NeuroImage 197 (2019) 402–413

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Peak p-values and false discovery rate inference in neuroimaging

Armin Schwartzman *, Fabian Telschow

Division of Biostatistics, University of California, San Diego, USA

A B S T R A C T

Peaks are a mainstay of neuroimage analysis for reporting localization results. The current peak detection procedure in SPM12 requires a pre-threshold for approximating p-values and a false discovery rate (FDR) nominal level for inference. However, the pre-threshold is an undesirable feature, while the FDR level is meaningless if the null hypothesis is not properly defined. This article provides: 1) a peak height distribution for smooth Gaussian error fields, which does not require a screening pre-threshold; 2) a signal-plus-noise model where FDR of peaks can be controlled and properly interpreted. Matlab code for calculation of p-values using the exact peak height distribution is available as an SPM extension.

1. Introduction

Since their introduction to neuroimaging (Poline et al., 1997) and identification as topological markers (Chumbley et al., 2010), peaks have become a mainstay of neuroimage analysis for statistical inference and localization. Peaks are critical for communicating and reporting neuroimaging results, particularly in task fMRI; tools like BrainMap and Neurosynth.org aggregate peak coordinates studies, providing the basis for meta-analyses. Peaks have also become essential for practical power analysis in neuroimaging (Durnez et al., 2014, 2016). And despite the appeal of cluster inference, peaks remain the most reliable form of topological inference given the inaccuracy of the current cluster null distribution approximations derived from Gaussian random field theory (Eklund et al., 2016).

The paradigm for peak inference, proposed by Chumbley et al. (2010) and currently implemented in SPM12, was formally studied in Schwartzman et al. (2011) and Cheng and Schwartzman (2017). Described there in general as the "smoothing and testing of maxima" (STEM) algorithm, it consists of the following steps:

1. *Kernel smoothing*: to increase the signal-to-noise ratio (SNR).
2. *Candidate peaks*: find local maxima of the smoothed field above a pre-threshold.
3. *P-values*: computed at each local maximum under the null hypothesis of no signal in a local neighborhood.
4. *Multiple testing*: apply a multiple testing procedure and declare as detected peaks those local maxima whose p-values are significant.

This general recipe relies on two critical elements for proper inference: 1) calculation of valid p-values in Step 3, and 2) interpretation of the nominal error rate in Step 4. The main goal of this article is to formally address these two issues in the context of neuroimaging. The discussion in this paper refers to SPM because FSL and AFNI do not provide peak inference.

1.1. Peak p-values

To calculate p-values, SPM12 relies on a screening pre-threshold, so that only high enough peaks are included in the analysis. However, a pre-threshold is generally not a desirable feature. How to choose it is not entirely clear, as it should be high enough to enter the regime of the tails of the distribution, but not so high that it eliminates potential discoveries. More importantly, changing the pre-threshold affects the analysis, which not only reduces confidence in the results but opens the door to manipulation of the pre-threshold until desirable results are obtained.

The main reason for requiring a pre-threshold has been one of necessity, as it has been the only known avenue to compute peak p-values (Adler, 1981; Chumbley et al., 2010). Recently, however, an exact formula for the peak height distribution of isotropic Gaussian fields has been obtained (Cheng and Schwartzman, 2015a; b, 2018). The formula is exact in the sense that it gives the true probability distribution of peak heights for all significance thresholds under the prescribed isotropic Gaussian noise model.

When the field has unit variance, this formula depends only on the shape of the autocorrelation function through a single parameter, not its spatial scaling. Thus we conjecture here that its validity is not restricted to isotropic fields but holds more generally for certain classes of nonstationary Gaussian fields, such as anisotropic fields (affine

* Corresponding author. Division of Biostatistics and Halicioğlu Data Science Institute University of California, San Diego, USA.
E-mail addresses: armins@ucsd.edu (A. Schwartzman), ftelschow@ucsd.edu (F. Telschow).

https://doi.org/10.1016/j.neuroimage.2019.04.041
Received 28 June 2018; Received in revised form 12 April 2019; Accepted 14 April 2019
Available online 24 April 2019

transformations of isotropic fields) and locally stationary fields (where the stationarity property holds locally). In this paper, we evaluate this claim via simulations for a variety of such fields, both Gaussian and *t*. A method is proposed for estimating the required shape parameter from data.

To connect to current practice, we also consider the overshoot distribution of peak height above a pre-threshold. A formula for this distribution is obtained directly from the peak height distribution and is exact in the sense that it gives the true probability distribution of peak heights above the pre-threshold under the prescribed isotropic Gaussian noise model. The various existing approximation methods (Adler, 1981; Chumbley et al., 2010) are only valid for high pre-thresholds and become better as the pre-threshold increases. We use simulations to assess the validity of the exact formula and evaluate the accuracy of the approximations under stationary and non-stationary scenarios. As an example, both the peak height distribution with no pre-threshold and the overshoot distributions and its approximations are illustrated in fMRI data from Moran et al. (2012).

We remark that both the exact and approximate peak p-value formulas are uncorrected and do not depend on the search volume; the dependence on search volume comes in through the multiple testing procedure, discussed next.

1.2. False discovery rate (FDR) inference for peaks

The second critical issue in peak inference is the interpretation of the error rate when declaring statistical significance. Based on the premise (Chumbley and Friston, 2009) that the signal is present everywhere in the brain, Chumbley et al. (2010) propose peaks as a form of topological inference and declare peaks as true or false positives based on their distance from a true peak. Proper error control, however, requires p-values to be valid under the null hypothesis that defines the false positives. In their case, this corresponds to being far from a true peak with a non-zero signal. Instead, p-values based on peak height are related to image intensity and are designed to be valid under the null hypothesis of no signal at the location of the peak.

As a consequence, the procedure in Chumbley et al. (2010) cannot be interpreted as controlling the FDR at the prescribed nominal level; if the procedure is applied with a nominal FDR of, say, 0.05, one cannot conclude that 5% of the detected peaks are expected to be false positives. Instead, the FDR level becomes more of a regularization parameter that controls the procedure's detection performance. One may think of the procedure as separating more interesting peaks from less interesting ones in the sense of Efron (2004). This approach was considered by Schwartzman et al. (2009) in the context of voxelwise inference, but it is more difficult to translate to the context of peak inference and we do not pursue it here.

One way to achieve meaningful FDR inference is to allow the signal to be zero in some regions and nonzero in others. The need for such a distinction is recognized by Li et al. (2018) in relation to peak and cluster inference. Specifically for peaks, Cheng and Schwartzman (2017) propose a more formal model where the signal is nonzero only on a set of unimodal functions with finite support, so that each signal component is topologically represented by its own mode. This model allows a precise definition of false positives as significant peaks in the zero-region. Moreover, if the noise field is stationary and ergodic, Cheng and Schwartzman (2017) prove asymptotic FDR control under this model as both the sample size and the search domain increase. Because their simulations are limited to 2D fields, in this paper we perform simulations to evaluate FDR control and detection power in 3D fields, more representative of neuroimaging, and under a larger variety of noise field conditions.

The rest of the paper addresses the calculation of p-values for peaks in Section 2 and the FDR inference in Section 3. The specific implementation for fMRI is discussed in Section 4 and some concluding remarks are provided in Section 5. Matlab code for calculation of p-values using the exact peak height distribution is available as an SPM extension in the SPM12 Extensions website (http://www.fil.ion.ucl.ac.uk/spm/ext//#STEM). All code and data used in this paper are available in GitHub (https://github.com/fjete88/STEM).

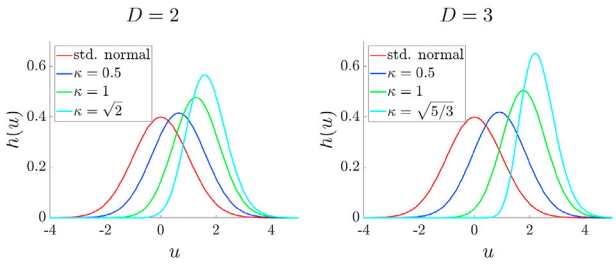

Fig. 1. Density functions $h(x;\kappa)$ of the height distribution of local maxima for dimensions $D = 2$ (left) and $D = 3$ (right). The largest value of κ allowed is $\kappa =$

2. Peak p-values

2.1. Peak height distribution

Suppose that $z(s)$ is a smooth Gaussian random field, with mean 0 and variance 1, indexed by a continuous space location index s in a Euclidean space of dimension D. Smoothness here refers to the field being 3-times differentiable.[1] This holds if the smoothing filter is itself 3-times differentiable, such as a Gaussian or a quartic filter.

The peak height distribution of a local maximum of $z(s)$, as a function of a height threshold u, is the probability that $z(s)$ is greater than u given that a local maximum has occurred at s. This is *not* the same as the marginal probability

$$\mathrm{P}[z(s) > u] = 1 - \Phi(u), \tag{1}$$

where $\Phi(u)$ is the cumulative distribution function (cdf) of a standard normal random variable. Marginal peak p-values using (1) are too liberal because they do not take into account the fact that peaks are selected for being local maxima. Peaks are, by definition, higher than their neighboring voxels, and thus likely to be higher than any voxel chosen at random. This distinction is important particularly when reading the output of SPM12, which reports marginal p-values for peaks calculated according to (1); those p-values are misleadingly too low.

The peak height distribution is properly defined as the *conditional* probability

$$F(u) = \mathrm{P}\big[z(s) > u \;\big|\; \nabla z(s) = 0 \text{ and } \nabla^2 z(s) \prec 0\big], \tag{2}$$

where $\nabla z(s)$ is the gradient of z at s, $\nabla^2 z(s)$ is the Hessian of z at s, and the symbol $\prec$ indicates negative definiteness, specifying s as a local maximum. The conditioning terms in (2) specify the selection criteria and make this distribution stochastically greater than the standard Gaussian distribution (1). Fig. 1 illustrates the difference between the two distributions in the case of an isotropic field. Note that the peak distribution is always shifted toward higher values with respect to the standard normal. The peak distribution depends on a single parameter κ, explained next.

2.2. Peak height distribution for isotropic Gaussian fields

When the field $z(s)$ is isotropic, i.e. its covariance function is invariant under translations and rotations, then the height distribution of local maxima has a closed-form expression for dimension up to $D = 3$ (Cheng

[1] To be precise, 3-times differentiable is not necessary but is sufficient, while 2-times differentiable is necessary but not sufficient. We settle on 3-times differentiable for simplicity.

and Schwartzman, 2015b, 2018). Specifically, the tail probability (2) can be written as

$$F(u) = F(u;\kappa) = \int_u^\infty h(x;\kappa)\, dx, \tag{3}$$

where $h(x;\kappa)$ is the probability density function (pdf) of the height of local maxima, depending on a single parameter κ. The explicit densities for $D = 2$ and $D = 3$ are given by equations (25) and (26) in Appendix A.1 and plotted in Fig. 1 for various values of κ. Note that increasing κ shifts the distribution toward higher values.

The single parameter κ contains all the necessary information about the covariance function of the field and is defined as follows. Recall that $z(s)$ is assumed to have mean 0 and variance 1. If we write the autocorrelation function (acf) of the field as $R(t) = \mathrm{E}[z(s)z(s+t)] = \rho(||t||^2)$ for an appropriate function ρ, then κ is the unit-less quantity

$$\kappa = -\rho' \Big/ \sqrt{\rho''} > 0, \tag{4}$$

where $\rho' = \rho'(0) < 0$ and $\rho'' = \rho''(0) > 0$ are the first and second derivatives of the function ρ evaluated at zero (Cheng and Schwartzman, 2015a). Common examples are the following:

- *Squared exponential or Gaussian acf*: Here $R(t) = \exp(-c\|t\|^2)$ with $\rho = e^{-cr}$, $c > 0$. Thus $\rho' = -c$, $\rho'' = c^2$ and $\kappa = 1$.
- *Inverse squared polynomial or Cauchy acf*: Here $R(t) = (1 + c||t||^2)^{-b}$ with $\rho(r) = (1 + cr)^{-b}$, $c > 0$, $b > 0$. Thus $\rho' = -cb$, $\rho'' = c^2 b(b+1)$ and $\kappa = \sqrt{b/(b+1)} < 1$.

Note that the value of κ only depends on the shape of the acf near the origin. In general, the largest value that κ may take is $\kappa = \sqrt{(D+2)/D}$, obtained in the special case of a field satisfying Helmholtz's equation $\mathrm{tr}[\nabla^2 z(s)] = 2D\rho' z(s)$ (Cheng and Schwartzman, 2018).

An important property of the peak height distribution is that it is invariant under scalings of the space, i.e. it is the same for the field $z(s)$ as for the field $z(as)$, for any fixed real number $a \neq 0$. This is not surprising, as scaling of the space does not affect peak height. Formally, this property stems from the spatial scale invariance of the parameter κ. It follows directly from definition (4) and explains why the values of κ in the two examples above do not depend on the constant c. We exploit this scale-invariant property in the simulations below.

2.3. Simulations of isotropic Gaussian fields

To confirm the 3D peak height formula (26) in a finite resolution setting, we simulated 1000 random fields of size $50 \times 50 \times 30$ voxels, each field obtained as the convolution of white Gaussian noise with an isotropic filter $w(s)$. To ensure validity of the convolution at the margins, the white noise was generated on a larger domain adding a filter support length in each direction, and the convolved field was then trimmed to the desired domain size. Further, the generated fields were normalized by the factor $w = [\int w^2(s)\, ds]^{1/2}$ to ensure unit variance.

Fig. 2 shows the results using a Gaussian filter with full-width half max (FWHM) equal to $\sqrt{8\log 2} \approx 2.355$ times the standard deviation of the filter. The top left panel shows that the distribution of peak heights closely matches the theoretical density (26) with $\kappa = 1$. To zoom in on the right tail of the distribution, the top right panel shows the empirical distribution of the p-values. As the filter size increases, the empirical peak height distribution becomes closer to the theory and the empirical p-value distribution becomes closer to the uniform distribution. Because of the discrete grid, the measured peak heights are slightly smaller than they would be in a continuous field, implying that the p-value distribution is stochastically larger than the uniform. Larger p-values are still valid for inference, albeit conservative.

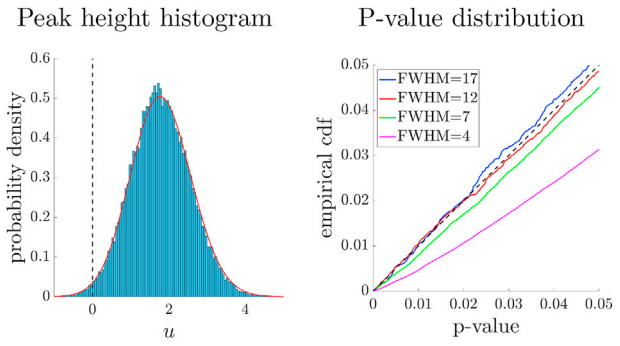

Fig. 2. Peak height simulations for 1000 instances of a 3D isotropic Gaussian field generated with a Gaussian filter. *Left column:* Histogram of peak heights for FWHM = 7 voxels; superimposed is the theoretical peak height density with $\kappa = 1$. *Right column:* Empirical p-value distributions for various values of FWHM.

To confirm formula (26) in the case of a non-Gaussian autocorrelation function, we repeated the above simulations (1000 random fields of size $50 \times 50 \times 30$ voxels), using instead an isotropic 3D quartic (biweight) filter

$$w(s) = \left(1 - s^2/a^2\right)^2, \qquad s < a, \tag{5}$$

with FWHM equal to $a\sqrt{4 - 2\sqrt{2}} \approx 1.082a$.

Fig. 3 (top row) confirms the validity of the distribution for this filter when κ is estimated from the data for the appropriate FWHM (estimation of κ is explained in the next section). Fig. 3 (bottom row) shows the p-value distributions when the value of κ is misspecified. On the left, the nominal value of $\kappa = 0.9049$ for a continuous process (see Appendix A.2) is too small, yielding liberal p-values. On the right, the nominal Gaussian

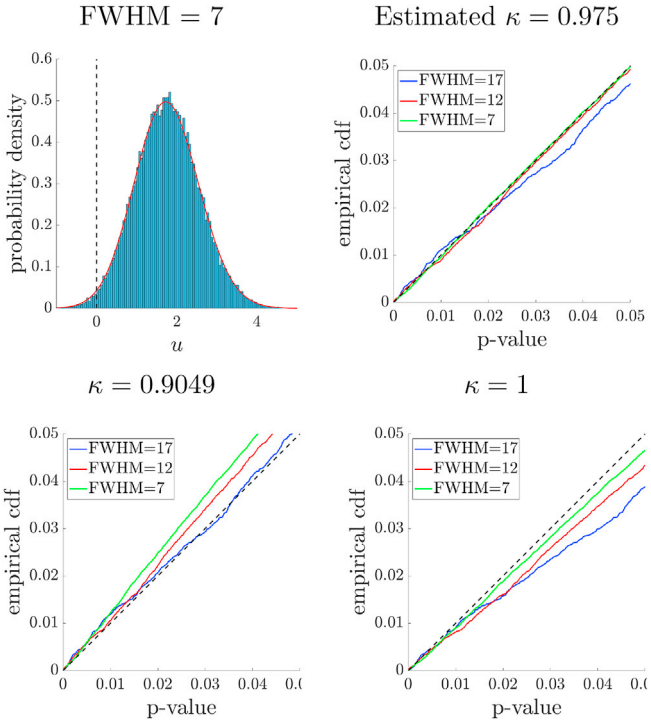

Fig. 3. Peak height simulations for 1000 instances of a 3D isotropic quartic filter. *Top left:* Histogram of peak heights for FWHM = 7 voxels; superimposed is the theoretical peak height density with estimated $\kappa = 0.975$. *Top right:* Empirical p-value distributions for estimated values of κ at each FWHM. *Bottom left:* Empirical p-value distributions for limiting theoretical value $\kappa = 0.9049$. *Bottom right:* Empirical p-value distributions for $\kappa = 1$.

value $\kappa = 1$ is too large, yielding conservative p-values.

2.4. Unknown κ

In the simulations above, the value of κ was taken as known, determined by the shape of the smoothing filter. If the shape of the correlation function cannot be assumed a priori, then the parameter κ can be estimated from the data as follows. Suppose that the observed noise field is $\tilde{z}(s) = \sigma z(s)$ with mean 0 and variance σ^2. Following definition (4), it is shown by Cheng and Schwartzman (2017) that

$$\rho' = -\frac{\lambda_2}{2\sigma^2}, \qquad \rho'' = \frac{\lambda_4}{12\sigma^2} \qquad \Rightarrow \qquad \kappa = \frac{\sqrt{3}\lambda_2}{\sqrt{\sigma^2\lambda_4}}, \tag{6}$$

where λ_2 and λ_4 are the spectral moments of the field $\tilde{z}$, defined as the variance of the first and second spatial derivatives, respectively. Assuming isotropy, the variances σ^2, λ_2 and λ_4 can be estimated empirically from the numerical derivatives in all three spatial directions and at all spatial locations. If multiple instances of the field are available, then they are all included in the calculation of the variances. Plugging in (6) yields an estimate of κ.

Table 1 shows the estimation accuracy of the above method when the field is produced as a convolution of white Gaussian noise with a Gaussian or a quartic filter. As the size of the smoothing filter increases, the estimated value slowly decreases toward the nominal value for a continuous process. For both filters, the discreteness of the grid biases the numerical derivatives and results in values that are larger than the nominal, more appropriate in such more realistic situations. In all cases, the standard error decreases with the sample size, as expected.

2.5. Simulations of non-isotropic fields

The scale-invariant property indicates that the peak height distribution only depends on the shape of the covariance function. This suggests that the height distribution may still be valid if the scaling factor is different in every direction in space, i.e. if the field is anisotropic, or if it changes slowly over space, i.e. if the field is mildly nonstationary.

To evaluate these claims, we performed simulations as above (1000 random fields of size $50 \times 50 \times 30$ voxels) under an anisotropic scenario and a nonstationary scenario. In the anisotropic scenario, the filter was Gaussian anisotropic with three different values of FWHM in the three spatial directions. In the nonstationary scenario, half of the white noise field (before smoothing) was binned into blocks of 2×2 voxels. Then the entire noise field was convolved with a Gaussian filter, producing a smooth field with effectively a different acf in the two halves. The fields were normalized to have unit variance at every voxel.

Fig. 4 shows the peak height and p-value distributions. The peak height distribution again gets closer to the theory as the filter FWHM increases.

2.6. Application to t-fields

In practice, the null distribution of the test statistic map is typically t rather than Gaussian. In Fig. 5, t-fields where created by voxelwise calculation of a t-statistic using

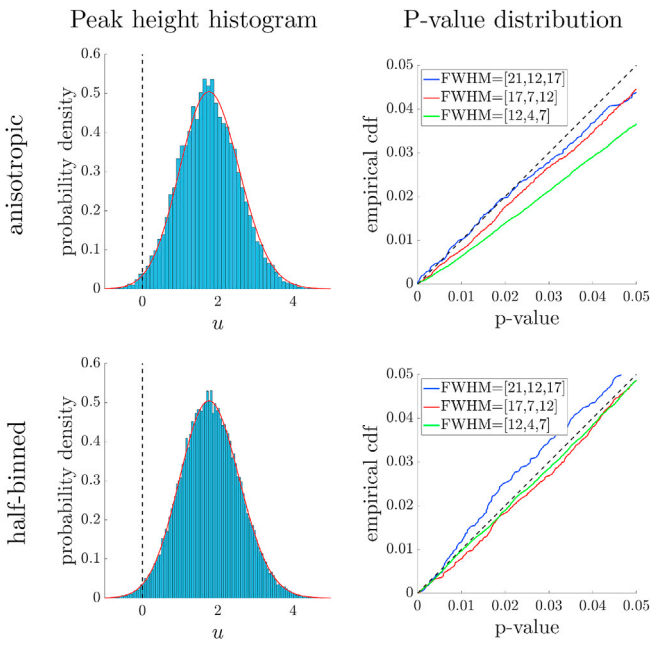

Figure 4. Peak height simulations for 1000 instances of a 3D Gaussian anisotropic field (top row) and half-binned nonstationary field (bottom row), generated with a Gaussian filter. *Left column:* Histogram of peak heights over 1000 simulated fields for FWHM = [12,4,7] (top row) and FWHM = 7 (bottom row); superimposed is the theoretical peak height density with $\kappa = 1$. *Right column:* Empirical p-value distributions for various values of FWHM.

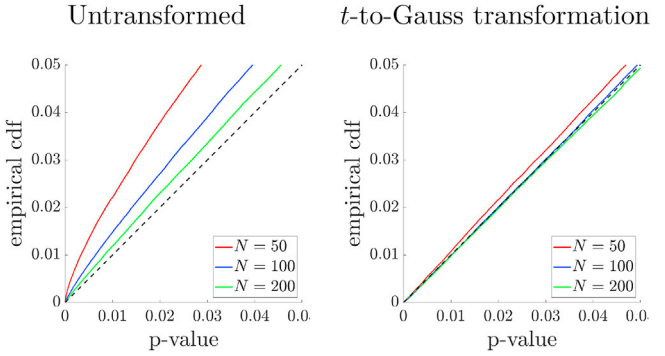

Fig. 5. Peak height simulations for a t field with N d.f., generated from Gaussian fields with FWHM = 7 voxels (1000 simulations). *Left:* Direct application of the Gaussian peak height distribution formula. *Right:* Application of the Gaussian peak height distribution formula after a t-to-Gaussian quantile transformation.

$$T(s) = \frac{\varepsilon(s)}{\sqrt{\sum_{i=1}^{N} Z_i^2 / N}}, \tag{7}$$

where $Z_1, \ldots, Z_N$ and $\varepsilon(s)$ are i.i.d. isotropic Gaussian fields such as those simulated in Fig. 2 with FWHM = 7. Because the exact height distribution

Table 1
Estimation of the parameter κ as a function of sample size N for Gaussian and quartic smoothing filters. Simulation standard errors are in parentheses. The rightmost column shows the nominal value for a continuous process.

Filter	N	FWHM = 7	FWHM = 12	FWHM = 17	Continuous
Gaussian	50	1.0091 (0.0042)	1.0039 (0.0091)	1.0038 (0.0153)	1
	100	1.0089 (0.0029)	1.0037 (0.0061)	1.0035 (0.0103)	
	200	1.0091 (0.0021)	1.0040 (0.0043)	1.0039 (0.0072)	
Quartic	50	0.9755 (0.0044)	0.9537 (0.0078)	0.9392 (0.0110)	0.9049
	100	0.9754 (0.0031)	0.9537 (0.0055)	0.9396 (0.0078)	
	200	0.9755 (0.0022)	0.9533 (0.0039)	0.9394 (0.0055)	

is designed for Gaussian fields, not t-fields, the p-value distribution in Fig. 5(left) is too liberal, although it improves as the number of d.f. increases.

To improve the accuracy for t-fields, we transform the t-fields to be marginally Gaussian by means of the quantile transformation

$$Z = -\Phi^{-1}[F_{t,N}(-T)], \tag{8}$$

where $F_{t,N}$ is the cdf of the t distribution with N d.f. The minus signs ensure the accuracy of the transformation at the upper tail of the distribution, where it matters most. We then apply the exact Gaussian height distribution to the transformed t-fields. Fig. 5(right) shows that the p-value distribution is now very accurate, especially if the number of d.f. is higher than 100.

2.7. Overshoot distribution

The overshoot distribution (Adler, 1981) is defined as the probability that a peak is higher than u given that it is already higher than a pre-threshold $v < u$:

$$F(u,v) = \mathrm{P}\big[z(s) > u \;\big|\; \nabla z(s) = 0, \nabla^2 z(s) \prec 0 \text{ and } z(s) > v\big]. \tag{9}$$

Using conditional probability rules, the overshoot distribution (9) can be readily obtained as

$$F(u,v) = \frac{F(u)}{F(v)}, \tag{10}$$

where $F(u)$ is the peak height distribution (2). Thus, for isotropic Gaussian fields, the overshoot distribution can be calculated exactly replacing $F(u) = F(u;\kappa)$ as given by (24).

For more general smooth fields, approximations have been obtained by two different approaches, all becoming more accurate as the pre-threshold v increases. Working directly from the Kac-Rice formula, Adler obtained the approximation (Adler, 1981; Adler et al., 2010)

$$F(u,v) \approx \frac{u^{D-1}e^{-u^2/2}}{v^{D-1}e^{-v^2/2}}, \tag{11}$$

for stationary Gaussian fields, although Cheng and Schwartzman (2015a) showed it to be asymptotically valid also for nonstationary Gaussian fields.

As an alternative approach, the peak p-values in SPM12 are calculated according to the approximation

$$F(u,v) \approx \frac{\mathrm{EEC}(u)}{\mathrm{EEC}(v)}, \tag{12}$$

where $\mathrm{EEC}(u)$ is the expected Euler characteristic at level u, given in its usual form by the Gaussian kinematic formula (GKF) using random field theory for Gaussian or t-fields (Worsley et al., 1996, 2004). We refer to this approximation below as 'GKF Ratio'.

In Chumbley et al. (2010), (12) is simplified by taking only the highest order (volume) term in the GKF expansion. This approximation, which we refer to below as 'GKF1 Ratio', takes the form

$$F(u,v) \approx \frac{H_{D-1}(u)e^{-u^2/2}}{H_{D-1}(v)e^{-v^2/2}} \tag{13}$$

for a Gaussian field, where $H_{D-1}(x)$ is the Hermite polynomial of order $D-1$, and

$$F(u,v) \approx \frac{(1+u^2/\nu)^{-(\nu-1)/2}[u^2(\nu-1)/\nu - 1]}{(1+v^2/\nu)^{-(\nu-1)/2}[v^2(\nu-1)/\nu - 1]}. \tag{14}$$

for a t-field with ν degrees of freedom. The GKF1 Ratio approximation (13) was also derived independently by Cheng and Schwartzman (2015a) using the Kac-Rice formula and was shown rigorously there to be accurate for stationary smooth Gaussian fields. However, currently there is no formal proof of validity for the overshoot distributions (12), (13) or (14) for nonstationary fields. An exception is (13) in $D = 2$, which becomes the same as (11).

Fig. 6 (left column) compares the various approximations above in the Gaussian isotropic case for a fixed pre-threshold of $v = 2.5$. For $D = 2$, the GKF1 Ratio approximation (13) and Adler's approximation (11) are the same. For $D = 3$, Adler's approximation is almost exact. In general, the GKF1 Ratio approximation (13) is conservative and the GKF Ratio approximation (12) is liberal, although the effect is most noticeable for $D = 3$. As a better reference to the simulations below, Fig. 6 (right column) shows the corresponding p-value distributions based on the exact overshoot distribution, so that the latter becomes uniform. Notice that the GKF Ratio approximation (12) depends on the Lipschitz-Killing curvatures (LKCs), or resel counts, of the field over the domain, and it gets worse as the filter FWHM increases. We offer an explanation for this phenomenon in Section 5 below.

To separate the estimation problem from the form of the overshoot approximation, in the following simulations we use the exact theoretical values of the required parameters, namely κ for the exact overshoot and the LKCs for the GKF Ratio approximation.

2.8. Overshoot distribution simulations

To compare the various options for the overshoot distribution, we performed simulations as above (random fields of size $50 \times 50 \times 30$ voxels) under an isotropic scenario. Compared to the peak height simulations above, the overshoot threshold results in a smaller number of peaks from which empirical distributions can be estimated. To compensate, we here increase the number of repetitions to $10,000$.

The first set of simulations evaluates the effect of the pre-threshold and the amount of smoothing. Here, white Gaussian noise was convolved with an isotropic Gaussian filter as in Fig. 2. Fig. 7 shows that the exact overshoot distribution (10) is precise for any pre-threshold v if the filter FWHM is high enough, otherwise it is conservative. Consistent with Fig. 6, the GKF1 Ratio and Adler approximations in Fig. 7 are always conservative, but get better as the pre-threshold and the FWHM increase. The GKF Ratio, on the other hand, becomes liberal if the FWHM is too high. This is consistent with the dependence on FWHM observed in Fig. 6.

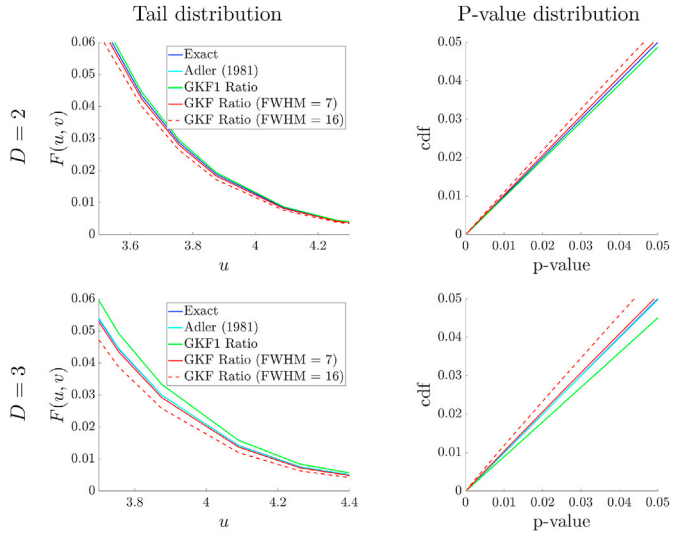

Fig. 6. Overshoot probabilities $F(u,v)$ of an isotropic Gaussian field with $\kappa = 1$ for dimensions $D = 2$ (top row) and $D = 3$ (bottom row), pre-threshold $v = 2.5$. *Left column:* Tail probabilities. *Right column:* P-value distribution with respect to the exact distribution (uniform); same legend as left column.

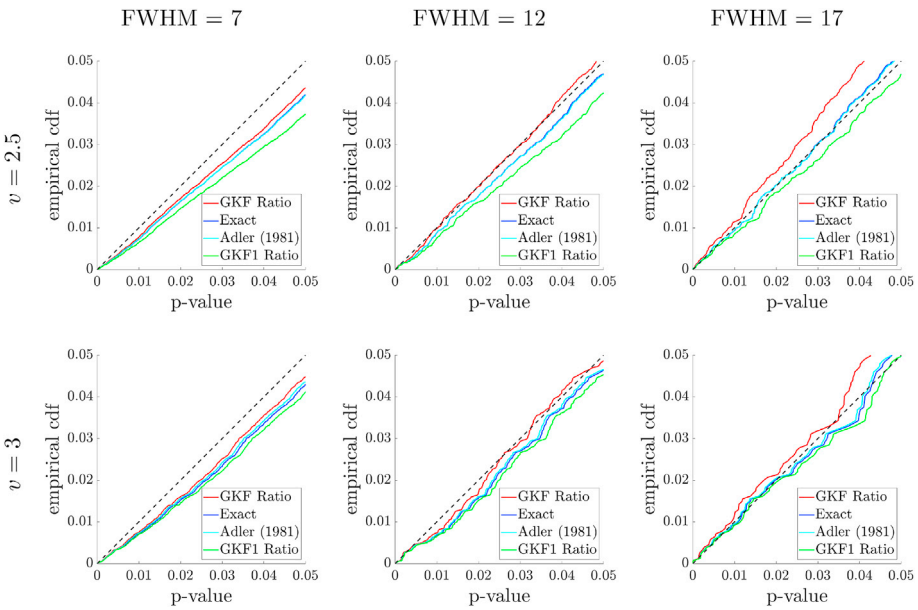

Fig. 7. Overshoot p-value distributions for an isotropic Gaussian field obtained with a Gaussian filter (10,000 simulations), as a function of the pre-threshold v and the filter FWHM.

Fig. 8 (top row) compares the overshoot distribution approximations for t-fields with N d.f. for a fixed pre-threshold $v = 2.5$. Here, t-fields where created by voxelwise calculation of a t-statistic from isotropic Gaussian fields as those simulated in Fig. 7 with fixed FWHM = 7 voxels. The exact distribution (10) is inaccurate because it is designed for Gaussian fields, but improves as the number of d.f. increases. Consistent with the top left panel of Fig. 7, the GKF Ratio approximation for t-fields (12) works well here, but the GKF1 Ratio approximation for t-fields (14) is conservative and does not improve with increasing d.f.

In contrast, Fig. 8 (bottom row) shows the results when the t-fields are transformed to marginally Gaussian by the quantile transformation (8). Applied to the Gaussianized fields, the exact overshoot distribution is slightly conservative at this value of FWHM, again improving with increasing number of d.f. The Gaussian version of GKF1 Ratio is always conservative. The Gaussian version of the GKF Ratio approximation is in this case very similar to the exact distribution. However, similar to Fig. 7, it becomes liberal if the FWHM increases further (results not shown).

3. FDR inference

3.1. Modeling assumptions

Cheng and Schwartzman (2017) show that rigorous peak FDR control is possible under the following modeling assumptions. They work under the signal + noise model

$$y(s) = \mu(s) + \sigma z(s), \qquad s \in S, \tag{15}$$

where $\mu(s)$ represents the signal and $\sigma z(s)$ the noise with $\sigma > 0$, both defined on a search region $S \subset \mathbb{R}^D$. As shown in Section 4 below,

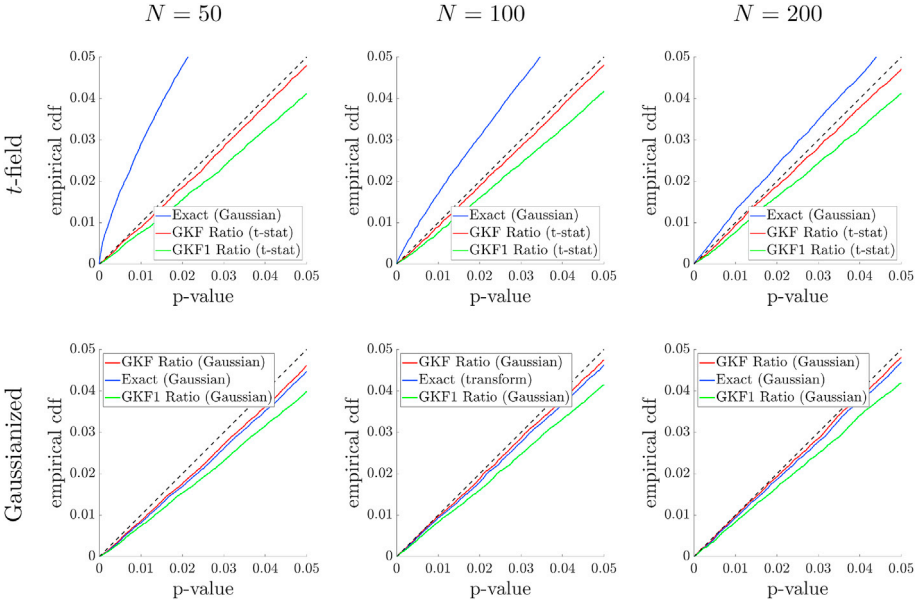

Fig. 8. *Top row:* Overshoot p-value distributions for an isotropic t-field with N d.f. (10,000 simulations), generated from isotropic Gaussian fields (Gaussian filter with FWHM = 7 voxels). The pre-threshold is fixed at $v = 2.5$. *Bottom row:* Same as top row, after t-field is transformed by a t-to-Gaussian quantile transformation.

coefficient estimates from the general linear model in fMRI can be written in this form.

The signal $\mu(s)$ is modeled as a linear composition of J bump functions

$$\mu(s) = \sum_{j=1}^{J} a_j h_j(s), \qquad a_j > 0. \tag{16}$$

Each bump function $h_j(s)$ is positive, unimodal and has compact connected support, representing a peak to be detected. For identifiability of the coefficient a_j, each bump function $h_j(s)$ is required to have unit action, i.e. integral equal to 1. In addition, bump functions are required to be twice differentiable within their support and have no critical points other than their mode. These functions need not be concave within their support nor their level sets need to be convex, allowing for a large variety of possibilities for the signal $\mu(s)$.

For the noise term $z(s)$, in addition to it being a smooth Gaussian field as in Section 2, here it is required that the noise field be stationary ergodic. Ergodicity refers to the property that statistics of the field over many realizations (e.g. moments or excursion probabilities) can be estimated from a single realization of the field by empirical calculations over space if the domain is large enough. Ergodicity usually holds if the autocorrelation function of the field decays fast enough, that is, if the values of the field at different locations tend to be less correlated as the locations are farther apart from each other. It does not hold if correlation reaches far in space, for example, if the noise is a periodic function over space.

3.2. FDR control

Because the location of truly detected peaks may shift as a result of noise, a significant local maximum is called a true positive in Cheng and Schwartzman (2017) if it falls anywhere inside the support of a true peak; otherwise, it is called a false positive. This allows for a formal definition of FDR as the expected proportion of falsely detected peaks among significant peaks. The detection power is defined as the expected fraction of truly discovered peaks out of all the J existing peaks. This detection power turns out to be the same as the average probability of detection for each of the existing peaks.

After filtering the data, local maxima are found, a p-value is attached to each one of them, and an multiple testing correction procedure is applied to the list of peak p-values to determine a significance threshold. This approach, similar to Chumbley et al. (2010), is called the "smoothing and testing of maxima" (STEM) algorithm in Cheng and Schwartzman (2017). Specifically, when applying the Benjamini-Hochberg (BH) procedure (Benjamini and Hochberg, 1995) to the list of peak p-values, Cheng and Schwartzman (2017) formally show that:

1. The FDR tends to $\alpha M/(M+J)$, where J is the number of true peaks and M is the expected number of local maxima over the null region.
2. The detection power tends to 1.

Both these results hold asymptotically as the search space and the signal-to-noise ratio (SNR) increase. The SNR is only required to increase faster than the logarithm of the size of the search space, so the size of the search space may be almost exponentially larger than the SNR. This type of scaling relationship is typical of high dimensional data problems (e.g. Candes and Tao (2007); Zhang (2010)).

The reasons for requiring both the search space and the SNR to increase are twofold. First, increasing search space is required to apply ergodicity. As the search space increases, more peaks appear, both true and false. By ergodicity, the empirical distributions of both true and false peak heights involved in the BH procedure converge to the theoretical distributions, allowing nominal FDR control.

Second, increasing SNR is necessary to avoid error inflation from smoothing and to guarantee power consistency. As pointed out by Chumbley and Friston (2009) and Chumbley et al. (2010), convolution with a filter artificially extends the signal beyond its original domain. This additional extent is called *transition region* in Cheng and Schwartzman (2017). Different solutions to this problem have been offered in the literature. For example, Pacifico et al. (2007) suggested "shaving" the extra signal, although they found their procedure difficult to calibrate so that it would not shave too much. In Cheng and Schwartzman (2017), the assumptions of unimodality and lack of other critical points for the peak functions in (16) guarantee that, as the SNR increases, the probability of observing any peaks in the transition region goes to 0. In other words, asymptotically, peaks occur either in the null region where the signal is zero, or very close to true peaks, eliminating the error inflation. As an additional benefit, the detection power goes to 1.

The limiting FDR above depends on the expected number of local maxima M over the null region. This quantity can be calculated as (Cheng and Schwartzman, 2015b)

$$M = (\text{Area of the null region}) \times \frac{1}{\sqrt{3}\pi}\left(\frac{\rho''}{-\rho'}\right) \tag{17}$$

for $D = 2$ and

$$M = (\text{Volume of the null region}) \times \frac{29\sqrt{6}-36}{36\pi^2}\left(\frac{\rho''}{-\rho'}\right)^{3/2} \tag{18}$$

for $D = 3$, where ρ' and ρ'' are the same quantities used in the calculation of κ in (4). Formulas (17) and (18) do not depend on the variance of the field and remain the same if the noise field does not have variance 1.

3.3. Simulations of FDR control and detection power

The simulations in Cheng and Schwartzman (2017) were limited to 2D images. To evaluate the methods in a situation closer to neuroimaging, we present here simulation results in a 3D setting. As signal, five non-overlapping signal bump functions were placed in a volume of size $50 \times 50 \times 30$ voxels. The signal bumps were generated as 3D outer products of 1D quartic filters with varying support sizes corresponding to FWHM between 4 and 8 voxels, heights within 25% of each other, and scaled by a global constant A to control the SNR (Fig. 9).

The first simulation setup considers Gaussian noise. Here the noise field was produced by convolution of white noise with a Gaussian filter using the SPM12 function spm_smooth and renormalizing to unit variance. The SNR is controlled by scaling the bumps so that the ratio (height of second highest peak)/(SD of noise field) ranges from 2 to 7.

Fig. 10 (left column) shows the realized FDR and detection power as a function of the SNR and the noise FWHM for a nominal FDR level $\alpha = 0.05$ when using the exact height distribution. As predicted by the theory, the FDR is below the theoretical FDR bound of $\alpha M/(M+J)$. The difference between the theoretical FDR bounds is due to the noise acf: a larger noise FWHM implies a smaller number of noise peaks M, and thus a smaller fraction $\alpha M/(M+J)$. Here $J = 5$ and M was calculated using (18). As the SNR increases (bottom left panel), the detection power

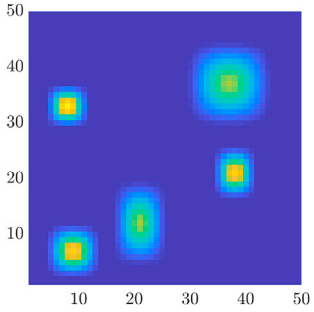

Fig. 9. Projection of the signal along the axis perpendicular to the page; in reality the five bumps belong to different slices.

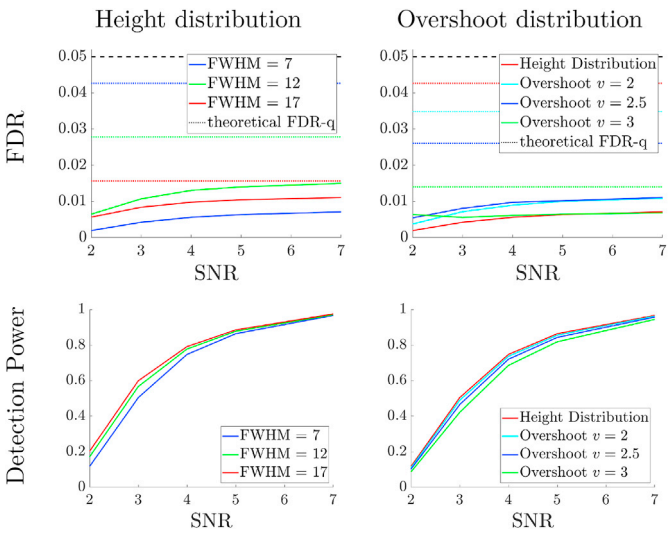

Fig. 10. Realized FDR and average detection power (10,000 simulations). Simulated fields are isotropic Gaussian with a Gaussian filter: *Left column:* no pre-threshold; *Right column:* FWHM = 7 voxels).

increases to 1. Because each FWHM is more or less matched to at least one signal peak, all values of FWHM here give about the same power. The power is slightly higher for larger FWHM because the peak distribution is more accurate.

Fig. 10 (right column) compares the realized FDR and detection power when using the exact height distribution versus the exact overshoot distribution in the calculation of peak p-values. Increasing the screening pre-threshold results in a smaller number of noise peaks and thus a smaller theoretical FDR bound, but also a smaller detection power. Using the exact height distribution with no pre-threshold results in the highest detection power while still keeping the FDR below the nominal level.

Fig. 11 compares the various approximations to the overshoot p-value distribution in terms of the realized FDR and detection power. As before, a higher pre-threshold yields a higher theoretical FDR bound. Because detection occurs at low p-values, where the agreement between the methods is high, all methods give similar realized FDR and detection power.

The second set of simulations considers *t*-fields. The only difference

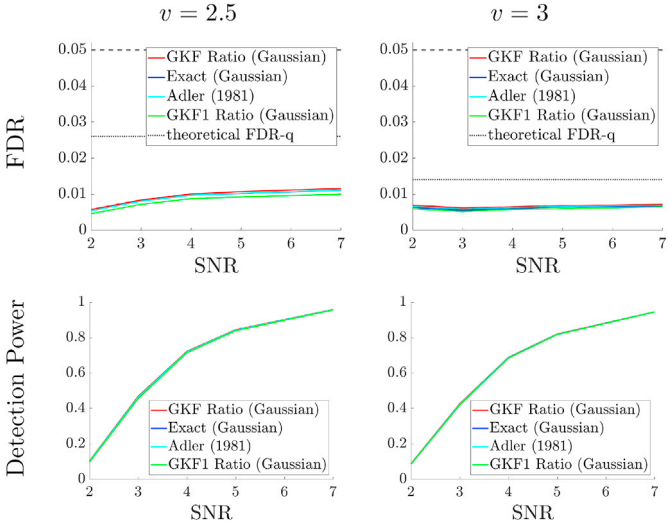

Fig. 11. Realized FDR and average detection power (10,000 simulations) for different overshoot p-value methods. Simulated fields are isotropic Gaussian (Gaussian filter with FWHM = 7 voxels).

with the Gaussian scenario above is that, for each Monte Carlo simulation, we generate $n+1$ independent Gaussian fields $Z_1(s), \ldots, Z_n(s)$ and $\varepsilon(s)$ as before, but try to detect peaks of the non-central *t*-fields with N d.f.

$$T(s) = \frac{A \cdot \mu(s) + \varepsilon(s)}{\sqrt{\sum_{i=1}^{N} Z_i^2 / N}}, \tag{19}$$

where again A is calibrated to achieve the correct SNR for the second highest bump. The construction (19) simulates a situation where the signal is estimated from an average of N independent field and detected using a *t* statistic.

Fig. 12 compares the realized FDR and detection power, as a function of the number of d.f. of the *t*-field, for various methods. For proper comparison, here we use the *t* versions of the GKF Ratio and GKF1 Ratio, while the exact overshoot distribution is used with marginally Gaussianized fields. Again, because detection occurs at low p-values, where the agreement between the methods is high, all methods give similar realized FDR and detection power.

4. Application to fMRI analysis

4.1. Voxelwise general linear model

Let

$$Y(s) = X\beta(s) + \varepsilon(s) \tag{20}$$

denote the general lineal model (GLM) at voxel s, where the entries of the noise vector $\varepsilon(s)$ have mean zero and variance $\sigma^2(s)$. In a single-subject analysis, this model may describe the fMRI signal as a function of the task stimulus, drift and other covariates encoded in the design matrix X. In a multiple-subject analysis on a common template, model (20) may describe a voxelwise effect from single-subject analyses as a function of subject-level covariates.

For a fixed vector c, the least-squares estimate of a contrast of interest $\eta(s) = c^{\mathrm{T}}\beta(s)$ is

$$\widehat{\eta}(s) = c^{\mathrm{T}}\widehat{\beta}(s) = c^{\mathrm{T}}(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}Y(s). \tag{21}$$

Substituting the model equation (20) and rewriting

$$\widehat{\eta}(s) = \eta(s) + c^{\mathrm{T}}(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}\varepsilon(s), \tag{22}$$

we see that this expression has the same form as the signal-plus-noise model (15), where $\eta(s)$ plays the role of the signal, the noise is the field $c^{\mathrm{T}}(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}\varepsilon(s)$, and the observation is the estimated field $\widehat{\eta}(s)$. Thus, we may use the methods described above to detect peaks in the parameter surface $\eta(s)$. To test the null hypothesis $H_0 : \eta(s) = 0$ at each location s, we may use the Wald statistic

$$\tilde{\eta}(s) = \frac{\widehat{\eta}(s)}{\widehat{\mathrm{se}}(\widehat{\eta}(s))} = \frac{c^{\mathrm{T}}\widehat{\beta}(s)}{\sqrt{\widehat{\sigma}^2(s)c^{\mathrm{T}}(X^{\mathrm{T}}X)^{-1}c}}, \tag{23}$$

where $\widehat{\sigma}^2(s) = y(s) - X\widehat{\beta}(s)^2/n'$ and n' is the number of degrees of freedom.

Following the prescribed recipe for peak detection, a test statistic map is first produced and the local maxima of this map are found. Each peak is then assigned a p-value and the BH procedure is applied to the list of p-values.

4.2. Data example

As a real data example, we use the fMRI data set from Moran et al. (2012), obtained from the public repository OpenfMRI (openfmri.org). Technical details on the full experiment and imaging protocol may be found in the original publication. Among the various experiments

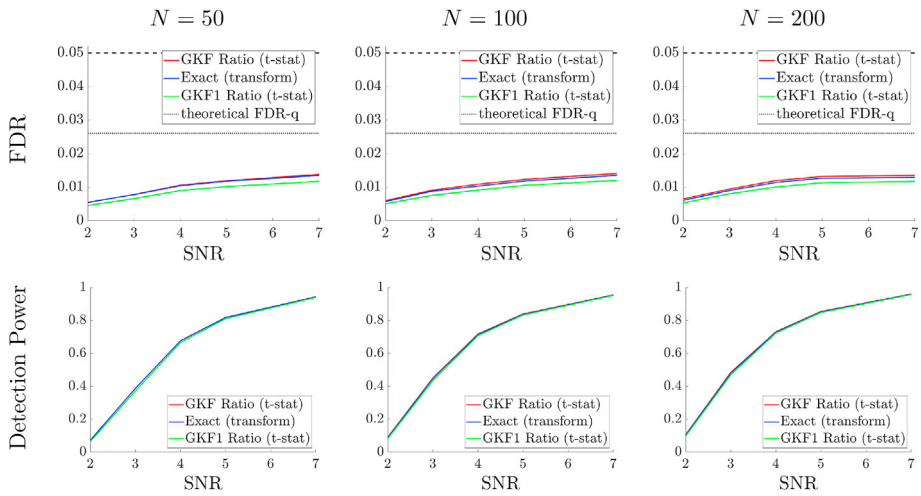

Fig. 12. Realized FDR and average detection power (10,000 simulations). Simulated fields are isotropic t with N d.f., generated from isotropic Gaussian fields (Gaussian filter with FWHM = 7 voxels). The pre-threshold is fixed at $\nu = 2.5$.

conducted in that study, we here focus on the "false-belief task", whose goal is to find brain regions that are active when processing other people's false beliefs about reality, in comparison with similar purely physical false realities. Briefly, subjects read short stories corresponding to either a person's false belief about reality or false realities not involving people. The effect sought after is the contrast between the neural activity in those two states.

As a single-subject analysis, we show here results for subject #49, the same subject analyzed in Cheng and Schwartzman (2017). After pre-processing, including motion correction and removal of missing data, the scan consisted of 179 time points with image size $71 \times 72 \times 36$ voxels, each voxel measuring 3 mm in all three directions. The data was convolved with a 3D isotropic Gaussian filter with standard deviation of 1.6 voxels, corresponding to a FWHM of 3.768 voxels or 11.3 mm. For computational purposes, the filter was truncated at 2.5 SDs from the mode, yielding a filter support of $8 \times 8 \times 8$ voxels. After convolution, only the valid portion of the image was retained, i.e. those voxels whose values were computed from neighborhood voxels strictly contained in the original image, yielding a valid image of size $64 \times 65 \times 29$ voxels. A linear model was then fitted against the stimulus design matrix and a t-statistic map of the form (23) was computed using the appropriate contrast corresponding to the task of interest mentioned above.

Fig. 13 shows the thresholded test statistic map using the exact peak height distribution after a t-to-Gaussian quantile transformation of the form (8). The significance threshold here is 2.99, equal to the height of the smallest significant peak after BH. Assuming isotropy, the estimate of κ for these data is 0.607, yielding the map on the left and containing 30 significant peaks. While the residual fields in these data give the impression of spatial isotropy, justifying the estimation of κ, the right panel shows the result when κ is fixed and equal to 1, corresponding to a Gaussian autocorrelation function and a noise field that is not necessarily isotropic or stationary. Not surprisingly, forgoing the isotropy assumption yields a more conservative result, with significance threshold 3.89 and 10 significant peaks.

To compare with other methods, Fig. 14 shows the activation maps obtained using the overshoot distribution with pre-threshold fixed at $\nu = 3$. The GKF Ration and GKF1 Ratio p-values were calculated most favorably using their t versions with no quantile transformation. In this particular analysis, the GKF Ratio yielded the same results as the left panel of Fig. 14 and the GKF1 Ratio yielded the same results as the right panel.

Table 2 summarizes the significance thresholds and number of significant peaks for the various methods. In this particular analysis, all the overshoot methods give similar results. However, the height distribution with no screening pre-threshold yields a much larger number of significant peaks.

5. Discussion

In this paper, we have provided exact formulas for the distribution of peak height and peak overshoot of smooth isotropic Gaussian fields, where the former does not need a screening pre-threshold. We have shown that the corresponding calculation of p-values is valid if the field is mildly non-isotropic or non-stationary, and conservative if the FWHM of the smoothing filter is not large enough. We have also provided a signal-plus-noise model composed of true unimodal peaks with finite support to be detected against a noisy background, and argued that it provides meaningful control of peak FDR.

5.1. Peak height vs. overshoot

The main reason for using the overshoot distribution instead of the peak height distribution is historical, as only approximations to the overshoot distribution were available. In this paper we have provided an exact formula for the height distribution of peaks of isotropic fields, not requiring a pre-threshold. This formula has been proven theoretically to be correct for isotropic fields and was shown here by simulations to also perform well when the field is anisotropic or mildly non-stationary. The exact height distribution for Gaussian fields is also a good approximation for t-fields with high number of d.f., if the t-field is transformed to Gaussian by a marginal quantile transformation.

When comparing the exact peak height distribution to the exact overshoot distribution, we have seen that the former results in a higher detection power. A pre-threshold reduces the number of features to be tested, ameliorating the multiple testing problem. However, screening removes potential true peaks to be found, and this effect appears to prevail. This phenomenon has been highlighted by Li et al. (2018).

Despite the advantages of the peak height distribution, the overshoot distribution may be preferable as a conservative approach if the stationarity of the noise field is highly in doubt. Cheng and Schwartzman (2015a) showed that Adler's approximation is asymptotically valid for high screening pre-thresholds in the case of nonstationary fields. Cheng and Schwartzman (2017) showed that when the approximate overshoot distribution is used, the pre-threshold can be optimized to balance accuracy against detection power. The GKF1 Ratio and Adler approximations may be advantageous too from a computational point of view, since the exact formulas require numerical integration.

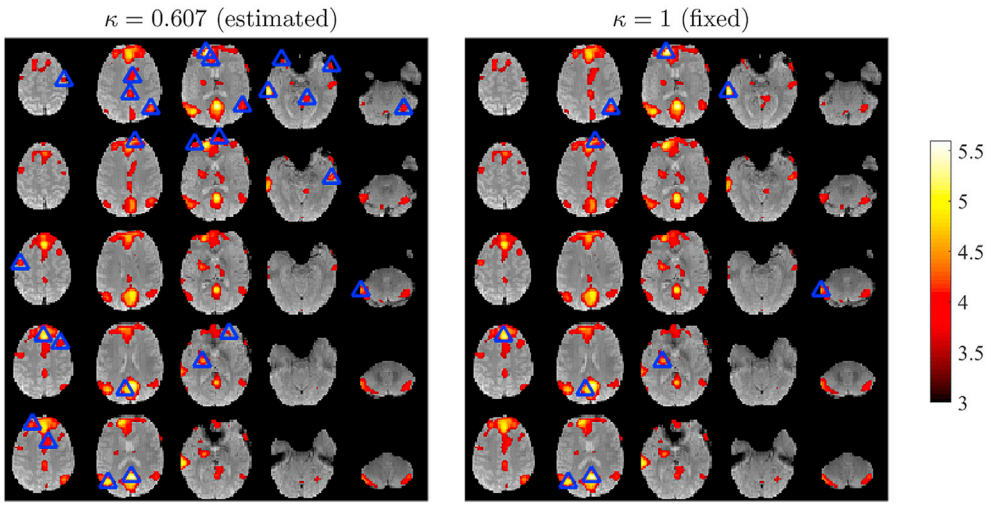

Figure 13. Analysis of the Moran et al. (2012) data: thresholded maps at FDR level 0.05 using the exact height distribution of local maxima, after t-to-Gaussian quantile transformation, for isotropic noise with estimated $\kappa = 0.605$ (left panel) and fixed $\kappa = 1$ (right panel). In each panel, montage shows the brain volume as transverse slices from the top of the brain (top left) to the bottom of the brain (bottom right). Significant local maxima are marked by blue triangles. Colored regions indicate the smoothed Wald statistic field above $v = 3$. Results are superimposed on an anatomical brain image (gray) for reference.

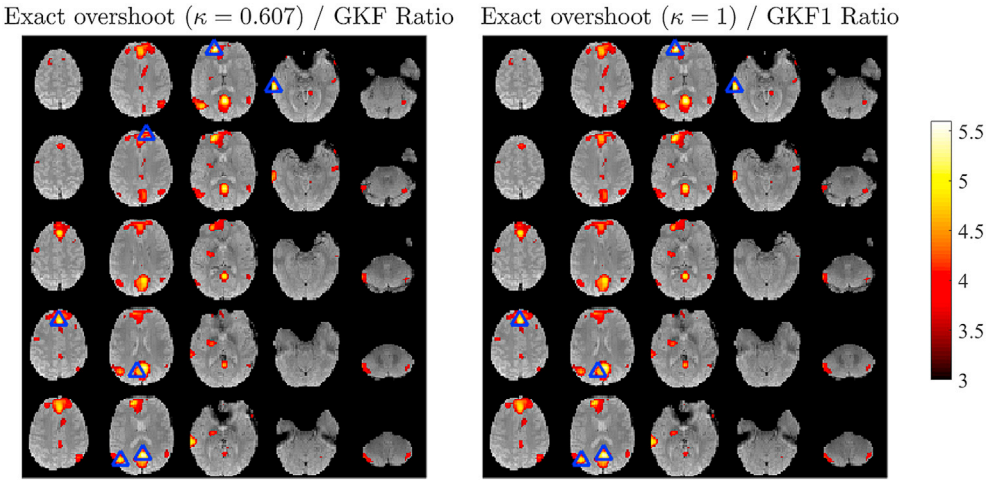

Fig. 14. Analysis of the Moran et al. (2012) data: thresholded maps at FDR level 0.05 using the exact overshoot distribution of local maxima, after t-to-Gaussian quantile transformation, for isotropic noise with estimated $\kappa = 0.607$ (left panel)and fixed $\kappa = 1$ (right panel). The screening pre-threshold is $v = 3$. The approximate overshoot distribution for t-fields gives the same result as the left panel for GKF Ratio, and the same result as the right panel for GKF1 Ratio. Montages follow the same format as in Fig. 13.

Table 2
Analysis of the Moran et al. (2012) data: height significance thresholds and number of significant peaks (in parentheses) for five different methods.

Distribution	Height	Overshoot	
		$v = 2.5$	$v = 3$
Exact, $\kappa = 0.607$ (estimated)	2.99 (30)	4.56 (6)	4.14 (7)
Exact, $\kappa = 1$ (fixed)	3.89 (10)	4.56 (6)	4.56 (6)
Adler (1981)	–	4.14 (7)	4.04 (8)
GKF1 Ratio	–	4.56 (6)	4.56 (6)
GKF Ratio	–	4.56 (6)	4.14 (7)

5.2. Estimation of parameters

Both the exact peak height and overshoot distribution formulas need an estimate of the shape parameter κ. We have seen in the simulations and data analysis that this is not a problem if one may take advantage of the isotropy of the field to obtain a global estimate. One may also simply take $\kappa = 1$, making the inference conservative. On the other hand, the GKF Ratio approximation, currently implemented in SPM12, depends on estimation of the LKCs or resel counts. Estimation of these parameters introduces additional variance and computation time, which we have not considered in this paper.

5.3. The role of smoothness

All the p-value calculation methods considered in this article are based on smooth field approximations and therefore require certain amount of filtering to be precise on discrete grids. In our simulations, a FWHM of about 7 voxels gives fairly accurate p-values. A smaller FWHM makes the inference conservative, but still valid. In our data analysis, we used a filter with FWHM of 3.768 voxels, but the data already had some autocorrelation in it. Using Table 2 in Worsley et al. (2004), we computed the equivalent FWHM to obtain the same autocorrelation as the filtered data from white noise. The obtained FWHM value of 5.608 voxels is smaller than 7, indicating that our data analysis may be more conservative than our simulation results.

We saw in the simulations that increasing the FWHM makes the GKF Ratio approximation liberal, producing invalid p-values. To understand this, note that the LKCs or resel counts of order greater than zero are inversely proportional to the FWHM (e.g. see Table 2 in Worsley et al. (2004)). As the FWHM increases, the resel counts gets smaller and the dominant term in the GKF becomes the zero-order term. In other words, the GKF Ratio converges to the marginal distribution $[1 - \Phi(u)]/[1 - \Phi(v)]$. As mentioned in Section 2.1, this distribution is stochastically smaller than the true distribution and gives liberally invalid p-values. In contrast, the GKF1 Ratio approximates the overshoot distribution by the highest order term in the GKF instead. Thus it does not depend on the FWHM except for the discrete grid issue mentioned above.

The filter FWHM also plays a role in detection power, defining a trade-off between noise variance and signal strength. As the filter size increases, the noise variance decreases. But if the filter is too large, it smears out the peaks and decreases signal strength. By the matched filter theorem, and also shown in Cheng and Schwartzman (2017), the detection power is largest when the size of the filter matches the size of the signal bumps to be detected.

5.4. Extensions

While we have presented an example of a single-subject analysis, we emphasize that the peak inference method presented in this paper can also be applied to multiple-subject analyses. This requires spatial registration of the subjects to a common template and computation of a voxelwise statistic that captures the effect of interest. If the residual fields from the voxlewise GLMs have the proper autocorrelation structure, then they can be used to estimate the parameter κ. Otherwise, a fixed value of $\kappa = 1$ will give valid but conservative results.

Finally, while we have described a way of providing proper inference for peaks, peak inference may be unsatisfying in the sense that it collapses all the activity into single points. Peaks serve as topological representatives of the activation regions surrounding them, but do not give the spatial extent of that activation. In future work, we plan to investigate this issue by describing the spatial uncertainty associated with activation peaks.

Acknowledgments

The authors wish to thank Thomas Nichols from the University of Oxford for his valuable advice and discussions, as well as the referees for their useful suggestions. This work was partially supported by NIH grants R01 CA157528 and R01 EB026859.

A Appendix.

A.1 Explicit peak height distributions for isotropic fields

When the field $z(s)$ is isotropic, the tail probability (2) can be obtained as (Cheng and Schwartzman, 2018, 2015b)

$$F(u) = F(u;\kappa) = \int_u^\infty h(x;\kappa)\, dx, \tag{24}$$

where $h(x;\kappa)$ is the probability density function (pdf) of the height of local maxima, depending on a single parameter κ. When $D = 2$, the pdf is given by

$$\begin{aligned} h(x;\kappa) &= \sqrt{3}\kappa^2\left(x^2-1\right)\phi(x)\Phi\left(\frac{\kappa x}{\sqrt{2-\kappa^2}}\right) + \frac{\kappa x\sqrt{3(2-\kappa^2)}}{2\pi}e^{-\frac{x^2}{2-\kappa^2}} \\ &+ \frac{\sqrt{6}}{\sqrt{\pi(3-\kappa^2)}}e^{-\frac{3x^2}{2(3-\kappa^2)}}\Phi\left(\frac{\kappa x}{\sqrt{(3-\kappa^2)(2-\kappa^2)}}\right). \end{aligned} \tag{25}$$

When $D = 3$, the pdf is given by

$$\begin{aligned} h(x;\kappa) = \frac{144\phi(x)}{29\sqrt{6}-36}\Bigg\{ &\left[\frac{\kappa^2\left[\left(1-\kappa^2\right)^3+6\left(1-\kappa^2\right)^2+12\left(1-\kappa^2\right)+24\right]}{4\left(3-\kappa^2\right)^2}x^2\right. \\ &\left.+\frac{2\left(1-\kappa^2\right)^3+3\left(1-\kappa^2\right)^2+6\left(1-\kappa^2\right)}{4(3-\kappa^2)}+\frac{3}{2}\right]\frac{e^{-\frac{\kappa^2x^2}{2(3-\kappa^2)}}}{\sqrt{2(3-\kappa^2)}}\Phi\left(\frac{2\kappa x}{\sqrt{(3-\kappa^2)(5-3\kappa^2)}}\right) \\ &+\left[\frac{\kappa^2\left(2-\kappa^2\right)}{4}x^2-\frac{\kappa^2\left(1-\kappa^2\right)}{2}-1\right]\frac{e^{-\frac{\kappa^2x^2}{2(2-\kappa^2)}}}{\sqrt{2(2-\kappa^2)}}\Phi\left(\frac{\kappa x}{\sqrt{(2-\kappa^2)(5-3\kappa^2)}}\right) \\ &+\left[7-\kappa^2+\frac{(1-\kappa^2)\left[3\left(1-\kappa^2\right)^2+12\left(1-\kappa^2\right)+28\right]}{2(3-\kappa^2)}\right]\frac{\kappa x e^{-\frac{3\kappa^2x^2}{2(5-3\kappa^2)}}}{4\sqrt{\pi}(3-\kappa^2)\sqrt{5-3\kappa^2}} \\ &+\frac{\sqrt{\pi}\kappa^3}{4}x\left(x^2-3\right)\left[\Phi_{\Sigma_1}\left(0,\kappa x/\sqrt{2}\right)+\Phi_{\Sigma_2}\left(0,\kappa x/\sqrt{2}\right)\right]\Bigg\} \end{aligned} \tag{26}$$

where Φ_{Σ_1} and Φ_{Σ_2} denote bivariate normal cdfs corresponding to the covariance matrices

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

respectively.

A.2 Numerical calculation of the parameter κ for a generic filter

This section describes how to numerically calculate the parameter κ in the case of white noise convolved with an isotropic filter $w(s)$.

As stated in definition (4), calculation of the parameter κ requires the derivatives of the function $\rho(r)$, which is an expression of the acf $R(s)$ as a

function of $r = ||s||^2$. For a field generated as white noise convolved with an isotropic filter $w(s)$, the acf $R(s)$ is the convolution of the filter with itself. However, computing this convolution analytically and then writing the result as a function of $||s||^2$ may be difficult. E.g., in the case of the quartic filter, the convolution is an 8th degree polynomial.

Instead, the following method may be used to calculate the parameter κ numerically. First, compute the 3D convolution numerically over a high-resolution grid and extract the values from the center radially in any direction, say, the x axis. Matching this vector to the vector $r = x^2$ gives a discrete version of the function $\rho(r)$. The derivatives ρ' and ρ'' are then estimated numerically by differentiation with respect to the vector r.

Applying this method to the quartic filter (5) with $a = 40$ yields the value $\kappa \approx 0.92$. Recall that κ does not depend on scaling (Section 2.1). Choosing a high value of a here helps increase smoothness and thus accuracy in the numerical calculations.

References

Adler, R.J., 1981. The Geometry of Random Fields, vol.62. SIAM.

Adler, R.J., Taylor, J.E., Worsley, K.J., 2010. Applications of random fields and geometry: foundations and case studies. http://webee.technion.ac.il/Sites/People/adler/hrf.pdf.

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

Candes, E., Tao, T., 2007. The Dantzig selector: statistical estimation when p is much larger than n. Ann. Stat. 35 (6), 2313–2351.

Cheng, D., Schwartzman, A., 2015a. Distribution of the height of local maxima of Gaussian random fields. Extremes 18 (2), 213–240.

Cheng, D., Schwartzman, A., 2015b. On the explicit height distribution and expected number of local maxima of isotropic Gaussian random fields. ArXiv:1503.01328.

Cheng, D., Schwartzman, A., 2017. Multiple testing of local maxima for detection of peaks in random fields. Ann. Stat. 45 (2), 529–556.

Cheng, D., Schwartzman, A., 2018. Expected number and height distribution of critical points of smooth isotropic Gaussian random fields. Bernoulli 24 (4B), 3422–3446.

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

Chumbley, J.R., Worsley, K., Flandin, G., Friston, K.J., 2010. Topological FDR for neuroimaging. Neuroimage 49, 3057–3064.

Durnez, J., Degryse, J., Moerkerke, B., Seurinck, R., Sochat, V., Poldrack, R., Nichols, T., 2016. Power and Sample Size Calculations for fMRI Studies Based on the Prevalence of Active Peaks. bioRxiv, p. 049429.

Durnez, J., Moerkerke, B., Nichols, T.E., 2014. Post-hoc power estimation for topological inference in fMRI. Neuroimage 49, 45–64.

Efron, B., 2004. Large-scale simultaneous hypothesis testing: the choice of a null hypothesis. J. Am. Stat. Assoc. 99 (465), 96–104.

Eklund, A., Nichols, T.E., Knutsson, H., 2016. Cluster failure: why fMRI inferences for spatial extent have inflated false-positive rates. Proc. Natl. Acad. Sci. Unit. States Am. 113 (28), 7900–7905.

Li, J., Gahm, J.K., Shi, Y., Toga, A.W., 2018. Topological false discovery rates for brain mapping based on signal height. Neuroimage 167, 478–487.

Moran, J.M., Jolly, E., Mitchell, J.P., 2012. Social-cognitive deficits in normal aging. J. Neurosci. 3 (16), 5553–5561.

Pacifico, M.P., Genovese, C., Verdinelli, I., Wasserman, L., 2007. Scan clustering: a false discovery approach. J. Multivar. Anal. 98 (7), 1441–1469.

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

Schwartzman, A., Dougherty, R.F., Lee, J., Ghahremani, D., Taylor, J.E., 2009. Empirical null and false discovery rate analysis in neuroimaging. Neuroimage 44 (1), 71–82.

Schwartzman, A., Gavrilov, Y., Adler, R.J., 2011. Multiple testing of local maxima for detection of peaks in 1d. Ann. Stat. 39 (6), 3290–3319.

Worsley, K.J., Marrett, S., Neelin, P., Evans, A.C., 1996. Searching scale space for activation in PET images. Hum. Brain Mapp. 4, 74–90.

Worsley, K.J., Taylor, J.E., Tomaiuolo, F., Lerch, J., 2004. Unified univariate and multivariate random field theory. Neuroimage 23, S189–S195.

Zhang, C.-H., 2010. Nearly unbiased variable selection under minimax concave penalty. Ann. Stat. 38 (2), 894–942.