ and that the power tends to 1 as the signal strength a increases. To compare with the usual STEM algorithm with optimal γ, Table 2 reports a maximal power over scales is akin to the multiscale power definition of [26]. The multiscale procedure has more power, but yields a slightly slower rate of FDR control, especially when the signal strength is very weak ($a = 35$).

5.4. Detection of peaks by the overshoot distribution.

Figure 4 shows the realized FDR and power of the STEM algorithm with the BH procedure, using the exact overshoot distribution $F_\gamma(\cdot, v)$ to compute p-values instead of the exact height distribution. Here, the bandwidth was fixed to the optimal value. The theoretical FDR curve was evaluated according to the upper bound in (18), while the theoretical power curve was derived by plugging the asymptotic threshold $u^*_{\mathrm{BH}, \gamma}(v)$ (19) into the approximate power (21). As shown in Figure 4, as the pre-threshold v gets larger, the FDR becomes smaller and so does the power. This confirms the observation made after (20) that the case of $v = -\infty$ gives the best performance if the exact height distribution F_γ is known. However, when the signal is relatively strong, prethresholding does not weaken the power too much.

In Figure 5, the approximate overshoot distribution $K_\gamma(\cdot, v)$ is used to compute p-values instead of the exact overshoot distribution. Here again, the bandwidth was fixed to the optimal value. The theoretical FDR curve was evaluated according to the upper bound in (28), while the theoretical power curve was derived by plugging the asymptotic threshold $u^{*\,*}_{\mathrm{BH}, \gamma}(v)$ (30) into the approximate power (21). The simulation shows that the pre-threshold maximizing the realized power is close to the optimal pre-threshold $v_{\mathrm{opt}, \gamma}$ (31), which in this example is about $2\sigma_\gamma$. Moreover, the realized curves still fit the theoretical curves well for

small v. This is because the limit in Theorem 10 is in fact taken when the BH threshold is large.

5.5. Comparison with other methods.

We compare the STEM algorithm 1 to two other methods. First, because the signal strength $a \to \infty$ in condition (C2), one may think this makes the model relatively simple asymptotically (in fact not, because the null domain is also getting large), and wonder if some simple tests can also detect peaks with error control and power consistency. To check this, we try a simple quantile test as follows. Under (C2), the ratio of the number of true peaks and the number of all peaks of the field tends to

$$r_{1, \gamma} = \frac{A_1}{\mathrm{E}\left[\widetilde{m}_{0, \gamma}(U(1))\right]\left(1 - A_{2, \gamma}\right) + A_1}.$$

Thus, choosing the $100(1 - r_{1, \gamma})\%$ quantile of the heights of all candidate peaks as threshold guarantees that all true peaks will be detected asymptotically. Since $r_{1, \gamma}$ is unknown, we call this an "oracle quantile test". Alternatively, if $r_{1, \gamma}$ is replaced by an arbitrary fixed fraction independent of γ, say 0.05, we call it "95% quantile test". Figure 6 (left) shows the comparison between the STEM algorithm by height distribution and these two quantile tests; even if they are asymptotically consistent, the quantile tests cannot control the error. Second, Figure 6 (right) compares the STEM algorithm by approximate overshoot distribution to the method in Chumbley et al. [14] (CWFF) using Eqs. (1) and (2) therein for computing p-values. These two similar methods give almost the same results, especially when the pre-threshold is not small. However, CWFF requires estimation of additional parameters that the approximate overshoot method does not need.

6. Data example.

6.1. Data description.

The fMRI data was obtained from the public repository OpenfMRI (openfmri.org). It involves an experiment whose goal is to find brain regions that are active when processing other people's false beliefs about reality, in comparison with similar purely physical false realities. Details about the psychological motivation and experimental design of this so-called "false belief task" can be found in [23]. In brief, subjects read short stories corresponding to either a person's false belief about reality or false realities with no people involved. The effect sought after in the analysis is the contrast between the neural activity in those two states.

For simplicity, we focus here on data from a single subject (# 49). The data consist of a sequence of $n = 179$ brain fMRI images of size $71 \times 72 \times 36$ voxels (one row was removed from the original size of $72 \times 72 \times 36$ due to absence of data there). As standard pre-processing, algorithms were applied for motion correction over the recording period.

6.2. Regression analysis.

To analyze the data, we followed the usual general linear model (GLM) approach [39], where, after spatial registration of the n images to a common template, the n image intensities at each voxel were modeled as a linear function of the stimuli. Letting $Y(t)$ denote the $n \times 1$ vector of observed intensities at spatial location (voxel) t, the model is $Y(t) = X\beta(t) + \in(t)$, where X is an $n \times 3$ matrix whose columns contain the duration of the two types of stories as 0–1 step functions of time, in addition to a column for the intercept term. The vector $\beta(t)$ contains the regression coeffcients, while $\in(t)$ is a $n \times 1$ random vector whose entries are assumed to be i.i.d. with zero mean and variance $\sigma^2(t)$. The least-squares estimate of the coefficient vector $\beta(t)$ at each location t is $\hat{\beta}(t) = (X'X)^{-1}X'Y(t)$. The contrast of interest $\eta(t) = c'\beta(t)$, $c = (0,1,-1)'$, is the difference between the regression coeffcients corresponding to the two stimuli. Its estimate is $\hat{\eta}(t) = c'\hat{\beta}(t)$ with variance $\mathrm{Var}(\hat{\eta}(t)) = \sigma^2(t)c'(X'X)^{-1}c$..

To test the null hypothesis $\mathscr{H}_0: \eta(t) = 0$ at each location t, we used a Wald statistic, defined as the estimate $\hat{\eta}(t)$ divided by its estimated standard error:

$$\tilde{\eta}(t) = \frac{\hat{\eta}(t)}{\widehat{\mathrm{se}}(\hat{\eta}(t))} = \frac{c'(X'X)^{-1}X'\hat{Y}(t)}{\sqrt{\hat{\sigma}^2(t)c'(X'X)^{-1}c}}. \tag{32}$$

The noise variance estimate $\hat{\sigma}^2(t)$ above was obtained as $\hat{\sigma}^2(t) = \|e(t)\|^2/(n-3)$, where $e(t) = y(t) - X\hat{\beta}(t)$ is the vector of regression residuals. Because the number of degrees of freedom $n - 3 = 176$ is large, we may consider $\tilde{\eta}(t)$ to be an approximately Gaussian 3D random field with constant variance 1, playing the role of $y(t)$ in model (1).

6.3. Inference via the STEM algorithm.

For the kernel smoothing step, we used a 3D isotropic Gaussian kernel as in Example 7 with bandwidth $\gamma = 1.6$ voxels. Given the voxel size of 3 mm, this corresponds to a bandwidth of 4.8 mm and a full width half maximum of 11.3 mm, optimizing the analysis for signal regions of about that size. To perform the convolution, the kernel was truncated at 2.5 standard deviations from the mode, yielding a kernel support of $8 \times 8 \times 8$ voxels. After convolution, only the "valid" portion of the image was retained, i.e. those voxels whose values were computed from neighborhood voxels strictly contained in the original image, yielding a valid image of size $64 \times 65 \times 29 = 120,640$.

We first assumed the noise field to be isotropic and applied the STEM algorithm using the exact height distribution of local maxima to compute p-values, as in Section 3.5. To estimate the required parameters σ_γ^2 and κ_γ of the height distribution, we used a volume of $27 \times 72 \times 36 = 69,984$ voxels consisting of the top 14 and bottom 13 rows of the data outside the skull, containing only noise. The sample variance within this region and across time yielded the estimate $\hat{\sigma}_\gamma^2 = 0.0058$. Similarly, under the isotropy assumption, the sample variance of the first and second numerical derivatives yielded estimates of ρ_γ' and ρ_γ'' by (23), leading to the estimate $\hat{\kappa}_\gamma = -\hat{\rho}_\gamma'/\sqrt{\hat{\rho}_\gamma''} = 1.1463$.

The analysis results using the BH algorithm at FDR level 0.05 are shown in Figure 7. The 55 significant local maxima can be interpreted as representing significant brain regions. Most of these are located in the anterior and posterior ventral parts of the cortex, as well as the dorsal parts of the pre-frontal cortex, consistent with the results reported in Moran et al. [23]. For comparison, we also performed the analysis with $\kappa_\gamma = 1$, as in Example 7, avoiding estimation of this parameter. The results obtained were very similar.

To evaluate the effect of the distribution used to compute p-values, Table 3 compares the results using the exact height distribution or the approximate overshoot distribution. As seen in the table, computing p-values using the approximate overshoot distribution with pre-threshold $v = 2\sigma_\gamma$ (close to optimal according to Figure 5) yields less candidate peaks but about the same number of significant ones. This confirms the simulation results that using the approximate overshoot distribution yields similar power, while not requiring the isotropy assumption.

As an additional comparison, we performed the analysis using the CWFF method where the expected Euler characteristic parameters were estimated from the residuals outside the brain. As expected from the simulation results, the significance threshold and number of significant peaks were the same as with the approximate overshoot distribution, except that the latter did not need estimation of those extra parameters.

7. Discussion.

7.1. Asymptotic considerations.

Regarding condition (C2), it is not surprising that if the volume of the search space L^N increases, then the signal strength a should also increase in order for the detection procedure to have good power while the error is controlled. We do not find this assumption restrictive because the search space may grow exponentially faster. Our simulations have shown that the theoretical results provide good approximations in non-asymptotic conditions when the search space L^N and the signal strength a are large, the former much larger than the latter.

These conditions are realistic in applications. In data with n repeated observations, the signal strength a is proportional to $\sqrt{n}$ Setting $p = L^N$ as the dimensionality of the problem, the condition $(\log L^N)/a^2$ 0 becomes $(\log p)/n \to 0$, which is similar to the condition required for consistent model selection in high dimensional regression [8, 42]. In our fMRI data example, the search volume corresponding to L^3 was $64 \times 65 \times 29 = 120,640$ voxels and the sample size was 179, giving a ratio $(\log L^N)/a^2$ of about $\log(120,640)/179 = 0.0284$, which is close to zero.

The assumption $(\log L^N)/a^2 \to 0$ is in fact stronger than needed. From the proof of Theorem 5, it can be seen that if the signal strength grows sufficiently fast so that $(\log L^N)/a^2$ is bounded asymptotically by some positive constant, then the power will be asymptotically bounded from above by some constant less than 1. A future analysis of contiguous alternative hypotheses may help clarify this.

Efforts were made in this paper to include convergence rates in the results, which were not available in Schwartzman et al. [31]. Doing so required calculating the order of the variance of the number of local maxima of a stationary random field in N dimensions. Computing the moments of the number of local maxima of a random field is known to be an important and non-trivial problem in probability theory [4, 35].

7.2. Signal model considerations.

The signal modeling assumptions of unimodality and compact support enabled defining true and false detections. It should be noted, however, that the formal hypothesis test performed at each local maximum is not about the signal peak but about its support. In this sense, the problem may be called "detection of signal support". The justification for the name "peak detection" is that, by the signal unimodality, the detected local maxima asymptotically coincide with the local maxima of the signal, thus detecting the peaks in addition to their support.

In future work, the assumptions of unimodality and compact support could be relaxed. As suggested by simulations in [31], we believe that the STEM algorithm can detect multimodal peaks because local maxima of the observed field in the troughs between the true modes are unlikely to occur. However, a more careful definition of power would be required. Siding with Chumbley et al. [14], it may be possible to assume that the signal is non-zero everywhere by defining true detections to occur within a given distance of a true mode, rather than outside the domain.

Except for the unimodality and compact support assumptions, the signal model was kept general in nonparametric form. Using a single smoothing bandwidth in the analysis presumes that the signal peaks have similar shapes and supports. To handle peaks of different spatial extents, a multi-scale procedure was proposed that enables peaks to be optimally detected by different bandwidths. A more principled approach may consider performing the inference on the continuous scale-rotation space produced by varying the smoothing bandwidth and spatial orientation of the smoothing kernel [1]. However, this approach remains challenging because, even under the complete null hypothesis, the scale-rotation field is not stationary. We leave this possibility for future work.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments.

References.

[1]. Adler Robert J., Subag Eliran, and Taylor Jonathan E.. Rotation and scale space random fields and the Gaussian kinematic formula. Ann Statist, 40(6):2910–2942, 2012.

[2]. Adler Robert J. and Taylor Jonathan E.. Random Fields and Geometry. Springer, New York, 2007.

[3]. Adler Robert J., Taylor Jonathan E., and Worsley Keith J.. Applications of Random Fields and Geometry: Foundations and Case Studies. URL http://webee.technion.ac.il/people/adler/publications.html. 2010.

[4]. Azais Jean-Marc and Wschebor Mario. Level Sets and Extrema of Random Processes and Fields. John Wiley & Sons, Hoboken, NJ, 2009.

[5]. Benjamini Yoav and Heller Ruth. False discovery rates for spatial signals. J Amer Statist Assoc, 102(480):1272–1281, 2007.

[6]. Benjamini Yoav and Hochberg Yosef. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Statist Soc B, 57(1):289–300, 1995.

[7]. Brutti Pierpaolo, Genovese Christopher R., Miller Christopher J., Nichol Robert C., and Wasserman Larry. Spike hunting in galaxy spectra. Technical report, Libera Università Internazionale degli Studi Sociali Guido Carli di Roma, 2005 URL http://www.stat.cmu.edu/tr/tr828/tr828.html.

[8]. Candes Emmanuel and Tao Terence. The dantzig sselector: statistical estimation when p is much larger than n. Ann Statist, 35(6):2313–2351, 2007.

[9]. Cheng Dan and Schwartzman Armin. Distribution of the height of local maxima of Gaussian random fields. Extremes, 18(2):213–240, 2015. [PubMed: 26478714]

[10]. Cheng Dan and Schwartzman Armin. Expected number and height distribution of critical points of smooth isotropic Gaussian random fields. arXiv:151106835, 2015.

[11]. Cheng Dan and Schwartzman Armin. On the explicit height distribution and expected number of local maxima of isotropic Gaussian random fields. arXiv:150301328, 2015.

[12]. Cheng Dan and Schwartzman Armin. Online Supplement to "Multiple Testing of Local Maxima for Detection of Peaks in Random Fields" DOI:????, 2016.

[13]. Chumbley Justin R. and Friston Karl J.. False discovery rate revisited: FDR and topological inference using Gaussian random fields. Neuroimage, 44(1):62–70, 2009. [PubMed: 18603449]

[14]. Chumbley Justin R., Worsley Keith, Flandin Guillaume, and Friston Karl J.. Topological FDR for neuroimaging. Neuroimage, 49:3057–3064, 2010. [PubMed: 19944173]

[15]. Harald Cramér and M Ross Leadbetter Stationary and Related Stochastic Processes. Wiley, New York, 1967.

[16]. Egner Alexander, Geisler Claudia, Claas von Middendorff Hannes Bock, Wenzel Dirk, Medda Rebecca, Andresen Martin, Stiel Andre C., Jakobs Stefan, Eggeling Christian, Andreas Schönle, and Stefan W. Hell. Fluorescence nanoscopy in whole cells by asynchronous localization of photoswitching emitters. Biophysical Journal, 93:3285–3290, 2007. [PubMed: 17660318]

[17]. Fithian William, Sun Dennis, and Taylor Jonathan. Optimal inference after model selection. arXiv:14102597, 2015.

[18]. Fyodorov Yan V.. Complexity of random energy landscapes, glass transition, and absolute value of the spectral determinant of random matrices. Phys. Rev. Lett, 92: 240601, 2004. [PubMed: 15245072]

[19]. Geisler C, Schönle A, von Middendorff C, Bock H, Eggeling C, Egner A, and Hell SW. Resolution of λ/10 in fluorescence microscopy using fast single molecule photo-switching. Applied Physics A, 88:223–226, 2007.

[20]. Genovese Christopher R., Lazar Nicole A., and Nichols Thomas E.. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. Neuroimage, 15:870–878, 2002. [PubMed: 11906227]

[21]. Heller Ruth, Stanley Damian, Yekutieli Daniel, Rubin Nava, and Benjamini Yoav. Cluster-based analysis of fMRI data. Neuroimage, 33(2):599–608, 2006. [PubMed: 16952467]

[22]. Lee Jason D., Sun Dennis L., Sun Yuekai, and Taylor Jonathan E.. Exact post-selection inference, with application to the lasso. arXiv:13116238, 2015.

[23]. Moran Joseph M., Jolly Eshin, and Mitchell Jason P.. Social-cognitive deficits in normal aging. J Neurosci, 32(16):5553–5561, 2012. [PubMed: 22514317]

[24]. Nichols TE and Hayasaka S. Controlling the familywise error rate in functional neuroimaging: A comparative review. Statistical Methods in Medical Research, 12: 419–446, 2003. [PubMed: 14599004]

[25]. Perone Pacifico M, Genovese C, Verdinelli I, and Wasserman L. False discovery control for random fields. J Amer Statist Assoc, 99(468):1002–1014, 2004.

[26]. Perone Pacifico M, Genovese C, Verdinelli I, and Wasserman L. Scan clustering: A false discovery approach. J Multivar Anal, 98(7):1441–1469, 2004.

[27]. Piterbarg Vladimir. Asymptotic Methods in the Theory of Gaussian Processes and Fields Translations of Mathematical Monographs, Vol. 148, American Mathematical Society, Providence, RI 1996.

[28]. Poline J-B, Worsley KJ, Evans AC, and Friston KJ Combining spatial extent and peak intensity to test for activations in functional imaging. Neuroimage, 5:83–96, 1997. [PubMed: 9345540]

[29]. Pratt William K.. Digital Image Processing. Wiley, New York, 1991.

[30]. Schwartzman Armin, Dougherty Robert F., and Taylor Jonathan E.. False discovery rate analysis of brain diffusion direction maps. Ann Appl Statist, 2(1):153–175, 2008.

[31]. Schwartzman Armin, Gavrilov Yulia, and Adler Robert J.. Multiple testing of local maxima for detection of peaks in 1D. Ann Statist, 39(6):3290–3319, 2011.

[32]. Siegmund DO, Zhang N and Yakir B. False discovery rate for scanning statistics. Biometrika, 98(4):979–985, 2011.

[33]. Simon Marvin. Digital Communication Techniques: Signal Design and Detection. Prentice Hall, Englewood Cliffs, NJ, 1995.

[34]. Smith Stephen M. and Nichols Thomas E.. Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. Neuroimage, 44:83–98, 2009. [PubMed: 18501637]

[35]. Subag Eliran. The complexity of spherical p-spin models – a second moment approach. arXiv:150402251, 2015.

[36]. Sun Wenguang, Reich Brian, Cai Tony, Guindani Michele and Schwartzman Armin. False discovery control in large-scale spatial multiple testing. J R Statist Soc B, to appear.

[37]. Taylor Jonathan E., Loftus Joshua, and Tibshirani Ryan J.. Inference in adaptive regression via the Kac-Rice formula. arXiv:13083020, 2014.

[38]. Taylor Jonathan E. and Worsley Keith J.. Detecting sparse signals in random fields, with an application to brain mapping. J. Am. Statist. Assoc, 102:913–928, 2007.

[39]. Worsley Keith J., Marrett S, Neelin P, and Evans AC. Searching scale space for activation in PET images. Human Brain Mapping, 4:74–90, 1996. [PubMed: 20408187]

[40]. Worsley Keith J., Marrett S, Neelin P, Vandal AC, Friston Karl J., and Evans AC. A unified statistical approach for determining significant signals in images of cerebral activation. Human Brain Mapping, 4:58–73, 1996. [PubMed: 20408186]

[41]. Worsley Keith J., Taylor Jonathan E., Tomaiuolo F, and Lerch J. Unified univariate and multivariate random field theory. Neuroimage, 23:S189–195, 2004. [PubMed: 15501088]

[42]. Zhang Cun-Hui. Nearly unbiased variable selection under minimax concave penalty. Ann Statist, 38(2):894–942, 2010.

[43]. Zhang Hui, Nichols Thomas E., and Johnson Timothy D.. Cluster mass inference via random field theory. Neuroimage, 44:51–61, 2009. [PubMed: 18805493]
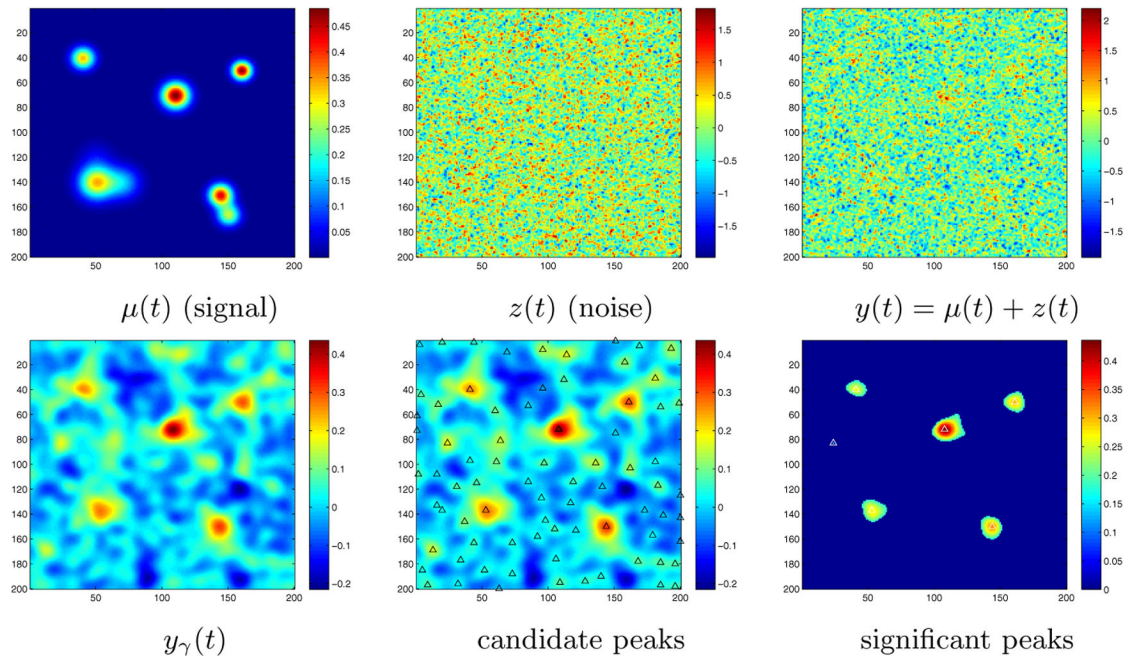
$\mu(t)$ (signal) $z(t)$ (noise) $y(t) = \mu(t) + z(t)$

$y_\gamma(t)$ candidate peaks   significant peaks

Fig 1.

Raw signal μ(t) with six true peaks of different shapes and simulated Gaussian noise z(t) produce the observed field y(t) and smoothed field $y_\gamma(t)$. Out of 77 local maxima of $y_\gamma(t)$ (candidate peaks), the BH detection threshold at FDR level 0.2 selects six (significant peaks), one of which is a false positive. In this case, five out of six true peaks are detected.
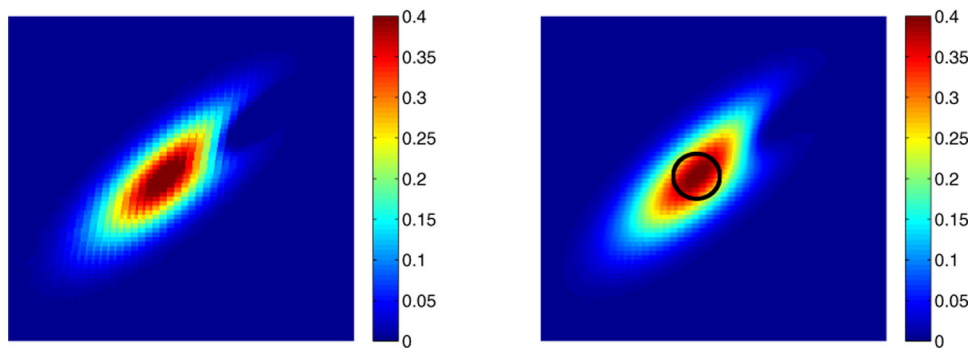
Fig 2.
An artificial example of unimodal functions $h_j(t)$ (left) and $h_{j,\gamma}(t)$ (right) allowed by our signal model. Here, $h_{j,\gamma}(t)$ is concave within $I_{j,\gamma}^{\mathrm{mode}}$ (interior of the black circle) and has no critical points outside $I_{j,\gamma}^{\mathrm{mode}}$, yet its level curves are not all convex.
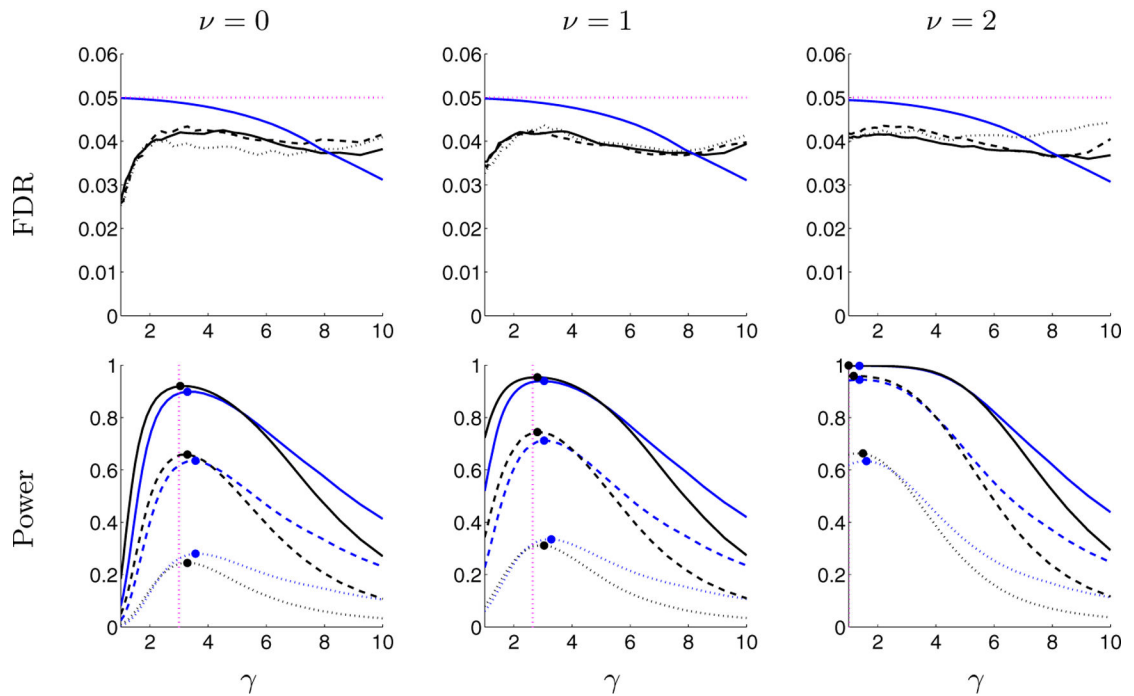
Fig 3.
Realized (black) and "theoretical" (blue) FDR, Realized (black) and "theoretical" (blue) power of the BH procedure by the exact height distribution F_γ (i.e., v = - ∞) for a = 55 (solid), a = 45 (dashed) and a = 35 (dotted). The maxima of the curves (solid circles) approach the optimal bandwidth (vertical dashed).
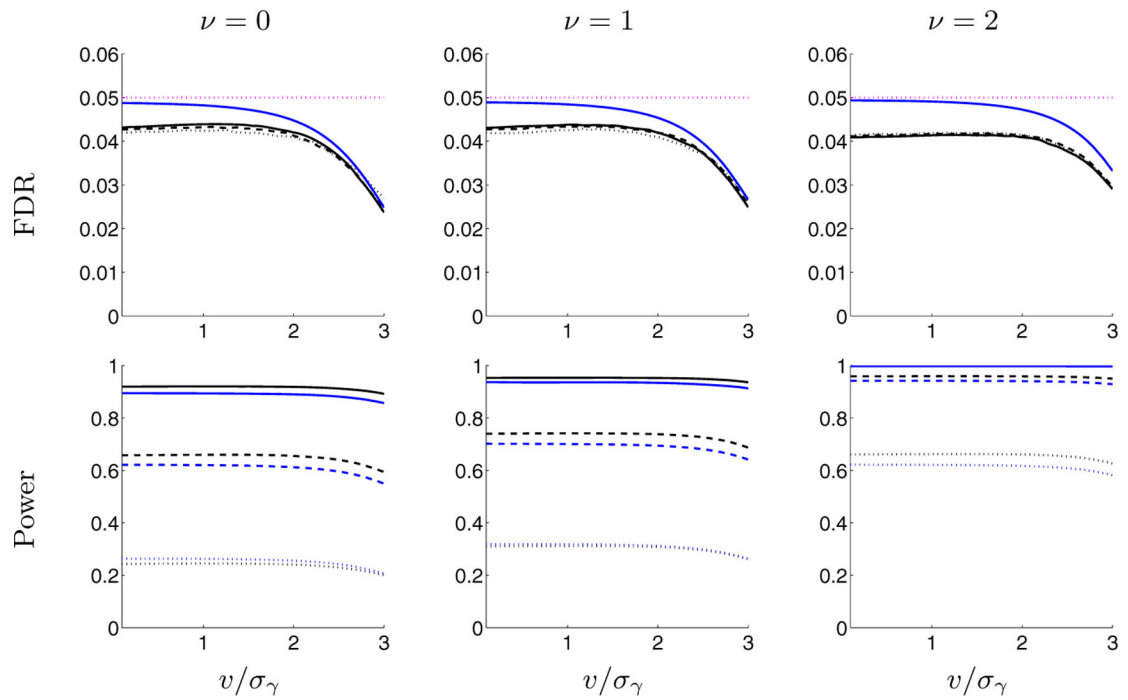
Fig 4.
Realized (black) and "theoretical" (blue) FDR, Realized (black) and "theoretical" (blue) power of the BH procedure by the exact overshoot distribution $F_\gamma(\cdot, v)$ for $a = 55$ (solid), $a = 45$ (dashed) and $a = 35$ (dotted).
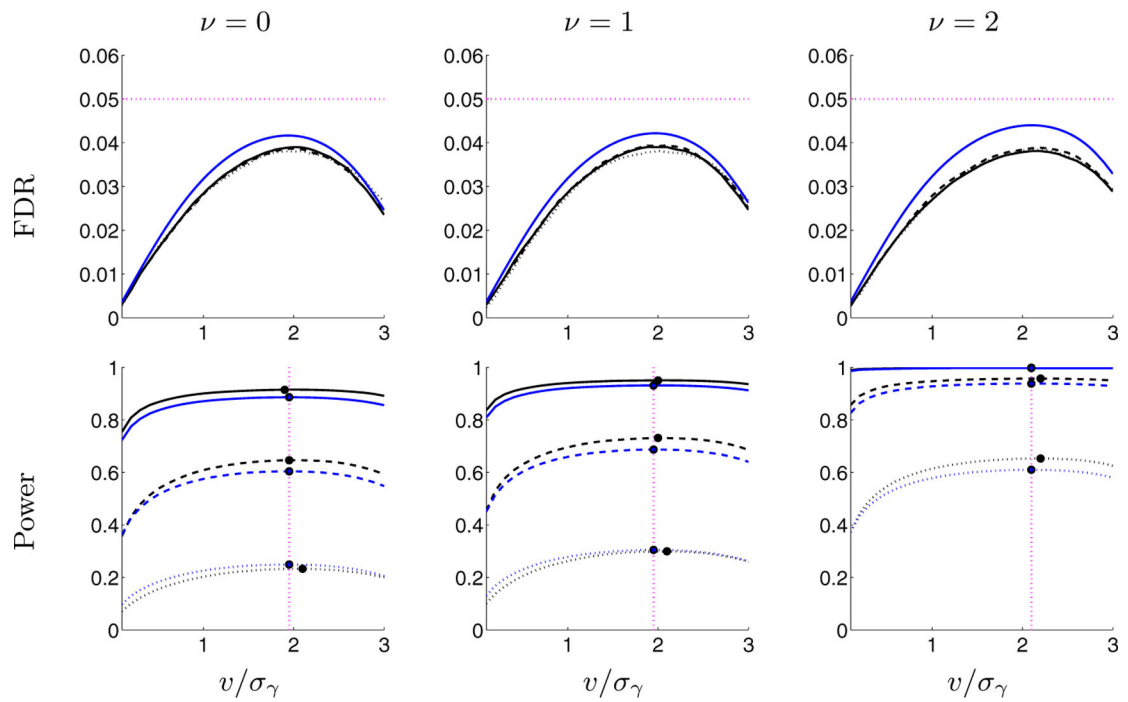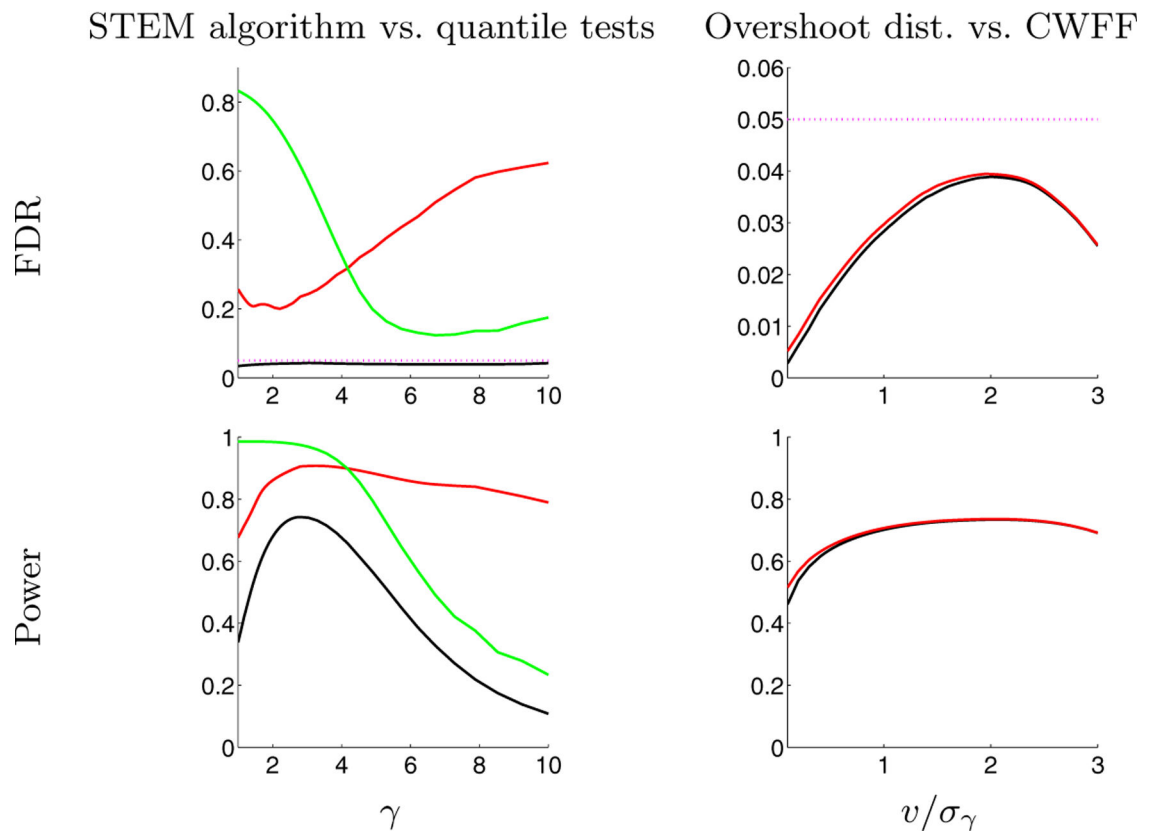
Fig 5.

Realized (black) and "theoretical" (blue) FDR, Realized (black) and "theoretical" (blue) power of the BH procedure by the approximated overshoot distribution $K_\gamma(\ \cdot\ , v)$ for $a = 55$ (solid), $a = 45$ (dashed) and $a = 35$ (dotted). The maxima of the curves (solid circles) approach the optimal pre-threshold $v_{opt,\gamma}$ (vertical dashed).

Fig 6.

Left panel: FDR and power comparison among the STEM algorithm by height distribution (black), the "oracle quantile test" (red) and the "95% quantile test" (green). Right panel: FDR and power comparison between the STEM algorithm by approximate overshoot distribution (black) and CWFF (red), the method in Chumbley et al. [14]. In both panels, $\nu = 1$ (moderate autocorrelation) and $a = 45$ (moderate signal strength).
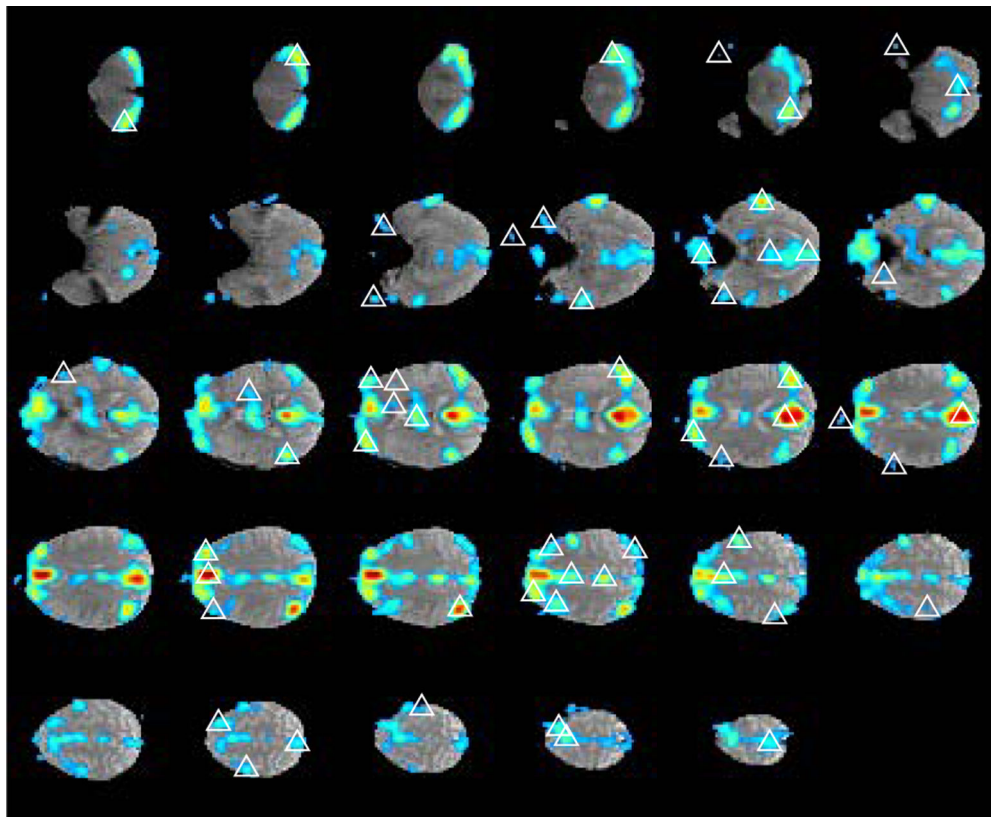
Fig 7.
fMRI data analysis results using the exact height distribution of local maxima for isotropic noise and FDR level 0.05. Montage shows the brain volume as transverse slices from the bottom of the brain (top left panel) to the top of the brain (bottom right panel). The 55 significant local maxima are marked by white triangles. Colored regions indicate the smoothed Wald statistic field above the height of the smallest significant local maximum. Results are superimposed on an anatomical brain image (gray) for reference.

Table 1

Simulated FDR by multiscale procedure (left) and simulated maximal FDR by STEM algorithm in Figure 3 (right); the numbers in parentheses are theoretical FDR's by multiscale procedure.

	$v = 0$	$v = 1$	$v = 2$
$a = 35$	0.0739; 0.0409 (0.0495)	0.0634; 0.0437 (0.0492)	0.0484; 0.0443 (0.0486)
$a = 45$	0.0498; 0.0434 (0.0495)	0.0495; 0.0423 (0.0492)	0.0454; 0.0435 (0.0486)
$a = 55$	0.0440; 0.0425 (0.0495)	0.0433; 0.0422 (0.0492)	0.0427; 0.0416 (0.0486)

Table 2

Simulated power by multiscale procedure (left) and the maximal simulated power by STEM algorithm in Figure 3 (right).

	$v = 0$	$v = 1$	$v = 2$
$a = 35$	0.3106; 0.2449	0.3823; 0.3118	0.7068; 0.6631
$a = 45$	0.7167; 0.6590	0.7950; 0.7444	0.9674; 0.9587
$a = 55$	0.9414; 0.9204	0.9670; 0.9541	0.9989; 0.9984

Table 3

Comparison of different methods for computing p-values.

	BH height threshold	# tests	# significant tests
Height distribution	$3.9156\sigma_\gamma$	334 peaks	55 peaks
Overshoot distribution	$3.8337\sigma_\gamma$	159 peaks	56 peaks