The expected behaviour of random fields in high dimensions: contradictions in the results of Bansal and Peterson (2018)

Samuel Davenport*and Thomas E. Nichols†‡§

January 21, 2021

Abstract

Bansal and Peterson (2018) found that in simple stationary Gaussian simulations Random Field Theory incorrectly estimates the number of clusters of a Gaussian field that lie above a threshold. Their results contradict the existing literature and appear to have arisen due to errors in their code. Using reproducible code we demonstrate that in their simulations Random Field Theory correctly predicts the expected number of clusters and therefore that many of their results are invalid.

The use of Random Field Theory (RFT) to control the false positive rate in neuroimaging is a well established testing framework. RFT consists of a set of theoretical results (originally due to Adler (1981)) that have been translated for use in neuroimaging (Worsley et al. (1992),Friston et al. (1994), Worsley et al. (1996)) and have been rigorously tested in the context of simulations (Hayasaka and Nichols (2003), Nichols and Hayasaka (2003). In a recent manuscript, Bansal and Peterson (2018) (henceforth BP) found that in simple stationary Gaussian simulations RFT incorrectly estimates the number of clusters of a Gaussian field that lie above a threshold. However, these results contradict the existing literature and our own previous findings, and exhibited highly counter-intuitive features. In particular, BP's results showed the expected number of clusters falling then growing as a function of smoothness. Instead, as smoothness increases the expected cluster count should simply fall. BP's authors helpfully provided some of their code and we identified some errors due to boundary wrap-around; the authors recomputed their results which then changed substantially but still did not align with previous results (i.e. there was still a great mismatch between theory and simulation). Unfortunately this implies that their conclusions regarding the performance of RFT are invalid.

Cluster size inference using Random Field Theory is based on the properties of null mean-zero homogeneous random fields thresholded at a given cluster defining threshold (CDT) to produce an excursion set (i.e. a collection of suprathreshold clusters). In this setting the expected Euler characteristic provides an approximation to the expected number of clusters. Then, under the assumption that clusters are elliptic paraboloids (true at high CDTs), results originally due to Nosko (1969), Wilson and Adler (1982) and Wilson (1988) provide a distribution for the extent of each cluster (see also Adler (1981) p158, Adler et al. (2010) chapter 6 and Cao (1999)). This was used to derive a form for

*Department of Statistics, University of Oxford, Oxford, OX1 3LB, UK

†Oxford Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, Nuffield Department of Population Health, University of Oxford, Oxford, OX3 7LF, UK

‡Wellcome Centre for Integrative Neuroimaging, FMRIB, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, OX3 9DU, UK

§Department of Statistics, University of Warwick, Coventry, CV4 7AL, UK

the maximum cluster size above the CDT allowing control of the familywise error rate over clusters (Friston et al. (1994), Friston et al. (1996)). At low CDTs resting state data validation has shown that that false positives are not correctly controlled using this approach (Eklund et al., 2016). They showed that this improves as the CDT is increased, in line with theory, and using high CDTs ensures good false positive rate control (Gong et al., 2018). However, RFT has long been known to perform well in simulations where the assumptions are satisfied.

The purpose of this note is to demonstrate with freely available and reproducible code the evaluations considered by BP; showing simulation and theory align. We find that RFT correctly estimates the expected number of clusters over a range of parameter settings. Analysis (random field simulation and thresholding) was performed using the RFTtoolbox, available at `https://github.com/sjdavenport/RFTtoolbox`, and SPM. Scripts to run this analysis are available at `https://github.com/sjdavenport/BansalResponse`. All code was run in MATLAB R2019b.

1 Methods

1.1 Random Field Generation

We generate Gaussian random fields (GRFs) using the function noisegen.m from the RFTtoolbox. Given a number of dimensions D and a grid of size $c_1 \times \cdots \times c_D \subset \mathbb{Z}^D$ and a given FWHM f, this generates i.i.d standard Gaussian random variables on a grid of size $(c_1+2C)\times\cdots\times(c_N+2C)$ where $C = \lceil 1.7f \rceil$ is a buffer to avoid boundary effects[1]. It then smooths it with a Gaussian kernel with FWHM f and takes the central $c_1 \times \cdots \times c_D$ subset as the output random field. Finally, the image is divided by the square root of the sum-of-squares of the kernel to ensure that the resulting field is unit-variance.

For our simulations we use the same settings as in BP. In particular for $D = 1$ we generate 1D GRFs consisting of 10000 voxels. For $D = 2$, we generate GRFs on a lattice of size 250×250 and for $D = 3$ we generate GRFs on a lattice of size $90 \times 90 \times 90$. For each dimension, each CDT $= 1.5, 2, 2.5, 3$ and each FWHM $= 5, 10, 15, 20, 25$ we generate 100000 GRFs (an increase on the 50 used in each setting in BP). This allows us to compute the average (Monte Carlo) number of clusters above the threshold for each setting in order to compare to theoretical predictions.

1.2 Expected Number of Clusters

In the RFT framework, for large enough thresholds, the Euler characteristic of the excursion set is equal to the number of its connected components (Worsley et al., 1992). As such the expected Euler characteristic can be used to estimate the expected number of clusters above the threshold.

BP use the original formula (due to Adler (1981)) for the expected Euler characteristic. For a D-dimensional stationary random field X on a set $S \subset \mathbb{R}^D$ with standard deviation σ, a cluster defining threshold u and an excursion set A_u above the threshold

[1] For $x \in \mathbb{R}$, $\lceil x \rceil$ denotes the smallest integer than is greater than or equal to x.

this gives

$$\mathbb{E}[\chi(A_u)] = \mu(S)(2\pi)^{(D+1)/2}|\Lambda|^{1/2}\sigma^{-(2D-1)}e^{-u^2/2\sigma^2}\sum_{j=0}^{(D-1)/2}(-1)^j a_j \sigma^{2j}\binom{D-1}{2j}u^{D-1-2j} \tag{1}$$

where Λ is the covariance matrix of the partial derivatives, $\mu(S)$ is the volume of S and $a_j = (2j)!/j!2^j$. Unfortunately BP appears to have miscalculated this (see Section 3). However this version is also out of date as it fails to account for local maxima on the boundary. The updated formula due to (Taylor (2006)) is

$$\mathbb{E}[\chi(A_u)] = \sum_{d=0}^{D}\mathcal{L}_d\rho_d(u) \tag{2}$$

where for $d = 1, \ldots, D$, ρ_d are known functions (see Worsley et al. (1996) for their explicit forms) and $\mathcal{L}_d$ are known as the Lipshitz Killing curvatures (LKCs) which are determined by the covariance function of the random fields and the shape of the parameter set on which the fields are defined. When the random fields are stationary these can be calculated (Worsley et al., 1996) and this is the version of RFT that has been implemented in current software packages (such as SPM and FSL).

For stationary fields the LKCs are determined by Λ and S. In particular, for a D-dimensional field consisting of i.i.d Gaussian noise and smoothed with an isotropic Gaussian kernel with FWHM s (as in our and BP's simulations), Λ is a diagonal matrix with diagonal entries $\Lambda_{ii} = 4\log(2)/s^2$ for $i = 1, \ldots, D$ (see Holmes (1994), Worsley et al. (1992)) and so the LKCs are known. Expansion (2) is in fact also valid under non-stationarity Taylor (2006). However in 3D the LKCs are rather difficult to compute and, while there has been promising recent work on this (Adler et al. (2017), Telschow et al. (2019)), it has not yet implemented in neuroimaging software packages.

2 Results

Rerunning the simulations and theory correctly allows us to reproduce BP's Table 1 in tables listed in Figure 4. We present these results for 2D and 3D in graphical form in Figures 1 and 2 where we plot the theoretical and average number of clusters. We show that there is large discrepancy between our results (which match previous work) and the results found by BP. The figures demonstrate that in these simulation settings RFT provides a close estimate of the expected Euler characteristic (at all applied smoothness levels) contradicting their findings.

These figures additionally show that, as the smoothness increases, the number of clusters above the threshold decreases. The BP results instead show the number of clusters first decreasing and then increasing as the smoothness increases (this issue appears to affect both their calculations of the expected cluster size and of the average number of clusters above the threshold). For further details, and a comparison of the results, see the discussion.

The corresponding 1D results are shown graphically in Figure 3. Unlike the 2D and 3D results, in this setting, at low CDTs BP's results closely match ours. However, as the CDT is increased, their calculation of the average number of clusters no longer matches. This could partially be due to noise as they only used 50 simulations in each setting.

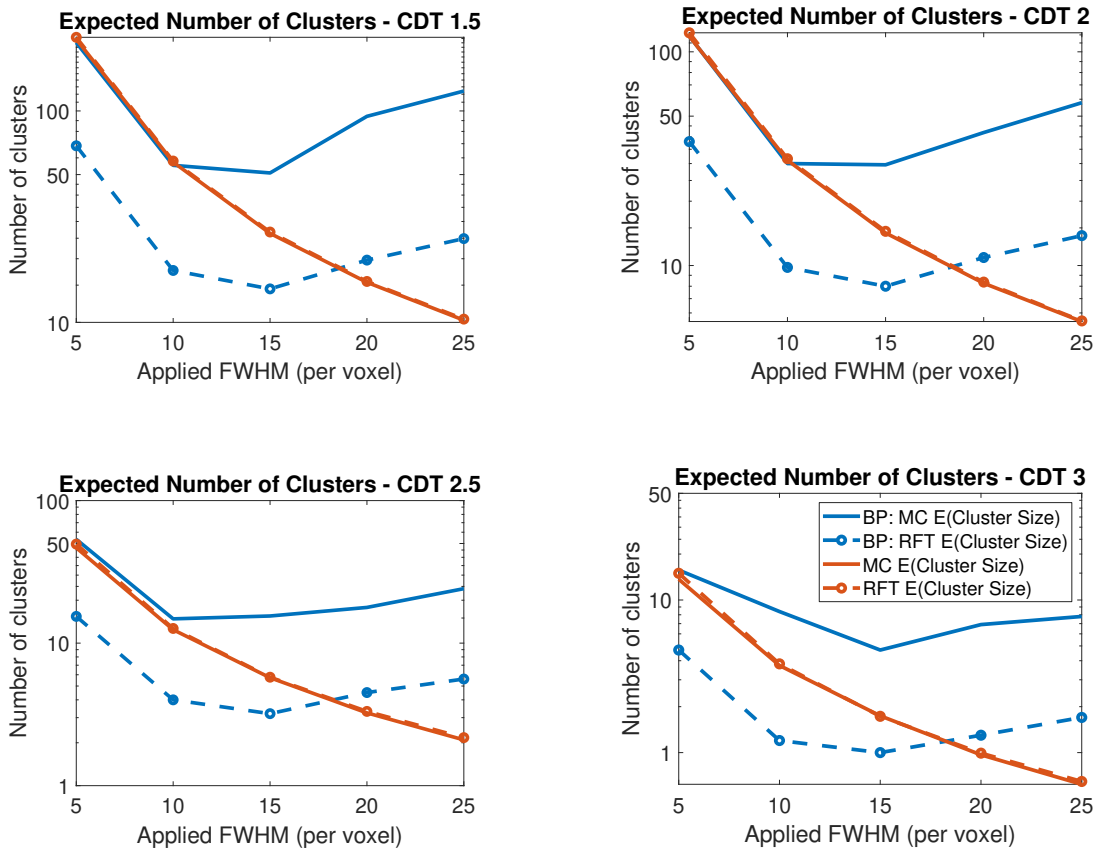

Figure 1: Comparison of Monte Carlo and theoretical expected number of clusters in 2D. In this setting the expected and average cluster size (shown in red) closely match and are very different from those found in BP.

3 Discussion

Our results show that in these simulations the expected Euler characteristic closely matches the average number of clusters as predicted by theory. Unfortunately it appears that BP have miscalculated the Euler characteristics and have simulated their random fields incorrectly. Their work predicts large differences between theory and simulations but this is unfounded. We have regenerated corrected versions of their Table (Figure 4) and have presented these graphically. Our results are very different from those of BP and show that the expected Euler characteristic clearly matches the average number of clusters, providing a validation of the theory in line with previous work (e.g. Hayasaka and Nichols (2003), Worsley et al. (1996)).

To immediately see that their results are incorrect, note the non-monotonicity of the blue curves shown in Figures 1 and 2. This non-monotonicity is present in both their theoretical and their Monte-Carlo calculations. However this indicates a significant error in their calculations because as the applied smoothing increases the number of clusters above the threshold should decrease. Indeed one only needs to study Equation (1), to see that as the FWHM increases the expected Euler characteristic should decrease.

It is difficult to identify the precise source of the error in the results BP. One issue is that they use Equation 1 (which does not account for boundary effects) to calculate the expected Euler characteristic instead of the updated formula (2) which is implemented (under stationarity) in neuroimaging software. Another source of error can be seen by

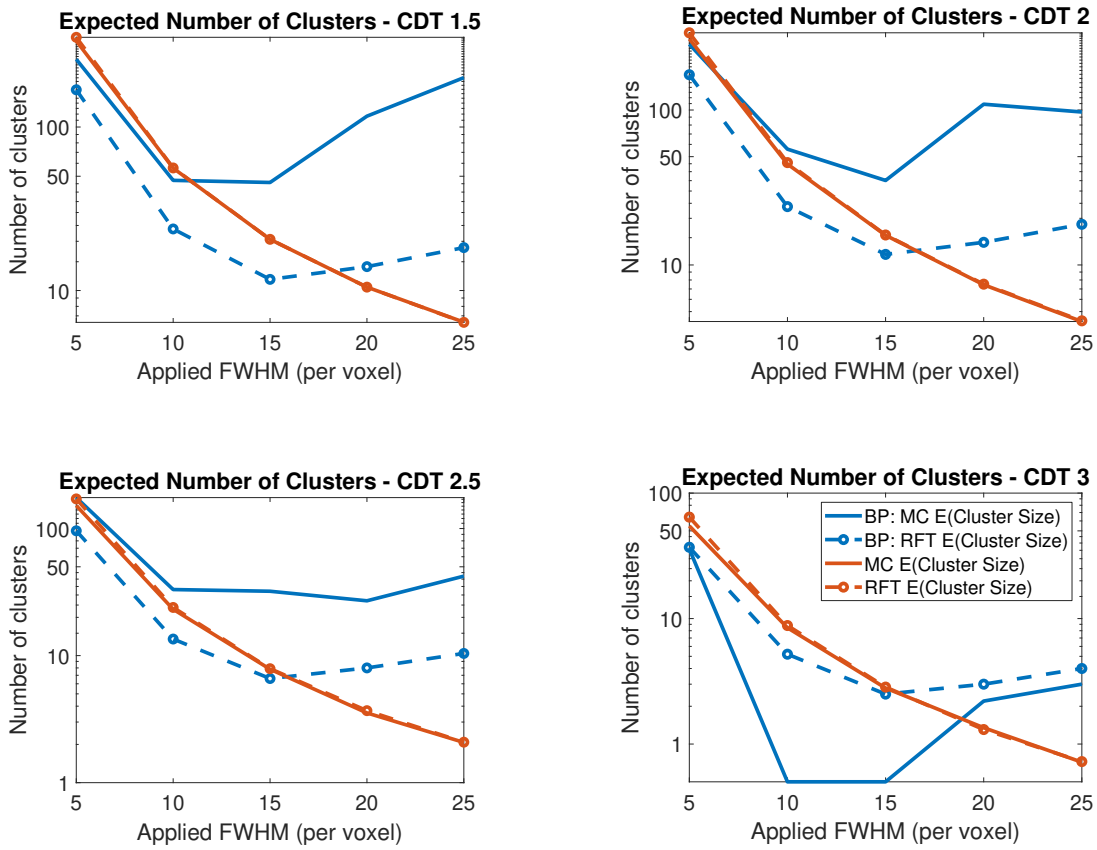

Figure 2: Comparison of Monte Carlo and theoretical expected number of clusters in 3D. In this setting the expected and average cluster size (shown in red) closely match and are very different from those found in BP.

examining Figure 1 from their paper. Close study of this Figure suggests that the data has been simulated on a 2D torus rather than on a grid as the clusters on the edge of the image can be matched (in correspondence with the Authors we confirmed that this is how they simulated their random fields). This is a likely cause of the non-monotonicity because clusters at the boundary would be counted twice and boundary clusters are more likely to occur at higher smoothness levels. However, this setting is not representative of the datasets used in fMRI and in particular equations (1) and (2) for the expected Euler characteristic of the excursion set are only valid for a closed and bounded subset of $\mathbb{R}^D$ and not for the torus. The theory for a torus would be completely different and so these formulae cannot be applied. This is thus a reasonable explanation for some of the disagreement, unfortunately it means that the results and many of the conclusions of their paper are invalid. We suspect that there are a number of further underlying issues with their code since, for instance, they also appear to have miscalculated the expected Euler characteristics.

BP make a number of claims which contradict the rest of the literature. They appear to have incorrectly generated their random fields and to have miscalculated the expected Euler characteristic, meaning that their conclusions should not be relied upon. A quick look at their Table 1 shows that they predict huge discrepancies between their simulations and well-established and validated theory. As such we find it difficult to understand how this paper got past the reviewing process.

4 Appendix

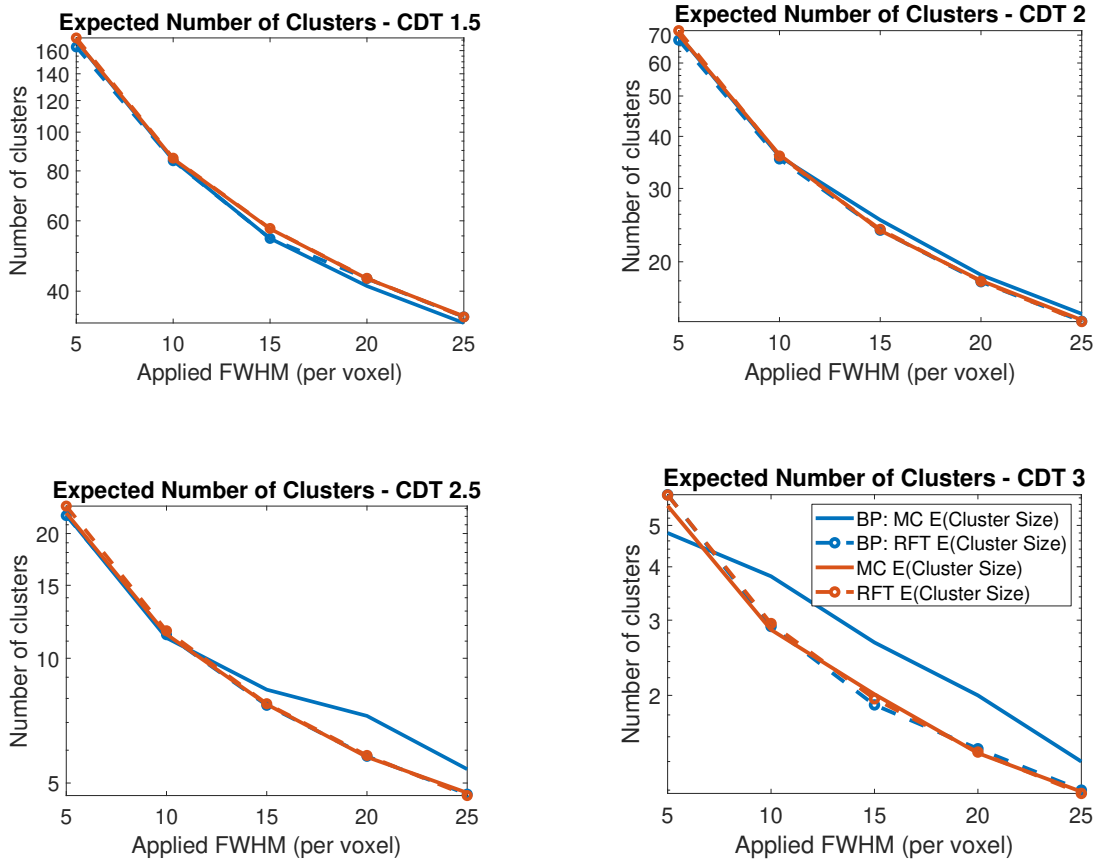

Figure 3: Comparison of Monte Carlo and theoretical expected number of clusters in 1D. At low CDTs our results match those of BP, however at high CDTs the average number of clusters that they find is slightly inflated but the theoretical number of clusters they find matches our calculations.

FWHM	Cluster Defining Threshold			
	1.5	2.0	2.5	3.0
5	$168.75 \pm 0.07(172.14)$	$69.79 \pm 0.05(71.75)$	$22.42 \pm 0.03(23.29)$	$5.59 \pm 0.01(5.89)$
10	$85.66 \pm 0.05(86.10)$	$35.62 \pm 0.04(35.89)$	$11.52 \pm 0.02(11.65)$	$2.92 \pm 0.01(2.95)$
15	$57.30 \pm 0.04(57.42)$	$23.84 \pm 0.03(23.93)$	$7.73 \pm 0.02(7.77)$	$1.95 \pm 0.01(1.96)$
20	$43.02 \pm 0.04(43.08)$	$17.92 \pm 0.03(17.96)$	$5.81 \pm 0.01(5.83)$	$1.47 \pm 0.01(1.47)$
25	$34.42 \pm 0.03(34.48)$	$14.36 \pm 0.02(14.37)$	$4.65 \pm 0.01(4.66)$	$1.17 \pm 0.01(1.18)$

(a) 1D simulations versus expected Euler characteristic.

FWHM	Cluster Defining Threshold			
	1.5	2.0	2.5	3.0
5	$218.59 \pm 0.07(222.99)$	$118.05 \pm 0.06(122.73)$	$46.59 \pm 0.04(49.51)$	$13.72 \pm 0.02(14.96)$
10	$57.34 \pm 0.03(57.95)$	$31.11 \pm 0.03(31.60)$	$12.42 \pm 0.02(12.67)$	$3.70 \pm 0.01(3.82)$
15	$26.54 \pm 0.02(26.75)$	$14.29 \pm 0.02(14.45)$	$5.69 \pm 0.01(5.77)$	$1.71 \pm 0.01(1.73)$
20	$15.48 \pm 0.02(15.61)$	$8.28 \pm 0.02(8.36)$	$3.29 \pm 0.01(3.32)$	$0.98 \pm 0.01(0.99)$
25	$10.29 \pm 0.01(10.36)$	$5.46 \pm 0.01(5.50)$	$2.15 \pm 0.01(2.17)$	$0.64 \pm 0.00(0.65)$

(b) 2D simulations versus expected Euler characteristic.

FWHM	Cluster Defining Threshold			
	1.5	2.0	2.5	3.0
5	$336.36 \pm 0.10(353.79)$	$288.06 \pm 0.08(315.00)$	$151.90 \pm 0.07(171.14)$	$55.17 \pm 0.05(64.35)$
10	$55.97 \pm 0.04(56.19)$	$44.81 \pm 0.03(45.77)$	$23.25 \pm 0.03(23.93)$	$8.46 \pm 0.02(8.80)$
15	$20.59 \pm 0.02(20.57)$	$15.47 \pm 0.02(15.61)$	$7.77 \pm 0.02(7.89)$	$2.79 \pm 0.01(2.84)$
20	$10.52 \pm 0.02(10.49)$	$7.47 \pm 0.01(7.52)$	$3.66 \pm 0.01(3.69)$	$1.29 \pm 0.01(1.31)$
25	$6.41 \pm 0.01(6.39)$	$4.35 \pm 0.01(4.36)$	$2.08 \pm 0.01(2.08)$	$0.73 \pm 0.01(0.72)$

(c) 3D simulations versus expected Euler characteristic.

Figure 4: Comparing the average number of clusters above a threshold with the expected Euler characteristic. For each table each entry is of the form $A \pm s(\text{EEC})$ where A is the average number of clusters above the threshold, $\pm s$ provides a 95% confidence interval for this average and EEC is the expected Euler characteristic provided by theory. This shows that the EEC provides a good approximation of the expected number of clusters above the threshold.

5 Acknowledgements

TEN is supported by the Wellcome Trust, 100309/Z/12/Z and SD is funded by the EPSRC.

References

Robert J. Adler. *The Geometry of Random Fields.* 1981.

Robert J Adler, Jonathan E Taylor, and Keith J. Worsley. *Applications of random fields and geometry: Foundations and case studies.* 2010.

Robert J. Adler, Kevin Bartz, Sam C. Kou, and Anthea Monod. Estimating thresholding levels for random fields via Euler characteristics. 2017.

Ravi Bansal and Bradley S. Peterson. Cluster-level statistical inference in fMRI datasets: The unexpected behavior of random fields in high dimensions. *Magnetic Resonance Imaging*, (2017), 2018. ISSN 0730725X. doi: 10.1016/j.mri.2018.01.004. URL `http://linkinghub.elsevier.com/retrieve/pii/S0730725X18300109`.

J. Cao. The size of the connected components of excursion sets of $\chi 2$, t and F fields. *Advances in Applied Probability*, 31(3):579–595, 1999. ISSN 00018678. doi: 10.1239/aap/1029955192.

Anders Eklund, Thomas E Nichols, and Hans Knutsson. Cluster failure: Why fmri inferences for spatial extent have inflated false-positive rates. *Proceedings of the national academy of sciences*, 113(28):7900–7905, 2016.

K J Friston, K. J. Worsley, R S J Frackowiak, J C Mazziotta, and A C Evans. Assessing the significance of focal activations using their spatial extent. *Human Brain Mapping*, 1:214–220, 1994.

K.J. Friston, A. Holmes, J-B. Poline, C.J. Price, and C.D. Frith. Detecting Activations in PET and fMRI: Levels of Inference and Power. *NeuroImage*, 4(3):223–235, 1996. ISSN 10538119. doi: 10.1006/nimg.1996.0074. URL `http://linkinghub.elsevier.com/retrieve/pii/S1053811996900749`.

Weikang Gong, Lin Wan, Wenlian Lu, Liang Ma, Fan Cheng, Wei Cheng, Stefan Grünewald, and Jianfeng Feng. Statistical testing and power analysis for brain-wide association study. *Medical Image Analysis*, 47:15–30, 2018. ISSN 13618423. doi: 10.1016/j.media.2018.03.014.

Satoru Hayasaka and Thomas E. Nichols. Validating cluster size inference: Random field and permutation methods. *NeuroImage*, 20(4):2343–2356, 2003. ISSN 10538119. doi: 10.1016/j.neuroimage.2003.08.003.

Andrew P Holmes. Statistical Issues in Function Brain Imaging. 1994.

Thomas E. Nichols and Satoru Hayasaka. Controlling the familywise error rate in functional neuroimaging: A comparative review. *Statistical Methods in Medical Research*, 12(5):419–446, 2003. ISSN 0962-2802. doi: 10.1191/0962280203sm341ra.

V. P. Nosko. Local Structure of Gaussian Random Fields in the Vicinity of High-Level Light Sources, 1969.

Jonathan Taylor. A Gaussian kinematic formula. *Annals of Probability*, 34(1):122–158, 2006. ISSN 00911798. doi: 10.1214/009117905000000594.

Fabian Telschow, Armin Schwartzman, Dan Cheng, and Pratyush Pranav. Estimation of Expected Euler Characteristic Curves of Nonstationary Smooth Gaussian Random Fields. *arXiv e-prints*, art. arXiv:1908.02493, August 2019.

Richard J. Wilson. Model fields in crossing theory: A weak convergence perspective. *Advances in Applied Probability*, 20:756–774, 1988.

Richard J. Wilson and Robert J. Adler. The structure of gaussian fields near a level crossing. *Advances in Applied Probability*, 14:543–565, 1982.

K. J. Worsley, A C Evans, S Marrett, and P Neelin. A three-dimensional statistical analysis for CBF activation studies in human brain. *Journal of cerebral blood flow and metabolism.*, 12(6):900–18, 1992. ISSN 0271-678X. doi: 10.1038/jcbfm.1992.127.

Keith J. Worsley, Sean Marrett, Peter Neelin, Alain C Vandal, Karl J Friston, and Alan C Evans. A unified statistical approach for determining significant signals in images of cerebral activation. *Human brain mapping*, 4(1):58–73, 1996.