

## MULTIPLE TESTING OF LOCAL MAXIMA FOR DETECTION OF PEAKS IN 1D

BY ARMIN SCHWARTZMAN<sup>1</sup>, YULIA GAVRILOV<sup>1</sup> AND ROBERT J. ADLER<sup>2</sup>

*Harvard School of Public Health, Harvard School of Public Health and Technion, Israel Institute of Technology*

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 nonasymptotic 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 peak regions. 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 common in neuroimaging, was established by Keith Worsley [Taylor and Worsley (2007), Worsley et al. (1996b, 2004)], who exploited the Euler characteristic

---

Received May 2011; revised November 2011.

<sup>1</sup>Supported in part by the Claudia Adams Barr Program in Cancer Research, the William F. Milton Fund and NIH Grant P01-CA134294.

<sup>2</sup>Supported in part by US–Israel Binational Science Foundation, 2008262 and Israel Science Foundation, 853/10.

*MSC2010 subject classifications.* Primary 62H15; secondary 62M10.

*Key words and phrases.* False discovery rate, Gaussian process, kernel smoothing, matched filter, topological inference.

heuristic for approximating the distribution of the maximum of a random field [Adler and Taylor (2007), Adler, Taylor and Worsley (2010)]. Methods for controlling the false discovery rate (FDR) [Benjamini and Hochberg (1995)] are also applied routinely in this setting, but the spatial structure is difficult to incorporate and often ignored [Genovese, Lazar and Nichols (2002), Nichols and Hayasaka (2003), Schwartzman, Dougherty and Taylor (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), Heller et al. (2006)], 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, Nichols and Johnson (2009)]. Perone Pacifico et al. (2004, 2007) 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 [Perone Pacifico et al. (2004)], 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 [Perone Pacifico et al. (2007)]. 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 [Smith and Nichols (2009), Worsley et al. (1996a)].
- (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*: apply a multiple testing procedure and declare as detected peaks those local maxima whose  $p$ -values are significant.

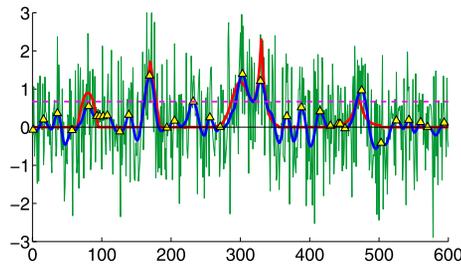


FIG. 1. Simulated observed sequence  $y(t)$  (green) and smoothed sequence  $y_\gamma(t)$  (blue) over five underlying true peaks of different shapes comprising  $\mu(t)$  (red). Out of 33 local maxima of  $y_\gamma(t)$  (yellow), the BH detection threshold at FDR level 0.2 (dashed magenta) selects five, one of which is a false positive. At this noise level, four out of five true peaks are detected. Note that this bandwidth is able to distinguish the overlapping peaks.

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 then relax these assumptions in the simulations. For Type I errors to be well defined, the signal is modeled as if composed of unimodal peak regions, 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 tractability, 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 [Adler, Taylor and Worsley (2010), Cramér and Leadbetter (1967)].

An interesting and challenging aspect of inference for local maxima is the fact that the number of tests, 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.

In order to achieve strong control of FWER and FDR, the asymptotics for large search space are combined with asymptotics for strong signal. The strong signal assumption asymptotically eliminates the false positives caused by the smoothed signal spreading into the null regions, by assuring that each signal peak region 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., Arzeno, Deng and Poon (2008), Baccus and Meister (2002), Brutti et al. (2005), Harezlak et al. (2008), Morris et al. (2006), Yasui et al. (2003)]. 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 [Li and Speed (2000, 2004), O'Brien, Sinclair and Kramer (1994), 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

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

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

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

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

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

where the smoothed signal and smoothed noise are defined as

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

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 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  $[-L/2, L/2]$

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

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

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

(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  $\alpha$ , declare significant all peaks whose  $p$ -values are smaller than  $\alpha/\tilde{m}$ . To apply BH at level  $\alpha$ , 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 1 are computed as

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

where

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

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 conditional distribution (7) is called a Palm distribution [Adler, Taylor and Worsley (2010), Chapter 6]. Unlike the marginal distribution of  $z_\gamma(t)$ , it is not Gaussian 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), Chapter 11] (equation 11.6.14), using the well-known Kac–Rice formula [Rice (1945), Adler and Taylor (2007), Chapter 11]. A direct application, borrowing notation from those sources, gives the following.

PROPOSITION 2. *Suppose the assumptions of Section 2.1 hold and that  $\mu(t) = 0, \forall t$ . Define the moments*

$$(8) \quad \sigma_\gamma^2 = \text{var}[z_\gamma(t)], \quad \lambda_{2,\gamma} = \text{var}[\dot{z}_\gamma(t)], \quad \lambda_{4,\gamma} = \text{var}[\ddot{z}_\gamma(t)].$$

Then the distribution (7) is given by

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

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

The quantities  $\sigma_\gamma^2, \lambda_{2,\gamma}$  and  $\lambda_{4,\gamma}$  in Proposition 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 3 (Gaussian autocorrelation model). Let the noise  $z(t)$  in (1) be constructed as

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

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

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

with moments (8) given by  $\sigma_\gamma^2 = \sigma^2/(2\sqrt{\pi}\xi)$ ,  $\lambda_{2,\gamma} = \sigma^2/(4\sqrt{\pi}\xi^3)$ ,  $\lambda_{4,\gamma} = 3\sigma^2/(8\sqrt{\pi}\xi^5)$ . The above expressions may be used as approximations if the kernel, required to have finite support, is truncated at  $t = \pm\gamma d$  for moderately large  $d$ , say  $d = 3$ .

2.4. *Error definitions.* Because truly detected peaks may be shifted with respect to the true peaks as a result of noise, we define a significant local maximum 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

$$(10) \quad \mathbb{S}_1 = \bigcup_{j=1}^J S_j \quad \text{and} \quad \mathbb{S}_0 = \left[-\frac{L}{2}, \frac{L}{2}\right] \setminus \left(\bigcup_{j=1}^J S_j\right).$$

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

$$(11) \quad \text{FWER}(u) = P\{V(u) \geq 1\} = P\left\{\tilde{T} \cap \mathbb{S}_0 \neq \emptyset \text{ and } \max_{t \in \tilde{T} \cap \mathbb{S}_0} y_\gamma(t) > u\right\}.$$

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

$$(12) \quad \text{FDR}(u) = E\left\{\frac{V(u)}{R(u) \vee 1}\right\}.$$

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 [Perone Pacifico et al. (2007)]. In contrast to (10), 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

$$(13) \quad \mathbb{S}_{1,\gamma} = \bigcup_{j=1}^J S_{j,\gamma} \quad \text{and} \quad \mathbb{S}_{0,\gamma} = \left[-\frac{L}{2}, \frac{L}{2}\right] \setminus \left(\bigcup_{j=1}^J S_{j,\gamma}\right),$$

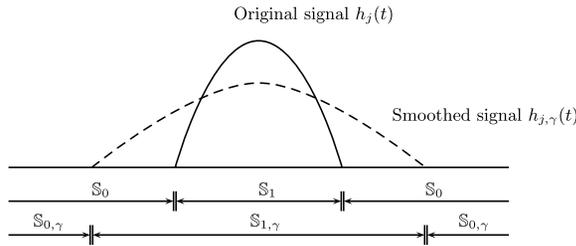


FIG. 2. Schematic signal and null regions, before and after smoothing, in the vicinity of one signal peak.

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

$$(14) \quad \mathbb{T}_\gamma = \mathbb{S}_{1,\gamma} \setminus \mathbb{S}_1 = \mathbb{S}_0 \setminus \mathbb{S}_{0,\gamma} = \bigcup_{j=1}^J T_{j,\gamma},$$

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

In general, a true peak may produce more than one significant local maximum, affecting the interpretation of definition (12) and the nonasymptotic validity of the FDR controlling procedure. However, as explained below, this multiplicity is unlikely to occur for strong signals, assuring validity at least asymptotically under that regime. The simulations of Section 3.1 show it not to be problematic in nonasymptotic situations for moderate signals and appropriate smoothing.

2.5. *Strong control of FWER.* In Algorithm 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

$$(15) \quad p_\gamma(t) < \frac{\alpha}{\tilde{m}} \iff y_\gamma(t) > \tilde{u}_{\text{Bon}} = F_\gamma^{-1}\left(\frac{\alpha}{\tilde{m}}\right),$$

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

**THEOREM 4.** *Suppose that Algorithm 1 is applied with the Bonferroni threshold  $\tilde{u}_{\text{Bon}}$  (15). Then, under conditions (C1) and (C2),*

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

The proof of Theorem 4 is given in Section 6.2. 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  $E[\tilde{m}/L]$  for large  $L$ . Thus the Bonferroni procedure with random threshold (15) has asymptotically the same error control properties as if the threshold were deterministic and equal to

$$(16) \quad u_{\text{Bon}}^* = F_\gamma^{-1}\left(\frac{\alpha}{E[\tilde{m}]}\right) \approx F_\gamma^{-1}\left(\frac{\alpha/L}{A_1 + E[\tilde{m}_{0,\gamma}(0, 1)]}\right),$$

where

$$(17) \quad E[\tilde{m}_{0,\gamma}(0, 1)] = \frac{1}{2\pi} \sqrt{\frac{\lambda_{4,\gamma}}{\lambda_{2,\gamma}}}$$

is the expected number of local maxima of  $z_\gamma(t)$  in the unit interval  $(0, 1)$  [Cramér and Leadbetter (1967), Chapter 10].

The strong signal assumption in (C2) implies (Lemma 10 in Section 6.1) that, with probability tending to 1, no local maxima are obtained in the transition region  $\mathbb{T}_\gamma$  (14), and exactly one local maxima is obtained for each signal peak in  $\mathbb{S}_1$ . This avoids the error inflation due to smoothing and provides the approximation in (16). The proof of Lemma 10 shows that 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 1. For a fixed  $\alpha \in (0, 1)$ , let  $k$  be the largest index for which the  $i$ th 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

$$(18) \quad p_\gamma(t) < \frac{k\alpha}{\tilde{m}} \iff y_\gamma(t) > \tilde{u}_{\text{BH}} = F_\gamma^{-1}\left(\frac{k\alpha}{\tilde{m}}\right),$$

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

**THEOREM 5.** *Suppose that Algorithm 1 is applied with the BH threshold  $\tilde{u}_{\text{BH}}$  (18). Then, under conditions (C1) and (C2),*

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

The proof of Theorem 5 is given in Section 6.3. The asymptotic assumptions (C2), imply that the BH procedure with random threshold (18) has asymptotically the same error control properties as if the threshold were deterministic and equal to

$$(19) \quad u_{\text{BH}}^* = F_\gamma^{-1}\left(\frac{\alpha A_1}{A_1 + E[\tilde{m}_{0,\gamma}(0, 1)](1 - \alpha)}\right),$$

where  $E[\tilde{m}_{0,\gamma}(0, 1)]$  is given by (17). The threshold (18) 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, Lazar and Nichols (2002)]. Taking the limit of that equation as  $L$  gets large yields the solution (19).

As before, the strong signal assumption in (C2) implies that there exists exactly one significant local maximum at each true peak with probability tending to 1 (Lemma 10 in Section 6.1), avoiding error inflation in the transition region and justifying the interpretation of definition (12) as the expected proportion of falsely discovered peaks. Again, 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  $u_{\text{Bon}}^*$  (16) which grows unbounded with increasing  $L$ , the asymptotic BH threshold  $u_{\text{BH}}^*$  (19) 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  $S_1$ . We define the power of Algorithm 1 as the expected fraction of true discovered peaks

$$\begin{aligned}
 \text{Power}(u) &= \mathbb{E} \left[ \frac{1}{J} \sum_{j=1}^J \mathbf{1} \left( \tilde{T} \cap S_j \neq \emptyset \text{ and } \max_{\tilde{t} \in \tilde{T} \cap S_j} y_\gamma(\tilde{t}) > u \right) \right] \\
 (20) \qquad &= \frac{1}{J} \sum_{j=1}^J \text{Power}_j(u),
 \end{aligned}$$

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

$$(21) \qquad \text{Power}_j(u) = \mathbb{P} \left\{ \tilde{T} \cap S_j \neq \emptyset \text{ and } \max_{t \in \tilde{T} \cap S_j} y_\gamma(t) > u \right\}.$$

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 10 in Section 6.1). The next result indicates that both the Bonferroni and BH procedures are asymptotically consistent. The proof is given in Section 6.4.

**THEOREM 6.** *Let the power be defined by (20), and let  $\tilde{u}_{\text{Bon}}$  and  $\tilde{u}_{\text{BH}}$  be the Bonferroni and BH thresholds (15) and (18), respectively. Under conditions (C1) and (C2),*

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

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 (16) and (19), if  $J \geq 1$ , the threshold  $u_{\text{Bon}}^*$  is higher than the threshold  $u_{\text{BH}}^*$ , 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 (20) under the true model. Because this maximization is analytically difficult, we resort to a less formal argument here. Lemma 10 in Section 6.1 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  $\tau_j$  with probability tending to 1. Thus the power for peak  $j$  (21) may be approximated as

$$(22) \quad \text{Power}_j(u) \approx \mathbb{P}\{y_\gamma(\tau_j) > u\} = \Phi\left[\frac{a_j h_{j,\gamma}(\tau_j) - u}{\sigma_\gamma}\right],$$

because  $y_\gamma(\tau_j) \sim N(a_j h_{j,\gamma}(\tau_j), \sigma_\gamma^2)$ . By Lemma 13 in Section 6.4, the asymptotically equivalent thresholds (16) and (19) for the Bonferroni and BH procedures satisfy  $u_{\text{Bon}}^*/a_j \rightarrow 0$  and  $u_{\text{BH}}^*/a_j \rightarrow 0$  for any  $j$ . Thus, for large  $a_j$ , the power (22) is maximized approximately by maximizing the SNR

$$(23) \quad \text{SNR}_\gamma = \frac{a_j h_{j,\gamma}(\tau_j)}{\sigma_\gamma} = \frac{a_j \int_{-\infty}^{\infty} w_\gamma(s) h_j(s) ds}{\sigma \sqrt{\int_{-\infty}^{\infty} w_\gamma^2(s) ds}},$$

where  $\sigma$  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 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 3. Ignoring the truncation,  $h_{j,\gamma}(t) = w_\gamma(t) * h_j(t)$  in (23) is the convolution of two Gaussian densities with variances  $\gamma^2$  and  $b_j^2$ , which is another Gaussian density with variance  $\gamma^2 + b_j^2$ . Using the moments from Example 3, we have that

$$(24) \quad \text{SNR}_\gamma = \frac{a_j h_{j,\gamma}(\tau_j)}{\sigma_\gamma} = \frac{a_j}{\sigma \pi^{1/4}} \left[ \frac{\gamma^2 + v^2}{(\gamma^2 + b_j^2)^2} \right]^{1/4}.$$

As a function of  $\gamma$ , the SNR is maximized at

$$(25) \quad \arg \max_\gamma \text{SNR}_\gamma = \begin{cases} \sqrt{b_j^2 - 2v^2}, & v < b_j/\sqrt{2}, \\ 0, & v > b_j/\sqrt{2}. \end{cases}$$

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  $\gamma$  is indeed close to (25).

**3. Simulation studies.**

3.1. *Nonasymptotic performance.* Simulations were used to evaluate the performance and limitations 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  $a_j h_j(t) = a/b\phi[(t - \tau_j)/b]\mathbf{1}[-cb, cb]$ ,  $j = 1, \dots, J$ , as in Example 7 with  $b = 3$ ,  $c = 3$  and varying  $a$ , were placed at uniformly spaced locations  $\tau_j = (j - 1/2)L/J$ ,  $j = 1, \dots, J$ , and sampled at integer values of  $t$ . The noise  $z(t)$  was constructed as in Example 3 with  $\sigma = 1$  and varying  $\nu$ . Algorithm 1 was carried out using as smoothing kernel a truncated Gaussian density  $w_\gamma(t) = (1/\gamma)\phi(t/\gamma)\mathbf{1}[-c\gamma, c\gamma]$  as in Example 3 with  $c = 3$  and varying  $\gamma$ . The noise parameters (8) were estimated independently as the empirical moments of smoothed sequences i.i.d. Gaussian noise of length 1000 and their first and second-order differences, using the same smoothing kernel. The Bonferroni and BH procedures were applied at level  $\alpha = 0.05$ .

Figure 3 shows the realized FWER and FDR levels of the Bonferroni and BH procedures, evaluated according to (11) and (12) with the expectations replaced by ensemble averages over 10,000 replications. Error rates are maintained below the nominal level  $\alpha = 0.05$  for all bandwidths and large enough signal strength  $a$ . The convergence is slower, however, when the bandwidth  $\gamma$  is much larger than the signal peak bandwidth  $b = 3$ . The increased error rates are the result of true

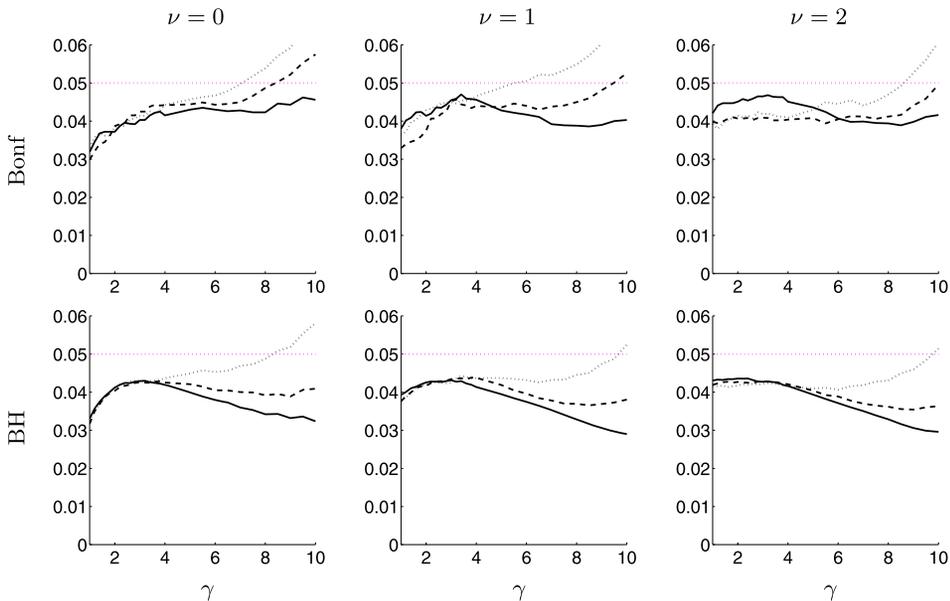


FIG. 3. FWER of the Bonferroni procedure (top row) and FDR of the BH procedure (bottom row) for  $a = 15$  (solid),  $a = 12$  (dashed) and  $a = 9$  (dotted). Nominal error level is 0.05.

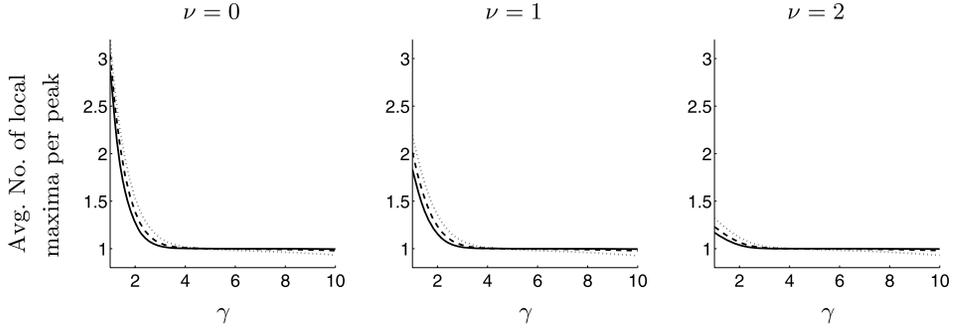


FIG. 4. Average number of local maxima for each true peak for  $a = 15$  (solid),  $a = 12$  (dashed) and  $a = 9$  (dotted).

peak maxima being shifted 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  $a$  because the probability of obtaining any local maxima in the transition region goes to zero asymptotically (Lemma 10 in Section 6.1).

As noted in Section 2.4, each true peak may contain more than one local maximum of the smoothed data  $y_\gamma(t)$ . Figure 4 shows that the expected number of local maxima per true peak decreases with increasing bandwidth, and is essentially equal to 1 for bandwidths equal to or greater than the optimal bandwidth. It also gets closer to 1 with increasing signal strength, consistent with the result of Lemma 10.

Figure 5 shows the realized power of the Bonferroni and BH procedures, evaluated according to (20) 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. The convergence is slower, however, when the bandwidth  $\gamma$  is far from the optimal value. To understand the dependence on bandwidth, superimposed is the theoretical approximate power (22) evaluated at the asymptotic thresholds  $u_{\text{Bon}}^*$  (16) and  $u_{\text{BH}}^*$  (19) and plugging in the SNR (24). The “theoretical” power curves largely capture the shape of the realized ones, but are lower because the asymptotic thresholds are more conservative. The curve shape is mostly determined by the SNR (24) as a function of  $\gamma$ . The bandwidth  $\gamma$  producing the largest power is always larger than the theoretical optimal bandwidth (25), but it approaches it from the right as  $a$  increases.

**3.2. Unequal peaks.** By assumption (Section 2.1), the signal peaks need not be equal. As in Figure 1,  $J = 5$  unequal peaks (Epanechnikov, triangular and truncated Gaussian, Laplace and Cauchy, with average half-support 24) were corrupted with white standard normal noise. Algorithm 1 was applied using a quartic smoothing kernel  $w_\gamma(t) = 15/(16\gamma)[1 - (t/\gamma)^2]^2\mathbf{1}[-\gamma, \gamma]$  with varying  $\gamma$ , the

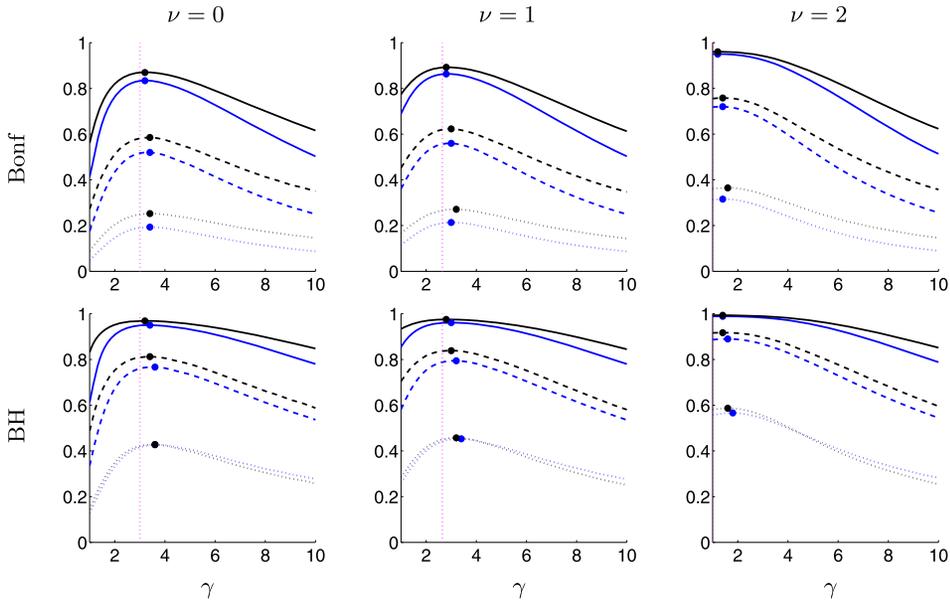


FIG. 5. Realized (black) and “theoretical” (blue) power of the Bonferroni (top row) and BH (bottom row) procedures for  $a = 15$  (solid),  $a = 12$  (dashed) and  $a = 9$  (dotted). The maxima of the curves (solid circles) approach the asymptotic optimal bandwidth (vertical dashed).

noise parameters estimated independently as in Section 3.1. For this configuration and 10,000 repetitions, the error was controlled below the nominal level 0.05 for values of  $\gamma$  up to 40, obtaining a maximum power of 0.81 and 0.88 for the Bonferroni and BH procedures at  $\gamma = 18$ . The maximizing bandwidth represents the average best match between the quartic smoothing kernel and the peaks present in the data.

3.3. *Overlapping peaks.* The theory of Section 2 assumed that the signal peaks had nonoverlapping supports. Simulations similar to those of Section 3.1 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, deceptively increased with increasing overlap. This is because definition (20) counts two overlapping peaks as detected even if only one significant local maximum is found in the overlapping region between them, as it belongs to both. Definition (20) does not measure the ability to distinguish between overlapping peaks.

3.4. *Comparison with pointwise testing.* To see the benefits of testing local maxima, Figure 6 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  $b = 3$  and  $\nu = 0$  were

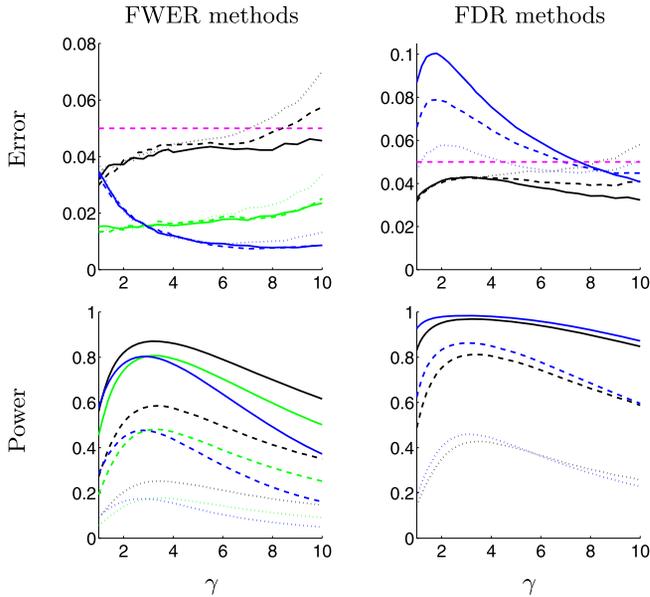


FIG. 6. Left panels: FWER and power of three FWER methods: STEM with Bonferroni (black), Bonferroni on all  $L$  locations (blue) and Supremum (green). Right panels: FDR and power of three FDR methods: STEM with BH (black), BH on all  $L$  locations (blue). Results in all panels are for  $a = 15$  (solid),  $a = 12$  (dashed) and  $a = 9$  (dotted). Nominal error level is 0.05.

smoothed with varying  $\gamma$ . For the pointwise Bonferroni and BH methods,  $p$ -values for testing  $H_0 : \mu(t) = 0$  at each  $t = 1, \dots, L = 1000$  were computed as  $p(t) = 1 - \Phi[y_\gamma(t)/\sigma_\gamma]$  and then corrected using Bonferroni and BH, respectively. The method ‘‘Supremum’’ was adapted from Worsley et al. (1996a) as follows. The 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)]

$$(26) \quad P\left(\sup_{t \in [0, T]} f(t) \geq u\right) \leq P[f(0) \geq u] + E[N_u],$$

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), page 194] gives that  $E[N_u] = L(\sqrt{\lambda_{2,\gamma}}/\sigma_\gamma)\phi(u/\sigma_\gamma)$ . The significance threshold is found as the largest  $u$  such that

$$(27) \quad P\left(\sup_{t \in [-L/2, L/2]} z_\gamma(t) \geq u\right) \leq 1 - \Phi\left(\frac{u}{\sigma_\gamma}\right) + L \frac{\sqrt{\lambda_{2,\gamma}}}{\sigma_\gamma} \phi\left(\frac{u}{\sigma_\gamma}\right) \leq \alpha.$$

Figure 6 indicates that the pointwise Bonferroni correction is too conservative. The Supremum method, despite accounting explicitly for the noise autocorrelation, performs only slightly better than pointwise Bonferroni, and not as well as Bonferroni performed on local maxima. The pointwise BH correction is designed

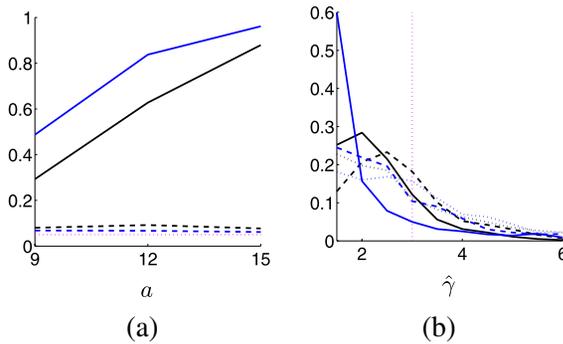


FIG. 7. (a) Power (solid) and realized error rate (dashed) for Bonferroni (black) and BH (blue) with automatic bandwidth selection as a function of signal strength  $a$ . Nominal error level is 0.05. (b) Proportion of automatically chosen smoothing bandwidth  $\hat{\gamma}$  over 1000 simulations for Bonferroni (black) and BH (blue); results are for  $a = 15$  (solid),  $a = 12$  (dashed) and  $a = 9$  (dotted). Nominal optimal bandwidth is  $\gamma = 3$ .

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 (12). Further simulations with  $\nu = 1$  and  $\nu = 2$  yielded similar results (not shown).

3.5. Automatic bandwidth selection. Rather than using a fixed smoothing bandwidth  $\gamma$ , the bandwidth may be chosen automatically 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  $b = 3$  and  $\nu = 0$ , the STEM algorithm was applied with  $\gamma$  ranging from  $\gamma = b/2 = 1.5$  to  $\gamma = 2b = 6$ , and results were retained for the bandwidth  $\hat{\gamma}$  that yielded the largest number of discoveries in each run. Figure 7(a) shows that this automatic criterion biases the results toward more detected peaks and therefore results in higher error rates (and power) than those obtained when  $\gamma$  is fixed (Figure 5). It also tends to select bandwidths that are smaller than the nominal optimal value  $\gamma = b$  [Figure 7(b)], with averages ranging between about 2.1 and 2.9.

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 seconds of recordings of live cells in the dark.
- (2) Training set 1: 60 seconds of recordings of live cells in the dark.
- (3) Training set 2: 60 seconds of recordings after the retina was allowed to die.

Each period of 60 seconds corresponds to  $L = 6 \times 10^5$  samples. The goal of the analysis was to detect neuronal spikes in the test set (Figure 8, top left).

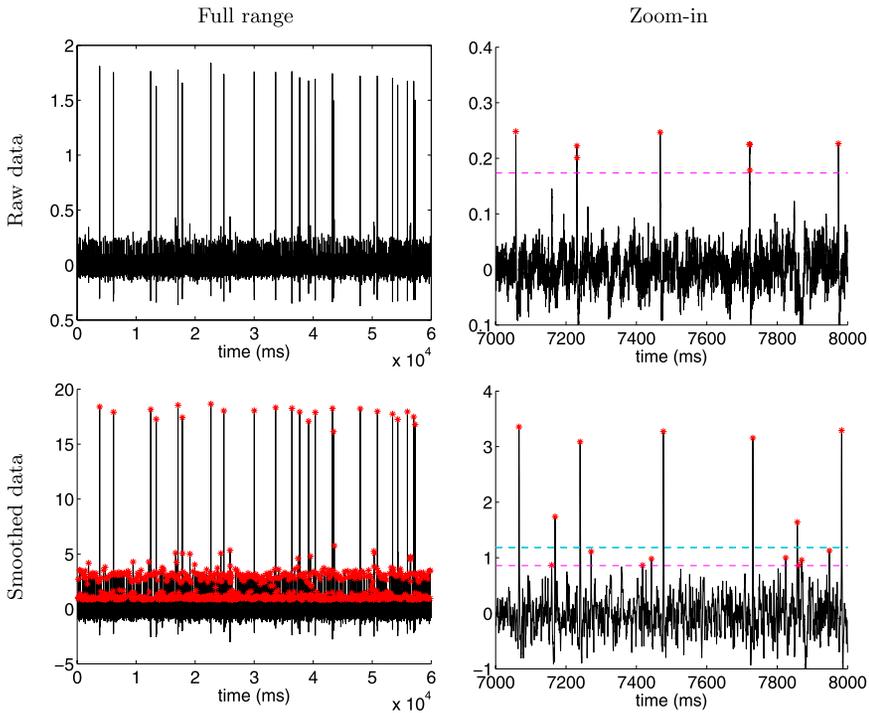


FIG. 8. *Top row: the neural spike data (test set); the stars in the right panel indicate peaks that are higher than 4 standard deviations of the raw data (dashed line), as suggested by Segev et al. (2004). Bottom row: the data smoothed using the estimated peak shape as kernel; the stars indicate significant local maxima higher than the BH threshold (magenta dashed line) at level 0.01. The Bonferroni threshold is indicated by the cyan dashed line.*

Assuming that neuronal action potentials have similar shapes, to maximize the SNR (23), 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 9(a)]. 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 coefficients 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.

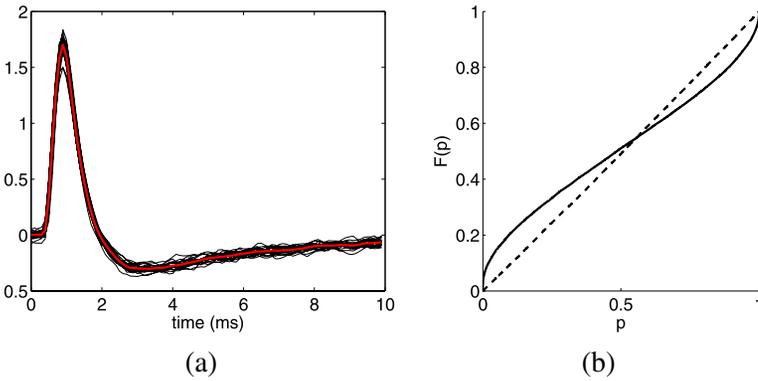


FIG. 9. (a) 23 strongest spikes aligned by their maximum (black); their average (red) is the estimated template. (b) Empirical cdf of  $p$ -values for the test set (solid) and training set 2 (dashed).

Convolving training set 2 with the template of Figure 9(a) produced smoothed noise with spectral moments  $\hat{\sigma}_\gamma^2 = 4.22 \times 10^{-4}$ ,  $\hat{\lambda}_{2,\gamma}^2 = 1.20 \times 10^{-4}$  and  $\hat{\lambda}_{4,\gamma}^2 = 1.96 \times 10^{-4}$ , 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 standard error of these estimates is negligible.

Algorithm 1 was applied to the test set (Figure 8, top left) by convolving it with the template of Figure 9(a), producing the smoothed process in Figure 8 (bottom left). In  $L = 6 \times 10^5$  samples,  $\tilde{m} = 30,426$  local maxima were found and their  $p$ -values were computed according to (6) and (9), plugging in the estimates  $\hat{\sigma}_\gamma^2$ ,  $\hat{\lambda}_{2,\gamma}^2$  and  $\hat{\lambda}_{4,\gamma}^2$  found above. The empirical cdf of the  $p$ -values [Figure 9(b)] shows a large fraction of nonnull  $p$ -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 9(b), the excess of large  $p$ -values from the test set is due to the negative portions of the smoothing function [Figure 9(a)]. These produce small negative anti-spikes whose  $p$ -values are large when tested for positiveness.

Applying the BH procedure to the  $\tilde{m} = 30,426$   $p$ -values obtained from the test set at FDR level 0.01 resulted in a  $p$ -value threshold of  $2.76 \times 10^{-4}$  and  $R = 843$  significant local maxima. These are indicated in Figure 8 (bottom left), showing three levels of spike strengths. Figure 8 (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  $3.29 \times 10^{-7}$  and only 411 detected spikes.

For comparison, Figure 8 (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 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 3.4, computing  $p$ -values at each location as  $p(t) = 1 - \Phi(y_\gamma(t)/\hat{\sigma}_\gamma)$ ,  $t = 1, \dots, L$ , and applying a global Bonferroni at level 0.01 was more conservative, resulting in a height threshold of 1.235 (comparable to Figure 8 bottom right) and detecting only 393 spikes. Similarly, the ‘‘Supremum’’ method, applied by replacing  $\hat{\sigma}_\gamma$  and  $\hat{\lambda}_\gamma$  in (27) 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 3.4, 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 for testing local maxima when the signal spreads over the entire domain, but in that case every positive is a true positive, making the inference unclear. On the other hand, 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. The unimodality assumption made local maxima good representatives of true peaks, being unique for medium to large bandwidths and asymptotically for increasing signal strength.

The strong signal assumption in condition (C2) was introduced to remove the excess error produced by the smoothed signal spreading into the neighboring null regions, thereby enabling asymptotic error control. The assumption is not restrictive in the sense that the search space may grow exponentially faster. Similar conditions are common for high-dimensional data. If the data are pointwise test statistics based on a sample of size  $n$ , with SNR increasing as  $a = \sqrt{n} \rightarrow \infty$ , then the condition  $(\log L)/a^2 \rightarrow 0$  becomes  $(\log L)/n \rightarrow 0$ . This is similar to the condition  $(\log p)/n \rightarrow 0$  required for consistent model selection in high-dimensional regression under sparsity where  $p$  is the number of features [Candes and Tao (2007), Zhang (2010)]. Our results, however, do not require sparsity. Condition (C2) is easy to state but stronger than needed; upon close inspection of the proof of Lemma 10 in Section 6.1, the limit of  $(\log L)/a^2$  need not be zero but need only be bounded by a constant that depends on the signal and noise first and second derivatives.

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. The asymptotic results were found to hold in practice for a wide range of bandwidths and strong enough signal. However, the convergence rate was slower for bandwidths less than half or more than double the optimal value. The matched filter principle suggests that the smoothing kernel should be chosen to be as close as possible in an  $L^2$  sense to the peaks to be detected. In the neuronal data analyzed, the peak shape and width were estimated from the data, dictating the best smoothing kernel. 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. *Supporting results.*

LEMMA 8. *Let  $\tilde{m}_{0,\gamma} = \#\{t \in \tilde{T} \cap \mathbb{S}_{0,\gamma}\}$  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*

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

*in probability as  $L \rightarrow \infty$ , 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

$$(28) \quad \frac{V_\gamma(u)}{\tilde{m}_{0,\gamma}} = \frac{\#\{t \in \tilde{T} \cap \mathbb{S}_{0,\gamma} : z_\gamma(t) > u\}/L}{\#\{t \in \tilde{T} \cap \mathbb{S}_{0,\gamma}\}/L}$$

converges to [Cramér and Leadbetter (1967)]

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

But also by ergodicity, ratio (28) converges to the conditional probability  $P[z_\gamma(t) > u | t \in \tilde{T} \cap \mathbb{S}_{0,\gamma}] = F_\gamma(u)$  by Definition (7). The two limits must be equal.  $\square$

LEMMA 9. 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  $|h_{j,\gamma}(t)|$  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  $I_j^{\text{mode}}$ .

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

$$\begin{aligned}
 & \mathbb{P}(\#\{t \in \tilde{T} \cap I_j^{\text{side}}\} = 0) \\
 & \geq 2\Phi\left(\frac{a_j C_j}{\sqrt{\lambda_{2,\gamma}}}\right) - 1 - |I_j^{\text{side}}| \sqrt{\frac{\lambda_{4,\gamma}}{\lambda_{2,\gamma}}} \phi\left(\frac{a_j C_j}{\sqrt{\lambda_{2,\gamma}}}\right), \\
 & \mathbb{P}(\#\{t \in \tilde{T} \cap I_j^{\text{mode}}\} = 1) \\
 & \geq \Phi\left(\frac{a_j D_j}{\sqrt{\lambda_{4,\gamma}}}\right) - |I_j^{\text{mode}}| \sqrt{\frac{\lambda_{6,\gamma}}{\lambda_{4,\gamma}}} \phi\left(\frac{a_j D_j}{\sqrt{\lambda_{4,\gamma}}}\right) - 2\Phi\left(\frac{-a_j C_j}{\sqrt{\lambda_{2,\gamma}}}\right), \\
 & \mathbb{P}(\#\{t \in \tilde{T} \cap I_j^{\text{mode}} : y_\gamma(t) > u\} = 1) \\
 & \geq \Phi\left(\frac{a_j D_j}{\sqrt{\lambda_{4,\gamma}}}\right) - |I_j^{\text{mode}}| \sqrt{\frac{\lambda_{6,\gamma}}{\lambda_{4,\gamma}}} \phi\left(\frac{a_j D_j}{\sqrt{\lambda_{4,\gamma}}}\right) - \Phi\left(\frac{u - a_j M_j}{\sigma_\gamma}\right),
 \end{aligned}
 \tag{29}$$

where  $\sigma_\gamma, \lambda_{2,\gamma}$  and  $\lambda_{4,\gamma}$  are given by (8) and  $\lambda_{6,\gamma} = \mathbb{E}[\ddot{z}_\gamma(t)]$ .

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

$$\begin{aligned}
 & \mathbb{P}\left(\inf_{I_j^{\text{left}}} \dot{y}_\gamma(t) > 0\right) \geq \mathbb{P}\left(\inf_{I_j^{\text{left}}} \dot{z}_\gamma(t) > -\inf_{I_j^{\text{left}}} \dot{\mu}_\gamma(t)\right) \\
 & = 1 - \mathbb{P}\left(\sup_{I_j^{\text{left}}} [-\dot{z}_\gamma(t)] > a_j C_j^{\text{left}}\right),
 \end{aligned}
 \tag{30}$$

where  $C_j^{\text{left}} > 0$  is the smallest value of  $\dot{h}_{j,\gamma}(t)$  in  $I_j^{\text{left}}$ . Inequality (26) applies above to the stationary Gaussian process  $-\dot{z}_\gamma(t)$ . The Kac–Rice formula [Cramér and Leadbetter (1967), page 194] gives in this case that  $\mathbb{E}[N_u] = |I_j^{\text{left}}| \sqrt{\lambda_{4,\gamma}} / \sqrt{\lambda_{2,\gamma}} \phi(u / \sqrt{\lambda_{2,\gamma}})$ . Thus (30) has the lower bound

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

A similar calculation for  $I_j^{\text{right}}$  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  $I_j^{\text{right}}$ . Putting the two together, the required probability  $P(\#\{t \in \tilde{T} \cap I_j^{\text{side}}\})$  that there are no local maxima in  $I_j^{\text{left}}$  nor  $I_j^{\text{right}}$  is bounded as in the first row of (29).

(2) The probability that  $y_\gamma(t)$  has no local maxima in  $I_j^{\text{mode}}$  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  $I_j^{\text{mode}}$  is bounded above by

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

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  $I_j^{\text{mode}}$  is less than the probability that  $\ddot{y}_\gamma(t) > 0$  for some  $t$  in  $I_j^{\text{mode}}$ . This probability is

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

where  $D_j < 0$  is the largest value of  $\ddot{\mu}_\gamma(t) < 0$  in  $I_j^{\text{mode}}$ . Applying (26) to the process  $\ddot{z}_\gamma(t)$  gives the further upper bound

$$(32) \quad P(\#\{t \in \tilde{T} \cap I_j^{\text{mode}}\} \geq 1) \leq 1 - \Phi\left(\frac{a_j D_j}{\sqrt{\lambda_{4,\gamma}}}\right) + |I_j^{\text{mode}}| \sqrt{\frac{\lambda_{6,\gamma}}{\lambda_{4,\gamma}}} \phi\left(\frac{a_j D_j}{\sqrt{\lambda_{4,\gamma}}}\right).$$

Putting (31) and (32) together gives the bound in the second row of (29).

(3) The probability that no local maxima of  $y_\gamma(t)$  in  $I_j^{\text{mode}}$  exceed the threshold  $u$  is less than the probability that  $y_\gamma(t)$  is below  $u$  anywhere in  $I_j^{\text{mode}}$ , 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  $I_j^{\text{mode}}$  exceed  $u$  is less than the probability that there exist more than one local maximum, which is bounded above by (32). Putting the two together gives the bound in the third row of (29).  $\square$

LEMMA 10. 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.*

$$P(\#\{t \in \tilde{T} \cap \mathbb{T}_\gamma\} \geq 1) \rightarrow 0.$$

(2) *The probability to get exactly  $J$  local maxima in the set  $\mathbb{S}_{1,\gamma}$ ,*

$$P(\tilde{m}_{1,\gamma} = J) = P(\#\{t \in \tilde{T} \cap \mathbb{S}_{1,\gamma}\} = J) \rightarrow 1.$$

(3) *The probability to get exactly  $J$  local maxima in the set  $\mathbb{S}_{1,\gamma}$  that exceed any fixed threshold  $u$ ,*

$$P[W_\gamma(u) = J] = P(\#\{t \in \tilde{T} \cap \mathbb{S}_{1,\gamma} : y_\gamma(t) > u\} = J) \rightarrow 1.$$

(4)  *$\tilde{m}_{1,\gamma}/L \rightarrow A_1$  in probability.*

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

PROOF. (1) Write  $\mathbb{T}_\gamma = \bigcup_{j=1}^J T_{j,\gamma}$ , where  $T_{j,\gamma} = S_{j,\gamma} \setminus S_j$  is the transition region for peak  $j$  (Figure 2). Under the assumptions of Lemma 9,  $T_{j,\gamma}$  is a subset of  $I_j^{\text{side}}$  because  $I_j^{\text{left}}$  or  $I_j^{\text{right}}$  may include points inside  $S_j$ . Using (29), the required probability  $P(\#\{t \in \tilde{T} \cap \mathbb{T}_\gamma\} \geq 1)$  that  $y_\gamma(t)$  has any local maxima in the transition region  $\mathbb{T}_\gamma$  is bounded above by

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

where  $a > 0$  is the infimum of the  $a_j$ 's and  $C > 0$  is the infimum of the  $C_j$ 's, that is, the infimum of  $|\dot{h}_{j,\gamma}(t)|$  for  $t \in \bigcup_{j=1}^J I_j^{\text{side}}$  [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 under condition (C2) because, for any  $K > 0$ ,

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

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

(2) The required probability to obtain exactly  $J$  local maxima in the set  $\mathbb{S}_{1,\gamma} = \bigcup_{j=1}^J S_{j,\gamma}$  is greater than the probability of obtaining exactly one local maximum in each interval  $I_j^{\text{mode}} \subset S_j$  and none in  $I_j^{\text{side}}$  for any  $j$ . Thus, using (29), the

required probability is bounded below by

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

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  $\mathbb{S}_{1,\gamma} = \bigcup_{j=1}^J S_{j,\gamma}$  that exceed  $u$  is greater than the probability that exactly one local maximum exceeds  $u$  in each interval  $I_j^{\text{mode}}$ . This probability is bounded below by

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

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

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

since  $\tilde{m}_{1,\gamma}$  and  $J$  are integers. But the probability to get exactly  $J$  local maxima goes to 1 by part (2) of this lemma.

(5) By part (2) of this lemma,  $\mathbb{P}[W_\gamma(u) = J] \rightarrow 1$  in probability; therefore, using the same arguments as in part (4) of this lemma, we get  $W_\gamma(u)/J \rightarrow 1$ . Now,

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

But  $\tilde{m}_{1,\gamma}/J \rightarrow 1$  by part (3) of this lemma.  $\square$

6.2. Strong control of FWER.

LEMMA 11. Let  $\tilde{m}_{0,\gamma}$  be the number of local maxima in  $\mathbb{S}_{0,\gamma}$  as in Lemma 8. Define the thresholds  $\tilde{v}_{\text{Bon}} = F_\gamma^{-1}(\alpha/\tilde{m}_{0,\gamma})$ , random, and  $v_{\text{Bon}}^* = F_\gamma^{-1}(\alpha/E[\tilde{m}_{0,\gamma}])$ , deterministic. Then  $|\tilde{v}_{\text{Bon}} - v_{\text{Bon}}^*| \rightarrow 0$  in probability as  $L \rightarrow \infty$ .

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

$$(33) \quad \left| \frac{\tilde{m}_{0,\gamma}}{L} - E[\tilde{m}_{0,\gamma}(0, 1)] \right| \rightarrow 0$$

in probability as  $L \rightarrow \infty$ , where  $E[\tilde{m}_{0,\gamma}(0, 1)] = E[\tilde{m}_{0,\gamma}]/L$ , given by (17), does not depend on  $L$  [Cramér and Leadbetter (1967)]. Since  $\log(\cdot)$  is continuous, the continuous mapping theorem gives that

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

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  $L \rightarrow \infty$ ,

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

PROOF OF THEOREM 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 11, and let  $\tilde{v}_{\text{Bon}} = F_\gamma^{-1}(\alpha/\tilde{m}_{0,\gamma}) \leq \tilde{u}_{\text{Bon}}$ . Then  $\text{FWER}(\tilde{u}_{\text{Bon}}) \leq \text{FWER}(\tilde{v}_{\text{Bon}})$ . Further, the bound  $\text{FWER}(\tilde{v}_{\text{Bon}})$  is the probability of obtaining at least one local maximum greater than  $\tilde{v}_{\text{Bon}}$  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  $\tilde{v}_{\text{Bon}}$  in  $\mathbb{S}_{0,\gamma}$  or at least one local maximum in  $\mathbb{T}_\gamma$ .

$$(34) \quad \text{FWER}(\tilde{u}_{\text{Bon}}) \leq P[V_\gamma(\tilde{v}_{\text{Bon}}) \geq 1] + P(\#\{t \in \tilde{T} \cap \mathbb{T}_\gamma\} \geq 1),$$

where  $V_\gamma(u) = \#\{t \in \tilde{T} \cap \mathbb{S}_{0,\gamma} : y_\gamma(t) > u\}$  as in Lemma 8.

The second probability in (34) goes to zero by Lemma 10, part (1). To bound the first probability in (34), write

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

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

$$(35) \quad \begin{aligned} P[V_\gamma(\tilde{v}_{\text{Bon}}) \geq 1] &\leq P[V_\gamma(v_{\text{Bon}}^* - \varepsilon) \geq 1] \\ &\quad + P\{\tilde{T} \cap \mathbb{S}_{0,\gamma} \neq \emptyset \text{ and } |\tilde{v}_{\text{Bon}} - v_{\text{Bon}}^*| > \varepsilon\}. \end{aligned}$$

The second summand goes to 0 in probability as  $L \rightarrow \infty$  by Lemma 11. For the first summand, Lemma 8 with level  $v_{\text{Bon}}^* - \varepsilon$  gives that

$$\begin{aligned} \mathbb{P}[V_\gamma(v_{\text{Bon}}^* - \varepsilon) \geq 1] &\leq \mathbb{E}[V_\gamma(v_{\text{Bon}}^* - \varepsilon)] = \mathbb{E}[\tilde{m}_{0,\gamma}]F_\gamma(v_{\text{Bon}}^* - \varepsilon) \\ &= \alpha \frac{F_\gamma(v_{\text{Bon}}^* - \varepsilon)}{F_\gamma(v_{\text{Bon}}^*)}, \end{aligned}$$

but the last fraction goes to 1 as  $L \rightarrow \infty$ . Replacing in (35) and (34) gives the result.  $\square$

6.3. Control of FDR.

LEMMA 12. For any nonnegative integer random variables  $V, W$  and fixed positive integer  $J$ ,

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

PROOF.

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

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

PROOF OF THEOREM 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  $\tilde{u}_{\text{BH}}$  (18) satisfies  $\alpha \tilde{G}(\tilde{u}_{\text{BH}}) = k\alpha / \tilde{m} = F_\gamma(\tilde{u}_{\text{BH}})$ , so  $\tilde{u}_{\text{BH}}$  is the largest  $u$  that solves the equation

$$(36) \quad \alpha \tilde{G}(u) = F_\gamma(u).$$

The strategy is to solve equation (36) in the limit when  $L, a \rightarrow \infty$ . 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 8 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

$$(37) \quad \tilde{G}(u) = \frac{R_\gamma(u)}{\tilde{m}} = \frac{V_\gamma(u)}{\tilde{m}_{0,\gamma}} \frac{\tilde{m}_{0,\gamma}}{\tilde{m}_{0,\gamma} + \tilde{m}_{1,\gamma}} + \frac{W_\gamma(u)}{\tilde{m}_{1,\gamma}} \frac{\tilde{m}_{1,\gamma}}{\tilde{m}_{0,\gamma} + \tilde{m}_{1,\gamma}}.$$

By the weak law of large numbers (33) and Lemma 10, part (3),

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

as  $L \rightarrow \infty$ , where the expectation is given by (17). In addition we have the results of Lemma 8 and Lemma 10, parts (4) and (5). Replacing these three limits in (37), we obtain

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

Now replacing  $\tilde{G}(u)$  by its limit in (36), and solving for  $u$  gives the deterministic solution

$$(38) \quad F_\gamma(u_{\text{BH}}^*) = \frac{\alpha A_1}{A_1 + E[\tilde{m}_{0,\gamma}(0, 1)](1 - \alpha)}.$$

The FDR at the threshold  $u_{\text{BH}}^*$  is bounded by Lemma 12 by

$$(39) \quad \begin{aligned} \text{FDR}(u_{\text{BH}}^*) &\leq P(W(u_{\text{BH}}^*) \leq J - 1) + \frac{E[V(u_{\text{BH}}^*)]}{E[V(u_{\text{BH}}^*)] + J} \\ &= P(W(u_{\text{BH}}^*) \leq J - 1) \\ &\quad + \frac{E[V_\gamma(u_{\text{BH}}^*)] + E[\#\{t \in \tilde{T} \cap \mathbb{T}_\gamma : y_\gamma(t) > u_{\text{BH}}^*\}]}{E[V_\gamma(u_{\text{BH}}^*)] + E[\#\{t \in \tilde{T} \cap \mathbb{T}_\gamma : y_\gamma(t) > u_{\text{BH}}^*\}] + J}, \end{aligned}$$

where we have split  $V_\gamma(u_{\text{BH}}^*)$  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 10, part (1), gives

$$(40) \quad 0 \leq E[\#\{t \in \tilde{T} \cap \mathbb{T}_\gamma : y_\gamma(t) > u_{\text{BH}}^*\}] \leq E[\#\{t \in \tilde{T} \cap \mathbb{T}_\gamma\}] \rightarrow 0.$$

By Lemma 8, the remaining terms of the last fraction in (39) can be written as

$$\begin{aligned} \frac{E[V_\gamma(u_{\text{BH}}^*)]}{E[V_\gamma(u_{\text{BH}}^*)] + J} &= \frac{F_\gamma(u_{\text{BH}}^*)E[\tilde{m}_{0,\gamma}(0, 1)]L}{F_\gamma(u_{\text{BH}}^*)E[\tilde{m}_{0,\gamma}(0, 1)]L + J} \\ &= \frac{F_\gamma(u_{\text{BH}}^*)E[\tilde{m}_{0,\gamma}(0, 1)]}{F_\gamma(u_{\text{BH}}^*)E[\tilde{m}_{0,\gamma}(0, 1)] + J/L}. \end{aligned}$$

Since  $u_{\text{BH}}^*$  solves (38), for  $L \rightarrow \infty$  such that  $J/L \rightarrow A_1$ , the above expression tends to

$$(41) \quad \frac{\alpha E[\tilde{m}_{0,\gamma}(0, 1)]}{\alpha E[\tilde{m}_{0,\gamma}(0, 1)] + A_1 + (1 - \alpha)E[\tilde{m}_{0,\gamma}(0, 1)]} = \alpha \frac{E[\tilde{m}_{0,\gamma}(0, 1)]}{E[\tilde{m}_{0,\gamma}(0, 1)] + A_1} \leq \alpha.$$

Combining equations (40), (41) and Lemma 10, part (3), in (39), we obtain  $\limsup \text{FDR}(u_{\text{BH}}^*) \leq \alpha$ .

Recall that the BH threshold  $\tilde{u}_{\text{BH}}$  solves equation (36), and  $u_{\text{BH}}^*$  satisfies (38), where the empirical marginal distribution,  $\tilde{G}(u)$ , is replaced by its limit. Since  $F_\gamma(t)$  is continuous,  $F_\gamma(\tilde{u}_{\text{BH}}) \rightarrow F_\gamma(u_{\text{BH}}^*)$ , leading to  $\limsup \text{FDR}(\tilde{u}_{\text{BH}}) \leq \alpha$ .  $\square$

6.4. Power.

LEMMA 13. 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),

$$u_{\text{Bon}}^*/[a_j h_{j,\gamma}(t)] \rightarrow 0, \quad u_{\text{BH}}^*/[a_j h_{j,\gamma}(t)] \rightarrow 0$$

in probability, where  $u_{\text{Bon}}^*$  and  $u_{\text{BH}}^*$  are given by (16) and (19), respectively.

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

$$(42) \quad \frac{C_1}{2} \phi\left(\frac{u}{\sigma_\gamma}\right) < F_\gamma(u) < (C_1 + 1) \phi\left(\frac{u}{\sigma_\gamma}\right), \quad C_1 = \sqrt{\frac{2\pi \lambda_{2,\gamma}^2}{\lambda_{4,\gamma} \sigma_\gamma^2}},$$

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 in (42) we obtain

$$(43) \quad 2\sigma_\gamma^2 \left( \log \frac{C_1}{2\sqrt{2\pi}} - \log v \right) < u^2 < 2\sigma_\gamma^2 \left( \log \frac{C_1 + 1}{\sqrt{2\pi}} - \log v \right).$$

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's 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 (42) and (16),

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

which goes to zero by the lemma's conditions.

(2) The FDR threshold  $u_{\text{BH}}^*$  (19) is bounded, so the result is immediate.  $\square$

PROOF OF THEOREM 6. For any threshold  $u$ , the detection power  $\text{Power}(u)$  (20) is greater than  $E[W_\gamma(u)]/J \geq P[W_\gamma(u) = J]$ . But this probability goes to 1 by Lemma 10, part (3), particularly for the deterministic thresholds  $u_{\text{Bon}}^*$  and  $u_{\text{BH}}^*$ . It was shown in the proofs of Theorems 4 and 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.  $\square$

**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. The authors also thank the Editor, Associate Editor and referees for their handling of the manuscript and their useful suggestions.

## REFERENCES

- ADLER, R. J. and TAYLOR, J. E. (2007). *Random Fields and Geometry*. Springer, New York. MR2319516
- ADLER, R. J., TAYLOR, J. E. and WORSLEY, K. J. (2010). Applications of random fields and geometry: Foundations and case studies. Available at <http://webee.technion.ac.il/people/adler/publications.html>.
- ARZENO, N. M., DENG, Z.-D. and POON, C.-S. (2008). Analysis of first-derivative based QRS detection algorithms. *IEEE Trans. Biomed. Eng.* **55** 478–484.
- BACCUS, S. A. and MEISTER, M. (2002). Fast and slow contrast adaptation in retinal circuitry. *Neuron* **36** 909–919.
- BENJAMINI, Y. and HELLER, R. (2007). False discovery rates for spatial signals. *J. Amer. Statist. Assoc.* **102** 1272–1281. MR2412549
- BENJAMINI, Y. and HOCHBERG, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* **57** 289–300. MR1325392
- BRUTTI, P., GENOVESE, C. R., MILLER, C. J., NICHOL, R. C. and WASSERMAN, L. (2005). Spike hunting in galaxy spectra. Technical report, Libera Univ. Internazionale degli Studi Sociali Guido Carli di Roma. Available at <http://www.stat.cmu.edu/tr/tr828/tr828.html>.
- CANDES, E. and TAO, T. (2007). The Dantzig selector: Statistical estimation when  $p$  is much larger than  $n$ . *Ann. Statist.* **35** 2313–2351. MR2382644
- CHUMBLEY, J. R. and FRISTON, K. J. (2009). False discovery rate revisited: FDR and topological inference using Gaussian random fields. *Neuroimage* **44** 62–70.
- CHUMBLEY, J. R., WORSLEY, K., FLANDIN, G. and FRISTON, K. J. (2010). Topological fdr for neuroimaging. *Neuroimage* **49** 3057–3064.
- CRAMÉR, H. and LEADBETTER, M. R. (1967). *Stationary and Related Stochastic Processes. Sample Function Properties and Their Applications*. Wiley, New York. MR0217860
- GENOVESE, C. R., LAZAR, N. A. and NICHOLS, T. E. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage* **15** 870–878.
- HAREZLAK, J., WU, M. C., WANG, M., SCHWARTZMAN, A., CHRISTIANI, D. C. and LIN, X. (2008). Biomarker discovery for arsenic exposure using functional data. Analysis and feature learning of mass spectrometry proteomic data. *J. Proteome Res.* **7** 217–224.
- HELLER, R., STANLEY, D., YEKUTIELI, D., RUBIN, N. and BENJAMINI, Y. (2006). Cluster-based analysis of fMRI data. *Neuroimage* **33** 599–608.
- LI, L. and SPEED, T. P. (2000). Parametric deconvolution of positive spike trains. *Ann. Statist.* **28** 1279–1301. MR1805784
- LI, L. M. and SPEED, T. P. (2004). Deconvolution of sparse positive spikes. *J. Comput. Graph. Statist.* **13** 853–870. MR2109055
- MORRIS, J. S., COOMBES, K. R., KOOMEN, J., BAGGERLY, K. A. and KOBAYASHI, R. (2006). Feature extraction and quantification for mass spectrometry in biomedical applications using the mean spectrum. *Bioinformatics* **21** 1764–1775.
- NICHOLS, T. and HAYASAKA, S. (2003). Controlling the familywise error rate in functional neuroimaging: A comparative review. *Stat. Methods Med. Res.* **12** 419–446. MR2005445
- O’BRIEN, M. S., SINCLAIR, A. N. and KRAMER, S. M. (1994). Recovery of a sparse spike train time series by  $l_1$  norm deconvolution. *IEEE Trans. Signal Process.* **42** 3353–3365.

- PERONE PACIFICO, M., GENOVESE, C., VERDINELLI, I. and WASSERMAN, L. (2004). False discovery control for random fields. *J. Amer. Statist. Assoc.* **99** 1002–1014. [MR2109490](#)
- PERONE PACIFICO, M., GENOVESE, C., VERDINELLI, I. and WASSERMAN, L. (2007). Scan clustering: A false discovery approach. *J. Multivariate Anal.* **98** 1441–1469. [MR2364129](#)
- POLINE, J. B., WORSLEY, K. J., EVANS, A. C. and FRISTON, K. J. (1997). Combining spatial extent and peak intensity to test for activations in functional imaging. *Neuroimage* **5** 83–96.
- PRATT, W. K. (1991). *Digital Image Processing*. Wiley, New York.
- RICE, S. O. (1945). Mathematical analysis of random noise. *Bell System Tech. J.* **24** 46–156. [MR0011918](#)
- SCHWARTZMAN, A., DOUGHERTY, R. F. and TAYLOR, J. E. (2008). False discovery rate analysis of brain diffusion direction maps. *Ann. Appl. Stat.* **2** 153–175. [MR2415598](#)
- SEGEV, R., GOODHOUSE, J., PUCHALLA, J. and BERRY, M. J. II (2004). Recording spikes from a large fraction of the ganglion cells in a retinal patch. *Nature Neuroscience* **7** 1155–1162.
- SIMON, M. (1995). *Digital Communication Techniques: Signal Design and Detection*. Prentice Hall, Englewood Cliffs, NJ.
- SMITH, S. M. and NICHOLS, T. E. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage* **44** 83–98.
- TAYLOR, J. E. and WORSLEY, K. J. (2007). Detecting sparse signals in random fields, with an application to brain mapping. *J. Amer. Statist. Assoc.* **102** 913–928. [MR2354405](#)
- TIBSHIRANI, R., SAUNDERS, M., ROSSET, S., ZHU, J. and KNIGHT, K. (2005). Sparsity and smoothness via the fused lasso. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **67** 91–108. [MR2136641](#)
- WORSLEY, K. J., MARRETT, S., NEELIN, P. and EVANS, A. C. (1996a). Searching scale space for activation in PET images. *Human Brain Mapping* **4** 74–90.
- WORSLEY, K. J., MARRETT, S., NEELIN, P., VANDAL, A. C., FRISTON, K. J. and EVANS, A. C. (1996b). A unified statistical approach for determining significant signals in images of cerebral activation. *Human Brain Mapping* **4** 58–73.
- WORSLEY, K. J., TAYLOR, J. E., TOMAIUOLO, F. and LERCH, J. (2004). Unified univariate and multivariate random field theory. *Neuroimage* **23** S189–195.
- YASUI, Y., PEPE, M., THOMPSON, M. L., BAO-LING, A., WRIGHT, J. G. L., YINSHENG, Q., POTTER, J. D., WINGET, M., THORNQUIST, M. and ZIDING, F. (2003). A data-analytic strategy for protein biomarker discovery: Profiling of high-dimensional proteomic data for cancer detection. *Biostatistics* **4** 449–463.
- ZHANG, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *Ann. Statist.* **38** 894–942. [MR2604701](#)
- ZHANG, H., NICHOLS, T. E. and JOHNSON, T. D. (2009). Cluster mass inference via random field theory. *Neuroimage* **44** 51–61.

A. SCHWARTZMAN  
 Y. GAVRILOV  
 DEPARTMENT OF BIostatISTICS  
 HARVARD SCHOOL OF PUBLIC HEALTH  
 AND  
 DANA-FARBER CANCER INSTITUTE  
 450 BROOKLINE AVE., CLS 11007  
 BOSTON, MASSACHUSETTS 02446  
 USA  
 E-MAIL: [armins@hsph.harvard.edu](mailto:armins@hsph.harvard.edu)  
[yuliagavrilov@gmail.com](mailto:yuliagavrilov@gmail.com)

R. J. ADLER  
 DEPARTMENT OF ELECTRICAL ENGINEERING  
 TECHNION, ISRAEL INSTITUTE OF TECHNOLOGY  
 HAIFA 32000  
 ISRAEL  
 E-MAIL: [robert@ee.technion.ac.il](mailto:robert@ee.technion.ac.il)