

---

# PEAK INFERENCE FOR GAUSSIAN RANDOM FIELDS ON A LATTICE

---

A PREPRINT

**Tuo Lin**

Department of Biostatistics and UF Health Cancer Center  
University of Florida  
Gainesville, FL 32603  
tuolin@ufl.edu

**Armin Schwartzman**

Division of Biostatistics and Halıcıoğlu Data Science Institute  
University of California, San Diego  
San Diego, CA 92122  
armins@ucsd.edu

**Samuel Davenport**

Division of Biostatistics  
University of California, San Diego  
San Diego, CA 92122  
sdavenport@health.ucsd.edu

September 10, 2024

## ABSTRACT

In this work we develop a Monte Carlo method to compute the height distribution of local maxima of a stationary Gaussian or Gaussian derived random field that is observed on a regular lattice. We show that our method can be used to provide valid peak based inference in order to detect and localize signals in random fields, under minimal assumptions. *This allows peak inference to be applied to datasets with low levels of smoothness where existing formulae derived for continuous domains are not accurate.* We also extend the methods in Worsley (2005) and Taylor et al. (2007) to compute the *peak* height distribution and compare them with our approach. Lastly, we apply our method to a task fMRI dataset to show how it can be used in practice.

**Keywords** peak inference · peak height · fMRI · connectivity · discrete lattice · local maxima

## 1 Introduction

Statistical parametric mapping (SPM) is widely used as a tool to conduct statistical inference on neuroimaging data (Friston et al., 1989; Worsley et al., 1992, 1996). Recently, Eklund et al. (2016, 2019) investigated the validity of cluster size and voxelwise inference based on random field theory (RFT) and found that a number of the assumptions that have been traditionally made do not hold in practice. *An important one of these assumptions*, which we address in this work, is that the data is sufficiently smooth so that it can be treated as a continuous random field. Inference based on peaks or local maxima, recognized as topological features of the statistical summary maps (Chumbley and Friston, 2009; Chumbley et al., 2010; Friston et al., 1989; Schwartzman et al., 2011; Cheng and Schwartzman, 2017) strongly relies on this assumption. In this paper, we circumvent this assumption and develop a method for performing peak inference that is valid for data observed on a regular lattice.

The traditional approach to obtaining peak  $p$ -values in fMRI analysis has been to assume that the data is distributed as a smooth stationary Gaussian random field. Given this, Nosko (1969); Adler (1981); Cheng and Schwartzman (2015a) showed that the distribution of the height of peaks above a peak-defining threshold  $u \in \mathbb{R}$  is asymptotically exponential (as  $u \rightarrow \infty$ ). The choice of  $u$  is somewhat arbitrary and this result only holds in practice for reasonably large choices of  $u$ . Recently, Cheng and Schwartzman (2015a) obtained a more general formula to calculate the exact height distribution of local maxima in an isotropic Gaussian random field, that is valid for all peak heights and does not require a pre-threshold. This distribution can be used to compute a  $p$ -value at each local maximum based on its height.

The formula has a single parameter  $\kappa$ , which only depends on the shape of the auto-correlation function near the origin, and is invariant under spatial scaling. While elegant, the formula is only accurate when the Gaussian random field is sampled on a continuous domain, instead of a discrete lattice grid, which in practice can require a high level of applied smoothing. To give context, Schwartzman and Telschow (2019) suggest that this formula is imprecise when data has an intrinsic FWHM that is lower than 7 voxels. However, since the typical smoothing kernel in an fMRI study has an FWHM of 3 voxels (Eklund et al., 2016), using this formula provides conservative  $p$ -values in practice. Moreover, the isotropic assumption is rather strong and is unlikely to hold in practice. Thus it is desirable to directly calculate the height distribution of local maxima sampled on a discrete lattice, which we shall refer to as discrete local maxima (DLM).

In order to address the difference between a discrete lattice and a smooth random field, Worsley (2005) and Taylor et al. (2007) introduced a method that targets the distribution of the global maximum on a lattice in order to provide control of the voxelwise family-wise error rate. Although this method aims to infer on the global maximum, it can also be used, after some modifications which we develop here, to compute the height distribution of local maxima. However, their approach is limited in that it is only valid for a narrow class of Gaussian random fields, namely the ones that arise as the result of convolving Gaussian white noise with a separable kernel. In addition, they require local maxima to be defined as those voxels with height values larger than *their* immediate neighbors along the coordinate axes (i.e. excluding diagonally adjacent neighbors). Figure 1 gives a rough idea of why this assumption is restrictive in practice by comparing density plots from peak height distributions calculated from both Worsley and Taylor's analytical DLM approach (which we shall refer to as ADLM) and Cheng and Schwartzman's continuous RFT approach.

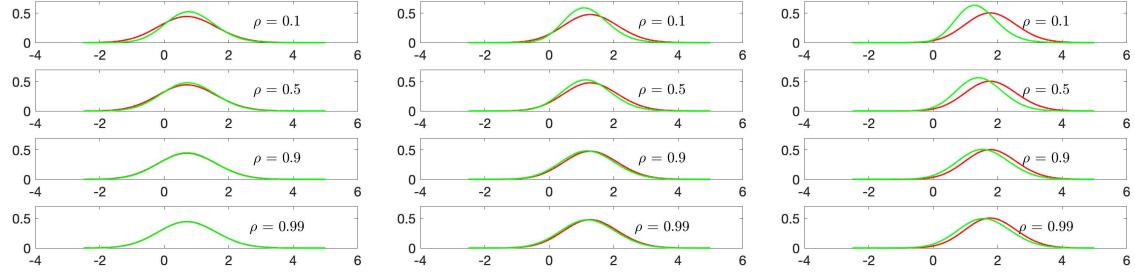


Figure 1: Theoretical peak height density function for local maxima. Left: 1D, Middle: 2D, Right: 3D. Each row is calculated using a different correlation between adjacent voxels. In each plot the green curve is from ADLM and the red curve is from the continuous RFT method. In the 1D case, when  $\rho$  is small, the ADLM density is narrower, but as  $\rho$  increases, the discrepancy disappears and the two methods converge. In 2D and 3D, the differences between the two methods remain for high  $\rho$ , with the ADLM density shifted to the left. This occurs because ADLM does not consider diagonal voxels as neighbors, so distribution of local maxima obtained from this method consists of smaller height values and is thus left shifted relative to the continuous method as dimension increases. Note that in 1D there are no diagonal voxels and so convergence occurs. [SD: Maybe you can get rid of the  $\rho = 0.9$  plot as it seems to tell the same story as the  $\rho = 0.90$  one.]

To address these issues, we propose a computation-based method called Monte Carlo DLM (MCDLM) that works for any stationary Gaussian random field under arbitrary connectivity (i.e. where local maxima are defined with respect to any desired neighborhood). This improves upon ADLM in that it allows the accurate computation of the height distribution of local maxima on a lattice without assuming a separable covariance function and permitting a range of neighbourhood structures. Our approach works by calculating the joint covariance of a voxel and its neighbors either theoretically or via empirical estimation, and then generating random samples from a multivariate Gaussian distribution with the calculated covariance and storing the samples that have larger height values than their neighbors. This provides an empirical cdf for the height of local maxima via numerical integration. A  $p$ -value for an observed peak in data can be obtained by interpolating the cdf [SD: What does this sentence mean?]. We also extend this approach to calculate the height distribution of local maxima of  $t$ -fields, by generating the random samples from a multivariate  $t$ -distribution.

The structure of this paper is as follows. Section 2 provides details about how to calculate the peak height distribution using continuous RFT, the ADLM and the MCDLM method. Section 3 and 4 apply MCDLM to isotropic Gaussian random fields,  $t$ -fields and stationary Gaussian fields with unknown covariance and compares its performance with ADLM and continuous RFT. Section 3 provides the simulation setup and Section 4 includes all the simulation results. Section 5 discusses a real example of applying the proposed methods. Section 6 gives the concluding remarks. Code to reproduce the results of this paper are available on GitHub (<https://github.com/tuolin123/DLM-Code>) and in the RFTtoolbox (<https://github.com/sjdavenport/RFTtoolbox>).

## 2 Theory and methods for calculating the height distribution of local maxima

Let  $\{Z(s), s \in \mathcal{L}\}$  be a real-valued stationary Gaussian random field parametrized on a  $D$ -dimensional set  $\mathcal{L}$ , where  $D \in \mathbb{Z}^+$ . We assume that  $\mathcal{L}$  is a regularly spaced discrete lattice, in particular that

$$\mathcal{L} \subset \left\{ \sum_{d=1}^D v_d e_d : v_d \in \mathbb{Z} \text{ for } 1 \leq d \leq D \right\},$$

where  $(e_d)_{1 \leq d \leq D}$  is the standard basis in  $\mathbb{R}^D$  and  $v_d$  represents the step size in the  $d$ th direction. Our interest lies in calculating the peak height distribution, which for  $u \in \mathbb{R}$ , is defined as

$$P(Z(s) > u | s \text{ is a local maximum}) = P[Z(s) > u | Z(t) < Z(s), \forall t \in \mathcal{N}(s)], \quad (1)$$

where  $\mathcal{N}(s)$  denotes the set of neighbors of  $s \in \mathcal{L}$  in the discrete lattice. The most relevant neighborhoods are the partially connected (PC) and fully connected (FC) ones that are respectively defined as

$$\mathcal{N}_{PC}(s) = \{s + k_d v_d e_d : k_d \in \{-1, 1\} \text{ for } 1 \leq d \leq D\} \text{ and} \quad (2)$$

$$\mathcal{N}_{FC}(s) = \left\{ s + \sum_{d=1}^D k_d v_d e_d : k_d \in \{-1, 0, 1\} \text{ for } 1 \leq d \leq D \right\} \setminus \{s\}. \quad (3)$$

Figure 2 illustrates the two types of neighborhoods for a point,  $s_5$ , on a 2D regular lattice. If  $s_5$  is partially connected to the adjacent pixels in the horizontal and vertical directions, then  $\mathcal{N}_{PC}(s_5) = \{s_2, s_4, s_6, s_8\}$ , shown in the left of Figure 2. If  $s_5$  is fully connected, meaning it is connected to pixels in the horizontal, vertical and diagonal directions, then  $\mathcal{N}_{FC}(s_5) = \{s_1, s_2, s_3, s_4, s_6, s_7, s_8, s_9\}$ , shown in the right of Figure 2.



Figure 2: Local pixel neighborhood in 2D. The partially and fully connected neighborhoods are shown on the left and right respectively. The point  $s_5$  in red is considered a local maximum if its value is larger than its neighbors.

### 2.1 Analytical DLM method

The DLM approach of Worsley (2005) and Taylor et al. (2007) provides closed form expressions for the family-wise error rate in testing the signals in data *where the noise is assumed to arise from Gaussian white noise smoothed with a separable* Gaussian kernel. They do not explicitly focus on the peak height distribution, however in inferring on the global maximum they calculate probabilities of the form  $P[\{Z(s) > u\} \cap_{t \in \mathcal{N}_{PC}(s)} \{Z(t) < Z(s)\}]$ . These probabilities can be used to calculate a peak height distribution for the partially connected neighbourhood since (1) can be written as

$$P[Z(s) > u | Z(t) < Z(s), \forall t \in \mathcal{N}_{PC}(s)] = \frac{P[\{Z(s) > u\} \cap_{t \in \mathcal{N}_{PC}(s)} \{Z(t) < Z(s)\}]}{P[Z(t) < Z(s), \forall t \in \mathcal{N}_{PC}(s)]}. \quad (4)$$

Using *the* results of Worsley (2005) and Taylor et al. (2007), we *can* expand the left hand side of (4) as  $\int_u^\infty f_{DLM}(z) dz$ , where  $f_{DLM}(z)$  is the density function of the height distribution. Under the assumption that  $Z$  *has a separable Gaussian covariance function*,  $f_{DLM}(z)$  has the form

$$f_{DLM}(z) = \frac{\prod_{d=1}^D Q(\rho_d, z) \phi(z)}{\int_{-\infty}^\infty \left( \prod_{d=1}^D Q(\rho_d, z) \right) \phi(z) dz}, \quad (5)$$

where

$$h_d = \sqrt{\frac{1 - \rho_d}{1 + \rho_d}}, \alpha_d = \sin^{-1} \left( \sqrt{(1 - \rho_d^2)/2} \right), z^+ = \max(z, 0)$$

$$Q(\rho_d, z) = 1 - 2\Phi(h_d z^+) + \frac{1}{\pi} \int_0^{\alpha_d} \exp \left( -\frac{1}{2} h_d^2 z^2 / \sin^2 \theta \right) d\theta,$$

and  $\rho_d$  is the correlation between two voxels along each axis direction  $d$ , given by  $\rho_d = \rho(s, s + v_d e_d)$ , where  $\rho(\cdot, \cdot)$  is introduced in (6). This approach also allows for the calculation of the height of local maxima on the boundary of the image or a mask by substituting  $Q(\rho_d, z)$  with  $1 - \Phi(h_d z)$  if a voxel on the boundary only has one neighbor, and with 1 if it has no neighbors. Further details regarding the derivation of (5) are provided in Appendix A.2. Since this *approach* provides a closed form density function, *we shall refer to it as the* analytical DLM (ADLM) method.

One critical assumption of ADLM is that the correlation function has a specific separable structure. Under this assumption things simplify because conditioned on the center voxel, the distribution of the height of neighboring voxels along a given axis are conditionally independent of the distribution of the height at neighboring voxels along the other (perpendicular) axis directions *as described formally* in the proposition below.

**Proposition 1.** *Given data  $\{Z(s), s \in \mathcal{L}\}$  such that the spatial correlation between  $s$  and  $t$  is*

$$\rho(s, t) = \exp[-(s - t)' \Lambda(s - t)/2], \quad (6)$$

*where  $\Lambda = \text{diag}(1/(2\eta_1^2), \dots, 1/(2\eta_D^2))$ , and  $(\eta_d)_{d=1,\dots,D}$  is the standard deviation of the Gaussian kernel in the  $d$ th direction. Then given  $d_1, d_2 \in \{1, \dots, D\}$  such that  $s \pm v_{d_1} e_{d_1}, s \pm v_{d_2} e_{d_2} \in S$  and letting  $\perp\!\!\!\perp$  denote independence, it follows that,*

$$\left( \begin{array}{c} Z(s - v_{d_1} e_{d_1}) \\ Z(s + v_{d_1} e_{d_1}) \end{array} \right) \perp\!\!\!\perp \left( \begin{array}{c} Z(s - v_{d_2} e_{d_2}) \\ Z(s + v_{d_2} e_{d_2}) \end{array} \right) \mid Z(s).$$

The correlation function in (6) arises, for example, from integration of continuous white noise against a Gaussian kernel. The result in Proposition 1 is stated [SD: briefly] in Taylor et al. (2007) and we provide a short *formal* proof in Appendix A.1. *To help understand this result with a visual example note that for the neighbourhood structures shown in Figure 2, under the required assumptions, we have*

$$\left( \begin{array}{c} Z(s_4) \\ Z(s_6) \end{array} \right) \perp\!\!\!\perp \left( \begin{array}{c} Z(s_2) \\ Z(s_8) \end{array} \right) \mid Z(s_5).$$

This conditional independence result holds along the horizontal and vertical axes and allows for an expansion for the distribution of partially connected local maxima. However, it does not imply independence when the diagonals are included, i.e.,

$$(Z(s_1), Z(s_3), Z(s_7), Z(s_9))^\top \not\perp\!\!\!\perp (Z(s_2), Z(s_4), Z(s_6), Z(s_8))^\top \mid Z(s_5).$$

Thus, this ADLM method can only be used to calculate the height distribution of peaks that are greater than their directly adjacent neighbors.

[SD: Before you said under stationarity but that was wrong right, I've fixed it below, we can discuss if you like.] Under 6, the correlation  $\rho_d$  between two adjacent voxels along each lattice axis  $d$  can be written as

$$\rho_d = \rho(s, s + v_d e_d) = \exp \left[ -\frac{1}{2} \left( \frac{v_d^2}{2\eta_d^2} \right) \right] = \exp \left[ -\frac{v_d^2}{4\eta_d^2} \right], \quad (7)$$

and if we assume  $Z(s)$  is isotropic with a common standard deviation of the Gaussian kernel  $\eta_d = \eta$  and  $v_d = 1$ ,  $\rho_d$  can be further simplified to  $\rho_d = \rho = \exp[-1/4\eta^2]$ , a function which does not depend on  $d$ .

The ADLM approach allows the calculation of the height distribution of local maxima on a discrete lattice. However, the method makes restrictive assumptions and its validity is limited to partial connectivity.

## 2.2 The correlation function on the lattice

The methods which we will develop in what follows strongly rely on the correlation function. In this section we provide some explicit expansions of this function under the assumption that the fields are derived by smoothing i.i.d white noise with a kernel (we will relax this assumption later on).

Define the correlation function  $\rho(s, t) : S \times \mathcal{L} \rightarrow \mathbb{R}$  to be the function that maps  $s, t \in \mathcal{L}$  to  $\text{corr}(Z(s), Z(t))$ . As a step toward our goal of calculating peak  $p$ -values for a Gaussian random field on a regular discrete lattice, we shall calculate the spatial correlation *analytically for fields generated by smoothing noise, before extending to the more general setting in Section 2.3*. Assume that  $W : \mathcal{L} \rightarrow \mathbb{R}$  is a Gaussian random field consisting of i.i.d. unit variance white noise and for some kernel  $K : \mathbb{R}^D \rightarrow \mathbb{R}$ ,

$$Z(s) = \sum_{l \in \mathcal{L}} K(s - l) W(l) \quad \text{for each } s \in \mathcal{L}.$$

The correlation function  $\rho(s, t)$  is then

$$\begin{aligned} \rho(s, t) &= \frac{E \left[ \sum_{l \in \mathcal{L}} K(s - l) W(l) \sum_{l' \in \mathcal{L}} K(t - l') W(l') \right]}{\sqrt{\text{Var} \left[ \sum_{l \in \mathcal{L}} K(s - l) W(l) \right] \text{Var} \left[ \sum_{l' \in \mathcal{L}} K(t - l') W(l') \right]}} \\ &= \frac{E \left[ \sum_{l \in \mathcal{L}} \sum_{l' \in \mathcal{L}} K(s - l) K(t - l') W(l) W(l') \right]}{\text{Var} \left[ \sum_{l \in \mathcal{L}} K(s - l) W(l) \right]} \\ &= \frac{\sum_{l \in \mathcal{L}} K(s - l) K(t - l)}{\sum_{l \in \mathcal{L}} K(s - l)^2} \end{aligned}$$

since  $E[W(l)W(l')] = 0$  for  $l \neq l'$  and  $EW(l)^2 = 1$  for all  $l$ . In particular when  $K$  is an isotropic Gaussian kernel, i.e.,  $K(s) = \frac{1}{\eta^D} \phi_D \left( \frac{\|s\|}{\eta} \right)$ , for some  $\eta > 0$  and each  $s, t \in S$ ,

$$\rho(s, t) = \frac{\sum_{l \in S} \frac{1}{\eta^{2D}} \phi_D \left( \frac{\|\mathcal{L}-l\|}{\eta} \right) \phi_D \left( \frac{\|t-l\|}{\eta} \right)}{\sum_{l \in \mathcal{L}} \frac{1}{\eta^{2D}} \left[ \phi_D \left( \frac{\|s-l\|}{\eta} \right) \right]^2}, \quad (8)$$

where  $\phi_D$  is the density function for the  $D$  dimensional standard Gaussian distribution. As is common in fMRI analysis we will typically refer to this kernel using its full width at half maximum (FWHM) which is defined as  $\text{FWHM} = 2\sqrt{2 \ln 2\eta}$ . Using (8) as the correlation function to ADLM defined in (5) improves the performance of the ADLM approach relative to using (6) - the correlation function used in Worsley (2005) - as discussed in Section D.2.

More generally if  $K$  is an elliptical Gaussian kernel, i.e.,  $K(s) = \prod_{j=1}^D \frac{1}{\eta_j} \phi_1 \left( \frac{[s]_j}{\eta_j} \right)$ , then

$$\rho(s, t) = \frac{\sum_{l \in \mathcal{L}} \prod_{j=1}^D \frac{1}{\eta_j^2} \phi_1 \left( \frac{[s-l]_j}{\eta_j} \right) \phi_1 \left( \frac{[t-l]_j}{\eta_j} \right)}{\sum_{l \in \mathcal{L}} \prod_{j=1}^D \frac{1}{\eta_j^2} \phi_1 \left( \frac{[s-l]_j}{\eta_j} \right)^2}, \quad (9)$$

where  $[s - w]_j$  and  $[s + v - w]_j$  refer to the  $j$ th *elements of the vectors*  $s - w$  and  $s + v - w$  respectively.

### 2.3 Monte Carlo DLM method

In this section we introduce a new method based on Monte Carlo simulation to calculate the height distribution of local maxima on a discrete lattice. Our approach is based on the observation that the probability that  $s \in \mathcal{L}$  is a local maximum based entirely on the distribution of  $s$  with its neighbors, *c.f.* (1). Define

$$\mathbf{Z}(s) = (Z(s), Z(n_1(s)), \dots, Z(n_k(s)))^\top,$$

where we *write* the neighborhood of  $s$  as  $\mathcal{N}(s) = \{n_1(s), \dots, n_k(s)\} \subseteq \mathcal{N}_{FC}(s)$  for some  $1 \leq k \leq |\mathcal{N}_{FC}(s)|$ . *For the* partially connected neighborhood  $k = 2D$  and for the fully connected neighborhood,  $k = 3^D - 1$ . *Under stationarity*,  $\mathbf{Z}(s) \sim N(\mathbf{0}, \Sigma)$  for each  $s$  where the covariance,  $\Sigma = \text{Cov}(\mathbf{Z}(s))$ , is constant over the domain. The covariance matrix  $\Sigma$  can be derived analytically under certain assumptions or estimated from the data, see below. Given  $\Sigma$ , the method calculates the peak height distribution via Monte Carlo experiments as described in Algorithm 1.

**Algorithm 1** MCDLM

**Require:** The number of iterations  $M \in \mathbb{N}$ , and the  $k \times k$  covariance matrix  $\Sigma$

```

1: procedure SIMULATELOCMAX( $M, \Sigma$ )
2:   for  $m = 1, \dots, M$  do
3:     Generate  $X_m \sim N(0, \Sigma)$ 
4:     if  $X_{m1} > \max_{2 \leq j \leq k} X_{mj}$  then
5:        $\mathbf{h} = [\mathbf{h}, X_{m1}]$ 
6:     end if
7:   end for
8:   return  $\mathbf{h}$ 
9: end procedure

```

After obtaining the vector  $\mathbf{h} = (h_1, h_2, \dots, h_N)^\top$ , for  $u \in \mathbb{R}$ , the MCDLM approximation to the peak height distribution can be calculated as

$$\hat{F}_N(u) = \frac{1}{N} \sum_{i=1}^N \mathbb{1}\{h_i \leq u\},$$

where  $\mathbb{1}\{\cdot\}$  denotes the indicator function and  $N$  is the length of  $\mathbf{h}$ . For an observed peak of height  $u$ , a peak  $p$ -value can be computed as  $1 - \hat{F}_N(u)$ . In order to make this empirical  $p$ -value as accurate as possible,  $N$  should be taken to be as large as possible.

Now we address the question of how to calculate  $\Sigma$ . We will first do this for the particular isotropic subcase, discussed in the previous sections, in which  $Z$  is obtained by integrating continuous white noise against a Gaussian kernel. In that case, assuming a fully connected neighborhood, it can be shown that, up to reindexing of  $\mathbf{Z}$ ,

$$\Sigma = \underbrace{A \otimes A \otimes \dots \otimes A}_{D \text{ terms of } A} = A^{\otimes D} \quad (10)$$

where

$$A = \begin{pmatrix} 1 & \rho & \rho^4 \\ \rho & 1 & \rho \\ \rho^4 & \rho & 1 \end{pmatrix}.$$

with  $\rho$  the correlation between adjacent voxels as defined in (7). *See Theorem 3 in Appendix B for formal statement and proof of this result. Technically I think this proof only holds for the corelation (6) and not (9), right? Even though we are using it in the simulations, I think we can ignore that point here though as I doubt anyone will catch it!*

Equation (10) holds under restrictive assumptions. For a general *mean-zero* stationary field we can instead use the data to estimate  $\Sigma$ . To do so, *without loss of generality, assume we observe and standardized* i.i.d Gaussian random fields  $Z_1, \dots, Z_n$  on  $\mathcal{L}$  [*SD: I changed it because you're not standardizing the fields your self right? Else they would become t-fields.*]. We wish to infer on the distribution of peaks of the mean  $\frac{1}{n} \sum_{i=1}^n Z_n$ . *We can then* estimate  $\Sigma$  from the data as follows. For each  $s, t \in \mathcal{L}$ ,

$$\text{Cov} \left( \frac{1}{n} \sum_{i=1}^n Z_i(s), \frac{1}{n} \sum_{i=1}^n Z_i(t) \right) = \frac{1}{n} \text{Cov}(Z_1(s), Z_1(t)), \quad (11)$$

as such, using the assumption of stationarity, we can estimate this covariance as

$$\widehat{\text{Cov}}(Z_1(s), Z_1(t)) = \frac{1}{n|L|} \sum_{i=1}^n \sum_{(s', t') \in L} Z_i(s') Z_i(t') \quad (12)$$

where  $L = \{(s', t') \in \mathcal{L} \times \mathcal{L} : s' - t' = s - t\}$ . *This allows us to build an estimate of  $\Sigma$  by calculating (12) between relevant pairs of voxels.* If *further* we assume that the fields are isotropic, then we can improve the accuracy of this estimate by taking  $L = \{(s', t') \in S \times S : \|s' - t'\| = \|s - t\|\}$ .

## 2.4 MCDLM for $t$ -fields

We can also use our MCDLM approach to calculate the height distribution of local maxima of a  $t$ -field. The  $t$ -fields are generated by voxelwise calculation of  $t$ -statistic using

$$T(s) = \frac{\varepsilon(s)}{\sqrt{\sum_{i=1}^N Z_i^2/N}}, s \in \mathcal{L}, \quad (13)$$

where  $Z_1, \dots, Z_N$  and  $\varepsilon(s)$  are i.i.d *stationary* Gaussian random fields observed on the lattice  $\mathcal{L}$ . In practice, the  $t$ -statistic is typically used as the test statistic for regression coefficients.

In this setting the local neighborhood has a multivariate  $t$ -distribution. Thus, given an estimate of the neighborhood covariance we can still apply MCDLM using Algorithm 1 by changing the simulation in line 3 from a multivariate Gaussian distribution to a multivariate  $t$ -distribution.

This approach works well in practice (see Section 4.2) however it is somewhat computational expensive (especially as  $\rho$  and the degrees of freedom increase). To get around this we consider a voxelwise Gaussianization transformation of the  $t$ -fields (as in Schwartzman and Telschow (2019)) which acts using the distribution function as follows:

$$Z(s) = -\Phi^{-1}[F_{t,\nu}(-T(s))], \quad (14)$$

where  $F_{t,\nu}$  is the cdf of the  $t$ -distribution with  $\nu$  degrees of freedom. We then apply the MCDLM method for the Gaussian field to the transformed  $t$ -field resulting in improved speed without significantly impacting performance, see Section D.4.

## 2.5 Continuous Gaussian random field theory

Historically (Chumbley and Friston, 2009; Chumbley et al., 2010; Schwartzman and Telschow, 2019) it has been common to use the results of continuous Gaussian random fields (Adler, 1981) to perform inference on the lattice. We compare to this approach in our simulations and so here briefly outline how this works and explain how it can be used to provide height distributions for local maxima in smooth random fields. *In order to apply Random Field Theory we must assume that the noise is sufficiently smooth for the good lattice assumption to hold (Davenport et al., 2023). Then given a domain  $S \supset \mathcal{L}$ , which is compact with non-empty interior  $\mathring{S}$ , the good lattice assumption implies that  $Z$  extends to a  $C^3$  random field  $Z$  on  $S$  the peaks of which can be identified with and have similar heights to the peaks of  $Z$  on the lattice  $\mathcal{L}$ .* Let

$$\nabla Z(s) = \left( \frac{\partial Z(s)}{\partial s_1}, \dots, \frac{\partial Z(s)}{\partial s_D} \right), \text{ and } \nabla^2 Z(s) = \left( \frac{\partial^2 Z(s)}{\partial s_{ij}} \right)_{1 \leq i,j \leq D}.$$

Then the local maxima of  $Z$  on  $S$  are defined to be the points  $s \in \mathring{S}$  such that  $\nabla Z(s) = 0$  and  $\nabla^2 Z(s) < 0$ .

The event that a local maximum is observed at a given  $s \in \mathring{S}$  has probability zero. As such, in order to obtain a conditional peak height distribution, Palm distributions must be used (see Cheng and Schwartzman (2015a) for details). For  $u \in \mathbb{R}$ , they provide formulae to calculate

$$\mathbb{P}[Z(s) > u \mid \nabla Z(s) = 0 \text{ and } \nabla^2 Z(s) < 0]. \quad (15)$$

In general these expressions are difficult to evaluate. However, under the assumption of isotropy, Cheng and Schwartzman (2015b) showed that they can be obtained explicitly. Recently, Cheng and Schwartzman (2020) extended these results to the case where the field arises as a diffeomorphic transformation of an isotropic field. Details of how to apply these methods to perform peak inference in fMRI data can be found in Schwartzman and Telschow (2019).

## 3 Simulation Setup

In this section we describe the different simulation settings we have considered in order to compare the performance of the three methods introduced in Section 2, i.e. ADLM, MCDLM continuous RFT method. For each simulation setting considered we generate a large number of stationary Gaussian random fields (or  $t$ -fields), collect the heights of the peaks across all fields and combine these to obtain a reference peak height distribution, which will allow us to test the validity of each of the methods.

For each method, we calculate a  $p$ -value at each peak. Local maxima were selected based on the criteria that their height values are larger than their neighbors - we *shall* consider both the fully connected and partially connected neighborhoods. We compare the validity and accuracy of these  $p$ -values using  $pp$  plots which compare the sorted

$p$ -values to the tail probability of the true peak height distribution. These are formally defined in Appendix C for clarity. The closer *the line in each* plot is to the identity function, the closer the approximation is to the true distribution. *Lines* lying below the identity correspond to conservative  $p$ -values and *lines* above the identity function correspond to liberal  $p$ -values. We use these  $pp$  plots to compare the performance of the three approaches in all of our simulation studies.

### 3.1 Isotropic Gaussian random fields

Our first set of simulations consists of isotropic Gaussian fields which are obtained by convolving Gaussian white noise with a Gaussian kernel with specified FWHM and normalizing so that the resulting fields have unit variance. To avoid any boundary effects the fields were initially generated on a  $D$ -dimensional large grid of size  $(50 + 2 \times \lceil 4 * \eta \rceil)$  at each direction and the central subset is taken, as described in Davenport and Nichols (2020). We do this in 2D and in 3D. The resulting 2D images are of size  $50 \times 50$  and the resulting 3D images are of size  $50 \times 50 \times 50$ . In the simulations we choose the FWHM so that the correlation between adjacent voxels in each perpendicular direction is equal to  $\rho \in \{0.01, 0.5, 0.99\}$  - which correspond to FWHM of 0.7, 1.5 and 11.7 voxels respectively. For now we assume that  $\rho$  is known and calculate the neighbourhood covariance  $\Sigma$ , needed for MCDLM, using (10). In each setting we generate 10,000 random fields and compare the different approaches using  $pp$  plots - as described above and in Appendix C. See Section 4.1 for the corresponding results. *In our examples here and in the following sections, when applying MCDLM, we choose  $M$  large enough to ensure that the number of resulting empirical samples is at least  $N = 10^6$  for  $\text{FWHM} < 11.7$  and at least  $N = 2 \times 10^5$  for  $\text{FWHM} = 11.7$ . We also generate a look-up table, which provides the same results under reduced computation time (for the case of Gaussian white noise smoothed with an isotropic Gaussian kernel). The results of using the lookup table are shown in Appendix D.3. [SD: I moved this last two sentences here from elsewhere as they fit better here].*

### 3.2 Isotropic $t$ -fields

*For our second of simulations*, we consider the performance of the different approaches when it comes to evaluating the height distribution of peaks of  $t$ -fields. To do so we generate fields with  $\nu$  degrees of freedom (taking  $\nu = 20, 50, 100$ ) by simulating i.i.d isotropic Gaussian fields  $Z_1, \dots, Z_\nu$  and  $\varepsilon(s)$  in (13) as in Section 3.1. In each setting we generate 10,000  $t$ -fields in both 2D and 3D and calculate peak-height  $p$ -values using the MCDLM approach for  $t$ -fields discussed in Section 2.5. We also calculate  $p$ -values using the continuous RFT approach; note that this is designed for Gaussian random fields so we would not expect it to work as well in this setting. *To help understand how close continuous RFT is to simulations from a Gaussian random field we compute a final set of  $p$ -values using simulated Gaussian random fields. As the number of degrees of freedom of the  $t$ -field increases the  $t$ -field converges to a Gaussian random field with the original covariance function so this comparison also allows us to understand whether the fields have converged. Note that in practice this final approach is not a viable measure for calculating a peak height distribution as it requires generating a large number of fields which is extremely computationally expensive but we include it in the plots as a reference. We do not compare to ADLM in this setting as it is only designed for Gaussian fields.* We compare the  $p$ -values obtained using these different approaches using  $pp$ -plots. The results are described in Section 4.2.

As discussed in Section 2.4, we also consider *the use of a* Gaussianization transformation of the  $t$ -field to improve the computational efficiency. We perform the same set of simulations but where each of the  $t$ -fields is Gaussianized, with the same  $t$ -field as described above. We then calculate  $p$ -values using the MCDLM for Gaussian fields and the continuous RFT approach. The results are described in Appendix D.4.

### 3.3 Stationary Gaussian fields with unknown covariance

*To study the performance of MCDLM, when the covariance is unknown and must therefore be estimated we consider several further simulation settings.* Firstly we use  $n$  fields ( $n = 20, 50, 100, 200$  for small  $\rho$  and  $n = 20, 50, 100, 200, 1000$  for large  $\rho$ ) to estimate the neighborhood covariance using the isotropic version of equation (12). We compare the performance of MCDLM with this estimated covariance across different sample sizes (*running 10,000 simulations* in each of the settings described in Section 3.1). Secondly we consider 2D non-isotropic Gaussian fields. To generate these we smooth Gaussian white noise with an elliptical Gaussian kernel with smoothing FWHM in each direction chosen such that the correlation between adjacent voxels in the vertical and horizontal directions is  $\rho_1$  and  $\rho_2$  respectively. We estimate the neighborhood covariance using equation (9) using  $n$  fields ( $n = 20, 50, 100, 200$ ) *and then apply MCDLM to generate peak  $p$ -values*. We consider two *2D non-isotropic scenarios*, one where  $\rho_1 = 0.01$  and  $\rho_2 = 0.5$  and a second where  $\rho_1 = 0.5$  and  $\rho_2 = 0.99$  (example realizations of these fields are *shown* in Figure 3). The results are shown in Section 4.3.

*For each of these simulations we estimate  $\Sigma$  from the data as described in Section 2.3. We take advantage of stationarity in order to do so as there is a lot of structure that can be taken advantage of when estimating  $\Sigma$ . In particular the*

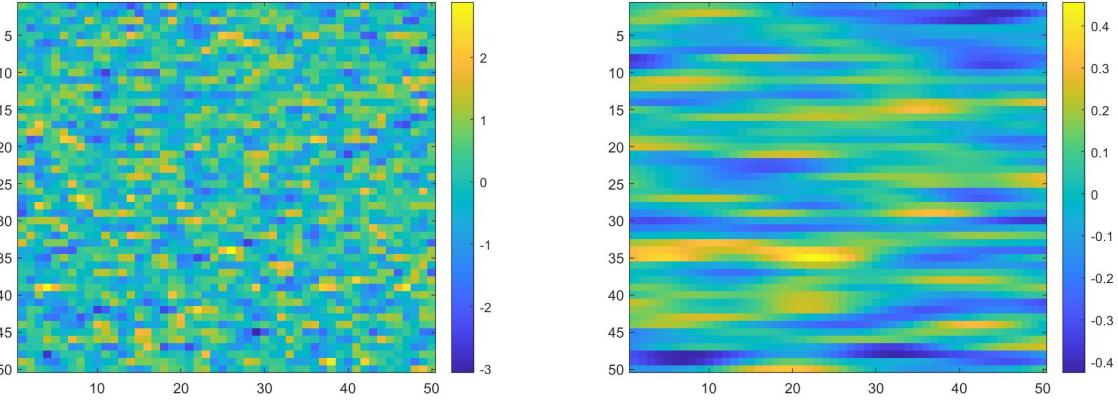


Figure 3: Examples of stationary Gaussian random fields which are obtained by convolving white noise with an elliptical Gaussian kernel with  $\rho_1 = 0.01$ ,  $\rho_2 = 0.5$  in the left plot and  $\rho_1 = 0.5$ ,  $\rho_2 = 0.99$  in the right plot.

*neighborhood covariance matrix has a block Toeplitz structure which makes it easier to estimate, see the examples in Appendix B.2. As a practical note, we observe that once the covariance function has been estimated we must ensure that it is positive semi-definite (p.s.d) in order to simulate from it. In order to ensure this we push the negative eigenvalues of estimated covariance matrix to a small positive value,  $1 \times 10^{-10}$ .*

## 4 Simulation Results

### 4.1 Results for isotropic Gaussian random fields

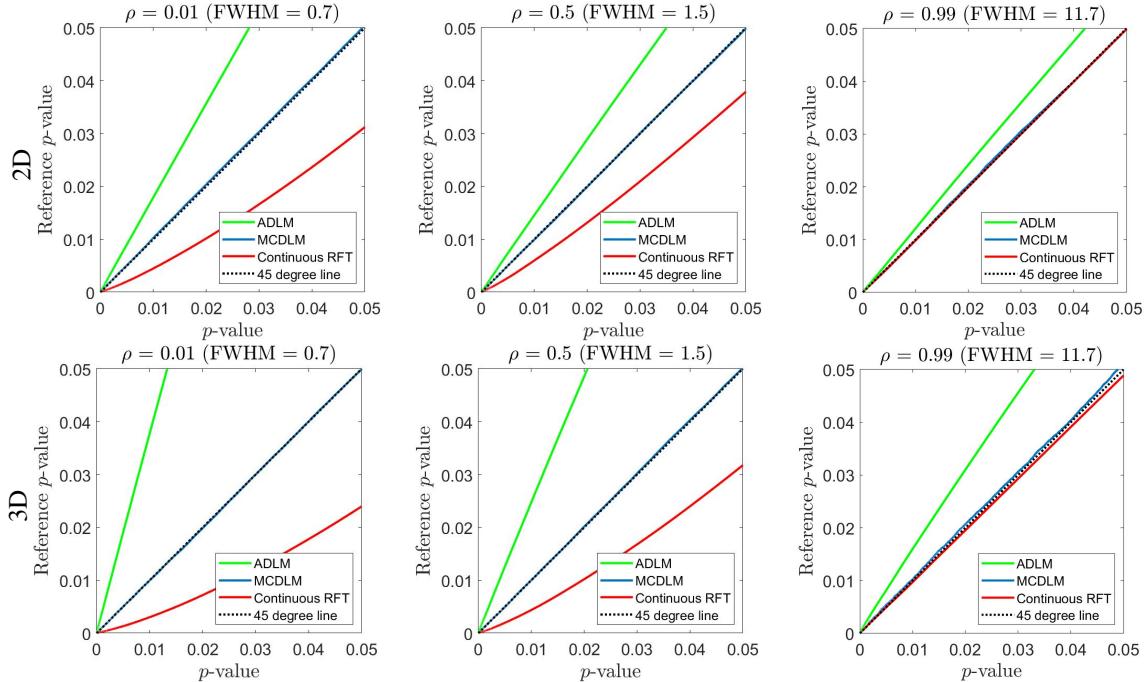


Figure 4:  $pp$  plots which compare the different methods of computing peak height  $p$ -values in the isotropic Gaussian random field scenario. 2D and 3D results are displayed in the first and second rows respectively. The correlations between adjacent voxels are  $\rho = 0.01, 0.5, 0.99$ . The plots compare the performance of ADLM, MCDLM and the continuous RFT approach.

Results comparing all three methods in the isotropic Gaussian random fields setting (described in Section 3.1) are presented in Figure 4. From this figure we see that the MCDLM method obtains  $p$ -values which are uniformly distributed and thus provides accurate and valid inference at all smoothness levels. The continuous approach is valid but conservative unless the data is very smooth, i.e., with a large smoothing FWHM. The ADLM method gives liberal  $p$ -values at all *considered* smoothness levels though the severity of this reduces when the smoothness is very large. In 3D the results are similar though at the highest smoothness level the curve corresponding to the MCDLM method is slightly rough, this could be made sharper if desired by increasing the number of Monte Carlo runs used. In these results we use the covariance function (8) for both MCDLM and ADLM distributions since it is the actual covariance of the data, as discussed in Section 2.2. The results are slightly worse if (7) is used instead (as was done in Worsley (2005); Taylor et al. (2007)), see Appendix D.2. Here we only consider the case where the neighborhood is fully connected; results for partially connected neighborhoods, in which the ADLM method performs much better than MCDLM method, are presented in Appendix D.1.

To quantify the difference of  $p$ -values from all three methods, we use the Root Mean Squared Error (RMSE) to calculate the difference between  $p$ -value from each of the method and the identity line, or  $45^\circ$  line. Our comparison focus on the region of  $p$ -value  $\leq 0.05$ . The RMSE results calculated from 2D isotropic Gaussian random fields with different  $\rho$  are shown in Table 1. From the table, MCDLM performs better than ADLM and continuous RFT approach in all low smoothness cases ( $\text{FWHM} \leq 5.2$ ). The MCDLM is outperformed by continuous RFT approach when smoothness level is high ( $\text{FWHM} = 8.3$  and  $11.7$ ).

Table 1: RMSE results from 2D isotropic Gaussian random fields for comparing the  $p$ -values from MCDLM, ADLM and continuous RFT approaches. The smallest value in each row is highlighted in red color.

	MCDLM	ADLM	Continuous RFT
$\rho = 0.01$ (FWHM = 0.7)	$1.71 \times 10^{-4}$	$5.77 \times 10^{-3}$	$8.18 \times 10^{-3}$
$\rho = 0.1$ (FWHM = 1)	$1.91 \times 10^{-4}$	$5.48 \times 10^{-3}$	$7.75 \times 10^{-3}$
$\rho = 0.3$ (FWHM = 1.2)	$6.16 \times 10^{-5}$	$4.83 \times 10^{-3}$	$6.53 \times 10^{-3}$
$\rho = 0.5$ (FWHM = 1.5)	$6.34 \times 10^{-5}$	$4.16 \times 10^{-3}$	$4.99 \times 10^{-3}$
$\rho = 0.7$ (FWHM = 2)	$1.22 \times 10^{-4}$	$3.58 \times 10^{-3}$	$3.01 \times 10^{-3}$
$\rho = 0.9$ (FWHM = 3.6)	$2.05 \times 10^{-4}$	$2.91 \times 10^{-3}$	$8.61 \times 10^{-4}$
$\rho = 0.95$ (FWHM = 5.2)	$1.12 \times 10^{-4}$	$2.56 \times 10^{-3}$	$5.04 \times 10^{-4}$
$\rho = 0.98$ (FWHM = 8.3)	$1.57 \times 10^{-4}$	$2.51 \times 10^{-3}$	$1.02 \times 10^{-4}$
$\rho = 0.99$ (FWHM = 11.7)	$2.87 \times 10^{-4}$	$2.54 \times 10^{-3}$	$1.29 \times 10^{-4}$

[SD: Could you add  $\rho = 0.97, 0.96$  to the table?]

## 4.2 Isotropic $t$ -fields

The  $pp$  plots comparing the different methods in the setting of isotropic  $t$ -fields (described in Section 3.2) are presented in Figure 5 (2D) and Figure 6 (3D). As shown in these two figures, MCDLM obtains  $p$ -values which are uniformly distributed at all smoothness levels and different degrees of freedom for 2D and lower smoothness levels for 3D. In the 3D case at high smoothness levels, the MCDLM approach becomes a rough approximation because the number of peaks generated is not sufficient. As such the height distribution computed is inaccurate, as shown in the noisy subfigure in the bottom right of Figure 6. The continuous method is designed for Gaussian fields rather than  $t$ -fields, so it is too liberal when  $\nu$  is small while too conservative when  $\nu$  is large and  $\rho$  is small. Although in the case that the number of degrees of freedom is large and FWHM large, the Gaussian field can approximate the  $t$ -field and continuous method has improved performance, the MCDLM method always outperforms the continuous method in all circumstances as long as we generate enough local maxima from Monte Carlo experiments.

As discussed in Section 3.2, a Gaussianization approach *can be used* to save computation time. The results for this are shown in Appendix D.4. They show that the MCDLM method performs well when the number of degrees of freedom is sufficiently large whereas continuous method requires both the degrees of freedom and FWHM to be large.

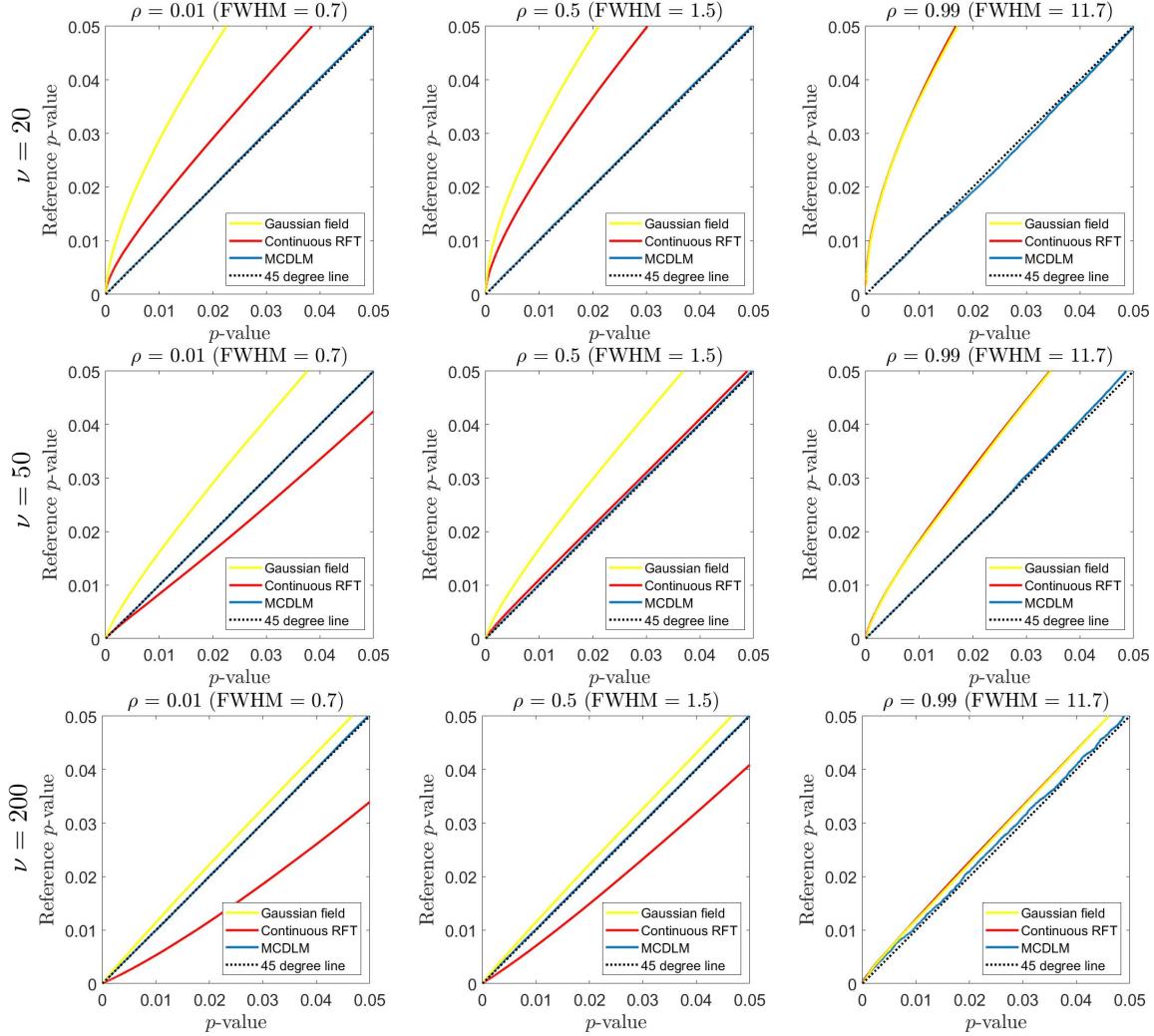


Figure 5: Comparing methods for calculating the peak height distribution of a 2D  $t$ -field with  $\nu$  degrees of freedom. From left to right: spatial correlation  $\rho = 0.01, 0.5, 0.99$ . From top to bottom:  $\nu = 20, 50, 200$ . The figure is generated based on the comparison of the  $p$ -values calculated using Gaussian field with same correlation  $\rho$ , continuous RFT and MCDLM approach. The reference is the true peak height distribution generated from the  $t$  field.

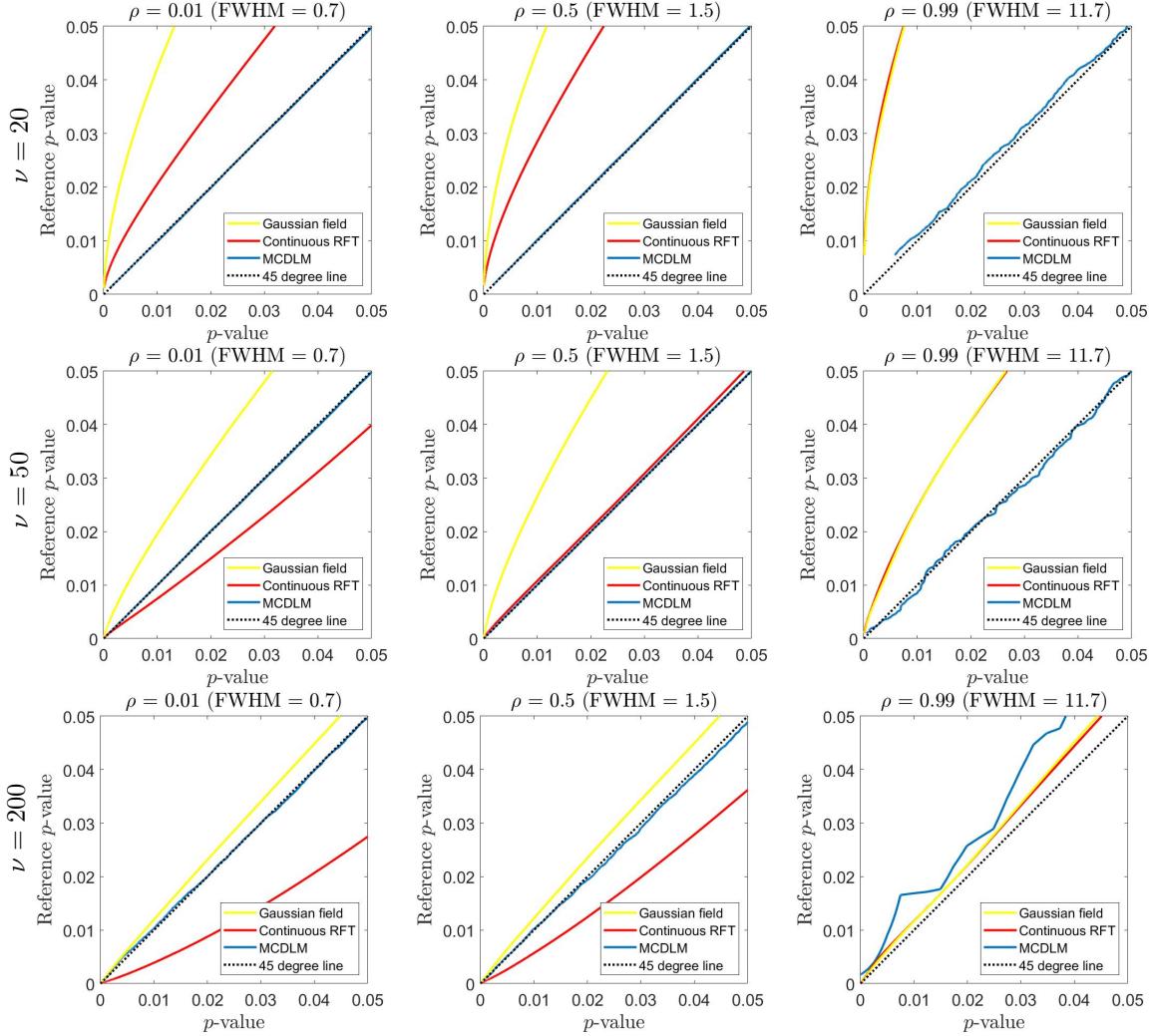


Figure 6: Comparing methods for calculating the peak height distribution of a 3D  $t$ -field with  $\nu$  degrees of freedom. From left to right: spatial correlation  $\rho = 0.01, 0.5, 0.99$ . From top to bottom:  $\nu = 20, 50, 200$ . The figure is generated based on the comparison of the  $p$ -values calculated using Gaussian field with same correlation  $\rho$ , continuous RFT and MCDLM approach. The reference is the true peak height distribution generated from the  $t$  field.

### 4.3 Stationary Gaussian fields with unknown covariance

In Figure 7, we compare the peak height distribution calculated from the MCDLM method using both theoretical (i.e. true) neighborhood covariance in (8) and estimated neighborhood covariance displayed in (12). Figure 7 shows that MCDLM with the estimated neighborhood covariance performs as well as MCDLM with the theoretical neighborhood covariance when  $\rho = 0.01$  and  $0.5$ . When  $\rho$  increases to  $0.99$ , the MCDLM method with estimated covariance function requires a large number of simulated peaks before it converges. This number decreases with the number of voxels in the image (which is why the performance of the 3D simulations are substantially better than the 2D ones), even when a very large sample size is used to estimate it. Since with  $\rho = 0.99$ , even with 1000 instances to estimate the neighborhood covariance the MCDLM method still performs poorly, we investigate more scenarios with  $\rho = 0.9, 0.93, 0.95$ . The detailed results are included in Appendix D.5. Based on the results, we recommend using the MCDLM method with estimated covariance function when  $\rho < 0.95$ , or  $\text{FWHM} < 5.2$  in practice.

The results for the second (non-isotropic) simulations discussed in Section 3.3 are shown in Figure 8. From this plot we see that the estimated version works well when  $\rho_1 = 0.01$  and  $\rho_2 = 0.5$ , and requires a larger number of realizations to converge when  $\rho_1 = 0.5$  and  $\rho_2 = 0.99$ .

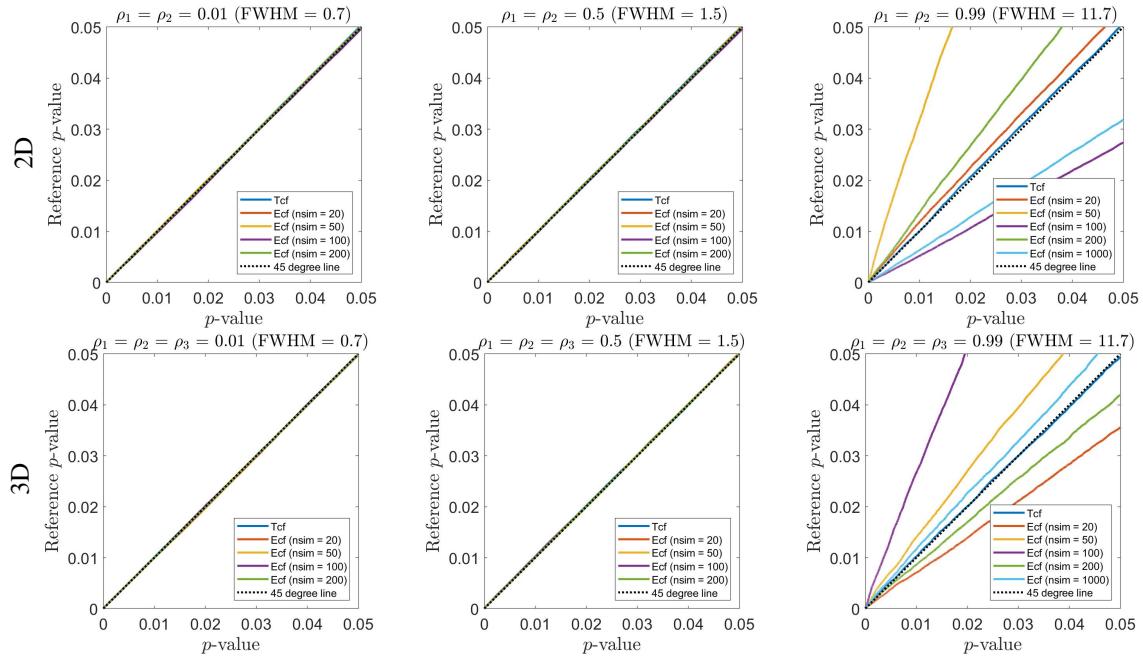


Figure 7: Comparison of the peak height distribution calculated from using MCDLM with different neighborhood covariance for 2D and 3D isotropic Gaussian fields. The covariance functions used here are theoretical covariance function (Tcf) and empirically estimated covariance function (Ecfn). The number of random fields used to estimate the covariance function is denoted using nsim. From left to right:  $\rho = 0.01, 0.5, 0.99$ .

## 5 HCP Task fMRI Results

We evaluate the methods on 80 unrelated subjects of the Human Connectome Project (HCP) by testing one of the working memory contrasts. Specifically, we focus on the average contrasts of 2-back and 0-back tasks (Davenport and Nichols, 2020). Since N-back task is commonly used to measure working memory, and N can be adjusted for task difficulty, such contrasts help to identify the brain regions supporting working memory (Kirchner, 1958).

We perform the peak inference by computing the one-sample  $t$ -statistics at voxel level and applying MCDLM for  $t$ -fields to calculate  $p$ -values for 1101 peaks that we detect. We use both MCDLM based on multivariate  $t$ -distribution and Gaussianization transformation method described in Section 2.4 for the calculation. The isotropy assumption is too strict in our study since we observe three different correlations, 1.30 (FWHM = 3.07), 1.39 (FWHM = 3.27) and 1.22 (FWHM = 2.88) in horizontal, vertical and longitudinal directions, respectively. More importantly, applying the continuous RFT method in this case will introduce bias due to the low smoothness of this data and the consequent inaccuracies in continuous approximation. Thus, the MCDLM with estimated covariance stands out as an ideal choice

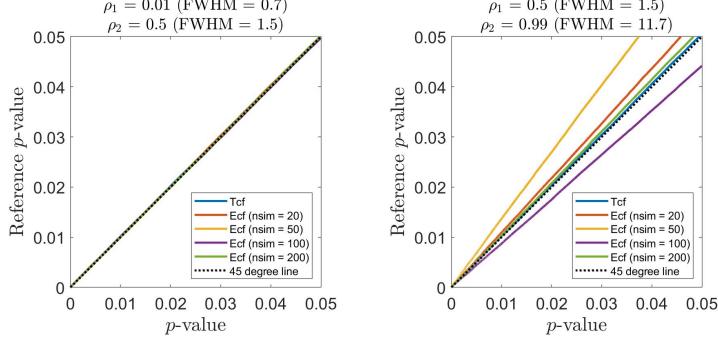


Figure 8: Comparison of the peak height distribution calculated from using MCDLM method with different covariance functions for 2D anisotropic stationary Gaussian fields. The left:  $\rho_1 = 0.01$  and  $\rho_2 = 0.5$  and the right:  $\rho_1 = 0.5$  and  $\rho_2 = 0.99$ .

for this real study. Additionally, the continuous RFT method has also been applied to approximate the discrete t-fields by continuous t-fields. The inference results from both approaches have been included in a comparative study.

The largest peak has a  $t$ -statistic value of 13.59. This peak is located within the medial prefrontal cortex, which is an area commonly associated with working memory (Euston et al., 2012; Perlstein et al., 2002). By MCDLM based on multivariate  $t$ -distribution, we calculate the  $p$ -value  $< 4.57 \times 10^{-7}$  and by Gaussianization transformation method, the  $p$ -value is  $< 9.42 \times 10^{-8}$ . Since MCDLM is a computation-based method, we do not have a closed-form function to calculate  $p$ -values for the whole real domain. When the  $t$ -statistic is extremely large, we are unable to calculate the exact  $p$ -value, and in such case we will conclude the  $p$ -value is smaller than the lower bound of a set of  $p$ -values based on the empirical distribution from our Monte Carlo experiments. Although the continuous RFT method provides the closed-form density function for 3D Gaussian fields, it still relies on numerical integration to calculate CDF and  $p$ -values. Therefore, in the extreme case, we will encounter a similar computational challenge unless we set the step size sufficiently small. The  $p$ -value for the largest peak calculated from the continuous method is  $5.33 \times 10^{-14}$ . Figure 9 shows slices in different directions through one-sample  $t$ -statistic corresponding to the largest peak.

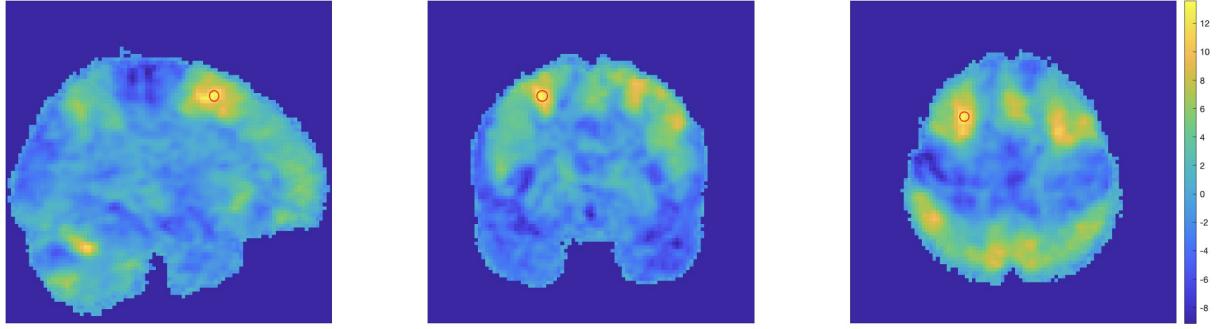


Figure 9: Slices in different directions through one-sample  $t$ -statistic corresponding to the largest peak from the average working memory contrasts from the Human Connectome Project. Red circles indicate the location of the largest peak.

The number of significant peaks ( $p$ -value  $< 0.05$ ) out of 1101 peaks from MCDLM based on multivariate  $t$ -distribution is 359, the MCDLM based on Gaussianization transformation method is 341, and the continuous RFT method is 325. We also implement the Benjamini-Hochberg procedure to control the false discovery rate (FDR) to an  $\alpha = 0.05$ . After such multiple comparisons adjustment, there are 293 significant peaks from MCDLM, 276 from MCDLM with Gaussianization, 270 from the continuous RFT method. Shown in Table 2 are 10 peaks that have  $p$ -values close to 0.05. MCDLM and continuous RFT method give different conclusions for all these 10 peaks, i.e., while all these 10 peaks are significant from MCDLM based on multivariate  $t$ -distribution, they are not significant from the continuous RFT method. To avoid making incorrect scientific conclusions, we recommend using the MCDLM method in this application

because (1) based on the FWHM we calculate, the continuous approximation might not be accurate ( $\text{FWHM} < 5.2$  in all directions), (2) even if the approximation is accurate, the continuous RFT method only works for Gaussian fields, so we still need to apply the Gaussianization for the  $t$ -fields, which might lead to biased inference when the degrees of freedom is small ( $\nu = 80$  in our case).

Table 2: The  $t$ -statistics and  $p$ -values from MCDLM based on multivariate  $t$ -distribution, MCDLM with Gaussianization and Continuous RFT method for 10 selected peaks that have  $p$ -values close to 0.05. The location indices are from the natural indexing within the 1 to 91, 1 to 109 and 1 to 91 image.

$t$ -statistics	Location	MCDLM	MCDLM w/ Gaussianization	Continuous RFT
3.151	(49,70,49)	0.033	0.045	0.060
3.146	(51,62,37)	0.033	0.045	0.061
3.144	(56,48,22)	0.033	0.045	0.061
3.130	(55,24,15)	0.034	0.047	0.063
3.124	(41,37,22)	0.035	0.047	0.063
3.108	(44,57,24)	0.036	0.049	0.066
3.104	(76,40,54)	0.036	0.049	0.066
3.088	(48,21,46)	0.038	0.051	0.069
3.081	(47,57,52)	0.038	0.052	0.070
3.069	(56,46,22)	0.039	0.053	0.071

## 6 Discussion

*[SD: This section needs work!!]* In this paper, we have proposed a new Monte Carlo method to calculate the distribution of the height of a peak of a discrete Gaussian random field which works under minimal assumptions. When inferring on the heights of the peaks of Gaussian fields, MCDLM performed well compared to other approaches. Historically, continuous RFT method has been used to calculate the distribution of the height of local maxima in a continuous random field. However, in practice we observe data on a lattice. As shown in Schwartzman and Telschow (2019) and the simulation studies in this paper, when the data is sufficiently smooth ( $\text{FWHM} \geq 7$  in Schwartzman and Telschow (2019) and  $\text{FWHM} > 5.2$  in this paper), the continuous formulae provide a good approximation to the height of local maxima, *but in many realistic situations ( $\text{FWHM} < 5.2$ ) the data is not sufficiently smooth and using the continuous formulae can lead to conservative inference. Furthermore, the continuous formulae only work for an isotropic field or a field that can be deformed to an isotropic field, which is a highly restrictive assumption and is not reasonable in practice. When applied to an isotropic  $t$ -field, continuous RFT method also relies heavily on convergence to Gaussianity, which is only valid when the degrees of freedom is large. Additionally, the peak height distribution will be different for points on the boundary of the domain which is difficult to account using continuous RFT but easy to do so using MCDLM by changing the neighbourhood structure.*

*We showed that continuous RFT was conservative while ADLM was liberal in a wide variety of settings. The conservativeness of continuous RFT is due to a failure of the good lattice assumption and is the same phenomenon observed in Telschow et al. (2023) and Davenport et al. (2023). Instead the anti-conservativeness of ADLM is due to the failure of its rather restrictive assumptions on the covariance and neighbourhood structure. MCDLM instead provided exact inference in most of the settings considered. We thus recommend using MCDLM to infer on peak height at low to medium levels of smoothness. However, when the data is very smooth and when it is reasonable to assume the data is isotropic, there may not be much gain relative to continuous RFT and since the latter is very efficient we recommend it in that setting as it provides a precise formula for the peak height distribution which can be quickly and accurately calculated. In this case the covariance matrix is nearly singular making it difficult to estimate, which causes problems when applying MCDLM.* A detailed running-time table for all three approaches, under different scenarios, is provided in Appendix E.

The proposed MCDLM method also works for  $t$ -fields, but it takes quite long time to implement this approach when the number of degrees of freedom is large. To improve the computational efficiency, we recommend using a Gaussianization transformation and then applying MCDLM to the Gaussianized field. In this setting continuous RFT works well when both smoothness and degrees of freedom are high, but even when degrees of freedom increases to 200, it is still outperformed by MCDLM. Our approach can also easily be extended to obtain the peak height distribution of two-sample  $t$ -statistic and  $F$ -fields observed on a lattice. The proposed method is limited to stationary Gaussian or Gaussian-derived random fields. However extensions to locally stationary and non-stationary fields are possible and are an interesting avenue for future research.

## References

- Adler, R. J. (1981). The geometry of random fields, vol. 62. *SIAM, Philadelphia*.
- Cheng, D. and Schwartzman, A. (2015a). Distribution of the height of local maxima of gaussian random fields. *Extremes*, 18(2):213–240.
- Cheng, D. and Schwartzman, A. (2015b). On the explicit height distribution and expected number of local maxima of isotropic gaussian random fields. *arXiv preprint arXiv:1503.01328*.
- Cheng, D. and Schwartzman, A. (2017). Multiple testing of local maxima for detection of peaks in random fields. *Annals of statistics*, 45(2):529.
- Cheng, D. and Schwartzman, A. (2020). On critical points of gaussian random fields under diffeomorphic transformations. *Statistics & Probability Letters*, 158:108672.
- Chumbley, J., Worsley, K., Flandin, G., and Friston, K. (2010). Topological FDR for neuroimaging. *Neuroimage*, 49(4):3057–3064.
- Chumbley, J. R. and Friston, K. J. (2009). False discovery rate revisited: FDR and topological inference using gaussian random fields. *Neuroimage*, 44(1):62–70.
- Davenport, S. and Nichols, T. E. (2020). Selective peak inference: Unbiased estimation of raw and standardized effect size at local maxima. *Neuroimage*, 209:116375.
- Davenport, S., Schwartzman, A., Nichols, T. E., and Telschow, F. J. (2023). Robust fwer control in neuroimaging using random field theory: Riding the surf to continuous land part 2. *arXiv preprint arXiv:2312.10849*.
- Eklund, A., Knutsson, H., and Nichols, T. E. (2019). Cluster failure revisited: Impact of first level design and physiological noise on cluster false positive rates. *Human brain mapping*, 40(7):2017–2032.
- Eklund, A., Nichols, T. E., and Knutsson, H. (2016). Cluster failure: Why fmri inferences for spatial extent have inflated false-positive rates. *Proceedings of the national academy of sciences*, 113(28):7900–7905.
- Euston, D. R., Gruber, A. J., and McNaughton, B. L. (2012). The role of medial prefrontal cortex in memory and decision making. *Neuron*, 76(6):1057–1070.
- Friston, K., Passingham, R., Nutt, J., Heather, J., Sawle, G., and Frackowiak, R. (1989). Localisation in PET images: direct fitting of the intercommissural (AC–PC) line. *Journal of Cerebral Blood Flow & Metabolism*, 9(5):690–695.
- Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of experimental psychology*, 55(4):352.
- Nosko, V. (1969). Local structure of gaussian random fields in vicinity of high-level shines. *Doklady Akademii Nauk SSSR*, 189(4):714.
- Perlstein, W. M., Elbert, T., and Stenger, V. A. (2002). Dissociation in human prefrontal cortex of affective influences on working memory-related activity. *Proceedings of the National Academy of Sciences*, 99(3):1736–1741.
- Schwartzman, A., Gavrilov, Y., and Adler, R. J. (2011). Multiple testing of local maxima for detection of peaks in 1D. *Annals of statistics*, 39(6):3290.
- Schwartzman, A. and Telschow, F. (2019). Peak p-values and false discovery rate inference in neuroimaging. *NeuroImage*, 197:402–413.
- Taylor, J. E., Worsley, K., and Gosselin, F. (2007). Maxima of discretely sampled random fields, with an application to ‘bubbles’. *Biometrika*, 94(1):1–18.
- Telschow, F., Davenport, S., and Schwartzman, A. (2023). Riding the surf to continuous land: precise fwer control for gaussian random fields. *Preprint*.
- Worsley, K. J. (2005). An improved theoretical P value for SPMs based on discrete local maxima. *Neuroimage*, 28(4):1056–1062.
- Worsley, K. J., Evans, A. C., Marrett, S., and Neelin, P. (1992). A three-dimensional statistical analysis for CBF activation studies in human brain. *Journal of Cerebral Blood Flow & Metabolism*, 12(6):900–918.
- Worsley, K. J., Marrett, S., Neelin, P., Vandal, A. C., Friston, K. J., and Evans, A. C. (1996). A unified statistical approach for determining significant signals in images of cerebral activation. *Human brain mapping*, 4(1):58–73.

## A Theory for ADLM

### A.1 Proof of Proposition 1

*Proof.* Without loss of generality, we assume that  $Z$  is mean zero and unit variance. For  $s \in S$ , taking  $\mathcal{N}(s)$  to be the partially connected neighborhood of  $s$ , for  $s_a, s_b \in \mathcal{N}(s)$ , we have

$$\begin{pmatrix} Z(s) \\ Z(s_a) \\ Z(s_b) \end{pmatrix} \sim N \left( \mathbf{0}, \begin{pmatrix} \Sigma_{AA} & \Sigma_{AB} \\ \Sigma_{BA} & \Sigma_{BB} \end{pmatrix} \right),$$

where  $\Sigma_{AA} = \text{Var}(Z(s)) = 1$ ,  $\Sigma_{AB} = \Sigma_{BA}^\top = [\text{Cov}(Z(s), Z(s_a)) \text{Cov}(Z(s), Z(s_b))] = [\rho(s, s_a) \rho(s, s_b)]$  and  $\Sigma_{BB} = \text{Var} \left[ \begin{pmatrix} Z(s_a) \\ Z(s_b) \end{pmatrix} \right] = \begin{pmatrix} 1 & \rho(s_a, s_b) \\ \rho(s_a, s_b) & 1 \end{pmatrix}$ .

For  $z \in \mathbb{R}$ , the covariance of  $Z(s_a)$  and  $Z(s_b)$  conditional on  $Z(s) = z$  is

$$\begin{aligned} \text{Var} \left[ \begin{pmatrix} Z(s_a) \\ Z(s_b) \end{pmatrix} \middle| Z(s) = z \right] &= \Sigma_{BB} - \Sigma_{BA}\Sigma_{AA}^{-1}\Sigma_{AB} \\ &= \Sigma_{BB} - \Sigma_{BA}\Sigma_{AB} \\ &= \begin{pmatrix} 1 & \rho(s_a, s_b) \\ \rho(s_a, s_b) & 1 \end{pmatrix} - \begin{pmatrix} \rho(s, s_a)^2 & \rho(s, s_a)\rho(s, s_b) \\ \rho(s, s_a)\rho(s, s_b) & \rho(s, s_b)^2 \end{pmatrix}. \end{aligned}$$

Under the assumption that  $Z$  is obtained by integrating Gaussian white noise against a separable Gaussian kernel, the off-diagonal entries are

$$\rho(s_a, s_b) - \rho(s, s_a)\rho(s, s_b) = \rho^{\|s_a - s_b\|^2} - \rho^{\|s_a - s\|^2}\rho^{\|s_b - s\|^2} = \rho^{\|s_a - s_b\|^2} - \rho^{\|s_a - s\|^2 + \|s_b - s\|^2},$$

where  $\rho$  is the correlation between two adjacent voxels and  $\|\cdot\|$  denotes the Euclidean norm. Thus, if  $\|s_a - s_b\|^2 = \|s_a - s\|^2 + \|s_b - s\|^2$ , then  $\text{Cov}(Z(s_a), Z(s_b)|Z(s)) = 0$ . Now  $(Z(s_a), Z(s_b))$  follows a bivariate Gaussian distribution so it follows that  $Z(s_a)$  and  $Z(s_b)$  is independent conditional on  $Z(s_5)$ . When  $s_a = s \pm e_{d_1}$  and  $s_b = s \pm e_{d_2}$ , where  $d_1$  and  $d_2$  denote different lattice directions,  $\|s_a - s_b\|^2 = \|e_{d_1} \pm e_{d_2}\|^2$  which equals  $\|s_a - s\|^2 + \|s_b - s\|^2 = \|e_{d_1}\|^2 + \|e_{d_2}\|^2$  and so the result follows.  $\square$

### A.2 Theoretical derivation of the probability density function of ADLM method

**Proposition 2.** Under the condition of 1. For each  $s \in S$  assume that  $\mathcal{N}_{PC}(s)$  is the partially connected neighbourhood defined in (2). Then

$$P[Z(s) > u | Z(t) < Z(s), \forall t \in \mathcal{N}_{PC}(s)] = \frac{\int_u^\infty \left( \prod_{d=1}^D Q(\rho_d, z)(z) \right) \phi(z) dz}{\int_{-\infty}^\infty \left( \prod_{d=1}^D Q(\rho_d, z)(z) \right) \phi(z) dz}$$

*Proof.* First, by the law of iterated expectations,

$$\begin{aligned} P \left[ \{Z(s) > u\} \cap_{t \in \mathcal{N}_{PC}(s)} \{Z(t) < Z(s)\} \right] &= \int_{-\infty}^\infty P \left[ \{z > u\} \cap \{Z(t) < z, \forall t \in \mathcal{N}_{PC}(s) | Z(s) = z\} \right] \phi(z) dz \\ &= \int_{-\infty}^\infty \mathbb{1}_{\{z > u\}} \cdot P \left[ \{Z(t) < z, \forall t \in \mathcal{N}_{PC}(s) | Z(s) = z\} \right] \phi(z) dz \\ &= \int_u^\infty P \left[ \{Z(t) < z, \forall t \in \mathcal{N}_{PC}(s) | Z(s) = z\} \right] \phi(z) dz. \end{aligned}$$

Next, by applying Proposition 1,

$$P \left[ \{Z(t) < z, \forall t \in \mathcal{N}_{PC}(s) | Z(s) = z\} \right] = P \left[ \bigcap_{d=1}^D (Z(t) < z, \forall t = s \pm v_d e_d | Z(s) = z) \right] = \prod_{d=1}^D Q(\rho_d, z),$$

where

$$Q(\rho_d, z) = P\{Z(t) < z, \forall t = s \pm v_d e_d | Z(s) = z\}.$$

Thus we have

$$P\left[\{Z(s) > u\} \cap_{t \in \mathcal{N}_{PC}(s)} \{Z(t) < Z(s)\}\right] = \int_u^\infty \left( \prod_{d=1}^D Q(\rho_d, z) \right) \phi(z) dz.$$

By Bayes Rule,

$$\begin{aligned} P[Z(s) > u | Z(t) < Z(s), \forall t \in \mathcal{N}_{PC}(s)] &= \frac{P[\{Z(s) > u\} \cap \{Z(t) < Z(s), \forall t \in \mathcal{N}_{PC}(s)\}]}{P[Z(t) < Z(s), \forall t \in \mathcal{N}_{PC}(s)]} \\ &= \frac{\int_u^\infty P[(Z(t) < z, \forall t \in \mathcal{N}_{PC}(s) | Z(x) = z)] \phi(z) dz}{\int_{-\infty}^\infty P[(Z(t) < z, \forall t \in \mathcal{N}_{PC}(s) | Z(x) = z)] \phi(z) dz} \\ &= \frac{\int_u^\infty \left( \prod_{d=1}^D Q(\rho_d, z) \right) \phi(z) dz}{\int_{-\infty}^\infty \left( \prod_{d=1}^D Q(\rho_d, z) \right) \phi(z) dz}. \end{aligned}$$

□

Thus, the peak height density is

$$f_{\text{DLM}} = \frac{\prod_{d=1}^D Q(\rho_d, z) \phi(z)}{\int_{-\infty}^\infty \left( \prod_{d=1}^D Q(\rho_d, z) \right) \phi(z) dz}.$$

Follow the proof from Taylor et al. (2007),

$$Q(\rho_d, z) = 1 - 2\Phi(h_d z^+) + \frac{1}{\pi} \int_0^{\alpha_d} \exp(-\frac{1}{2} h_d^2 z^2 / \sin^2 \theta) d\theta.$$

where

$$h_d = \sqrt{\frac{1 - \rho_d}{1 + \rho_d}}, \alpha_d = \sin^{-1} \left( \sqrt{(1 - \rho_d^2)/2} \right), z^+ = \max(z, 0).$$

## B The neighbourhood covariance matrix in the fully connected setting

### B.1 Theoretical derivation of the neighbourhood covariance for the integral convolution field

To prove (10), it suffices to establish following claim:

**Theorem 3.** Let  $Z$  be a  $D$  dimensional isotropic Gaussian random field on a discrete lattice  $S$  with mean zero and unit variance. Assume that  $Z$  is derived by integrating continuous white noise against a Gaussian kernel. Let  $s_0, t_0 \in S$ , we first define a special indexing of  $\mathcal{N}_{FC}(s_0)$  and  $\mathcal{N}_{FC}(t_0)$ , the elements of the fully connected neighbourhood of  $s_0$  and  $t_0$  defined in (3). We index the neighbours in the order of 0, 1, 2 in each direction. A 2D example of  $\mathcal{N}_{FC}(s_0)$  for  $s_0 = (1, 1)$  is given as

$$\begin{matrix} (0, 0) & (0, 1) & (0, 2) \\ (1, 0) & (1, 1) & (1, 2) \\ (2, 0) & (2, 1) & (2, 2). \end{matrix}$$

Let  $s = (s_1, \dots, s_D)^\top$  and  $t = (t_1, \dots, t_D)^\top$  be the vector that is one element inside of  $\mathcal{N}_{FC}(s_0)$  and  $\mathcal{N}_{FC}(t_0)$ . Then

$$\text{Cov}(Z(s), Z(t)) = [A \otimes A \otimes \dots \otimes A]_{m,n}, \quad (16)$$

where

$$A = \begin{pmatrix} 1 & \rho & \rho^4 \\ \rho & 1 & \rho \\ \rho^4 & \rho & 1 \end{pmatrix},$$

$\rho$  is the correlation between adjacent voxels, and  $m = \sum_{i=1}^D 3^{i-1} s_i + 1$  and  $n = \sum_{i=1}^D 3^{i-1} t_i + 1$ .

*Proof.* Because of the form of  $Z$ ,  $\text{Cov}(Z(s), Z(t)) = \rho^{\|s-t\|^2}$ . We have that  $[A \otimes A \otimes \dots \otimes A]_{m,n} = \prod_{i=1}^D A_{s_i t_i}$ . If  $\|s_i - t_i\| = 0$ ,  $A_{s_i t_i} = 1 = \rho^0$ ; if  $\|s_i - t_i\| = 1$ ,  $A_{s_i t_i} = 1 = \rho$  and if  $\|s_i - t_i\| = 2$ ,  $A_{s_i t_i} = \rho^4$ . Thus,

$$A_{s_i t_i} = \rho^{\|s_i - t_i\|^2},$$

and hence

$$\prod_{i=1}^D A_{s_i t_i} = \rho^{\sum_{i=1}^D \|s_i - t_i\|^2} = \rho^{\|s-t\|^2}.$$

As such both sides of (??) match and the result follows.

This proof relies on a special indexing of  $\mathbf{Z}$ . However, this is in practice not a restriction as a re-indexing of  $\mathbf{Z}$  can be treated as a linear transformation, i.e.  $\mathbf{Z} = T\mathbf{Z}_o$ , where  $\mathbf{Z}_o$  is the vector with the special indexing. In that case  $\Sigma$  then can be calculated by

$$\Sigma = T\text{Cov}(\mathbf{Z}_o)T^\top.$$

□

## B.2 Neighbourhood covariance matrix for a 3D stationary Gaussian random field

Figure 10 shows the theoretical covariance function for a 3D stationary field. The construction and indexing of this  $27 \times 27$  matrix follow the logic from Section B.1. The numbers  $r0, r1, \dots, r61$  denote all 62 distinct values of covariance between two voxels. We use the same color to denote all  $3 \times 3$  matrices with the same values. One useful conclusion from this figure is this covariance matrix is a block Toeplitz matrix with 9 blocks, and each block is still a block Toeplitz matrix with 9 sub-blocks. In addition, each sub-block is a Toeplitz matrix.

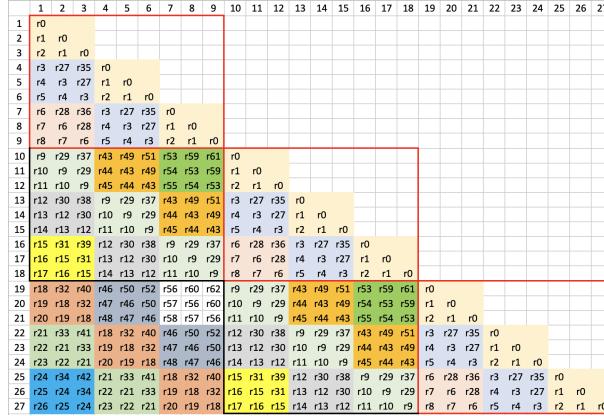


Figure 10: The theoretical covariance function for a 3D stationary field.

## C Introduction to the $pp$ plot

The  $pp$  plots are used throughout Section 3 and 4. In this section, we will formally define this  $pp$  plot.  $pp$  plots are used to compare the peak  $p$ -values computed using the ADLM, MCDLM and continuous methods we discussed in Section 2 to the reference tail probability of the true peak height distribution, i.e. the reference  $p$ -value. We obtain the latter from simulation. To do so, first we generate  $N$  i.i.d. random fields as described in each setting in Section 3. Let  $n$  be the number of obtained local maxima across all fields, where the local maxima are selected based on the criteria that their height values are larger than all their neighbours in the specified neighbourhood (we will consider both the fully connected and partially connected neighbourhoods). Let  $g_1, \dots, g_n$  be the heights of the recorded local maxima. For each peak this allows us to compute an estimation of the reference  $p$ -value as

$$p_i = \frac{1}{n} \sum_{j=1}^n \mathbb{1}[g_j > g_i], \quad 1 \leq i \leq n, \tag{17}$$

where  $\mathbb{1}[\cdot]$  denotes the indicator function,  $p_i$  is the  $p$ -value when observed value is  $g_i$ . As  $n \rightarrow \infty$ ,  $p_i$  calculated by (17) converges to the true tail probability. Moreover for each peak, we calculate a  $p$ -value for each of the three approaches. Next we plot the reference  $p$ -values against the  $p$ -values obtained using each method. Since the reference distribution converges to the true peak height distribution as the number of instances converges to infinity, the closer these plots are to the identity function, the closer the approximation to the true distribution. We use these  $pp$  plots to compare the performance of the three approaches in all of our simulation studies.

The idea of this  $pp$  plot is similar to the one used in Schwartzman and Telschow (2019). Although the two plotting mechanisms look different, the logic behind is the same, as we justify below.

Our two plotting mechanisms are

- Plot  $p(i)$  vs.  $i/n$  and  $q(i)$  vs.  $i/n$  (Schwartzman and Telschow (2019)),
- Plot  $p(i)$  vs.  $p(i)$  and  $q(i)$  vs.  $p(i)$  ( $pp$  plot).

Let  $F(z)$  be true cdf of  $z$ ,  $G(z)$  be one of the other cdfs of  $z$  used for comparison purpose. Next, we define

$$\begin{aligned} p &= 1 - F(z) \\ q &= 1 - G(z) \end{aligned}$$

Suppose that we now generate  $n$   $p$ -values for both the true distribution and the distribution for comparison purpose, i.e., we generate  $p_1, \dots, p_n$  and  $q_1, \dots, q_n$  as in (17). Denote  $p(i)$  and  $q(i)$  as the order statistics of  $p_1, \dots, p_n$  and  $q_1, \dots, q_n$ . Then under monotonicity,

$$\begin{aligned} \text{ecdf}(p(i)) &= \frac{1}{n} \sum_{i=1}^n I[p_i \leq p(i)] = \frac{i}{n} \\ \text{ecdf}(q(i)) &= \frac{1}{n} \sum_{i=1}^n I[q_i \leq q(i)] = \frac{i}{n} \end{aligned}$$

If  $z$  is distributed according to  $F$ ,  $p$ -values are uniformly distributed on  $[0, 1]$ . Thus, by the LLN,

$$\text{ecdf}(p) = \frac{1}{n} \sum_{i=1}^n I(p_i \leq p) \xrightarrow{P} P(p_i \leq p) = p$$

Thus, when  $p$  is from the true distribution and  $n$  is large enough, we expect the two plotting mechanisms provide similar plots. The comparison of two method is shown below in Figure 11.

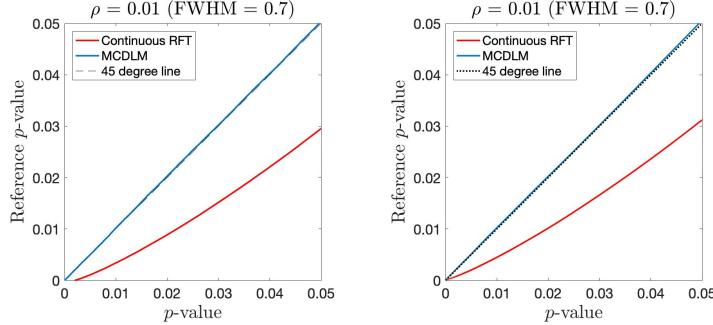


Figure 11: Comparison of two plotting mechanism using the height distribution generated from 2D isotropic Gaussian field with spatial correlation  $\rho = 0.01$ . Left is from the plotting mechanism in Schwartzman and Telschow (2019) and right is from the plotting mechanism in this paper

## D Additional results

### D.1 Partial connectivity case

In Section 2 we discussed some of the disadvantages of ADLM. In particular local separability (i.e. the result of Proposition 1) does not apply to diagonal neighbours and so ADLM cannot be used for peaks of the fully connected

neighbourhood. However (when its assumptions hold) it can be used to infer on the peak height of peaks in the partially connected neighbourhood defined in ((5)). We illustrate this here with in the same simulation setting of Section 3.1.

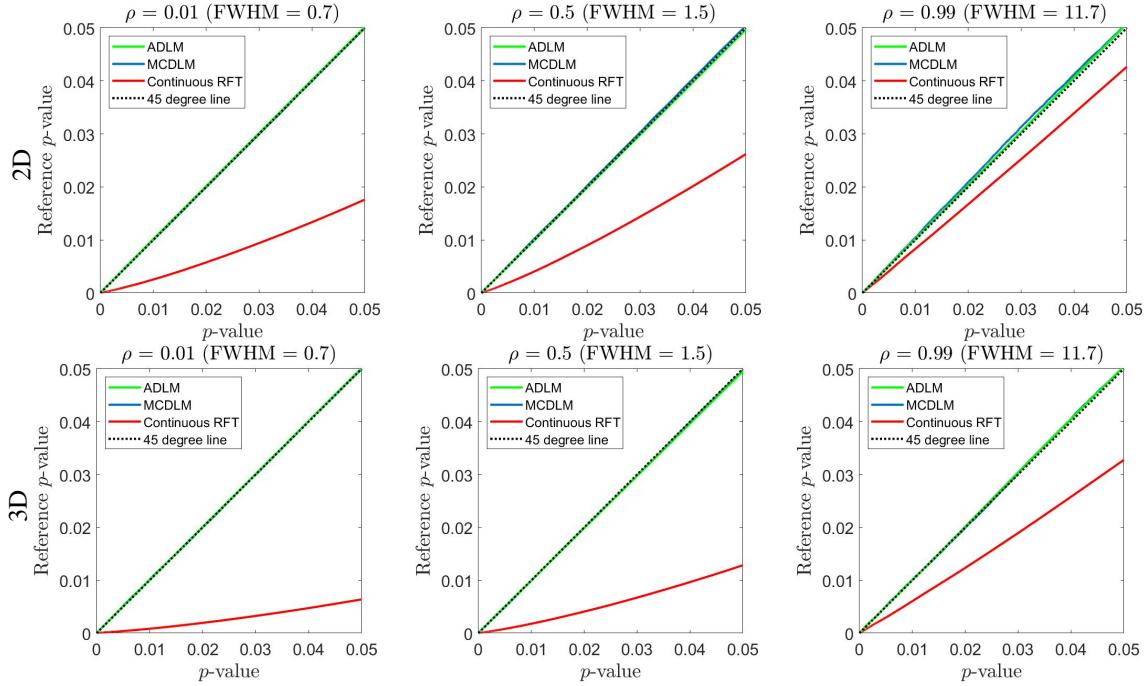


Figure 12: Comparison of the peak height distribution for peaks in a partially connected neighbourhood calculated via the different methods for 2D and 3D isotropic Gaussian fields.

Figure 12 shows the comparison between MCDLM, ADLM and the continuous RFT method in 2D and 3D. In all three scenarios, the  $p$ -value distribution of ADLM and MCDLM match and are close to a uniform distribution. As in the main text the continuous RFT approach is conservative.

## D.2 Applying the neighborhood covariance function in (7)

In this section we perform the same simulations as in Section 3.1 but with the neighbourhood covariance of Worsley (2005) (given in (7)) instead of the actual neighbourhood covariance (which we derived in (9)).

The results are shown in Figure 13. They are similar to those of Figure 4. However, as exemplified in the  $\rho = 0.5$  case the MCDLM approach is incorrect when this covariance function is used. This is because it is not in fact the correct neighbourhood covariance.

## D.3 Applying the look-up table

Since our method is Monte Carlo based method, it is desirable to reduce the computation time in certain settings. We list the computation time under different scenarios in Appendix E. In order to cut the time of running the simulation each time for a different height threshold  $u$  and spatial correlation  $\rho$ , it is possible to use a look-up table to pre-record the simulation results for  $10^5$  possible local maxima height values at different values of  $\rho$  when the Gaussian random field is isotropic. To do so we vary  $\rho$  from 0.01 to 0.99 with increments of 0.01. In order to calculate a look up table we do the following.

1. Loop through the array of different values of  $\rho$  and obtain  $10^5$  local maxima for each  $\rho$ .
2. From the union obtained in step 1, sample  $10^5$  local maxima, which consist a set of local maxima,  $u$ , that we want to evaluate the CDF,  $F(\cdot)$ , at.
3. Loop through the array of different values of  $\rho$  again. For each value of  $\rho$ , interpolate  $F(\cdot)$  at each of  $u$  we get in step 2. Record all the  $F(u; \rho)$  in a matrix of look-up table with row the  $\rho$  and column the  $u$ .

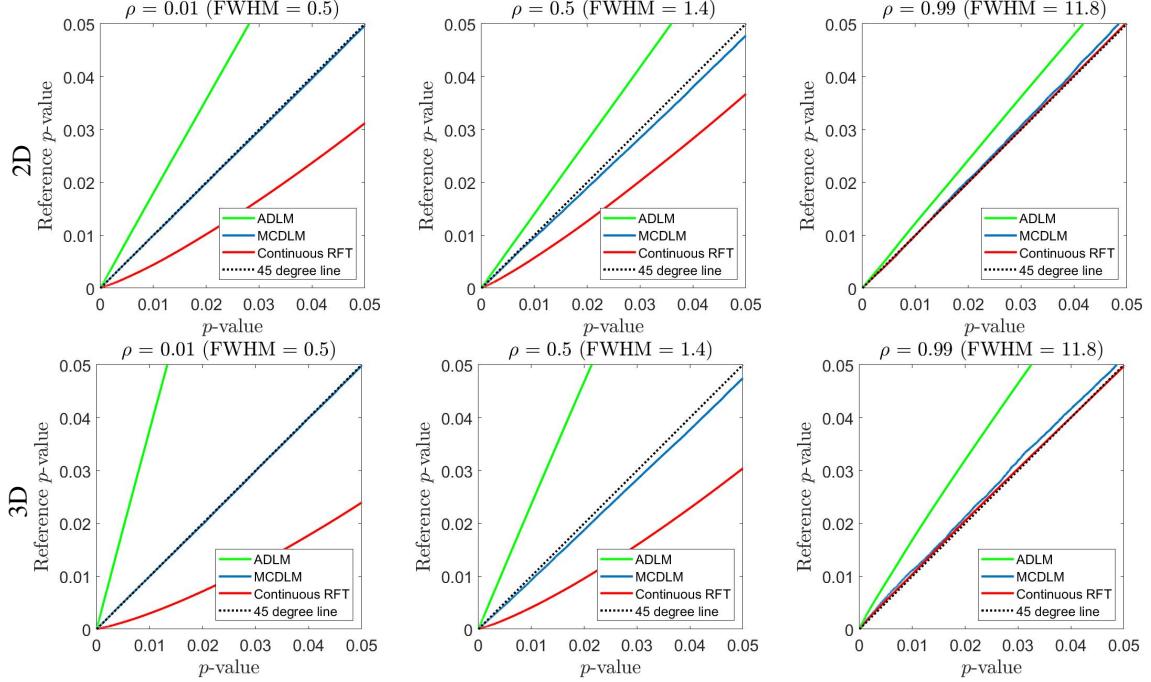


Figure 13: Comparison of peak height distribution calculated from different methods for 2D and 3D isotropic Gaussian field using neighborhood covariance function in (7).

To evaluate the  $p$ -value at a given threshold  $u$  and correlation  $\rho$ , we interpolate  $F(\cdot)$  at  $u$  and  $\rho$  through pre-recorded look-up table, and the  $p$ -value is calculated by  $1 - F(u; \rho)$ .

After generating a look-up table, we apply the cubic spline smoothing to smooth the noisy look-up table. The procedure of our smoothing is as follow:

1. Use the Cubic spline smoothing to smooth the matrix across  $\rho$ ;
2. Use the Cubic spline smoothing to smooth the matrix we smoothed in step 1 across  $u$ .

In doing so, we aim to reduce the violation of monotonicity across  $u$  but also retain the smoothness. The smoothing parameters in Cubic spline smoothing is selected by 5-fold cross validation in each scenario separately.

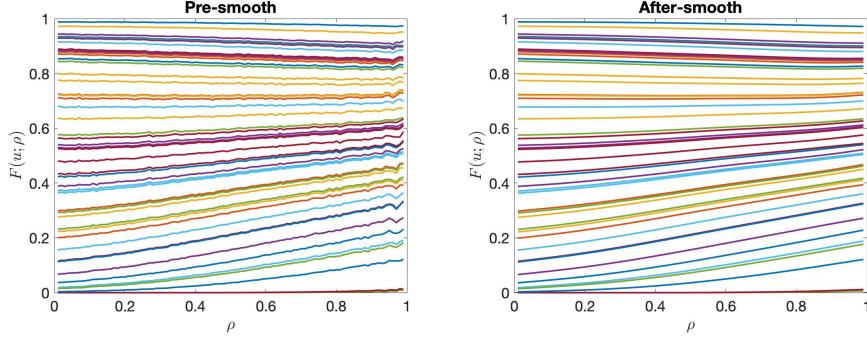


Figure 14: This figure shows  $F(u; \rho)$  of selected 50 samples across  $\rho$  from pre-smooth table (left) and after-smooth table (right). Same color is used for the same sample before and after-smoothing.

In Figure 14, we select 50 columns of both pre-smooth and after-smooth look-up table generated from 3D Gaussian random field and plot the CDF across different  $\rho$  values. Different samples selected within one look-up table are denoted by different colors, while same color means same sample between two tables. The pre-smooth plot shows that the look-up table we generated is very noisy across  $\rho$ , especially at some large values of  $\rho$ . This will cause big problems in

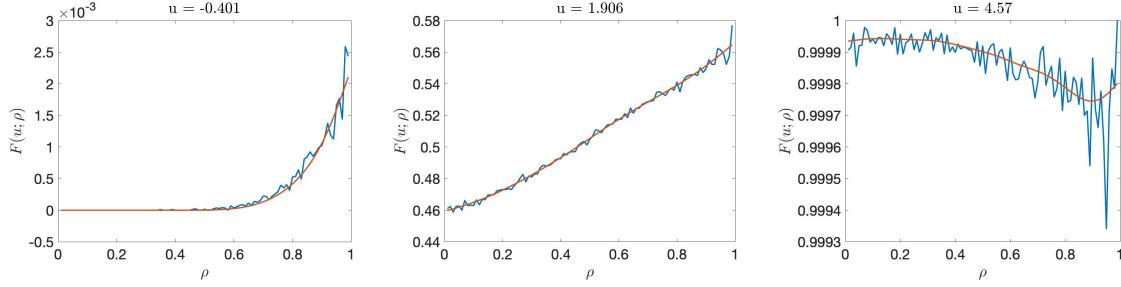


Figure 15: From left to right are the 10, 50000 and 99990 column of the pre-smooth table (red) and after-smooth table (blue).

computing  $p$ -values for different values of  $\rho$ . For example, if two are interested in calculating the  $p$ -values for a  $\rho$  close to 1, a slightly difference in picking the  $\rho$  will bring a significant change in  $p$ -values, and such ambiguity could finally result in an inconsistent interpretation of the scientific findings. From the after-smooth plot, we observe that the look-up table is smooth enough to provide consistent results.

In addition, we explore to what extent the smoothing works in figure 15. In this figure, we select three columns of both pre-smooth table and after-smooth table and then compare. The  $u$  that we select are -0.401, 1.906 and 4.57, which are in both ends and middle of the support. From these three plots, the noise before the smoothing is obvious, but the cubic spline smoothing fits the curve perfectly, in the sense that it preserves the general shape of the curve yet removes the noise.

#### D.4 Applying Gaussianization transformation of the $t$ -fields

As discussed in Section 3.2, to improve the computation efficiency for the height distribution of peaks of  $t$ -fields, we consider using Gaussianization transformation of the  $t$ -fields. In this section we perform the same simulations as in Section 3.2 but the simulated  $t$ -fields were Gaussianized by (14) and compared with the MCDLM for Gaussian field and continuous RFT approach.

The results are shown in Figure 16 (2D) and Figure 17 (3D). MCDLM works well when degrees of freedom is large and  $\rho$  is small. At high smoothness, MCDLM and continuous RFT work similarly. They are only correct when degrees of freedom is large since in this case the  $t$ -fields can approximate Gaussian fields.

#### D.5 Additional simulation results for MCDLM based on estimated neighborhood covariance

In Section 4.3 we discuss that the performance of MCDLM is worsen as  $\rho$  increases to 0.99. Since we observe when  $\rho = 0.5$  both 2D and 3D have good performance, we consider exploring 3 more scenarios in between,  $\rho = 0.9, 0.93, 0.95$ . In Figure 18, when  $\rho < 0.95$ , or  $\text{FWHM} < 5.2$ , MCDLM with estimated covariance works well for both 2D and 3D cases with enough number of random fields to estimate the covariance function ( $\text{nsim} \geq 50$ ). By comparing the neighborhood covariance matrices (18) (theoretical covariance) and (19) (estimated covariance) for  $\rho = 0.99$  in 2D case, the estimation is very close to the true covariance. However, the covariance matrix in this case is nearly singular, which causes problems when applied to our MCDLM method.

$$\begin{bmatrix} 1.0000 & 0.9900 & 0.9606 & 0.9900 & 0.9801 & 0.9510 & 0.9606 & 0.9510 & 0.9227 \\ 0.9900 & 1.0000 & 0.9900 & 0.9801 & 0.9900 & 0.9801 & 0.9510 & 0.9606 & 0.9510 \\ 0.9606 & 0.9900 & 1.0000 & 0.9510 & 0.9801 & 0.9900 & 0.9227 & 0.9510 & 0.9606 \\ 0.9900 & 0.9801 & 0.9510 & 1.0000 & 0.9900 & 0.9606 & 0.9900 & 0.9801 & 0.9510 \\ 0.9801 & 0.9900 & 0.9801 & 0.9900 & 1.0000 & 0.9900 & 0.9801 & 0.9900 & 0.9801 \\ 0.9510 & 0.9801 & 0.9900 & 0.9606 & 0.9900 & 1.0000 & 0.9510 & 0.9801 & 0.9900 \\ 0.9606 & 0.9510 & 0.9227 & 0.9900 & 0.9801 & 0.9510 & 1.0000 & 0.9900 & 0.9606 \\ 0.9510 & 0.9606 & 0.9510 & 0.9801 & 0.9900 & 0.9801 & 0.9900 & 1.0000 & 0.9900 \\ 0.9227 & 0.9510 & 0.9606 & 0.9510 & 0.9801 & 0.9900 & 0.9606 & 0.9900 & 1.0000 \end{bmatrix} \quad (18)$$

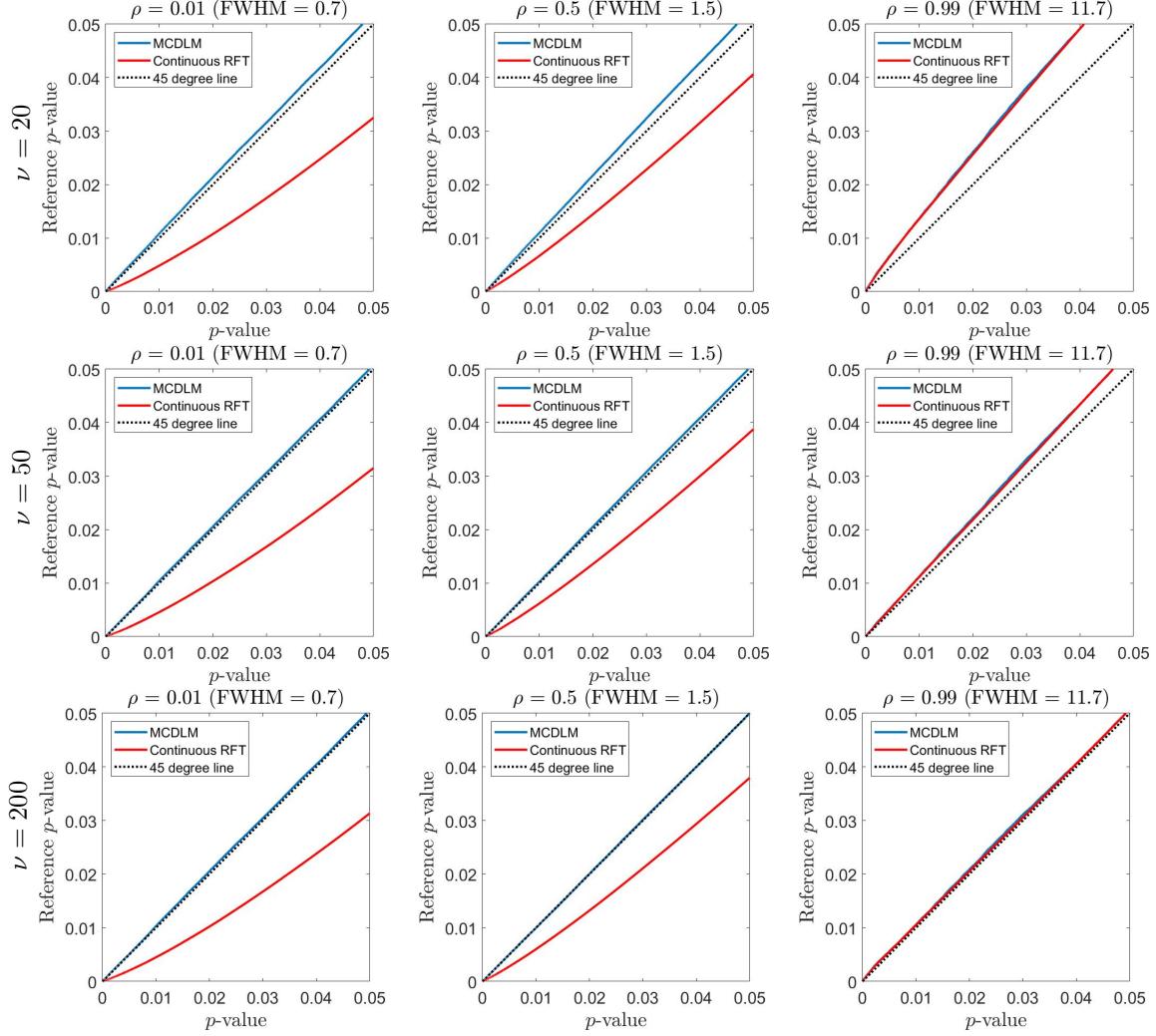


Figure 16: Comparison methods for calculating the peak height distribution of a Gaussianized 2D  $t$ -field with  $\nu$  degrees of freedom.

$$\begin{bmatrix} 1.0000 & 0.9906 & 0.9612 & 0.9903 & 0.9809 & 0.9516 & 0.9606 & 0.9513 & 0.9228 \\ 0.9906 & 1.0000 & 0.9906 & 0.9811 & 0.9903 & 0.9809 & 0.9518 & 0.9606 & 0.9513 \\ 0.9612 & 0.9906 & 1.0000 & 0.9521 & 0.9811 & 0.9903 & 0.9237 & 0.9518 & 0.9606 \\ 0.9903 & 0.9811 & 0.9521 & 1.0000 & 0.9906 & 0.9612 & 0.9903 & 0.9809 & 0.9516 \\ 0.9809 & 0.9903 & 0.9811 & 0.9906 & 1.0000 & 0.9906 & 0.9811 & 0.9903 & 0.9809 \\ 0.9516 & 0.9809 & 0.9903 & 0.9612 & 0.9906 & 1.0000 & 0.9521 & 0.9811 & 0.9903 \\ 0.9606 & 0.9518 & 0.9237 & 0.9903 & 0.9811 & 0.9521 & 1.0000 & 0.9906 & 0.9612 \\ 0.9513 & 0.9606 & 0.9518 & 0.9809 & 0.9903 & 0.9811 & 0.9906 & 1.0000 & 0.9906 \\ 0.9228 & 0.9513 & 0.9606 & 0.9516 & 0.9809 & 0.9903 & 0.9612 & 0.9906 & 1.0000 \end{bmatrix} \quad (19)$$

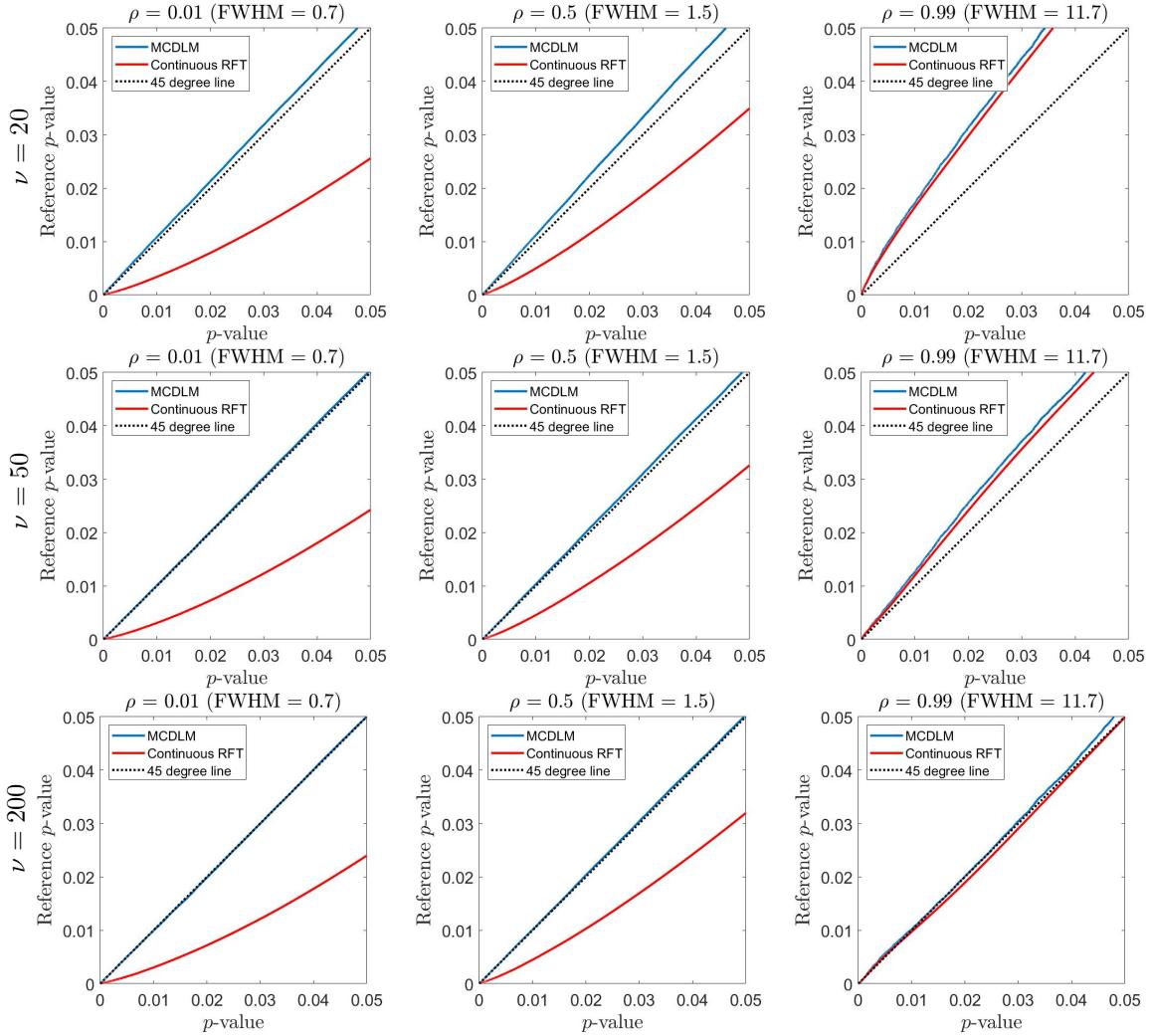


Figure 17: Comparison methods for calculating the peak height distribution of a Gaussianized 3D  $t$ -field with  $\nu$  degrees of freedom.

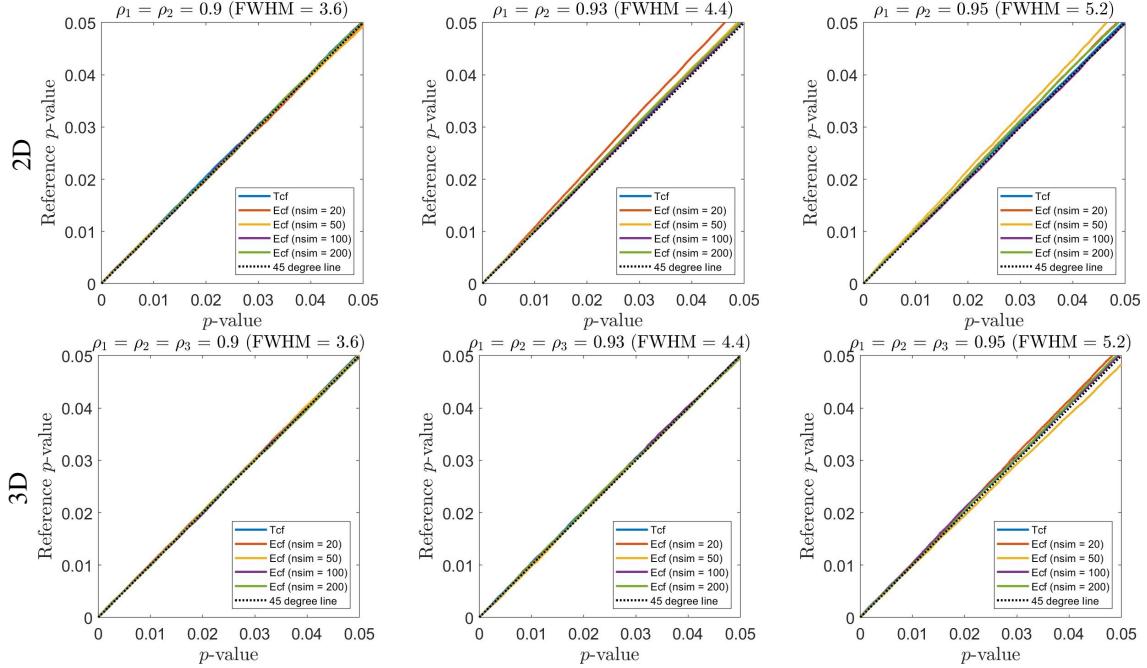


Figure 18: Comparison of the peak height distribution calculated from using MCDLM with different neighborhood covariance for 2D and 3D isotropic Gaussian fields. The covariance functions used here are theoretical covariance function (Tcf) and empirically estimated covariance function (Ecf). The number of random fields used to estimate the covariance function is denoted using nsim. From left to right:  $\rho = 0.9, 0.93, 0.95$ .

## E Computation time table

The simulation times under different scenarios are shown in Table 3. In the case of 2D Gaussian fields, applying MCDLM with full or partial connectivity is comparable in terms of running time, but the look up table is much faster (between 5 and 50 times faster), getting more efficient as the field correlation increases. In the case of 3D Gaussian fields, running times are 5 to 10 times larger than in 2D. However, the look up table method is just as fast as in 2D, making its use very worthwhile from a computational standpoint. In the case of 2D  $t$ -field, the running times increase about 10 to 15 times from degrees of freedom equals 20 to 200. In the case of 3D  $t$ -field, the running times are 5 to 10 times larger than in 2D, making the code stop at a pre-specified threshold when degrees of freedom increases to 200. The MCDLM with empirical covariance function has similar running times to MCDLM with theoretical covariance function, in the case of 2D Gaussian field with the number of fields to estimate the covariance small (200 or 1000). In 3D case, when the number of fields to estimate the covariance is 200, the running times are 2 times larger than the case of applying the theoretical covariance function. When the number of fields increases to 1000, the time to estimate the covariance further increases, leading to the running times 4-8 times larger than the case of applying the theoretical covariance function.

Table 3: Running time (in seconds) of our MCDLM method under different scenarios. For  $\rho = 0.01$  and  $\rho = 0.05$ , the code is targeted to obtain at least  $n = 1e6$  peak height values and for  $\rho = 0.99$ ,  $n = 2e5$  peak height values. In some extreme cases, the code stops at a pre-specified threshold with the number of instances generated recorded in parentheses.

	$\rho = 0.01 (n = 1e6)$	$\rho = 0.5 (n = 1e6)$	$\rho = 0.99 (n = 2e5)$
2D Gaussian field			
Full connectivity (continuous covariance function)	9.81	13.83	106.41
Full connectivity (discrete covariance function)	9.95	14.76	112.24
Partial connectivity (discrete covariance function)	7.43	12.15	110.86
Full connectivity (look up table)	1.67	1.77	1.88
2D $t$ -field			
$\nu = 20$	74.48	105.74	835.80
$\nu = 50$	217.63	307.56	1992.58
$\nu = 200$	1064.24	1356.55	1646.86 ( $n = 37289$ )
3D Gaussian field			
Full connectivity (continuous covariance function)	41.79	64.16	1395.14 ( $n = 1e5$ )
Full connectivity (discrete covariance function)	41.79	66.32	1377.69 ( $n = 1e5$ )
Partial connectivity (discrete covariance function)	13.87	29.18	1753.71
Full connectivity (look up table)	1.56	2.04	1.76
3D $t$ -field			
$\nu = 20$	659.09	1043.29	2199.84 ( $n = 11614$ )
$\nu = 50$	1719.52	2755.58	2611.89 ( $n = 11244$ )
$\nu = 200$	5325.24 ( $n = 74403$ )	5159.63 ( $n = 46409$ )	11914.56 ( $n = 402$ )
2D isotropic Gaussian field (empirical covariance case)			
number of fields = 200	9.56	15.13	208.33
number of fields = 1000	9.91	14.19	145.04
number of fields = 10,000	14.56	18.05	119.83
3D isotropic Gaussian field (empirical covariance case)			
number of fields = 200	82.94	105.66	1412.21
number of fields = 1000	320.91	290.76	1722.92