Harvard University
Harvard University Biostatistics Working Paper Series

Year 2011 *Paper* 131

Multiple Testing of Local Maxima for Detection of Unimodal Peaks in 1D

Armin Schwartzman* Yulia Gavrilov†
Robert J. Adler‡

*Harvard School of Public Health and Dana Farber Cancer Institute, armin@jimmy.harvard.edu

†Harvard School of Public Health

‡Israel Institute of Technology, robert@ee.technion.ac.il

Multiple Testing of Local Maxima for Detection of Unimodal Peaks in 1D

Armin Schwartzman[1], Yulia Gavrilov[1], Robert J. Adler[2]

[1] Department of Biostatistics, Harvard School of Public Health
[2] Department of Electrical Eng., Technion - Israel Institute of Technology

June 15, 2011

Abstract

A topological multiple testing scheme for one-dimensional domains is proposed where, rather than testing every spatial or temporal location for the presence of a signal, tests are performed only at the local maxima of the smoothed observed sequence. Assuming unimodal true peaks with finite support and Gaussian stationary ergodic noise, it is shown that the algorithm with Bonferroni or Benjamini-Hochberg correction provides asymptotic strong control of the family wise error rate and false discovery rate, and is power consistent, as the search space and the signal strength get large, where the search space may grow exponentially faster than the signal strength. Simulations show that error levels are maintained for non-asymptotic conditions, and that power is maximized when the smoothing kernel is close in shape and bandwidth to the signal peaks, akin to the matched filter theorem in signal processing. The methods are illustrated in an analysis of electrical recordings of neuronal cell activity.

1 Introduction

One of the most challenging aspects of multiple testing problems in spatial and temporal domains is how to account for the spatial or temporal structure in the underlying signal. The usual paradigm considers a separate test at each observed location. However, the interest is usually in detecting signal regions that span several neighboring locations. This paper considers a new multiple testing paradigm for spatial and temporal domains where tests are not performed at every observed location but only at the local maxima of the observed data, seen as representatives of underlying signal peaks. The proposed inference is not pointwise but topological, based on the observed local maxima as topological features.

In pointwise testing, the control of family-wise error rate (FWER), now commonplace in neuroimaging, was established by Keith Worsley (Worsley et al., 1996b, 2004; Taylor and Worsley, 2007), who exploited the Euler characteristic heuristic for approximating the distribution of the maximum of a random field (Adler and Taylor, 2007; Adler et al., 2010). Methods for controling the false discovery rate (FDR) (Benjamini and Hochberg, 1995) are also applied routinely in this setting, but the difficulty of incorporating the spatial structure causes it often to be ignored (Genovese et al., 2002; Nichols and Hayasaka, 2003; Schwartzman et al., 2008).

Despite pointwise testing being so common, the real interest is usually not in detecting individual locations but connected regions or clusters. This has prompted the adaptation

of discrete FDR methods to pre-defined clusters (Benjamini and Heller, 2007), and the use of Gaussian random field theory for computing p-values corresponding to the height, extent, and mass of clusters obtained by pre-thresholding the observed random field (Poline et al., 1997; Zhang et al., 2009). Pacifico et al. (2004a,b) proposed data-dependent thresholds so that FDR is controlled at the cluster level, using Gaussian random field theory to approximate the null distribution. However, the definition of Type I error for clusters requires a tolerance parameter for the overlap between a discovered cluster and the null region (Pacifico et al., 2004a), while spatial smoothing, which is often applied for improving signal-to-noise ratio (SNR), creates the need to remove the spread of the signal over the null region to avoid error inflation (Pacifico et al., 2004b). Chumbley and Friston (2009) have argued that current cluster methods are unsatisfactory because, just like marginal FDR procedures, they rely on the basic premise of having a test at each spatial location; instead, inference should be topological.

This article proposes a different multiple testing paradigm where tests are performed not at each spatial or temporal location but only at the local maxima of the smoothed data, seen as topological representatives of their neighborhood region or cluster. A similar idea was recently proposed independently by Chumbley et al. (2010), but they did not consider whether Type I error could be controlled. Here we extend the classical control of FWER via the global maximum to control of both FWER and FDR via local maxima. Because the distributional theory for local maxima of random fields is more difficult than that for global maxima, this paper only considers one-dimensional domains (spatial or temporal), where closed-form solutions exist, leaving the two- and three-dimensional cases for future work.

Our general proposed algorithm consists of the following steps:

1. *Kernel smoothing*, to increase SNR (Worsley et al., 1996a; Smith and Nichols, 2009).
2. *Candidate peaks*: find local maxima of the smoothed sequence.
3. *P-values*, computed at each local maximum under the complete null hypothesis of no signal anywhere.
4. *Multiple testing*: use a multiple testing procedure to find a global threshold and declare significant all peaks exceeding that threshold.

In this paper, the p-values in Step 3 are computed using theory of Gaussian processes. For Step 4, we consider two standard multiple testing procedures: Bonferroni to control FWER and Benjamini-Hochberg (BH) (Benjamini and Hochberg, 1995) to control FDR. The algorithm is illustrated by a simulated example in Figure 1.

We study the theoretical properties of the above algorithm under a specific signal-plus-noise model and later relax these assumptions in the simulation studies. For Type I errors to be well defined, the signal is modeled as composed of unimodal peaks, each considered detected if a significant local maximum occurs inside its finite support. For simplicity, we concentrate on positive signals and one-sided tests, but this is not crucial. For tractibility, the theory assumes that the observation noise follows a smooth stationary ergodic Gaussian process. This assumption permits an explicit formula for computing the p-values corresponding to local maxima of the observed process. The distribution of the height of a local maximum of a Gaussian process is not Gaussian but has a heavier tail, and its computation requires careful conditioning based on the calculus of Palm probabilities (Cramér and Leadbetter, 1967; Adler et al., 2010).

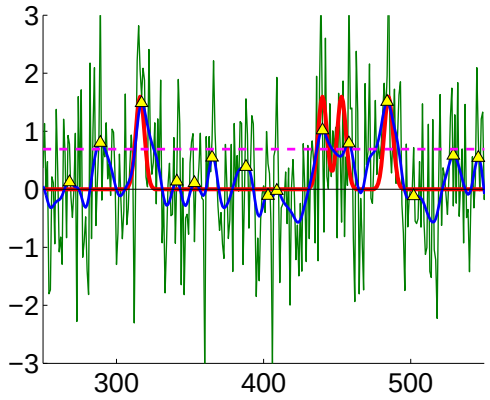

Figure 1: (a) Segment of a simulated observed sequence $y(t)$ (green) and smoothed sequence $y_\gamma(t)$ (blue) over four underlying true peaks as part of $\mu(t)$ (red). Out of 15 local maxima of $y_\gamma(t)$ (yellow), the BH detection threshold (dashed magenta) selects 5, 1 of which is a false positive. Note that this bandwidth is able to distinguish the overlapping peaks.

An interesting and challenging aspect of inference for local maxima is the fact that the number of tests, which is equal to the number of observed local maxima, is random. The multiple testing literature usually assumes the number of tests to be fixed. We overcome this difficulty with an asymptotic argument for large search space, so that by ergodicity, the error behaves approximately as it would if the number of tests were equal to its expected value.

The asymptotics for large search space are combined here with asymptotics for strong signal. The strong signal assumption solves the problem of the signal spreading into the null region as a consequence of smoothing, as it implies that each signal peak is represented by only one observed local maximum within the true domain with probability tending to one. The strong signal assumption is not restrictive in the sense that the search space may grow exponentially faster. Simulations show that error levels are maintained at finite search spaces and moderate signal strength.

Defining detection power as the expected fraction of true peaks detected, we prove that the algorithm is consistent in the sense that its power tends to one under the above asymptotic conditions. We 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 so-called matched filter theorem in signal processing (Pratt, 1991; Simon, 1995). This optimal bandwidth is much larger than the usual optimal bandwidth for nonparametric regression.

In one dimension, the problem of identifying significant local maxima is similar to that of peak detection in signal processing (e.g. (Baccus and Meister, 2002; Yasui et al., 2003; Morris et al., 2006; Harezlak et al., 2008; Brutti et al., 2005; Arzeno et al., 2008)). In this literature, though large, the detection threshold is predominantly chosen heuristically and conservatively. Our multiple testing viewpoint provides a formal mechanism for choosing the detection threshold, allowing detection under higher noise conditions. This view also eliminates the need to estimate an unknown number of peak location parameters, encountered in the signal estimation approach (O'Brien et al., 1994; Li and Speed, 2000, 2004; Tibshirani et al., 2005).

We illustrate our procedure with a data set of neural electrical recordings, where the objective is to detect action potentials representing cell activity (Baccus and Meister, 2002;

Segev et al., 2004). The noise parameters and signal peak shape are estimated from a training set and then applied to a test set for peak detection.

The data analysis and all simulations were implemented in `Matlab`.

2 Theory

2.1 The model

Consider the signal-plus-noise model

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

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

$$\mu(t) = \sum_{j=-\infty}^{\infty} a_j h_j(t), \qquad a_j > 0, \tag{2}$$

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 $w_\gamma(t) \geq 0$ with bandwidth barameter $\gamma > 0$ be a unimodal kernel with compact connected support and unit action. Convolving the process (1) with the kernel $w_\gamma(t)$ results in the smoothed process

$$y_\gamma(t) = w_\gamma(t) * y(t) = \int_{-\infty}^{\infty} w_\gamma(t-s) y(s)\, ds = \mu_\gamma(t) + z_\gamma(t), \tag{3}$$

where the smoothed signal and smoothed noise are defined as

$$\mu_\gamma(t) = w_\gamma(t) * \mu(t) = \sum_{j=-\infty}^{\infty} a_j h_{j,\gamma}(t), \qquad z_\gamma(t) = w_\gamma(t) * z(t). \tag{4}$$

For each j, the smoothed peak shape $h_{j,\gamma}(t) = w_\gamma(t) * h_j(t) \geq 0$ is unimodal 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 (but this is not required in practice, as shown via simulations in Section 3). The smoothed noise $z_\gamma(t)$ defined by (3) and (4) is assumed to be a zero-mean thrice differentiable stationary ergodic Gaussian process.

2.2 The STEM algorithm

Suppose we observe $y(t)$ defined by (1) in the segment $[-L/2, L/2]$, which contains J peaks. We call the following procedure STEM (Smoothing and TEsting of Maxima).

Algorithm 2.1 (STEM algorithm)**.**

1. *Kernel smoothing*: Construct the process (3), ignoring the boundary effects at $\pm L/2$.
2. *Candidate peaks*: Find the set of local maxima of $y_\gamma(t)$ in $[-L/2, L/2]$

$$\tilde{T} = \left\{ t \in \left[-\frac{L}{2}, \frac{L}{2}\right] : \quad \dot{y}_\gamma(t) = \frac{dy_\gamma(t)}{dt} = 0, \quad \ddot{y}_\gamma(t) = \frac{d^2 y_\gamma(t)}{dt^2} < 0 \right\}. \tag{5}$$

3. *P-values*: For each $t \in \tilde{T}$ compute the p-value $p_\gamma(t)$ for testing the hypothesis

$$\mathcal{H}_0(t):\ \mu(t) = 0 \quad \text{vs.} \quad \mathcal{H}_A(t):\ \mu(t) > 0, \qquad t \in \tilde{T}.$$

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

Steps 1 and 2 above are well defined under the model assumptions (for data on a grid, local maxima are defined as points higher than their neighbors). Step 3 is detailed in Section 2.3 below. For Step 4, we use the Bonferroni procedure to control FWER and the BH procedure to control FDR. To apply Bonferroni at level α, declare significant all peaks whose p-values are smaller than $\alpha/\tilde{m}$. To apply BH at level α, find the largest index k for which the i-th smallest p-value is smaller than $i\alpha/\tilde{m}$, and declare as significant the k peaks with smallest p-values. Notice that, in contrast to the usual application of the Bonferroni and BH procedures, the number of tests $\tilde{m}$ is random.

2.3 P-values

Given the observed heights $y_\gamma(t)$ at the local maxima $t \in \tilde{T}$, the p-values in Step 3 of Algorithm 2.1 are computed as

$$p_\gamma(t) = F_\gamma[y_\gamma(t)], \qquad t \in \tilde{T}, \tag{6}$$

where

$$F_\gamma(u) = \mathrm{P}\Big\{ z_\gamma(t) > u \ \Big|\ t \in \tilde{T} \Big\} \tag{7}$$

denotes the *right* cumulative distribution function (cdf) of $z_\gamma(t)$ at the local maxima $t \in \tilde{T}$, evaluated under the complete null hypothesis $\mu(t) = 0, \forall t$.

The distribution (7) is called a Palm distribution (Adler et al., 2010, Ch. 6) and is not Gaussian, as the conditioning may incorrectly suggest, but stochastically greater. This is because the point of evaluation $t \in \tilde{T}$ is not a fixed point $t \in \mathbb{R}$ but the random location of a local maximum of $z_\gamma(t)$. Moreover, the conditioning event has probability zero. The Palm distribution (7) has a closed-form expression, originally obtained by Cramér and Leadbetter (1967, Ch. 11) (Eq. 11.6.14), using the well known Kac-Rice formula (Rice, 1945), (Adler and Taylor, 2007, Ch. 11). A direct application, borrowing notation from those sources, gives the following.

Proposition 2.2. *Suppose the assumptions of Section 2.1 hold and that $\mu(t) = 0, \forall t$. Define the moments*

$$\sigma^2_\gamma = \mathrm{var}[z_\gamma(t)], \qquad \lambda_{2,\gamma} = \mathrm{var}[\dot{z}_\gamma(t)], \qquad \lambda_{4,\gamma} = \mathrm{var}[\ddot{z}_\gamma(t)]. \tag{8}$$

Then the distribution (7) *is given by*

$$F_\gamma(u) = 1 - \Phi\left(u\sqrt{\frac{\lambda_{4,\gamma}}{\Delta}}\right) + \sqrt{\frac{2\pi\lambda^2_{2,\gamma}}{\lambda_{4,\gamma}\sigma^2_\gamma}}\phi\left(\frac{u}{\sigma_\gamma}\right)\Phi\left(u\sqrt{\frac{\lambda^2_{2,\gamma}}{\Delta\sigma^2_\gamma}}\right), \tag{9}$$

where $\Delta = \sigma^2_\gamma\lambda_{4,\gamma} - \lambda^2_{2,\gamma}$, and $\phi(x)$, $\Phi(x)$, are the standard normal density and cdf, respectively.

The quantities σ_γ^2, $\lambda_{2,\gamma}$, and $\lambda_{4,\gamma}$ in Proposition 2.2 depend on the kernel $w_\gamma(t)$ and the autocorrelation function of the original noise process $z(t)$. Explicit expressions may be obtained, for instance, for the following Gaussian autocorrelation model, which we use later in the simulations.

Example 2.3 (Gaussian autocorrelation model). Let the noise $z(t)$ in (1) be constructed as

$$z(t) = \sigma \int_{-\infty}^{\infty} \frac{1}{\nu} \phi\left(\frac{t-s}{\nu}\right) dB(s), \qquad \sigma, \nu > 0,$$

where $B(s)$ is standard Brownian motion and $\nu > 0$. Convolving with a Gaussian kernel $w_\gamma(t) = (1/\gamma)\phi(t/\gamma)$ with $\gamma > 0$ as in (4) produces a zero-mean infinitely differentiable stationary ergodic Gaussian process

$$z_\gamma(t) = w_\gamma(t) * z(t) = \sigma \int_{-\infty}^{\infty} \frac{1}{\xi} \phi\left(\frac{t-s}{\xi}\right) dB(s), \qquad \xi = \sqrt{\gamma^2 + \nu^2},$$

with moments (8) given by

$$\sigma_\gamma^2 = \frac{\sigma^2}{2\sqrt{\pi}\xi}, \qquad \lambda_{2,\gamma} = \frac{\sigma^2}{4\sqrt{\pi}\xi^3}, \qquad \lambda_{4,\gamma} = \frac{3\sigma^2}{8\sqrt{\pi}\xi^5} \tag{10}$$

(details given in Section 6.1). The above expressions may be used as approximations if the kernel, required to have finite support, is truncated at $t = \pm c\gamma$ for moderately large c, say $c = 3$.

2.4 Error definitions

Since the smoothing operation (Step 1 in Algorithm 2.1) in the presence of noise can shift the location of a true peak, we define a detected peak to be a true positive if it falls anywhere inside the support of a true peak. Conversely, we define it to be a false positive if it falls outside the support of any true peak. Assuming the model of Section 2.1, define the signal region $\mathbb{S}_1$ and null region $\mathbb{S}_0$ respectively by

$$\mathbb{S}_1 = \bigcup_{j=1}^{J} S_j \quad \text{and} \quad \mathbb{S}_0 = \left[-\frac{L}{2}, \frac{L}{2}\right] \setminus \left(\bigcup_{j=1}^{J} S_j\right). \tag{11}$$

For a significance threshold u, the total number of detected peaks and the number of falsely detected peaks are

$$R(u) = \#\{t \in \tilde{T} : y_\gamma(t) > u\} \quad \text{and} \quad V(u) = \#\{t \in \tilde{T} \cap \mathbb{S}_0 : y_\gamma(t) > u\},$$

respectively. Both are defined as zero if $\tilde{T}$ is empty. The FWER is defined as the probability of obtaining at least one falsely detected peak

$$\text{FWER}(u) = \text{P}\{V(u) \geq 1\} = \text{P}\left\{\tilde{T} \cap \mathbb{S}_0 \neq \emptyset \text{ and } \max_{t \in \tilde{T} \cap \mathbb{S}_0} y_\gamma(t) > u\right\}. \tag{12}$$

The FDR is defined as the expected proportion of falsely detected peaks

$$\text{FDR}(u) = \text{E}\left\{\frac{V(u)}{R(u) \vee 1}\right\}. \tag{13}$$

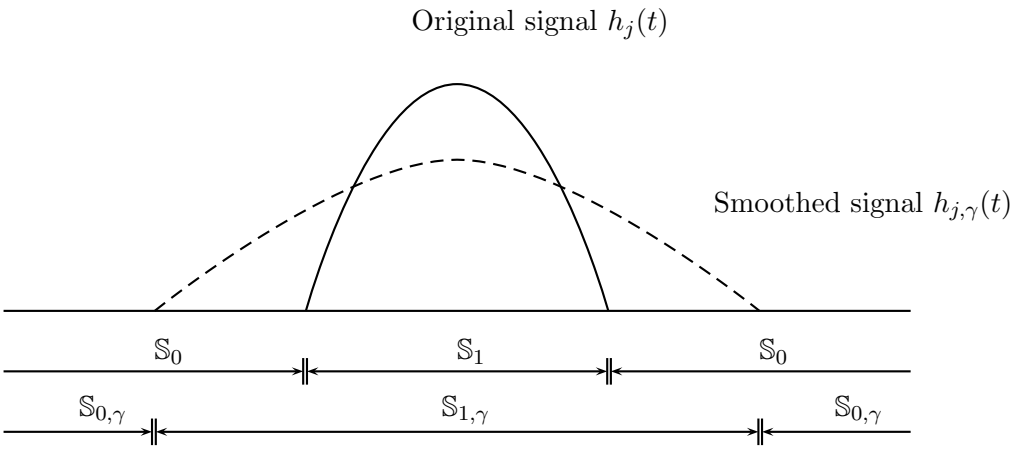

Figure 2: Schematic signal and null regions, before and after smoothing, in the vicinity of one signal peak.

Note that the above definitions are with respect to the original signal support $\mathbb{S}_1$, while the inference is carried out using the smoothed observed process $y_\gamma(t)$. Kernel smoothing enlarges the signal support and increases the probability of obtaining false positives in the null regions neighboring the signal (Pacifico et al., 2004b). In contrast to (11), the smoothed signal region $\mathbb{S}_{1,\gamma} \supset \mathbb{S}_1$ and smoothed null region $\mathbb{S}_{0,\gamma} \subset \mathbb{S}_0$ are

$$\mathbb{S}_{1,\gamma} = \bigcup_{j=1}^{J} S_{j,\gamma} \quad \text{and} \quad \mathbb{S}_{0,\gamma} = \left[-\frac{L}{2}, \frac{L}{2}\right] \setminus \left(\bigcup_{j=1}^{J} S_{j,\gamma}\right), \tag{14}$$

respectively (Figure 2). We call the difference between the expanded signal support and the true signal support the transition region

$$\mathbb{T}_\gamma = \mathbb{S}_{1,\gamma} \setminus \mathbb{S}_1 = \mathbb{S}_0 \setminus \mathbb{S}_{0,\gamma} = \bigcup_{j=1}^{J} T_{j,\gamma}, \tag{15}$$

where $T_{j,\gamma} = S_{j,\gamma} \setminus S_j$ is the transition region corresponding to each peak j.

2.5 Strong control of FWER

In Algorithm 2.1, Step 3 produces a list of $\tilde{m}$ p-values. If the Bonferroni correction is applied in Step 4 with level $\alpha \in (0, 1)$, then the null hypothesis $\mathcal{H}_0(t)$ at $t \in \tilde{T}$ is rejected if

$$p_\gamma(t) < \frac{\alpha}{\tilde{m}} \qquad \Longleftrightarrow \qquad y_\gamma(t) > \tilde{u}_{\text{Bon}} = F_\gamma^{-1}\left(\frac{\alpha}{\tilde{m}}\right), \tag{16}$$

where $\alpha/\tilde{m}$ is defined as 1 if $\tilde{m} = 0$. Recall that, in contrast to the usual Bonferroni algorithm, the number of p-values $\tilde{m}$ is random.

Define the conditions:

(C1) The assumptions of Section 2.1 hold.

(C2) $L \to \infty$ and $a = \inf_j a_j \to \infty$, such that $(\log L)/a^2 \to 0$ and $J/L \to A_1$ with $0 < A_1 < 1$.

Theorem 2.4. *Suppose that Algorithm 2.1 is applied with the Bonferroni threshold* $\tilde{u}_{\text{Bon}}$ (16)*. Then, under conditions (C1) and (C2),*

$$\limsup \text{FWER}(\tilde{u}_{\text{Bon}}) \leq \alpha.$$

The proof of Theorem 2.4 is given in Section 6.3. The large search space assumption in (C2) solves the problem of $\tilde{m}$ being random, implying that by the weak law of large numbers, the ratio $\tilde{m}/L$ is close to its expectation $\text{E}[\tilde{m}/L]$ for large L, and thus the Bonferroni procedure with random threshold (16) has asymptotically the same error control properties as if the threshold were deterministic and equal to

$$u^*_{\text{Bon}} = F_\gamma^{-1}\left(\frac{\alpha}{\text{E}[\tilde{m}]}\right) \approx F_\gamma^{-1}\left(\frac{\alpha/L}{A_1 + \text{E}[\tilde{m}_{0,\gamma}(0,1)]}\right) \tag{17}$$

where

$$\text{E}[\tilde{m}_{0,\gamma}(0,1)] = \frac{1}{2\pi}\sqrt{\frac{\lambda_{4,\gamma}}{\lambda_{2,\gamma}}} \tag{18}$$

is the expected number of local maxima of $z_\gamma(t)$ in the unit interval $(0,1)$ (Cramér and Leadbetter, 1967, Ch. 10). The strong signal assumption implies (Lemma 6.3 in Section 6.2) that the expected number of local maxima in the transition region $\mathbb{T}_\gamma$ (15) converges to 0, so that smoothing does not produce error inflation there. It also implies that the number of local maxima in the signal region $\mathbb{S}_1$ is J with probability tending to 1, providing the approximation in (17). The asymptotic rates are exponential and controlled partially by the smallest absolute derivative of the smoothed peak shape in the transition region and the curvature of the smoothed peak shape at the mode.

2.6 Control of FDR

Suppose the BH procedure is applied in Step 4 of Algorithm 2.1. For a fixed $\alpha \in (0,1)$, let k be the largest index for which the ith smallest p-value is less than $i\alpha/\tilde{m}$. Then the null hypothesis $\mathcal{H}_0(t)$ at $t \in \tilde{T}$ is rejected if

$$p_\gamma(t) < \frac{k\alpha}{\tilde{m}} \qquad \Longleftrightarrow \qquad y_\gamma(t) > \tilde{u}_{\text{BH}} = F_\gamma^{-1}\left(\frac{k\alpha}{\tilde{m}}\right), \tag{19}$$

where $k\alpha/\tilde{m}$ is defined as 1 if $\tilde{m} = 0$.

Theorem 2.5. *Suppose that Algorithm 2.1 is applied with the BH threshold* $\tilde{u}_{\text{BH}}$ (19)*. Then, under conditions (C1) and (C2),*

$$\limsup \text{FDR}(\tilde{u}_{\text{BH}}) \leq \alpha.$$

The proof of Theorem 2.5 is given in Section 6.4. The asymptotic assumptions (C2), imply that the BH procedure with random threshold (19) has asymptotically the same error control properties as if the threshold were deterministic and equal to

$$u^*_{\text{BH}} = F_\gamma^{-1}\left(\frac{\alpha A_1}{A_1 + \text{E}[\tilde{m}_{0,\gamma}(0,1)](1-\alpha)}\right), \tag{20}$$

where $\mathrm{E}[\tilde{m}_{0,\gamma}(0,1)]$ is given by (18). The threshold (19) can be viewed as the largest solution of the equation $\alpha G(u) = F_\gamma(u)$, where $G(u)$ is the empirical right cumulative distribution function of $y_\gamma(t),\ t \in \tilde{T}$ (Genovese et al., 2002). Thus the threshold (20) is obtained by taking the limit of that equation as L gets large, combined with the result that there exists exactly one significant local maximum at each true peak with probability tending to 1 for a large (Lemma 6.3 in Section 6.2). As before, the asymptotic rates are exponential and controlled partially by the smallest absolute derivative of the smoothed peak shape in the transition region and the curvature of the smoothed peak shape at the mode.

Notice that, in contrast to the asymptotic Bonferroni threshold u^*_{Bon} (17) which grows unbounded with increasing L, the asymptotic BH threshold u^*_{BH} (20) is finite.

2.7 Power

Recall from Section 2.4 that a significant local maximum is considered a true positive if it falls in the true signal region $\mathbb{S}_1$. We define the statistical power of Algorithm 2.1 as the expected fraction of true discovered peaks

$$\mathrm{Power}(u) = \mathrm{E}\left[\frac{1}{J}\sum_{j=1}^{J} 1\left(\tilde{T} \cap S_j \neq \emptyset \text{ and } \max_{\tilde{t}\in\tilde{T}\cap S_j} y_\gamma(\tilde{t}) > u\right)\right] = \frac{1}{J}\sum_{j=1}^{J}\mathrm{Power}_j(u), \quad (21)$$

where $\mathrm{Power}_j(u)$ is the probability of detecting peak j

$$\mathrm{Power}_j(u) = \mathrm{P}\left\{\tilde{T} \cap S_j \neq \emptyset \text{ and } \max_{t\in\tilde{T}\cap S_j} y_\gamma(t) > u\right\}. \quad (22)$$

The maximum operator above indicates that if more than one significant local maximum fall within the same peak support, only one is counted, so power is not inflated. However, 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 (Lemma 6.3 in Section 6.2). The next result indicates that both the Bonferroni and BH procedures are asymptotically consistent. The proof is given in Section 6.5.

Theorem 2.6. *Let the power be defined by* (21) *and let* $\tilde{u}_{\mathrm{Bon}}$ *and* $\tilde{u}_{\mathrm{BH}}$ *be the Bonferroni and BH thresholds* (16) *and* (19)*, respectively. Under conditions (C1) and (C2),*

$$\mathrm{Power}(\tilde{u}_{\mathrm{Bon}}) \to 1, \qquad \mathrm{Power}(\tilde{u}_{\mathrm{BH}}) \to 1.$$

For pointwise tests, if there exists a signal anywhere, the BH procedure is more powerful than the Bonferroni procedure (Benjamini and Hochberg, 1995). This is also true in our case. Comparing (17) and (20), if $J \geq 1$, the threshold u^*_{Bon} is higher than the threshold u^*_{BH}, promising a larger power for the BH procedure.

2.8 Optimal smoothing kernel

The best smoothing kernel $w_\gamma(t)$ is that which maximizes the power (21) under the true model. Because this maximization is analytically difficult, we resort to a less formal argument here. Lemma 6.3 in Section 6.4 shows that, under conditions (C1) and (C2), every true peak j is represented by exactly one significant local maximum located in a small

neighborhood containing the true peak mode τ_j with probability tending to 1. Thus the power for peak j (22) may be approximated as

$$\text{Power}_j(u) \approx \text{P}\left\{y_\gamma(\tau_j) > u\right\} = \Phi\left[\frac{a_j h_{j,\gamma}(\tau_j) - u}{\sigma_\gamma}\right], \tag{23}$$

because $y_\gamma(\tau_j) \sim N(a_j h_{j,\gamma}(\tau_j), \sigma_\gamma^2)$. By Lemma 6.6 in Section 6.5, the asymptotically equivalent thresholds (17) and (20) for the Bonferroni and BH procedures satisfy $u^*_{\text{Bon}}/a_j \to 0$ and $u^*_{\text{BH}}/a_j \to 0$ for any j. Thus, for large a_j, the power (23) is maximized approximately by maximizing the SNR

$$\text{SNR}_\gamma = \frac{a_j h_{j,\gamma}(\tau_j)}{\sigma_\gamma} = \frac{a_j \int_{-\infty}^{\infty} w_\gamma(s) h_j(s)\, ds}{\sigma \sqrt{\int_{-\infty}^{\infty} w_\gamma^2(s)\, ds}}, \tag{24}$$

where σ is the standard deviation of the observed process $y(t)$. The optimal smoothing kernel $w_\gamma(t)$ is that which is closest to $h_j(t)$ in an L_2 sense. This result is similar to the matched filter theorem for detecting a single signal peak of known shape at a fixed time location t (Pratt, 1991; Simon, 1995). The result only holds approximately in our case because the peak locations are unknown.

Example 2.7 (Gaussian autocorrelation model). Suppose the signal peak j is a truncated Gaussian density $h_j(t) = (1/b_j)\phi[(t-\tau_j)/b_j]\mathbf{1}[-c_j, c_j]$, $b_j, c_j > 0$, and let the noise be generated as in Example 2.3. Ignoring the truncation, $h_{j,\gamma}(t) = w_\gamma(t) * h_j(t)$ in (24) is the convolution of two Gaussian densities with variances γ^2 and b_j^2, which is another Gaussian density with variance $\gamma^2 + b_j^2$. Using (10), we have that

$$\text{SNR}_\gamma = \frac{a_j h_{j,\gamma}(\tau_j)}{\sigma_\gamma} = \frac{a_j}{\sigma \pi^{1/4}} \left[\frac{\gamma^2 + \nu^2}{(\gamma^2 + b_j^2)^2}\right]^{1/4}. \tag{25}$$

As a function of γ, the SNR is maximized at

$$\underset{\gamma}{\text{argmax}}\, \text{SNR}_\gamma = \begin{cases} \sqrt{b_j^2 - 2\nu^2}, & \nu < b_j/\sqrt{2} \\ 0, & \nu > b_j/\sqrt{2} \end{cases} \tag{26}$$

In particular, when $\nu = 0$, 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 (26).

3 Simulation studies

3.1 Non-asymptotic performance

Simulations were used to evaluate the performance of the STEM algorithm for finite range L and moderate signal strength a. In a segment of length $L = 1000$, $J = 10$ equal truncated Gaussian peaks $a_j h_j(t) = a/b\ \phi[(t-\tau_j)/b]\mathbf{1}[-cb, cb]$, $j = 1, \ldots, J$, as in Example 2.7 with $b = 3$, $c = 3$ and varying a, were placed at uniformly spaced locations $\tau_j = (j - 1/2)L/J$, $j = 1, \ldots, J$ and sampled at integer values of t. The noise $z(t)$ was constructed as in Example 2.3 with $\sigma = 1$ and varying ν. As smoothing kernel, we used a truncated Gaussian density

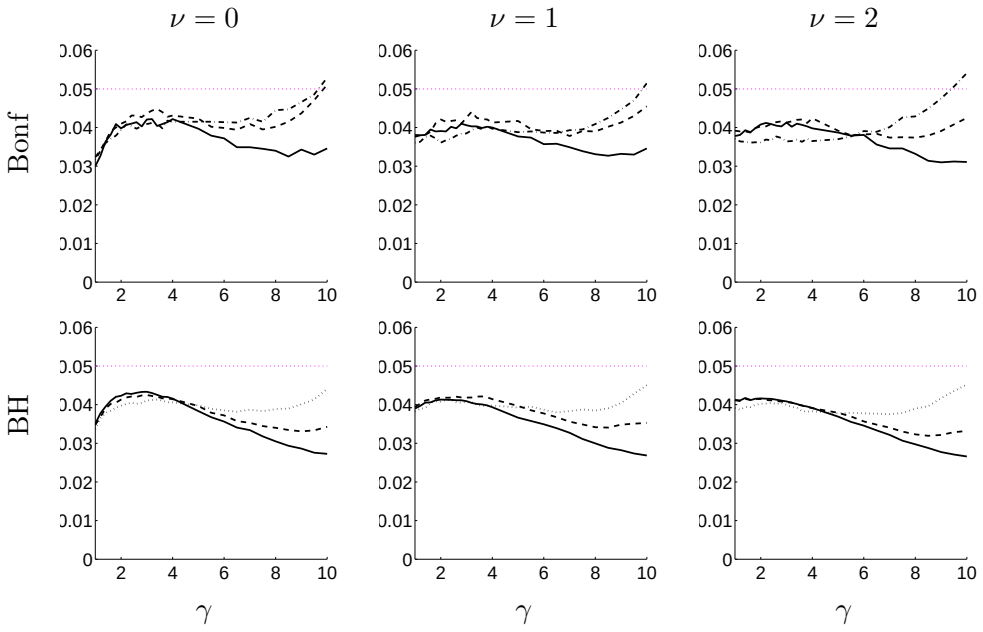

Figure 3: FWER of the Bonferroni procedure (top row) and FDR of the BH procedure (bottom row) for $a = 15$ (solid), $a = 12$ (dashed) and $a = 9$ (dot-dashed). Nominal level is 0.05.

$w_\gamma(t) = (1/\gamma)\phi(t/\gamma)\mathbf{1}[-c\gamma, c\gamma]$ as in Example 2.3 with $c = 3$ and varying γ. Algorithm 2.1 was carried out using the noise parameters (10), and the Bonferroni and BH procedures, both at level $\alpha = 0.05$.

Figure 3 shows the realized FWER and FDR levels of the Bonferroni and BH procedures, evaluated according to (12) and (13) with the expectations replaced by ensemble averages over 10,000 replications. Error rates are maintained below the nominal level $\alpha = 0.05$, except when the bandwidth γ is large compared to the signal peak bandwidth $b = 3$. The increased error rates are the result of true peak maxima being moved away from the original signal region $\mathbb{S}_1$ into the transition region $\mathbb{T}_\gamma$, where they are counted as false positives. This phenomenon disappears with increasing signal strength because the probability of obtaining any local maxima in the transition region goes to zero exponentially with increasing a (Lemmas 6.2 and 6.3).

Figure 4 shows the realized power of the Bonferroni and BH procedures, evaluated according to (21) with the expectations replaced by ensemble averages over the same 10,000 replications. In all cases, the power increases asymptotically to 1 with the signal strength for every fixed bandwidth, and is always larger for BH than it is for Bonferroni. To understand the dependence on bandwidth, superimposed is the theoretical approximate power (23) evaluated at the asymptotic thresholds u^*_{Bon} (17) and u^*_{BH} (20) and plugging in the SNR (25). The "theoretical" power curves largely capture the shape of the realized ones, but are lower because the asymptotic thresholds are larger than the actual thresholds for finite a. The curve shape is mostly determined by the SNR (25) as a function of γ. The bandwidth γ producing the largest power is always larger than the theoretical optimal bandwidth (26), but it approaches it from the right as a increases.

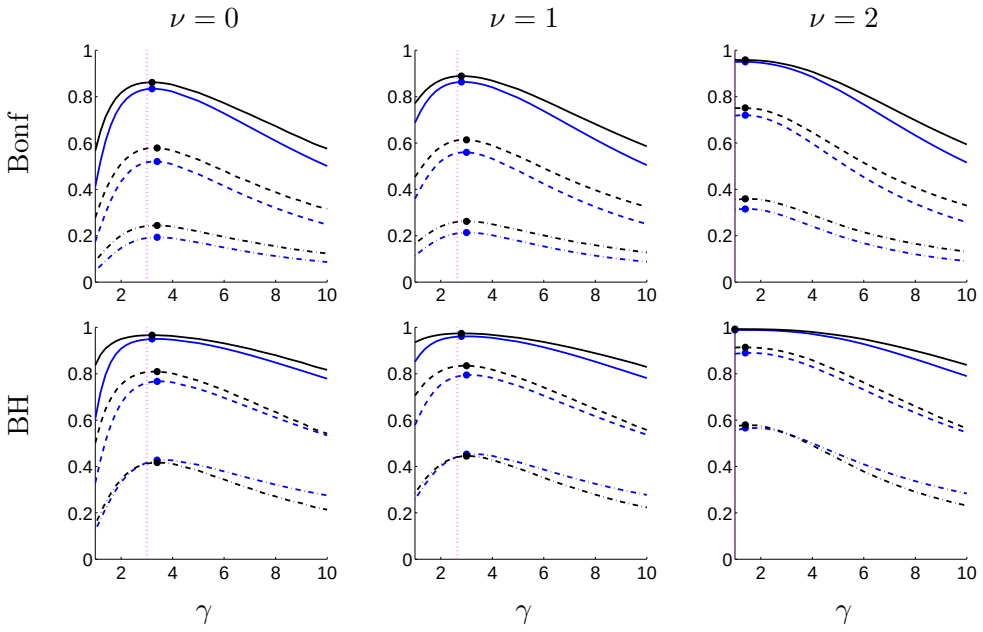

Figure 4: Realized (black) and "theoretical" (blue) power of the Bonferroni (top row) and BH (bottom row) procedures for $a = 15$ (solid), $a = 12$ (dashed) and $a = 9$ (dot-dashed). The maxima of the curves (solid circles) approach the asymptotic optimal bandwidth (vertical dashed).

3.2 Overlapping peaks

While the theory of Section 2 assumed that the signal peaks had non-overlapping supports, this is not necessary in practice. Further similar simulations with $J = 10$ partially overlapping peaks showed that the error rates were below the nominal level regardless of the amount of overlap between peaks. The detection power, however, increased with increasing overlap. This is because definition (21) counts two overlapping peaks as detected if a significant local maxima is found in the overlapping region between them, as it belongs to both. Definition (21) does not measure the ability to distinguish between overlapping peaks, which certainly decreases with increasing overlap.

3.3 Comparison with pointwise testing

To see the benefits of dimension reduction by testing local maxima, Table 1 compares the performance of the STEM algorithm (with Bonferroni and BH corrections) to three other methods that test at every single location. Simulated data sets as in Section 3.1 with $a = 12$, $b = 3$ and $\nu = 0$ were smoothed with $\gamma = 3$. For methods 'Bonf-all' and 'BH-all', p-values for testing $H_0 : \mu(t) = 0$ at each t were computed as $p(t) = 1 - \Phi[y_\gamma(t)/\sigma_\gamma]$, $t = 1, \ldots, L = 1000$ and then corrected using Bonferroni and BH, respectively. The STEM algorithm tested in average 65.3 p-values instead of 1000. Table 1 indicates that the global Bonferroni correction is too conservative. On the other hand, the global BH correction is designed to control FDR at the level of individual locations, and thus produces too many false positives when the FDR is measured in terms of detected peaks using (13).

The method 'Supremum' was adapted from Worsley et al. (1996a) as follows. The

Table 1: Performance of FWER methods (STEM with Bonferroni, Bonferroni on all L locations, Supremum) and FDR methods (STEM with BH, BH on all L locations) averaged over 10000 simulations. Nominal error rate is 0.05.

	STEM-Bonf	Bonf-all	Supremum	STEM-BH	BH-all
Error rate	0.0432	0.0148	0.0151	0.0429	0.0769
Power	0.6014	0.4979	0.5009	0.8246	0.8784

probability that the supremum of any differentiable random process $f(t)$ in the interval $[0, T]$ exceeds u is bounded by (Adler and Taylor, 2007)

$$\mathrm{P}\left(\sup_{t\in[0,T]} f(t) \geq u\right) \leq \mathrm{P}\left[f(0) \geq u\right] + \mathrm{E}[N_u], \tag{27}$$

where N_u is the number of up-crossings by $f(t)$ of the level u in $[0, T]$. For the stationary Gaussian process $z_\gamma(t)$, application of the Kac-Rice formula (Cramér and Leadbetter, 1967, p. 194) gives that $\mathrm{E}[N_u] = L(\sqrt{\lambda_{2,\gamma}}/\sigma_\gamma)\phi(u/\sigma_\gamma)$. The significance threshold is found as the largest u such that

$$\mathrm{P}\left(\sup_{t\in[-L/2,L/2]} z_\gamma(t) \geq u\right) \leq 1 - \Phi\left(\frac{u}{\sigma_\gamma}\right) + L\frac{\sqrt{\lambda_{2,\gamma}}}{\sigma_\gamma}\phi\left(\frac{u}{\sigma_\gamma}\right) \leq \alpha. \tag{28}$$

Table 1 indicates that, despite accounting explicitly for the noise autocorrelation, the 'Supremum' method performs only slightly better than global Bonferroni, and not as well as Bonferroni performed on local maxima.

3.4 Automatic bandwidth selection

If the signal bandwidth b is unknown, the bandwidth may be chosen from the data as the one that yields the largest number of discoveries for a fixed error level. For simulated data sets as in Section 3.1 with $a = 12$, $b = 3$ and $\nu = 0$, the STEM algorithm was applied with γ ranging from 1 to 10, and results were retained for the bandwidth $\hat{\gamma}$ that yielded the largest number of discoveries. The realized FWER and FDR over 1000 simulations were 0.069 for Bonferroni and 0.062 for BH, with respective realized powers 0.630 and 0.830. The power is about the same as that obtained when γ is fixed and chosen optimally (Figure 4), but the error rates are increased because of the variable bandwidth. Figure 5 shows that the chosen bandwidth is sometimes too large, shifting true peaks into the neighboring null region and mistaking them for false ones.

4 Data example

The data consists of recordings from a single electrode inserted in a salamander's retina, digitized at a sampling frequency of 10 KHz. Data of these kind are routinely collected in large amounts in neuroscience experiments (Baccus and Meister, 2002; Segev et al., 2004). For the purposes of this paper, three data sets were used:

1. Test set: 60 sec of recordings of live cells in the dark.
2. Training set 1: 60 sec of recordings of live cells in the dark.

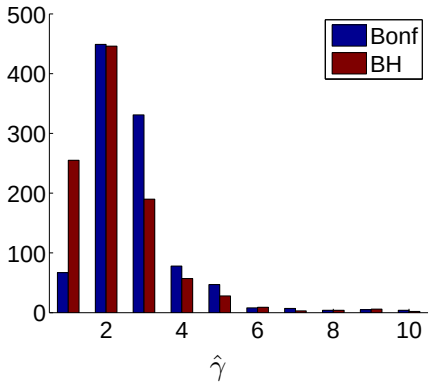

Figure 5: Histograms of automatically chosen smoothing bandwidth $\hat{\gamma}$ over 1000 simulations. Averages are 2.73 for Bonferroni and 2.30 for BH. True signal bandwidth is $b = 3$.

3. Training set 2: 60 sec of recordings after the retina was allowed to die.

Each period of 60 sec corresponds to $L = 6 \times 10^5$ samples. The goal of the analysis was to detect neuronal spikes in the test set (Figure 6, top left).

Assuming that neuronal action potentials have similar shapes, to maximize the SNR (24), the smoothing filter should be close in shape and bandwidth to that of the peaks to be detected. Training set 1 was used to estimate the peak shape. In training set 1, spikes with raw maximum exceeding 1 were selected and aligned by their maxima (Figure 7a). The peak shape template was obtained as the average of the 23 selected major spikes and truncated to a length of 100 samples.

Training set 2, recorded under pure noise conditions, was used to estimate the noise parameters. The noise in training set 2 can be well modeled by an AR(3) process with autoregressive coefficients -1.13, 0.42, and -0.13 estimated by the Yule-Walker algorithm, so that whitening with these coeffcients produces a process whose autocovariance function cannot be distinguished from that of white noise using a Bartlett's test. A similar analysis in segments of length $L/10$ showed that the estimated AR coefficients have a coefficient of variation of no more than 1% over the 10 segments, supporting the stationarity assumption. A Jarque-Bera test of normality for the entire sequence returned a p-value of 0.224, supporting the Gaussianity assumption.

Convolving training set 2 with the template of Figure 7a produced smoothed noise with spectral moments $\hat{\sigma}^2_\gamma = 4.22 \times 10^{-4}$, $\hat{\lambda}^2_{2,\gamma} = 1.20 \times 10^{-4}$ and $\hat{\lambda}^2_{4,\gamma} = 1.96 \times 10^{-4}$, estimated respectively by the empirical variances of the observed process, its first order difference and its second order difference. Given the length of the process, the variance of these estimates is negligible.

Algorithm 2.1 was applied to the test set (Figure 6, top left) by convolving it with the template of Figure 7a, producing the smoothed process in Figure 6 (bottom left). In $L = 6 \times 10^5$ samples, $\tilde{m} = 30426$ local maxima were found and their p-values were computed according to (6) and (9), plugging in the estimates $\hat{\sigma}^2_\gamma$, $\hat{\lambda}^2_{2,\gamma}$ and $\hat{\lambda}_{4,\gamma}$ found above. The empirical cdf of the p-values (Figure 7b) shows a large fraction of non-null p-values near 0. For comparison, the same procedure of smoothing, finding local maxima and computing their p-values was applied to training set 2. The empirical cdf of those p-values is virtually uniform, emphasizing that formula (9) for Gaussian noise is appropriate. Also in Figure 7b, the excess of large p-values from the test set is due to the negative portions of the

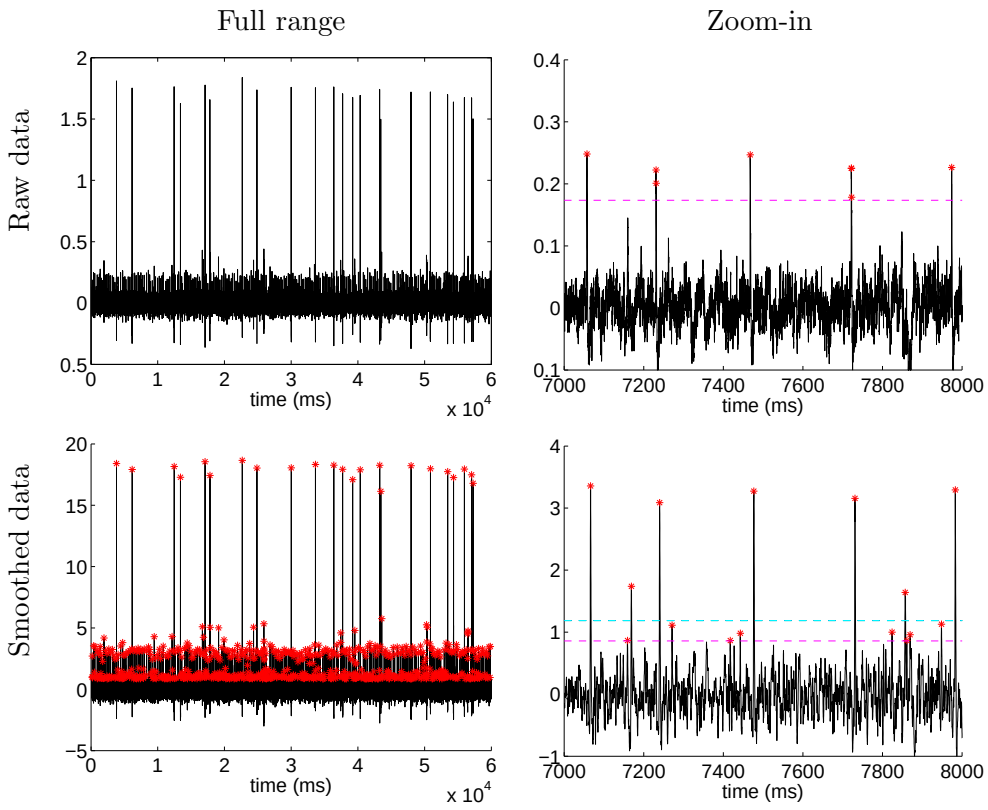

Figure 6: Top row: the neural spike data (test set); the stars in the right panel indicate peaks that are higher than 4 standard deviations of the raw data (dashed line), as suggested by Segev et al. (2004). Bottom row: the data smoothed using the estimated peak shape as kernel; the stars indicate significant local maxima higher than the BH threshold (magenta dashed line) at level 0.01. The Bonferroni threshold is indicated by the cyan dashed line.

smoothing function (Figure 7a). These produce small negative anti-spikes whose p-values are large when tested for positiveness.

Applying the BH procedure to the $\tilde{m} = 30426$ p-values obtained from the test set at FDR level 0.01 resulted in a p-value threshold of 2.76×10^{-4} and $R = 843$ significant local maxima. These are indicated in Figure 6 (bottom left), showing three levels of spike strengths. Figure 6 (bottom right) zooms in to show a few of the weaker spikes. Applying the Bonferroni procedure instead in Algorithm 1 resulted in a p-value threshold of 3.29×10^{-7} and only 411 detected spikes.

For comparison, Figure 6 (top right) shows the same segment of the raw data and the spikes selected using one of the recommended methods in the neuroscience literature, which is to threshold at 4 standard deviations of the raw data (Segev et al., 2004). Our method is able to identify a few more spikes at a low FDR level of 0.01, but more importantly, it attaches to the findings a significance level, expecting about 1% of the detected spikes to be false. The conventional method does not offer this useful statistical interpretation.

As in Section 3.3, computing p-values at each location as $p(t) = 1 - \Phi(y_\gamma(t)/\hat{\sigma}_\gamma)$, $t = 1, \ldots, L$, and applying a global Bonferroni at level 0.01 was more conservative, resulting in a height threshold of 1.235 (comparable to Figure 6 bottom right) and detecting only 393

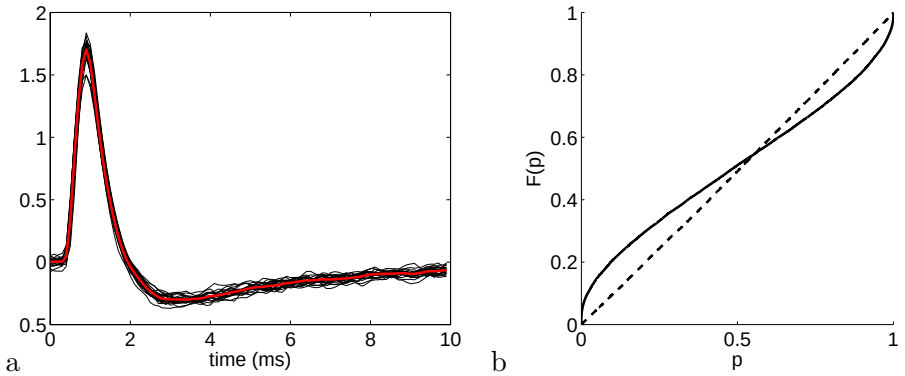

Figure 7: (a) 23 strongest spikes aligned by their maximum (black); their average (red) is the estimated template. (b) Empirical cdf of p-values for the test set (solid) and training set 2 (dashed).

spikes. Similarly, the 'Supremum' method, applied by replacing $\hat{\sigma}_\gamma$ and $\hat{\lambda}_\gamma$ in (28) at level 0.01, yielded a height threshold 1.229 and 394 detected spikes. Finally, applying the global BH procedure at level 0.01 with L p-values gave a height threshold of 0.780 detecting 1149 spikes, but as shown in Section 3.3, this result is too optimistic because the actual error rate for peaks is higher than 0.01.

5 Discussion

For the theoretical results, the most critical assumptions were that the noise process is stationary ergodic Gaussian and that the signal peaks are unimodal with compact support. The Gaussianity assumption was chosen because it enabled a closed formula for computing the p-values associated with the heights of local maxima. For non-Gaussian noise, p-values could be computed via Monte Carlo simulation.

The assumption of compact support for the signal peaks was necessary for true and false positives to be well defined. Chumbley et al. (2010) argued that testing local maxima is appropriate when the signal spreads over the entire domain, but in that case every positive is a true positive, making the inference unclear. The unimodality assumption made local maxima good representatives of true peaks, as the probability that a true peak is represented by exactly one observed local maximum tends to one as the signal strength increases. The convergence rates for this were controlled by the smallest absolute derivative of the smoothed peaks in the transition region and their smallest curvature near the mode. There was no assumption of sparsity of peaks in the theoretical results. The simulations showed that the error rates and power are almost unaffected if the peaks overlap. Agreeing with Chumbley and Friston (2009), applying BH globally resulted in inflated error rates for peaks, while applying Bonferroni or the Supremum method globally was too conservative.

While the theory was developed for continuous processes, in practice the observations are given in a discrete grid. In our simulations we found that the results were not reliable when the smoothing bandwidth was smaller than the grid spacing, as the theory for continuous random processes is no longer a good approximation in that case.

The asymptotic error control and power consistency did not require the peaks to have the same shape or width. However, the matched filter principle suggests that the smoothing

kernel should be chosen to match the peaks to be detected. If all peaks have the same shape and width, as in the neuronal data analyzed, then a single matching kernel suffices. If the peaks to be detected have different widths, then the bandwidth may be adapted to the width of each peak. We leave this possibility for future work, as well as the obliged extension of the proposed methods to two- and three-dimensional domains.

6 Technical details

6.1 Computations for the Gaussian autocorrelation model (Example 2.3)

Let $\phi_\xi(t) = (1/\xi)\phi(t/\xi)$. Since the k-th derivative of $z_\gamma(t)$ has zero mean, its variance is equal to the second moment

$$\mathrm{E}\left\{[z_\gamma^{(k)}(t)]^2\right\} = \sigma^2 \int_{-\infty}^{\infty} \left[\phi_\xi^{(k)}(t-s)\right]^2 ds = \frac{\sigma^2}{\xi^{2k+2}} \int_{-\infty}^{\infty} H_k^2\left(\frac{t}{\xi}\right) \phi^2\left(\frac{t}{\xi}\right) dt,$$

where $H_k(t)$ is the k-th Hermite polynomial: $H_0(x) = 1$, $H_1(x) = x$, $H_2(x) = x^2 - 1$, and so on. Replacing $\phi^2(t/\xi) = \phi\big(\sqrt{2}\, t/\xi\big)/\sqrt{2\pi}$ and making the change of variable $x = \sqrt{2}\, t/\xi$ gives

$$\mathrm{E}\left\{[z_\gamma^{(k)}(t)]^2\right\} = \frac{\sigma^2}{2\sqrt{\pi}\xi^{2k+1}} \int_{-\infty}^{\infty} H_k^2\left(\frac{x}{\sqrt{2}}\right) \phi(x)\, dx. \tag{29}$$

The expressions in (10) are obtained setting $k = 0$ for σ_γ^2, $k = 1$ for $\lambda_{2,\gamma}$, and $k = 2$ for $\lambda_{4,\gamma}$.

6.2 Supporting results

Lemma 6.1. *Let $\tilde{m}_{0,\gamma} = \#\{t \in \tilde{T} \cap \mathbb{S}_{0,\gamma}\}$ be be the number of local maxima of $y_\gamma(t)$ (or $z_\gamma(t)$) in $\mathbb{S}_{0,\gamma}$. Let $V_\gamma(u) = \#\{t \in \tilde{T} \cap \mathbb{S}_{0,\gamma} : y_\gamma(t) > u\}$ be the number of local maxima of $y_\gamma(t)$ (or $z_\gamma(t)$) in $\mathbb{S}_{0,\gamma}$ whose heights are above the level u. Then*

$$\frac{V_\gamma(u)}{\tilde{m}_{0,\gamma}} \to \frac{\mathrm{E}[V_\gamma(u)]}{\mathrm{E}[\tilde{m}_{0,\gamma}]} = F_\gamma(u)$$

in probability as $L \to \infty$, where $F_\gamma(u)$ is the Palm distribution (7).

Proof. Notice that $y_\gamma(t) = z_\gamma(t)$ for all $t \in \mathbb{S}_{0,\gamma}$, so the process $y_\gamma(t)$ has the same properties as the stationary process $z_\gamma(t)$ on the set $\mathbb{S}_{0,\gamma}$. By ergodicity, the weak law of large numbers applied to the numerator and denominator gives that

$$\frac{V_\gamma(u)}{\tilde{m}_{0,\gamma}} = \frac{\#\{t \in \tilde{T} \cap \mathbb{S}_{0,\gamma} : z_\gamma(t) > u\}/L}{\#\{t \in \tilde{T} \cap \mathbb{S}_{0,\gamma}\}/L} \tag{30}$$

converges to (Cramér and Leadbetter, 1967)

$$\frac{\mathrm{E}[\#\{t \in \tilde{T} \cap \mathbb{S}_{0,\gamma} : z_\gamma(t) > u\}]}{\mathrm{E}[\#\{t \in \tilde{T} \cap \mathbb{S}_{0,\gamma}\}]} = \frac{\mathrm{E}[V_\gamma(u)]}{\mathrm{E}[\tilde{m}_{0,\gamma}]}.$$

But also by ergodicity, the ratio (30) converges to the conditional probability $\mathrm{P}[z_\gamma(t) > u \mid t \in \tilde{T} \cap \mathbb{S}_{0,\gamma}] = F_\gamma(u)$ by definition (7). The two limits must be equal. □

Lemma 6.2. *Assume the model of Section 2.1. Let* $S_{j,\gamma} = I_j^{\text{left}} \cup I_j^{\text{mode}} \cup I_j^{\text{right}}$ *be a partition, where* $I_j^{\text{mode}} = [c_j, d_j] \subset S_j$ *is a fixed interval containing the mode of* $\mu_\gamma(t) = a_j h_{j,\gamma}(t)$ *in* S_j *as an interior point, such that* $\ddot{h}_{j,\gamma}(t) < 0$ *for* $t \in I_j^{\text{mode}}$, $\dot{h}_{j,\gamma}(t) > 0$ *for* $t \in I_j^{\text{left}}$ *and* $\dot{h}_{j,\gamma}(t) < 0$ *for* $t \in I_j^{\text{right}}$. *Let*

- M_j *be the largest value of* $|h_{j,\gamma}(t)|$ *in* $S_{j,\gamma}$;
- C_j *be the smallest value of* $|\dot{h}_{j,\gamma}(t)|$ *in* $I_j^{\text{side}} = I_j^{\text{left}} \cup I_j^{\text{right}}$;
- D_j *be the smallest value of* $|\ddot{h}_{j,\gamma}(t)|$ *in* I_j^{mode}.

For $\tilde{T}$ *given by* (5) *and any threshold* u,

$$\begin{aligned}
\mathrm{P}\left(\#\{t \in \tilde{T} \cap I_j^{\text{side}}\} = 0\right) &\geq 2\Phi\left(\frac{a_j C_j}{\sqrt{\lambda_{2,\gamma}}}\right) - 1 - \left|I_j^{\text{side}}\right| \sqrt{\frac{\lambda_{4,\gamma}}{\lambda_{2,\gamma}}} \phi\left(\frac{a_j C_j}{\sqrt{\lambda_{2,\gamma}}}\right), \\
\mathrm{P}\left(\#\{t \in \tilde{T} \cap I_j^{\text{mode}}\} = 1\right) &\geq \Phi\left(\frac{a_j D_j}{\sqrt{\lambda_{4,\gamma}}}\right) - \left|I_j^{\text{mode}}\right| \sqrt{\frac{\lambda_{6,\gamma}}{\lambda_{4,\gamma}}} \phi\left(\frac{a_j D_j}{\sqrt{\lambda_{4,\gamma}}}\right) - 2\Phi\left(\frac{-a_j C_j}{\sqrt{\lambda_{2,\gamma}}}\right), \\
\mathrm{P}\left(\#\{t \in \tilde{T} \cap I_j^{\text{mode}} : y_\gamma(t) > u\} = 1\right) &\geq \Phi\left(\frac{a_j D_j}{\sqrt{\lambda_{4,\gamma}}}\right) - \left|I_j^{\text{mode}}\right| \sqrt{\frac{\lambda_{6,\gamma}}{\lambda_{4,\gamma}}} \phi\left(\frac{a_j D_j}{\sqrt{\lambda_{4,\gamma}}}\right) - \Phi\left(\frac{u - a_j M_j}{\sigma_\gamma}\right),
\end{aligned} \tag{31}$$

where σ_γ, $\lambda_{2,\gamma}$ *and* $\lambda_{4,\gamma}$ *are given by* (8) *and* $\lambda_{6,\gamma} = \mathrm{E}[\dddot{z}_\gamma(t)]$.

Proof.

1. Consider first the compact interval I_j^{left}. The probability that there are no local maxima of $y_\gamma(t)$ in I_j^{left} is greater than the probability that $\dot{y}_\gamma(t) > 0$ for all t in the interval. This probability is equal to

$$\mathrm{P}\left(\inf_{I_j^{\text{left}}} \dot{y}_\gamma(t) > 0\right) \geq \mathrm{P}\left(\inf_{I_j^{\text{left}}} \dot{z}_\gamma(t) > -\inf_{I_j^{\text{left}}} \dot{\mu}_\gamma(t)\right) = 1 - \mathrm{P}\left(\sup_{I_j^{\text{left}}} [-\dot{z}_\gamma(t)] > a_j C_j^{\text{left}}\right) \tag{32}$$

where $C_j^{\text{left}} > 0$ is the smallest value of $\dot{h}_{j,\gamma}(t)$ in I_j^{left}. Inequality (27) applies above to the stationary Gaussian process $-\dot{z}_\gamma(t)$. The Kac-Rice formula (Cramér and Leadbetter, 1967, p. 194) gives in this case that $\mathrm{E}[N_u] = |I_j^{\text{left}}| \sqrt{\lambda_{4,\gamma}} / \sqrt{\lambda_{2,\gamma}} \phi(u/\sqrt{\lambda_{2,\gamma}})$. Thus (32) has the lower bound

$$\mathrm{P}\left(\#\{t \in \tilde{T} \cap I_j^{\text{left}})\} = 0\right) \geq \Phi\left(\frac{a_j C_j^{\text{left}}}{\sqrt{\lambda_{2,\gamma}}}\right) - \left|I_j^{\text{left}}\right| \sqrt{\frac{\lambda_{4,\gamma}}{\lambda_{2,\gamma}}} \phi\left(\frac{a_j C_j^{\text{left}}}{\sqrt{\lambda_{2,\gamma}}}\right).$$

A similar calculation for I_j^{right} gives a similar bound with the superscript "left" replaced by "right" and $C_j^{\text{right}} > 0$ being the smallest value of $|\dot{h}_{j,\gamma}(t)|$ in I_j^{right}. Putting the two together, the required probability $\mathrm{P}(\#\{t \in \tilde{T} \cap I_j^{\text{side}}\})$ that there are no local maxima in I_j^{left} nor I_j^{right} is bounded as in the first row of (31).

2. The probability that $y_\gamma(t)$ has no local maxima in I_j^{mode} is less than the probability that $\dot{y}_\gamma(c_j) \leq 0$ or $\dot{y}_\gamma(d_j) \geq 0$, for a positive derivative at c_j and a negative one at d_j would

imply the existence of at least one local maximum in I_j. Thus, the probability of no local maxima in I_j^{mode} is bounded above by

$$\begin{aligned} \mathrm{P}\left(\#\{t \in \tilde{T} \cap I_j^{\text{mode}})\} = 0\right) &\leq \mathrm{P}\left[\dot{y}_\gamma(c_j) \leq 0\right] + \mathrm{P}\left[\dot{y}_\gamma(d_j) \geq 0\right] \\ &= \Phi\left(\frac{-a_j \dot{h}_{j,\gamma}(c_j)}{\sqrt{\lambda_{2,\gamma}}}\right) + 1 - \Phi\left(\frac{-a_j \dot{h}_{j,\gamma}(d_j)}{\sqrt{\lambda_{2,\gamma}}}\right) \leq 2 - 2\Phi\left(\frac{a_j C_j}{\sqrt{\lambda_{2,\gamma}}}\right), \end{aligned} \tag{33}$$

because $\dot{y}_\gamma(t) \sim N(\dot{\mu}_\gamma(t), \lambda_{2,\gamma})$ and $\dot{h}_\gamma(c_j) > C_j > 0$ and $-\dot{h}_\gamma(d_j) > C_j > 0$.

On the other hand, the probability that $y_\gamma(t)$ has more than one local maxima in I_j^{mode} is less than the probability that $\ddot{y}_\gamma(t) > 0$ for some t in I_j^{mode}. This probability is

$$\mathrm{P}\left(\sup_{I_j^{\text{mode}}} \ddot{y}_\gamma(t) > 0\right) \leq \mathrm{P}\left(\sup_{I_j^{\text{mode}}} \ddot{z}_\gamma(t) > a_j D_j\right),$$

where $D_j < 0$ is the largest value of $\ddot{\mu}_\gamma(t) < 0$ in I_j^{mode}. Applying (27) to the process $\ddot{z}_\gamma(t)$ gives the further upper bound

$$\mathrm{P}\left(\#\{t \in \tilde{T} \cap I_j^{\text{mode}})\} \geq 1\right) \leq 1 - \Phi\left(\frac{a_j D_j}{\sqrt{\lambda_{4,\gamma}}}\right) + \left|I_j^{\text{mode}}\right| \sqrt{\frac{\lambda_{6,\gamma}}{\lambda_{4,\gamma}}} \phi\left(\frac{a_j D_j}{\sqrt{\lambda_{4,\gamma}}}\right). \tag{34}$$

Putting (33) and (34) together gives the bound in the second row of (31).

3. The probability that no local maxima of $y_\gamma(t)$ in I_j^{mode} exceed the threshold u is less than the probability that $y_\gamma(t)$ is below u anywhere in I_j^{mode}, so it is bounded above by $\Phi[(u - a_j M_j)/\sigma_\gamma]$. On the other hand, the probability that more than one local maxima of $y_\gamma(t)$ in I_j^{mode} exceed u is less than the probability that there exist more than one local maximum, which is bounded above by (34). Putting the two together gives the bound in the third row of (31). □

Lemma 6.3. *Assume the model of Section 2.1. For $\tilde{T}$ given by (5), let $\tilde{m}_{1,\gamma} = \#\{\tilde{T} \cap \mathbb{S}_{1,\gamma}\}$ be the number of local maxima in the set $\mathbb{S}_{1,\gamma}$ and recall that $W_\gamma(u) = \#\{t \in \tilde{T} \cap \mathbb{S}_{1,\gamma} : y_\gamma(t) > u\}$ is the number of local maxima in $\mathbb{S}_{1,\gamma}$ above threshold u. Under conditions (C1) and (C2),*

1. *The probability that $y_\gamma(t)$ has any local maxima in the transition region $\mathbb{T}_\gamma$ tends to 0:*
$$\mathrm{P}\left(\#\{t \in \tilde{T} \cap \mathbb{T}_\gamma\} \geq 1\right) \to 0.$$
2. *The probability to get exactly J local maxima in the set $\mathbb{S}_{1,\gamma}$,*
$$\mathrm{P}\left(\tilde{m}_{1,\gamma} = J\right) = \mathrm{P}\left(\#\{t \in \tilde{T} \cap \mathbb{S}_{1,\gamma}\} = J\right) \to 1.$$
3. *The probability to get exactly J local maxima in the set $\mathbb{S}_{1,\gamma}$ that exceed any fixed threshold u,*
$$\mathrm{P}\left[W_\gamma(u) = J\right] = \mathrm{P}\left[\#\{t \in \tilde{T} \cap \mathbb{S}_{1,\gamma} : y_\gamma(t) > u\} = J\right] \to 1.$$
4. *$\tilde{m}_{1,\gamma}/L \to A_1$ in probability.*

5. $W_\gamma(u)/\tilde{m}_{1,\gamma} \to 1$ in probability.

Proof.

1. Write $\mathbb{T}_\gamma = \cup_{j=1}^J T_{j,\gamma}$, where $T_{j,\gamma} = S_{j,\gamma} \setminus S_j$ is the transition region for peak j (Figure 2). Under the assumptions of Lemma 6.2, $T_{j,\gamma}$ is a subset of I_j^{side} because I_j^{left} or I_j^{right} may include points inside S_j. Using (31), the required probability $\mathrm{P}\left(\#\{t \in \tilde{T} \cap \mathbb{T}_\gamma\} \geq 1\right)$ that $y_\gamma(t)$ has any local maxima in the transition region $\mathbb{T}_\gamma$ is bounded above by

$$\sum_{j=1}^J \left[2 - 2\Phi\left(\frac{a_j C_j}{\sqrt{\lambda_{2,\gamma}}}\right) + \left|I_j^{\text{side}}\right| \sqrt{\frac{\lambda_{4,\gamma}}{\lambda_{2,\gamma}}} \phi\left(\frac{a_j C_j}{\sqrt{\lambda_{2,\gamma}}}\right)\right]$$
$$\leq 2\frac{J}{L} L\left[1 - \Phi\left(\frac{aC}{\sqrt{\lambda_{2,\gamma}}}\right)\right] + L\sqrt{\frac{\lambda_{4,\gamma}}{\lambda_{2,\gamma}}} \phi\left(\frac{aC}{\sqrt{\lambda_{2,\gamma}}}\right)$$

where $a > 0$ is the infimum of the a_j's and $C > 0$ is the infimum of the $C_j's$, i.e. the infimum of $|\dot{h}_{j,\gamma}(t)|$ for $t \in \cup_{j=1}^J I_j^{\text{side}}$ (recall that every peak $h_{j,\gamma}(t)$ has no critical points in the transition region for any j). But the expression above goes to zero as $a, L \to \infty$ such that $(\log L)/a^2 \to 0$ because, for any $K > 0$,

$$L\phi(Ka) = \frac{1}{\sqrt{2\pi}} \exp\left[a^2\left(\frac{\log L}{a^2} - \frac{K^2}{2}\right)\right] \to 0,$$

and $L[1 - \Phi(Ka)] \leq L\phi(Ka)/(Ka) \to 0$.

2. The required probability to obtain exactly J local maxima in the set $\mathbb{S}_{1,\gamma} = \cup_{j=1}^J S_{j,\gamma}$ is greater than the probability of obtaining exactly one local maximum in each interval $I_j^{\text{mode}} \subset S_j$ and none in I_j^{side} for any j. Thus, using (31), the required probability is bounded below by

$$\mathrm{P}\left[\cap_{j=1}^J \left(\#\{t \in \tilde{T} \cap I_j^{\text{mode}}\} = 1 \ \cap \ \#\{t \in \tilde{T} \cap I_j^{\text{side}}\} = 0\right)\right]$$
$$\geq 1 - \sum_{j=1}^J \left[1 - \mathrm{P}\left(\#\{t \in \tilde{T} \cap I_j^{\text{mode}}\} = 1 \ \cap \ \#\{t \in \tilde{T} \cap I_j^{\text{side}}\} = 0\right)\right]$$
$$\geq 1 - \sum_{j=1}^J \left[5 - 4\Phi\left(\frac{a_j C_j}{\sqrt{\lambda_{2,\gamma}}}\right) - \Phi\left(\frac{a_j D_j}{\sqrt{\lambda_{4,\gamma}}}\right) + \left|I_j^{\text{side}}\right| \sqrt{\frac{\lambda_{4,\gamma}}{\lambda_{2,\gamma}}} \phi\left(\frac{a_j C_j}{\sqrt{\lambda_{4,\gamma}}}\right) + \left|I_j^{\text{mode}}\right| \sqrt{\frac{\lambda_{6,\gamma}}{\lambda_{4,\gamma}}} \phi\left(\frac{a_j D_j}{\sqrt{\lambda_{4,\gamma}}}\right)\right]$$
$$\geq 1 - \frac{J}{L} L\left[5 - 4\Phi\left(\frac{aC}{\sqrt{\lambda_{2,\gamma}}}\right) - \Phi\left(\frac{aD}{\sqrt{\lambda_{4,\gamma}}}\right)\right] - L\sqrt{\frac{\lambda_{4,\gamma}}{\lambda_{2,\gamma}}} \phi\left(\frac{aC}{\sqrt{\lambda_{4,\gamma}}}\right) - L\sqrt{\frac{\lambda_{6,\gamma}}{\lambda_{4,\gamma}}} \phi\left(\frac{aD}{\sqrt{\lambda_{4,\gamma}}}\right).$$

But this bound goes to 1 under condition (C2) as in Part 1.

3. The required probability to obtain exactly J local maxima in the set $\mathbb{S}_{1,\gamma} = \cup_{j=1}^J S_{j,\gamma}$ that exceed u is greater than the probability that exactly one local maximum exceeds u in each interval I_j^{mode}. This probability is bounded below by

$$\mathrm{P}\left[\cap_{j=1}^J \left(\#\{t \in \tilde{T} \cap I_j^{\text{mode}} : y_\gamma(t) > u\} = 1 \ \cap \ \#\{t \in \tilde{T} \cap I_j^{\text{side}}\} = 0\right)\right]$$

but this goes to 1 by a similar argument as the one in Part 2 of this lemma.

4. Since $\tilde{m}_{1,\gamma}/L = (\tilde{m}_{1,\gamma}/J)(J/L)$, with $J/L \to A_1$, we need to show that $\tilde{m}_{1,\gamma}/J \to 1$ in probability. For any fixed $\varepsilon > 0$,

$$0 \leq \mathrm{P}\left(\left|\frac{\tilde{m}_{1,\gamma}}{J} - 1\right| \geq \varepsilon\right) = \mathrm{P}\left(|\tilde{m}_{1,\gamma} - J| \geq J\varepsilon\right) \leq \mathrm{P}\left(\tilde{m}_{1,\gamma} \neq J\right) = 1 - \mathrm{P}\left(\tilde{m}_{1,\gamma} = J\right)$$

since $\tilde{m}_{1,\gamma}$ and J are integers. But the probability to get exactly J local maxima goes to 1 by Part 1 of this lemma.

5. By Part 2 of this lemma $\mathrm{P}[W_\gamma(u) = J] \to 1$ in probability, therefore, using the same arguments as in Part 3 of this lemma, we get $W_\gamma(u)/J \to 1$. Now,

$$\frac{W_\gamma(u)}{\tilde{m}_{1,\gamma}} = \frac{W_\gamma(u)}{J}\frac{J}{\tilde{m}_{1,\gamma}}.$$

But $\tilde{m}_{1,\gamma}/J \to 1$ by Part 3 of this lemma. □

6.3 Strong control of FWER

Lemma 6.4. *Let $\tilde{m}_{0,\gamma}$ be the number of local maxima in $\mathbb{S}_{0,\gamma}$ as in Lemma 6.1. Define the thresholds $\tilde{v}_{\mathrm{Bon}} = F_\gamma^{-1}(\alpha/\tilde{m}_{0,\gamma})$, random, and $v^*_{\mathrm{Bon}} = F_\gamma^{-1}(\alpha/\mathrm{E}[\tilde{m}_{0,\gamma}])$, deterministic. Then $|\tilde{v}_{\mathrm{Bon}} - v^*_{\mathrm{Bon}}| \to 0$ in probability as $L \to \infty$.*

Proof. By ergodicity, the weak law of large numbers gives that

$$\left|\frac{\tilde{m}_{0,\gamma}}{L} - \mathrm{E}[\tilde{m}_{0,\gamma}(0,1)]\right| \to 0 \tag{35}$$

in probability as $L \to \infty$, where $\mathrm{E}[\tilde{m}_{0,\gamma}(0,1)] = \mathrm{E}[\tilde{m}_{0,\gamma}]/L$, given by (18), does not depend on L (Cramér and Leadbetter, 1967). Since $\log(\cdot)$ is continuous, the continuous mapping theorem gives that

$$\left|\log\frac{\tilde{m}_0}{L} - \log\frac{\mathrm{E}[\tilde{m}_{0,\gamma}]}{L}\right| = \left|\log\frac{\tilde{m}_0}{\alpha} - \log\frac{\mathrm{E}[\tilde{m}_{0,\gamma}]}{\alpha}\right| \to 0,$$

where we have used the additive property of the logarithm.

Define now the monotone increasing function $\psi_\gamma(x) = F_\gamma^{-1}(1-e^{-x})$. The function $\psi_\gamma(x)$ is Lipschitz continuous for all $x > 1$ because its derivative $d\psi_\gamma(x)/dx = e^{-x}/\dot{F}_\gamma[\psi_\gamma(x)]$ is bounded for all $x > 1$. Hence, as $L \to \infty$,

$$\left|\psi_\gamma\left(\log\frac{\tilde{m}_{0,\gamma}}{\alpha}\right) - \psi_\gamma\left(\log\frac{\mathrm{E}[\tilde{m}_{0,\gamma}]}{\alpha}\right)\right| = |\tilde{v}_{\mathrm{Bon}} - v^*_{\mathrm{Bon}}| \to 0.$$

□

Proof of Theorem 2.4.
Let $\tilde{m}_{0,\gamma} \leq \tilde{m}$ be the number of local maxima in the set $\mathbb{S}_{0,\gamma}$ as in Lemma 6.4, and let $\tilde{v}_{\mathrm{Bon}} = F_\gamma^{-1}(\alpha/\tilde{m}_{0,\gamma}) \leq \tilde{u}_{\mathrm{Bon}}$. Then $\mathrm{FWER}(\tilde{u}_{\mathrm{Bon}}) \leq \mathrm{FWER}(\tilde{v}_{\mathrm{Bon}})$. Further, the bound $\mathrm{FWER}(\tilde{v}_{\mathrm{Bon}})$ is the probability of obtaining at least one local maximum greater than $\tilde{v}_{\mathrm{Bon}}$ in $\mathbb{S}_0 = \mathbb{S}_{0,\gamma} \cup \mathbb{T}_\gamma$, which is less than the probability of obtaining at least one local maximum greater than $\tilde{v}_{\mathrm{Bon}}$ in $\mathbb{S}_{0,\gamma}$ or at least one local maximum in $\mathbb{T}_\gamma$:

$$\mathrm{FWER}(\tilde{u}_{\mathrm{Bon}}) \leq \mathrm{P}\left[V_\gamma(\tilde{v}_{\mathrm{Bon}}) \geq 1\right] + \mathrm{P}\left(\#\{t \in \tilde{T} \cap \mathbb{T}_\gamma\} \geq 1\right), \tag{36}$$

where $V_\gamma(u) = \#\{t \in \tilde{T} \cap \mathbb{S}_{0,\gamma} : y_\gamma(t) > u\}$ as in Lemma 6.1.

The second probability in (36) goes to zero by Lemma 6.3. To bound the first probability in (36), write

$$\mathrm{P}\left[V_\gamma(\tilde{v}_{\mathrm{Bon}}) \geq 1\right] = \mathrm{P}\left\{\tilde{T} \cap \mathbb{S}_{0,\gamma} \neq \emptyset \text{ and } \max_{t \in \tilde{T}} y_\gamma(t) > (\tilde{v}_{\mathrm{Bon}} - v^*_{\mathrm{Bon}}) + v^*_{\mathrm{Bon}}\right\},$$

where $v^*_{\mathrm{Bon}} = F_\gamma^{-1}\left(\alpha/\mathrm{E}[\tilde{m}_{0,\gamma}]\right)$ is deterministic. For any two random variables X, Y and any two constants c, ε: $P(X > Y + c) \leq P(X > c - \varepsilon) + P(|Y| > \varepsilon)$. Applying this inequality with $X = \max_{t \in \tilde{T}} y_\gamma(t)$, $Y = \tilde{v}_{\mathrm{Bon}} - v^*_{\mathrm{Bon}}$ and $c = v^*_{\mathrm{Bon}}$,

$$\mathrm{P}\left[V_\gamma(\tilde{v}_{\mathrm{Bon}}) \geq 1\right] \leq \mathrm{P}\left[V_\gamma(v^*_{\mathrm{Bon}} - \varepsilon) \geq 1\right] + \mathrm{P}\left\{\tilde{T} \cap \mathbb{S}_{0,\gamma} \neq \emptyset \text{ and } |\tilde{v}_{\mathrm{Bon}} - v^*_{\mathrm{Bon}}| > \varepsilon\right\}. \quad (37)$$

The second summand goes to 0 in probability as $L \to \infty$ by Lemma 6.4. For the first summand, Lemma 6.1 with level $v^*_{\mathrm{Bon}} - \varepsilon$ gives that

$$\mathrm{P}\left[V_\gamma(v^*_{\mathrm{Bon}} - \varepsilon) \geq 1\right] \leq \mathrm{E}[V_\gamma(v^*_{\mathrm{Bon}} - \varepsilon)] = \mathrm{E}[\tilde{m}_{0,\gamma}] F_\gamma(v^*_{\mathrm{Bon}} - \varepsilon) = \alpha \frac{F_\gamma(v^*_{\mathrm{Bon}} - \varepsilon)}{F_\gamma(v^*_{\mathrm{Bon}})},$$

but the last fraction goes to 1 as $L \to \infty$. Replacing in (37) and (36) gives the result. □

6.4 Control of FDR

Lemma 6.5. *For any non-negative integer random variables V, W, and fixed positive integer J,*

$$\mathrm{E}\left(\frac{V}{V+W}\right) \leq \mathrm{P}(W \leq J-1) + \frac{\mathrm{E}[V]}{\mathrm{E}[V] + J}.$$

Proof.

$$\begin{aligned}
\mathrm{E}\left(\frac{V}{V+W}\right) &= \sum_{v=0}^{\infty}\sum_{w=0}^{J-1}\left(\frac{v}{v+w}\right)\mathrm{P}\left(V=v, W=w\right) + \sum_{v=0}^{\infty}\sum_{w=J}^{\infty}\left(\frac{v}{v+w}\right)\mathrm{P}\left(V=v, W=w\right) \\
&\leq \sum_{w=0}^{J-1}\sum_{v=0}^{\infty}\mathrm{P}\left(V=v, W=w\right) + \sum_{v=0}^{\infty}\sum_{w=J}^{\infty}\left(\frac{v}{v+J}\right)\mathrm{P}\left(V=v, W=w\right) \\
&\leq \mathrm{P}\left(W \leq J-1\right) + \mathrm{E}\left(\frac{V}{V+J}\right) \leq \mathrm{P}\left(W \leq J-1\right) + \frac{\mathrm{E}(V)}{\mathrm{E}(V)+J}.
\end{aligned}$$

The last inequality holds by Jensen's inequality, since $V/(V+J)$ is a concave function of V for $V \geq 0$ and $J \geq 1$. □

Proof of Theorem 2.5. Let $\tilde{G}(u) = \#\{t \in \tilde{T} : y_\gamma(t) > u\}/\#\{t \in \tilde{T}\}$ be the empirical marginal right cdf of $y_\gamma(t)$ given $t \in \tilde{T}$. Then the BH threshold $\tilde{u}_{\mathrm{BH}}$ (19) satisfies $\alpha\tilde{G}(\tilde{u}_{\mathrm{BH}}) = k\alpha/\tilde{m} = F_\gamma(\tilde{u}_{\mathrm{BH}})$, so $\tilde{u}_{\mathrm{BH}}$ is the largest u that solves the equation

$$\alpha\tilde{G}(u) = F_\gamma(u). \quad (38)$$

The strategy is to solve equation (38) in the limit when $L, a \to \infty$. We first find the limit of $\tilde{G}(u)$. Letting $V_\gamma(u) = \#\{t \in \tilde{T} \cap \mathbb{S}_{0,\gamma} : y_\gamma(t) > u\}$ as in Lemma 6.1 and $W_\gamma(u) = \#\{t \in \tilde{T} \cap \mathbb{S}_{1,\gamma} : y_\gamma(t) > u\}$, so that $R_\gamma(u) = V_\gamma(u) + W_\gamma(u)$, write

$$\tilde{G}(u) = \frac{R_\gamma(u)}{\tilde{m}} = \frac{V_\gamma(u)}{\tilde{m}_{0,\gamma}}\frac{\tilde{m}_{0,\gamma}}{\tilde{m}_{0,\gamma}+\tilde{m}_{1,\gamma}} + \frac{W_\gamma(u)}{\tilde{m}_{1,\gamma}}\frac{\tilde{m}_{1,\gamma}}{\tilde{m}_{0,\gamma}+\tilde{m}_{1,\gamma}}. \quad (39)$$

By the weak law of large numbers (35) and Lemma 6.3, Part 3,

$$\frac{\tilde{m}_{0,\gamma}}{\tilde{m}_{0,\gamma}+\tilde{m}_{1,\gamma}} = \frac{\tilde{m}_{0,\gamma}/L}{\tilde{m}_{0,\gamma}/L+\tilde{m}_{1,\gamma}/L} \to \frac{\mathrm{E}[\tilde{m}_{0,\gamma}(0,1)]}{\mathrm{E}[\tilde{m}_{0,\gamma}(0,1)]+A_1},$$

as $L \to \infty$, where the expectation is given by (18). In addition we have the results of Lemma 6.1 and Lemma 6.3, Part 4. Replacing these three limits in (39), we obtain

$$\tilde{G}(u) \to F_\gamma(u)\frac{\mathrm{E}[\tilde{m}_{0,\gamma}(0,1)]}{\mathrm{E}[\tilde{m}_{0,\gamma}(0,1)]+A_1} + \frac{A_1}{\mathrm{E}[\tilde{m}_{0,\gamma}(0,1)]+A_1}.$$

Now replacing $\tilde{G}(u)$ by its limit in (38), and solving for u gives the deterministic solution

$$F_\gamma(u^*_{\mathrm{BH}}) = \frac{\alpha A_1}{A_1 + \mathrm{E}[\tilde{m}_{0,\gamma}(0,1)](1-\alpha)}. \tag{40}$$

The FDR at the threshold u^*_{BH} is bounded by Lemma 6.5 by

$$\begin{aligned}\mathrm{FDR}(u^*_{\mathrm{BH}}) &\le \mathrm{P}(W(u^*_{\mathrm{BH}}) \le J-1) + \frac{\mathrm{E}\left[V(u^*_{\mathrm{BH}})\right]}{\mathrm{E}\left[V(u^*_{\mathrm{BH}})\right]+J} \\ &= \mathrm{P}(W(u^*_{\mathrm{BH}}) \le J-1) + \frac{\mathrm{E}\left[V_\gamma(u^*_{\mathrm{BH}})\right] + \mathrm{E}\left[\#\{t\in\tilde{T}\cap\mathbb{T}_\gamma : y_\gamma(t) > u^*_{\mathrm{BH}}\}\right]}{\mathrm{E}\left[V_\gamma(u^*_{\mathrm{BH}})\right] + \mathrm{E}\left[\#\{t\in\tilde{T}\cap\mathbb{T}_\gamma : y_\gamma(t) > u^*_{\mathrm{BH}}\}\right]+J},\end{aligned} \tag{41}$$

where we have split $V_\gamma(u^*_{\mathrm{BH}})$ into the reduced null region $\mathbb{S}_{0,\gamma}$ and the transition region $\mathbb{T}_\gamma = \mathbb{S}_0 \setminus \mathbb{S}_{0,\gamma}$. Under condition (C2), Lemma 6.3 gives

$$0 \le \mathrm{E}\left[\#\{t\in\tilde{T}\cap\mathbb{T}_\gamma : y_\gamma(t) > u^*_{\mathrm{BH}}\}\right] \le \mathrm{E}\left[\#\{t\in\tilde{T}\cap\mathbb{T}_\gamma\}\right] \to 0. \tag{42}$$

By Lemma 6.1, the remaining terms of the last fraction in (41) can be written as

$$\frac{\mathrm{E}\left[V_\gamma(u^*_{\mathrm{BH}})\right]}{\mathrm{E}\left[V_\gamma(u^*_{\mathrm{BH}})\right]+J} = \frac{F_\gamma(u^*_{\mathrm{BH}})\mathrm{E}[\tilde{m}_{0,\gamma}(0,1)]L}{F_\gamma(u^*_{\mathrm{BH}})\mathrm{E}[\tilde{m}_{0,\gamma}(0,1)]L+J} = \frac{F_\gamma(u^*_{\mathrm{BH}})\mathrm{E}[\tilde{m}_{0,\gamma}(0,1)]}{F_\gamma(u^*_{\mathrm{BH}})\mathrm{E}[\tilde{m}_{0,\gamma}(0,1)]+J/L}.$$

Since u^*_{BH} solves (40), for $L\to\infty$ such that $J/L \to A_1$, the above expression tends to

$$\frac{\alpha\mathrm{E}[\tilde{m}_{0,\gamma}(0,1)]}{\alpha\mathrm{E}[\tilde{m}_{0,\gamma}(0,1)]+A_1+(1-\alpha)\mathrm{E}[\tilde{m}_{0,\gamma}(0,1)]} = \alpha\frac{\mathrm{E}[\tilde{m}_{0,\gamma}(0,1)]}{\mathrm{E}[\tilde{m}_{0,\gamma}(0,1)]+A_1} \le \alpha. \tag{43}$$

Combining equations (42), (43) and Lemma 6.3, Part 3, in (41), we obtain $\limsup \mathrm{FDR}(u^*_{\mathrm{BH}}) \le \alpha$.

Recall that the BH threshold $\tilde{u}_{\mathrm{BH}}$ solves equation (38), and u^*_{BH} satisfies (40), where the empirical marginal distribution, $\tilde{G}(u)$, is replaced by its limit. Since $F_\gamma(t)$ is continuous, $F_\gamma(\tilde{u}_{\mathrm{BH}}) \to F_\gamma(u^*_{\mathrm{BH}})$, leading to $\limsup \mathrm{FDR}(\tilde{u}_{\mathrm{BH}}) \le \alpha$. □

6.5 Power

Lemma 6.6. *For any $j = 1,\dots,J$, let t be any interior point of the support S_j of peak j. Under conditions (C1) and (C2),*

$$u^*_{\mathrm{Bon}}/[a_j h_{j,\gamma}(t)] \to 0, \qquad u^*_{\mathrm{BH}}/[a_j h_{j,\gamma}(t)] \to 0,$$

in probability, where u^*_{Bon} and u^*_{BH} are given by (17) and (20), respectively.

Proof of Lemma 6.6.

1. From (9), for $u > \sigma_\gamma$, $F_\gamma(u)$ is bounded above and below by

$$\frac{C_1}{2}\phi\left(\frac{u}{\sigma_\gamma}\right) < F_\gamma(u) < (C_1+1)\phi\left(\frac{u}{\sigma_\gamma}\right), \qquad C_1 = \sqrt{\frac{2\pi\lambda_{2,\gamma}^2}{\lambda_{4,\gamma}\sigma_\gamma^2}}, \tag{44}$$

where the lower bound was obtained using $\Phi(x) > 1/2$ for $x > 1$, and the upper bound used the fact that $\sqrt{\lambda_{4,\gamma}/\Delta} \geq 1/\sigma_\gamma$ and $1 - \Phi(x) < \phi(x)/x$ for $x > 1$. Let $v = F_\gamma(u)$. Inverting the bounds (44) we obtain

$$2\sigma_\gamma^2\left(\log\frac{C_1}{2\sqrt{2\pi}} - \log v\right) < u^2 < 2\sigma_\gamma^2\left(\log\frac{C_1+1}{\sqrt{2\pi}} - \log v\right). \tag{45}$$

Applying these inequalities to $v^* = F_\gamma(u^*_{\text{Bon}})$ and $w = F_\gamma[a_j h_{j,\gamma}(t)]$ gives that

$$0 \leq \frac{(u^*_{\text{Bon}})^2}{[a_j h_{j,\gamma}(t)]^2} < \frac{\log[(C_1+1)/\sqrt{2\pi}] - \log(v^*)}{\log[C_1/(2\sqrt{2\pi})] - \log(w)}.$$

Applying L'Hôpital rule, the limit of the above fraction when v^* and w go to zero is the same as the limit of w/v^*. But this limit is zero because, by the upper bound in (44) and (17),

$$\frac{F_\gamma[a_j h_{j,\gamma}(t)]}{F_\gamma(u^*_{\text{Bon}})} < (C_1+1)\frac{A_1 + \mathrm{E}[\tilde{m}_{0,\gamma}(0,1)]}{\alpha} L\phi\left(\frac{a_j h_{j,\gamma}(t)}{\sigma_\gamma}\right),$$

which goes to zero by the lemma's conditions.

2. The FDR threshold u^*_{BH} (20) is bounded, so the result is immediate. □

Proof of Theorem 2.6. For any threshold u, the detection power $\text{Power}(u)$ (21) is greater than $\mathrm{E}[W_\gamma(u)]/J \geq \mathrm{P}[W_\gamma(u) = J]$. But this probability goes to 1 by Lemma 6.3, particularly for the deterministic thresholds u^*_{Bon} and u^*_{BH}. It was shown in the proofs of Theorems 2.4 and 2.5 that the gap between the deterministic thresholds and the random thresholds $\tilde{u}_{\text{Bon}}$ and $\tilde{u}_{\text{BH}}$ narrows to zero asymptotically. Therefore the power for these thresholds goes to 1 as well. □

Acknowledgments

The authors thank Pablo Jadzinsky for providing the neural recordings data, as well as Igor Wigman, Felix Abramovich, and Yoav Benjamini for helpful discussions.

References

Robert J. Adler and Jonathan E. Taylor. *Random fields and geometry.* Springer, New York, 2007.

Robert J. Adler, Jonathan E. Taylor, and Keith J. Worsley. Applications of random fields and geometry: Foundations and case studies. URL `http://webee.technion.ac.il/people/adler/publications.html`. 2010.

Natalia M. Arzeno, Zhi-De Deng, and Chi-Sang Poon. Analysis of first-derivative based QRS detection algorithms. *IEEE Trans Biomed Eng*, 55(2):478–484, 2008.

Stephen A. Baccus and Markus Meister. Fast and slow contrast adaptation in retinal circuitry. *Neuron*, 36(5):909–919, 2002.

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

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

Pierpaolo Brutti, Christopher R. Genovese, Christopher J. Miller, Robert C. Nichol, and Larry Wasserman. 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`.

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

Justin R. Chumbley, Keith Worsley, Guillaume Flandin, and Karl J. Friston. Topological fdr for neuroimaging. *Neuroimage*, 49:3057–3064, 2010.

Harald Cramér and M. Ross Leadbetter. *Stationary and related stochastic processes*. Wiley, New York, 1967.

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

Jaroslaw Harezlak, Michael C. Wu, Mike Wang, Armin Schwartzman, David C. Christiani, and Xihong Lin. Biomarker discovery for arsenic exposure using functional data. analysis and feature learning of mass spectrometry proteomic data. *Journal of Proteome Research*, 7(1):217–224, 2008.

Lei M. Li and Terrence P. Speed. Parametric deconvolution of sparse positive spikes. *Ann Statist*, 28(5):1279–1301, 2000.

Lei M. Li and Terrence P. Speed. Deconvolution of sparse positive spikes. *J Comp Graph Statist*, 13(4):853–870, 2004.

Jeffrey S. Morris, Kevin R. Coombes, John Koomen, Keith A. Baggerly, and Ryuji Kobayashi. Feature extraction and quantification for mass spectrometry in biomedical applications using the mean spectrum. *Bioinformatics*, 21(9):1764–1775, 2006.

T. E. Nichols and S. Hayasaka. Controlling the familywise error rate in functional neuroimaging: A comparative review. *Statistical Methods in Medical Research*, 12:419–446, 2003.

Michael S. O'Brien, Anthony N. Sinclair, and Stuart M. Kramer. Recovery of a sparse spike train time series by l_1 norm deconvolution. *IEEE Trans Signal Proc*, 42(12):3353–3365, 1994.

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

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

J-B. Poline, K. J. Worsley, A. C. Evans, and K. J. Friston. Combining spatial extent and peak intensity to test for activations in functional imaging. *Neuroimage*, 5:83–96, 1997.

William K. Pratt. *Digital image processing*. Wiley, New York, 1991.

S. O. Rice. Mathematical theory of random noise. *Bell. System Tech. J*, 25:46–156, 1945.

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

Ronen Segev, Joe Goodhouse, Jason Puchalla, and Michael J Berry II. Recording spikes from a large fraction of the ganglion cells in a retinal patch. *Nature Neuroscience*, 7: 1155–1162, 2004.

Marvin Simon. *Digital communication techniques: signal design and detection*. Prentice Hall, Englewood Cliffs, NJ, 1995.

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

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

Robert Tibshirani, Michael Saunders, Saharon Rosset, Ji Zhu, and Keith Knight. Sparsity and smoothness via the fused lasso. *J R Statist Soc B*, 67(1):91–108, 2005.

Keith J. Worsley, S. Marrett, P. Neelin, and A. C. Evans. Searching scale space for activation in PET images. *Human Brain Mapping*, 4:74–90, 1996a.

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

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

Yutaka Yasui, Margaret Pepe, Mary Lou Thompson, Bao-Ling Adam, Jr. George L. Wright, Yinsheng Qu, John D. Potter, Marcy Winget, Mark Thornquist, and Ziding Feng. A data-analytic strategy for protein biomarker discovery: profiling of high-dimensional proteomic data for cancer detection. *Biostatistics*, 4(3):449–463, 2003.

Hui Zhang, Thomas E. Nichols, and Timothy D. Johnson. Cluster mass inference via random field theory. *Neuroimage*, 44:51–61, 2009.