

# Spatial multiresolution cluster detection method

LINGSONG ZHANG\* AND ZHENGYUAN ZHU†,‡

A novel multi-resolution cluster detection (MCD) method is proposed to identify irregularly shaped clusters in space. Multi-scale test statistic on a single cell is derived based on likelihood ratio statistic for Bernoulli sequence, Poisson sequence and Normal sequence. A neighborhood variability measure is defined to select the optimal test threshold. The MCD method is compared with single scale testing methods controlling for false discovery rate and the spatial scan statistics using simulation and f-MRI data. The MCD method is shown to be more effective for discovering irregularly shaped clusters, and the implementation of this method does not require heavy computation, making it suitable for cluster detection for large spatial data.

AMS 2000 SUBJECT CLASSIFICATIONS: Primary 60K35, 60K35; secondary 60K35.

KEYWORDS AND PHRASES: Multiresolution, Scan statistics, Scale space inference, Spatial test.

## 1. INTRODUCTION

Spatial cluster detection is an important problem in many different fields, such as epidemiology and image analysis. In this paper, we focus on the following spatial cluster detection problem: to detect an (irregular) spatial region which has a different signal intensity compared to the background. Specifically, we focus on a large spatial region, e.g. an image or a geographical map. We assume that when there is no signal in it, the data collected at all locations are independent and follow the same distribution with an unknown parameter. Under the alternative hypothesis, within an (unknown) sub-region, the data observed are from a different distribution. Let  $\mathcal{D}$  be the entire spatial region,  $s \in \mathcal{D}$  be the location in  $\mathcal{D}$ , and  $Y(s)$  be the observed data at location  $s$ . Let  $\mathcal{D}_A$  be a subregion in  $\mathcal{D}$ , i.e.,  $\mathcal{D}_A \subset \mathcal{D}$ . We assume that  $Y(s) \stackrel{iid}{\sim} f(\theta_0)$  under  $H_0$ . Under  $H_1$ ,  $Y(s) \stackrel{iid}{\sim} f(\theta_1)$ , when  $s \in \mathcal{D}_A$ ; and  $Y(s) \stackrel{iid}{\sim} f(\theta_0)$ , when  $s \in \mathcal{D}/\mathcal{D}_A$ . Typical examples include

1. Binomial distribution,  $f(\theta_0) = \text{Bin}(N, p_0)$ , and  $f(\theta_1) = \text{Bin}(N, p_1)$ , with  $p_1 > p_0$ .

2. Poisson distribution,  $f(\theta_0) = P(\lambda_0)$ , and  $f(\theta_1) = P(\lambda_1)$ , with  $\lambda_1 > \lambda_0$ .
3. Normal distribution,  $f(\theta_0) = N(\mu_0, \sigma^2)$  and  $f(\theta_1) = N(\mu_1, \sigma^2)$ , with  $\mu_1 > \mu_0$ .

Our main objective is to identify the unknown region  $\mathcal{D}_A$ . Here we do not have any assumptions on the shapes of  $\mathcal{D}$  and  $\mathcal{D}_A$ .

One popular method to identify such spatial cluster is the spatial scan statistics. Scan statistics was first developed in Naus (1965a,b). See comprehensive review in Glaz and Balakrishnan (1999); Glaz et al. (2001, 2009). Kulldorff extended scan statistics into multidimensional case which includes spatial problems (Kulldorff, 1999). See also in Costa and Kulldorff (2009); Kulldorff et al. (2006); Loh et al. (2008); Zhang and Lin (2009). Kulldorff also developed a software to apply spatial scan statistics to spatial temporal data (Kulldorff, 2010). The basic idea of scan statistics is to use scan windows of different sizes and locations and perform likelihood ratio tests on all the scan windows using simulation. Typical window shapes in Kulldorff (2010) include circle or ellipse. Other methods for detecting spatial clusters include modified Knox test (Baker, 2004), point process method (Diggle et al., 2005), Bayesian hierarchical method (Liang et al., 2009), methods based on  $K$  function (Loh, 2011; Wheeler, 2007). The spatial scan statistics methodology is shown to be very powerful compared to some other spatial cluster detection methods (Kulldorff et al., 2003). However, the power of the test is reduced when the shape of the true cluster is not circle or ellipsoid. It is also less useful when part of the inference objective is to identify the shape of the clusters.

Another commonly used approach is to perform test at each location in  $\mathcal{D}$ , and use multiple comparison methods to overcome the issue of large number of tests. One typical multiple comparison method is to control false discovery rate (FDR) (Benjamini and Hochberg, 1995; Storey, 2002). This method overcomes the fixed shape problem of the scan statistics, since it performs tests at every single individual location. However, it ignores the spatial information and the detected region is usually not spatially continuous. Recently this method has been extended by using an adjusted local false discovery rate method (Zhang et al., 2011) for spatial clustering detection.

The multi-resolution cluster detection (MCD) method we propose in this paper takes advantage of the spatial information to make the test more powerful, while still maintains the flexibility of detecting irregularly shaped spatial

arXiv: 1205.2106

\*Zhang's research is partially supported by Purdue Research Foundation International Travel Grant 51060311.

†Zhu's research is partially supported by NRCS-ISU cooperative agreement number 68-7482-11-534.

‡Corresponding author.

clusters. The MCD method is motivated by the scale-space inference ideas in functional estimation and image analysis (Chaudhuri and Marron, 1999, 2000; Lindeberg, 1993, 1994). Zhang et al. (2007) and Zhang et al. (2008) applied scale-space inference method to time series analysis with applications in Internet anomaly detection. The MCD method uses a similar idea as the scale-space inference by forming a test statistic based on multi-scale windows to effectively use the spatial information. A novel variability measure is used to identify the test threshold, which helps to maintain a balance between the sensitivity and specificity. The use of the variability measure overcomes the multiple comparison issues of the individual test, since it is a global procedure. Simulation studies and an application to fMRI data show that the MCD method compared favorably to the spatial scan statistic method and the multiple testing method based on FDR for identifying spatial clusters.

The remaining part of this paper is organized as follows. In Section 2 we describe the methodology and theory, which includes detailed derivation of the MCD method for binomial, Poisson and Normal distributions. The validity of the local variability measure is also shown in this section. Section 3 provides simulation studies to compare different approaches for Binomial distributions with different alternative probability of success and different signal regions. An application of the three methods to a real fMRI data is given in Section 4, and in Section 5 we discuss possible extensions.

## 2. METHODOLOGY

Let  $Y(s)$ ,  $s \in \mathcal{D}$ , be a sequence of independent random variables from the distribution  $\mathcal{F}_{\theta(s)}$ . Here  $s$  is the location in a two dimensional region  $\mathcal{D}$ , and  $\theta(s)$  are the population parameters. In the later sections, if  $s = (k, l)$ , we may use  $Y(s)$  and  $Y_{kl}$  interchangeably. Under the null hypothesis, we assume  $\theta(s) = \theta_0$  for all  $s \in \mathcal{D}$ . Under the alternative hypothesis, we assume  $\theta(s) = \theta_1$  for all  $s \in \mathcal{D}_A \subset \mathcal{D}$ , and  $\theta = \theta_0$  otherwise. Our objective is to identify  $\mathcal{D}_A$ . We consider three examples in this paper: 1)  $\mathcal{F}$  is  $Bin(N_0, p)$ , and  $\theta = p$ , the probability of success; 2)  $\mathcal{F}$  is  $Poisson(\lambda)$ , and  $\theta = \lambda$ ; 3)  $\mathcal{F}$  is  $N(\mu, \sigma_0^2)$ , and  $\theta = \mu$ . In the above three examples,  $N_0$  and  $\sigma_0^2$  are assumed to be known.

Our multi-resolution cluster detection (MCD) method has two steps: 1) At each location, we form a MCD test statistic which is expected to be large in the signal region, and small in the background region. 2) Determine a threshold for the test at each location which maintains a balance between specificity and sensitivity of the procedure. In Section 2.1, we describe the MCD test statistics and give detailed formula to compute the test statistics for the distributions mentioned above. In Section 2.2 we introduce the notion of neighborhood variability and describe a method to determine the detection threshold based on neighborhood variability. Some theoretical results are derived in Section 2.3 to justify our method for setting the threshold. In Section 2.4 we briefly discuss the choice of scales.

### 2.1 The MCD test statistic

The test statistic for MCD at each location is formed using ideas from scale-space inference. In this subsection, let us focus on a specific location  $s$ . For easy presentation, let  $\mathcal{D}$  be a two dimensional grid with  $n_r$  rows and  $n_c$  columns, and  $s = (i, j)$ , with  $i = 1, 2, \dots, n_r$ , and  $j = 1, 2, \dots, n_c$ . Define a sequence of local regions  $\{s\} = \mathcal{D}_1 \subset \mathcal{D}_2 \subset \dots \subset \mathcal{D}_M \subset \mathcal{D}$ . Note that our method does not require  $\mathcal{D}$  to be square, and the shapes of  $\mathcal{D}_i$  can be irregular as well. Let us define the following aggregation vector  $(X_{ij}^1, X_{ij}^2, \dots, X_{ij}^M)^T$ , where  $X_{ij}^1 = Y_{ij}$ , and

$$(1) \quad X_{ij}^r = \sum_{kl \in \mathcal{D}_r} Y_{kl}, \quad \text{for all } 1 \leq r \leq M.$$

In this section, since we focus on a single location, we further simplify our notation  $X_{ij}^r$  to be  $X_r$ . Thus, the aggregation vector is  $(X_1, X_2, \dots, X_M)^T$ . Assuming that  $(x_1, x_2, \dots, x_M)^T$  is the observed vector, the likelihood ratio test statistics can be written as

$$\begin{aligned} \Lambda &= \frac{\sup_{\Theta_0} L(\theta; x)}{\sup_{\Theta} L(\theta; x)} \\ &= \frac{\sup P(X_1 = x_1, X_2 = x_2, \dots, X_M = x_M | \Theta_0)}{\sup P(X_1 = x_1, X_2 = x_2, \dots, X_M = x_M | \Theta)}. \end{aligned}$$

Based on standard theory on likelihood ratio test, under appropriate regularity conditions, we have

$$-2 \log(\Lambda) \sim \chi^2(M).$$

Since  $\mathcal{D}_A$  is unknown, it is difficult to compute the denominator of the likelihood ratio statistic under the original alternative hypothesis. Instead we derive the formula for  $\Lambda$  under the alternative that the parameter  $\theta$  is a constant within each of the regions  $\mathcal{D}_1, \mathcal{D}_2 - \mathcal{D}_1, \dots$ , and  $\mathcal{D}_k - \mathcal{D}_{k-1}$  for three commonly used distributions, Binomial, Poisson and Normal, and use them as the test statistics. The results does not depend on the shapes of  $\mathcal{D}_r$ . Thus, the user can use their favorite shape in the MCD method. We provide a function in R to calculate the test statistic, which provides two regular shapes (square and circle) in the implementation. In this paper, we will use the square shape for illustration purpose.

All the calculation next uses the following fact:

$$\begin{aligned} P(X_1 = x_1, \dots, X_M = x_M) \\ &= P(X_1 = x_1, X_2 - X_1 = x_2 - x_1, \dots, \\ &\quad X_M - X_{M-1} = x_M - x_{M-1}). \end{aligned}$$

Note that since  $Y_{ij}$  are independent, the  $\{X_i - X_{i-1}\}$  are independent to each other. Using this formula, it is relatively straightforward to derive the results for three different distributions. In what follows we further introduce a notation  $m_r = |\mathcal{D}_r|$ , the cardinality of  $\mathcal{D}_r$ . We leave the details of the computation to the appendix.

### 2.1.1 Binomial distribution

Let  $Y_{ij} \stackrel{iid}{\sim} \text{Bin}(N_{ij}, p_0)$  under the null hypothesis. Under the alternative,  $Y_{ij} \stackrel{iid}{\sim} \text{Bin}(N_{ij}, p_k)$  for all  $(i, j) \in \mathcal{D}_k \setminus \mathcal{D}_{k-1}$ , with  $p_k > p_0$ , and  $\text{Bin}(N_{ij}, p_0)$  otherwise. It is relatively easy to get that, under the null hypothesis,

$$X_r = \sum_{kl \in \mathcal{D}_r} Y_{kl} \sim \text{Bin}(N_r, p_0),$$

where  $N_r = \sum_{kl \in \mathcal{D}_r} N_{kl}$ . It is straightforward that  $N_i \leq N_{i+1}$ . Let  $\hat{p}_{ij} = \frac{Y_{ij}+1}{N_{ij}+2}$  (note that this adjustment is to avoid 0 and 1 as an estimate, which is similar to the Wilson's estimate). Define the follows:

$$\begin{aligned} \hat{p}_0 &= \text{Median}_{\mathcal{D}}(\hat{p}_{ij}), \\ \hat{p}_1 &= \max(\text{Median}_{\mathcal{D}_1}(\hat{p}_{ij}), \hat{p}_0), \\ \hat{p}_r &= \max(\text{Median}_{\mathcal{D}_r \setminus \mathcal{D}_{r-1}}(\hat{p}_{ij}), \hat{p}_0). \end{aligned}$$

Here the maximization is to make sure the estimate under the alternative is not smaller than the estimate under the null hypothesis, and we use median instead of mean for robustness. Then we have the following formulae to compute the approximated likelihood ratio test statistic for Binomial distribution:

$$\begin{aligned} -2 \log(\Lambda) &= -2 \log \left( \frac{\sup_{\Theta_0} L(\theta; x)}{\sup_{\Theta} L(\theta; x)} \right) \\ &\approx -2[x_1(\log(\hat{p}_0) - \log(\hat{p}_1)) + (N_1 - x_1)(\log(1 - \hat{p}_0) \\ &\quad - \log(1 - \hat{p}_1)) + (x_2 - x_1) \times (\log(\hat{p}_0) - \log(\hat{p}_2)) \\ &\quad + ((N_2 - N_1) - (x_2 - x_1))(\log(1 - \hat{p}_0) - \log(1 - \hat{p}_2)) \\ &\quad + \dots + (x_k - x_{k-1})(\log(\hat{p}_0) - \log(\hat{p}_k)) \\ &\quad + ((N_k - N_{k-1}) - (x_k - x_{k-1})) \\ &\quad \times (\log(1 - \hat{p}_0) - \log(1 - \hat{p}_k))]. \end{aligned}$$

### 2.1.2 Poisson distribution

Let  $Y_{ij} \sim P(\lambda_0)$  under the null hypothesis, and  $Y_{ij} \sim P(\lambda_k)$  for  $(i, j) \in \mathcal{D}_k \setminus \mathcal{D}_{k-1}$ , with  $\lambda_k > \lambda_0$ . It is clear that, under the null hypothesis,

$$X_{ij}^k \sim P(m_k \lambda_0).$$

Let  $\hat{\lambda}_0 = \text{Median}_{\mathcal{D}}(Y_{ij})$ , and

$$\hat{\lambda}_i = \max \left( \frac{x_i - x_{i-1}}{m_i - m_{i-1}}, \hat{\lambda}_0 \right),$$

then we have the following formulae to compute approximate likelihood ratio test statistic for Poisson distribution:

$$\begin{aligned} -2 \log(\Lambda) &\approx -2[x_1(\log(\hat{\lambda}_0) - \log(\hat{\lambda}_1)) + (\hat{\lambda}_1 - \hat{\lambda}_0) \\ &\quad + (x_2 - x_1)(\log(\hat{\lambda}_0) - \log(\hat{\lambda}_2)) + m_2(\hat{\lambda}_2 - \hat{\lambda}_0)] \\ &= -2[(x_k - x_{k-1})(\log(\hat{\lambda}_0) - \log(\hat{\lambda}_k)) \\ &\quad + m_k(\hat{\lambda}_k - \hat{\lambda}_0)]. \end{aligned}$$

### 2.1.3 Normal distribution

We assume that under the null hypothesis,  $Y_{ij} \sim N(\mu, \sigma^2)$  and they are independent. Under the alternative,  $Y_{ij} \sim N(\mu_k, \sigma^2)$  when  $(i, j) \in \mathcal{D}_k \setminus \mathcal{D}_{k-1}$ , and  $\mu_k > \mu$ .

Here we assume that  $\sigma^2$  is given. For real application,  $\sigma^2$  is unknown, thus, we will use a robust estimate for the variance.

Let  $\hat{\mu}_0 = \text{Median}_{\mathcal{D}}(Y_{ij})$ , and

$$\hat{\mu}_i = \max \left( \frac{x_i - x_{i-1}}{m_i - m_{i-1}}, \hat{\mu}_0 \right).$$

We have

$$\begin{aligned} -2 \log(\Lambda) &\approx \frac{2x_1(\hat{\mu}_1 - \hat{\mu}_0)}{\sigma^2} + \frac{\hat{\mu}_0^2 - \hat{\mu}_1^2}{\sigma^2} \\ &\quad + \frac{2(x_2 - x_1)(\hat{\mu}_2 - \hat{\mu}_0)}{\sigma^2} + \frac{m_2(\hat{\mu}_0^2 - \hat{\mu}_2^2)}{\sigma^2} \\ &\quad + \dots + \frac{2(x_k - x_{k-1})(\hat{\mu}_k - \hat{\mu}_0)}{\sigma^2} \\ &\quad + \frac{m_k(\hat{\mu}_0^2 - \hat{\mu}_k^2)}{\sigma^2}. \end{aligned}$$

## 2.2 Test threshold

The likelihood ratio test theory suggests that  $\chi^2$  is the approximated distribution of the test statistic, and one way to determine the test threshold is to control for FDR based on this asymptotic distribution. However, such procedure turns out to have poor performance due to the fact that 1) the asymptotic distribution may not be a good approximation for small sample sizes; 2) the test statistics at different locations are correlated.

In this section, we introduce a neighborhood variability measure to determine the test threshold. Let  $(i, j)$  be the location to calculate the local variability. We choose the four nearest cells, where the  $(i, j)$  cell serves as the center in the region, as shown in Figure 1. The variance of these five data points is defined as the neighborhood variability.

Our assumption is that both signal region and the noise region have homogenous distribution within themselves. Thus, the local variability within these two regions should be small, while the local variability for points close to the cluster boundary should be relatively large. In order to identify the "optimal" detection threshold, we compute the average neighborhood variability on the region with test statistics between two consecutive detection thresholds for a sequence of monotone increasing test thresholds  $t_1, t_2, \dots, t_M$ . When the threshold is too small, most of the signal regions are identified as signal, along with many locations in the noise region. As the threshold increases, only the noise region are excluded, leading to a small average neighborhood variability. Similarly when the threshold is too large, the average neighborhood variability between two consecutive detection regions will also be small as they only include signal regions. We identify the two thresholds between which the average

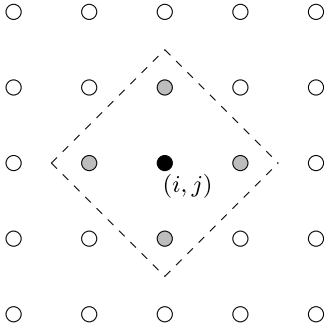


Figure 1. The diagram for the neighborhood variability. At each location  $(i, j)$ , the variance of the four nearest cells plus the  $(i, j)$  point will be calculated as the local variability. This helps to identify the edge of the detection boundary.

neighborhood variability is the highest and take the average of these two as the optimal detection threshold.

To summarize, our MCD method can be carried out as follows:

1. At each location  $s$ , compute the MCD test statistic  $T(s)$  at this location as described in Section 2.1 and the neighborhood variability measure  $V(s)$  as described in Section 2.2.
2. Let  $t_1 = \min_s T(s)$ ,  $t_M = \max_s T(s)$ , and define an arithmetic sequence  $t_1, t_2, \dots, t_M$ .
3. Let  $\mathcal{D}_k = \{s : T(s) > t_k\}$ , and  $V_k = \text{ave}\{V(s) : s \in \mathcal{D}_k, s \notin \mathcal{D}_{k+1}\}$ . Let  $k^* = \text{argmax}_k V_k$ . The optimal test threshold is defined as  $t^* = 0.5(t_{k^*} + t_{k^*+1})$ , and the final detected region is defined as  $\hat{\mathcal{D}}_{\mathcal{A}} = \{s : T(s) > t^*\}$ .

### 2.3 Validity of the test threshold

In this subsection, we illustrate that under certain conditions, the MCD method can identify the signal region almost surely. Let  $\mathcal{B}$  be the set of the boundary points,  $\mathcal{N}$  be the set of inner points in the noise region, and  $\mathcal{S}$  be the set of inner points in the signal region. For simplicity, we also use  $\mathcal{B}^c = \mathcal{N} \cup \mathcal{S}$ . Furthermore, let  $n_b$  be the number of boundary points,  $n_s$  be the number of (non-boundary) signal points, and  $n_n$  be the number of (non-boundary) noise points. Let  $N = n_b + n_n + n_s$  be the total number of data points in the spatial region  $\mathcal{D}$ . A boundary point is defined as the point where its local five observations (including the point itself) are not from the same group. See Figure 2 for illustrations of some boundary points.

For easy presentation, all the properties are shown under Normality assumption. We assume for any  $Y$  in the noise region,  $Y \sim N(0, 1)$ ; and for any  $Y$  in the signal region,  $Y \sim N(\delta, 1)$ , where  $\delta > 0$ . The parameter  $\delta$  measures the signal strength. In addition, the proportion of boundary point  $n_b/N$  is denoted as  $p_B$ . The following two theorems give some justification for the procedure we outlined in Section 2.2. The proofs of both theorems are in the appendix.

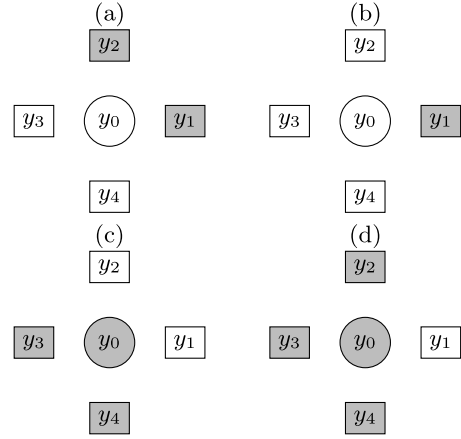


Figure 2. The diagram for four types of boundary points. The white cells are noise points, and the gray cells are signal points. Panel (a) shows a boundary point at the noise region with an angle; panel (b) shows a boundary point at the noise region without an angle; panel (c) shows a boundary point at the signal region with an angle; and panel (d) shows a boundary point at the signal region without an angle.

**Theorem 2.1.** The average test statistic  $T$  at  $\mathcal{N}$ ,  $\mathcal{B}$  and  $\mathcal{S}$  has the following relation: if  $N, n_b, n_n, n_s \rightarrow \infty$ , and  $\delta > 0$ ,

$$P \left( \text{Ave}_{Y \in \mathcal{N}} T(Y) < \text{Ave}_{Y \in \mathcal{B}} T(Y) < \text{Ave}_{Y \in \mathcal{S}} T(Y) \right) \rightarrow 1.$$

**Theorem 2.2.** If  $\frac{\delta^2 N p_B (1-p_B)}{8+4(1-p_B)} \rightarrow \infty$ , then

$$P \left( \text{Ave}_{Y \in \mathcal{B}} V(Y) - \text{Ave}_{Y \in \mathcal{B}^c} V(Y) > 0 \right) \rightarrow 1.$$

*Remark.* Theorem 2.1 shows that the MCD test statistic at points in the boundary region are on average smaller than those in the purely signal region and larger than those in the noise region. Theorem 2.2 shows that the average variability measure on the boundary set is the largest if either the signal strength is large or the points in both boundary and non-boundary regions are large. Thus by selecting the belt with the maximum average variability, we have a good chance of identifying the true boundary of the signal region.

### 2.4 Number of scales

To use the MCD method, one need to determine: 1) how many scales to be used, and 2) what is the largest scale to be used. The answer to these two questions are related to each other and depends on specific applications. Typically, when the maximum scale is the same, the test power of the MCD method with more scales tends to be larger, as shown in one dimensional situation in Zhang et al. (2007). For the two dimensional case, we compared two-scale vs. five-scale

and single-scale methods and found that even though the test power of the five-scale method is better than the other two methods, the improvement from two to five is not as significant as those from one to two. Thus, in this paper, we will only use a two-scale MCD method for all the simulation examples and real application.

By aggregating local values, the MCD method will have some dilution effect, i.e., the noise points which are close to the signal region are more likely to be identified as false signals. The larger the aggregation region, the higher chance of false detection for these points. Thus, we will not use a very large window as the maximum scale. For a typical  $100 \times 100$  image, we find that the largest aggregation scale around 10 (or a radius as 5) works well. In the following simulation and real examples, we will use window size 10 as our maximum scale.

### 3. SIMULATION AND COMPARISON

In this section, we use simulation to compare the MCD method with the single scale test method controlling FDR and the spatial scan statistics. All the examples are simulated in a  $100 \times 100$  map. To save space, in this manuscript, we only simulate binomial distribution. Poisson and Normal examples are provided in the author’s personal website. We use sensitivity and specificity to compare the performance of different methods. Under each simulation setting, we will repeat the same simulation 100 times, and also report the probability of identification at each spatial location.

#### 3.1 Simulation settings

Figure 3 shows the shapes of the simulations in this paper. We will study four different shapes of regions: *L*-shape, triangle, oval and *Y*-shape. Note that the triangle and oval are convex, while *Y* and *L*-shapes are not. Note that a data image contains 10,000 pixels. Table 1 provides the number and the percentage of the pixels in the simulated signal regions.

For each simulation run, the background of the image is simulated as independent identically distributed Binomial distribution  $Y(s) \sim Bin(100, 0.2)$ , if  $s \in \mathcal{D}/\mathcal{D}_s$ . The signal part,  $Y(s) \sim Bin(100, p_1)$ , if  $s \in \mathcal{D}_s$ , where  $p_1 > 0.2$ . In this paper, we focus on five different alternative probabilities of success,  $p_1 = 0.21, 0.22, 0.23, 0.24,$  and  $0.25$ . Note that when  $p_1 = 0.21$ , the detection should be very challenging. We also assume that the population is preknown. For real applications such as epidemiology, the population of a specific geographical location can be approximated by a recent survey. So this assumption is reasonable even for real applications.

#### 3.2 Simulation results

For each simulation setting, i.e., the same signal shape and the same alternative probability of success, we repeat the simulation 100 times. We use the average of several measures to compare the performance of our method with other

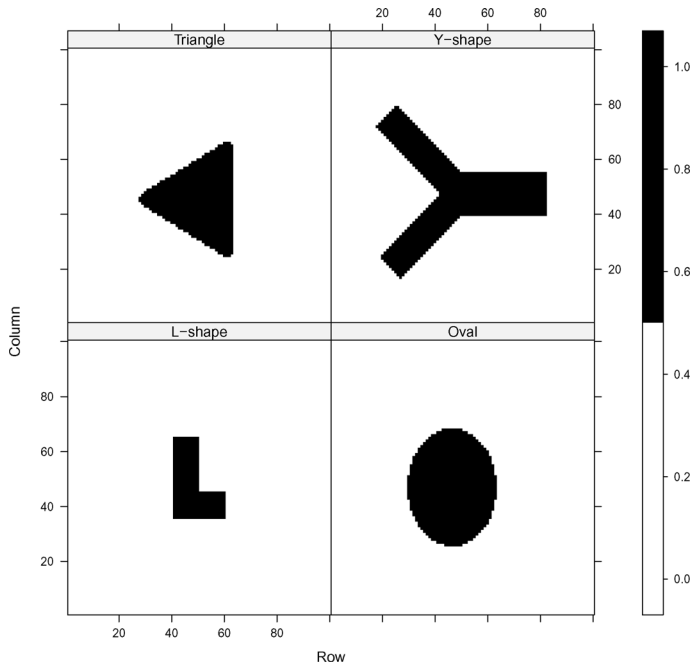


Figure 3. The simulation settings.

Table 1. Total true number of signal pixels

Shapes	<i>L</i> -shape	Oval	Triangle	<i>Y</i> -shape
Number of pixels	400	1142	864	1344
(percentage)	4%	11.42%	8.64%	13.44%

related methods. The measurements include specificity (the percentage of correctly identification of noise points), sensitivity (the percentage of correctly identification of signal points), and the probability (relative frequency) of identifications at each pixels.

Tables 2–4 provide the average specificities and the average sensitivities for these different simulation settings. The numbers in the parenthesis provides the standard deviation of the performance measures. It shows that when the alternative probability increases, both sensitivity and specificity of the MCD method increases. The variability also decreases as the alternative probability increases. Note that even when the alternative probability of success is 0.21, for all the shapes, our MCD method has an average sensitivity of 40%, which is quite impressive, although the variability is large. Note that when the alternative probability of success increases to 0.22, the specificity increases dramatically to over 90%, meanwhile, the sensitivity also increases.

Table 3 shows the specificity and sensitivity of the single scale method with test threshold selected to control false discovery rate. We use the direct false discovery rate method in Storey (2002). In order to have a sensible detection result, we use FDR level as 0.60, which is relatively high. Lower FDR level will lead to much smaller sensitivity. It shows that the specificity will decrease when the sensitivity increases.

Table 2. The average specificity/sensitivity for the MCD method by different simulation settings

P1	L-shape		Oval shape		Triangle shape		Y-shape	
	Specificity (Std)	Sensitivity (Std)	Specificity (Std)	Sensitivity (Std)	Specificity (Std)	Sensitivity (Std)	Specificity (Std)	Sensitivity (Std)
0.21	0.8415 (0.2899)	0.3818 (0.3081)	0.8462 (0.2865)	0.3972 (0.3353)	0.8273 (0.3142)	0.4036 (0.323)	0.8072 (0.3193)	0.406 (0.3558)
0.22	0.9401 (0.1516)	0.6252 (0.2574)	0.9309 (0.1621)	0.5125 (0.3865)	0.939 (0.1459)	0.5806 (0.3533)	0.9367 (0.1436)	0.5299 (0.3832)
0.23	0.9845 (0.0327)	0.7986 (0.2358)	0.9738 (0.0315)	0.7669 (0.3744)	0.9801 (0.0172)	0.8079 (0.3159)	0.9626 (0.0237)	0.8232 (0.3313)
0.24	0.987 (0.0043)	0.9387 (0.0999)	0.9769 (0.0108)	0.9003 (0.2776)	0.9774 (0.0079)	0.9455 (0.1827)	0.9484 (0.0963)	0.98 (0.0911)
0.25	0.9856 (0.0032)	0.9723 (0.0198)	0.9745 (0.0061)	0.9817 (0.1261)	0.9759 (0.0043)	0.9923 (0.0069)	0.96 (0.0118)	0.9588 (0.1309)

Table 3. The average specificity/sensitivity for single scale detection method by different simulation settings

P1	L-shape		Oval shape		Triangle shape		Y-shape	
	Specificity (Std)	Sensitivity (Std)	Specificity (Std)	Sensitivity (Std)	Specificity (Std)	Sensitivity (Std)	Specificity (Std)	Sensitivity (Std)
0.21	0.9386 (0.0203)	0.1003 (0.0317)	0.9575 (0.0164)	0.0709 (0.0276)	0.9475 (0.0207)	0.0845 (0.0297)	0.9274 (0.0262)	0.1148 (0.0359)
0.22	0.8726 (0.0333)	0.2613 (0.0536)	0.9428 (0.0218)	0.1382 (0.046)	0.9027 (0.0266)	0.212 (0.0458)	0.8532 (0.0348)	0.2904 (0.0539)
0.23	0.8167 (0.0427)	0.4322 (0.0621)	0.9305 (0.0265)	0.2277 (0.0612)	0.8653 (0.0323)	0.3556 (0.0566)	0.7787 (0.0459)	0.4854 (0.0614)
0.24	0.7632 (0.0451)	0.5976 (0.0566)	0.9126 (0.0222)	0.353 (0.0593)	0.826 (0.0386)	0.5122 (0.0598)	0.7099 (0.0597)	0.6568 (0.0772)
0.25	0.7208 (0.0434)	0.7272 (0.0437)	0.8989 (0.0238)	0.4755 (0.0551)	0.7954 (0.0367)	0.6464 (0.0462)	0.7342 (0.1309)	0.6584 (0.2286)

Table 4. The average specificity/sensitivity for the spatial scan method by different simulation settings

P1	L-shape		Oval shape		Triangle shape		Y-shape	
	Specificity (Std)	Sensitivity (Std)	Specificity (Std)	Sensitivity (Std)	Specificity (Std)	Sensitivity (Std)	Specificity (Std)	Sensitivity (Std)
0.21	0.9819 (0.0238)	0.4658 (0.2977)	0.9809 (0.0283)	0.8713 (0.0851)	0.9738 (0.0245)	0.734 (0.1682)	0.9012 (0.0996)	0.5461 (0.2354)
0.22	0.9746 (0.019)	0.792 (0.132)	0.9885 (0.0071)	0.9087 (0.0509)	0.9831 (0.0136)	0.7869 (0.0704)	0.8697 (0.0863)	0.791 (0.1106)
0.23	0.9755 (0.0142)	0.84 (0.0802)	0.9898 (0.0052)	0.9173 (0.0347)	0.9855 (0.0082)	0.7841 (0.054)	0.8788 (0.0782)	0.8342 (0.0838)
0.24	0.9783 (0.0109)	0.8486 (0.0832)	0.9904 (0.0038)	0.9192 (0.0278)	0.9865 (0.0073)	0.785 (0.0478)	0.8762 (0.0808)	0.8356 (0.0792)
0.25	0.9779 (0.0082)	0.8457 (0.068)	0.9902 (0.0035)	0.9249 (0.0244)	0.9865 (0.0072)	0.7874 (0.0492)	0.8677 (0.0793)	0.8471 (0.0697)

When the alternative probability is 0.22, the specificity of this method is similar to that of the MCD method, but the sensitivity is much smaller, which shows better performance of the MCD method over the single scale method.

Table 4 shows the specificity and sensitivity for the Spatial Scan method. It shows that when the data is oval or triangle, the performance is among the best of these three methods, even when the signal strength is 0.21. But for L-shape and Y-shape, the sensitivity is comparable to the

MCD method. When the signal strength increases, the MCD method performs similarly as the spatial scan method, even for the two convex shapes. The MCD method simultaneously have better specificity and sensitivity than the Spatial Scan method on the other two complicated shapes.

Figures 4–6 provide the probability map of the pixels being detected by each of these three methods. The probability is calculated by the relatively frequency of each pixels being detected by each method under the 100 runs under each

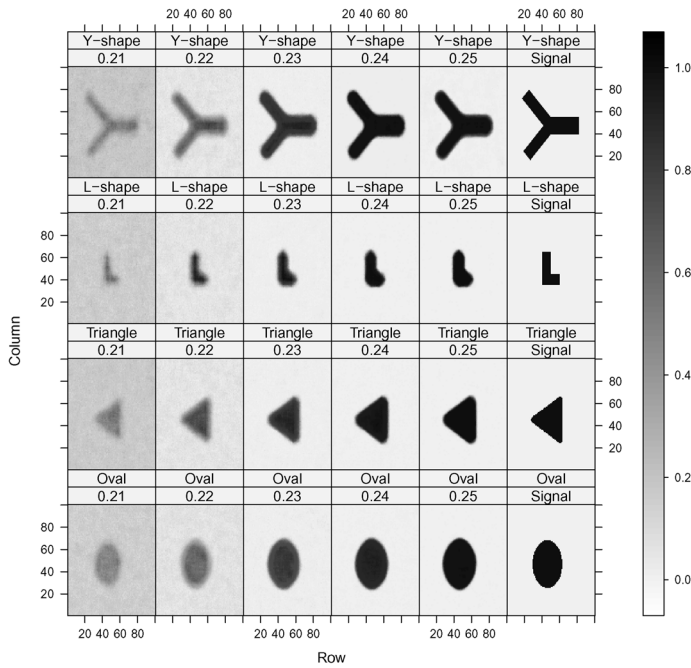


Figure 4. The probability map of identification from the MCD method by different alternatives.

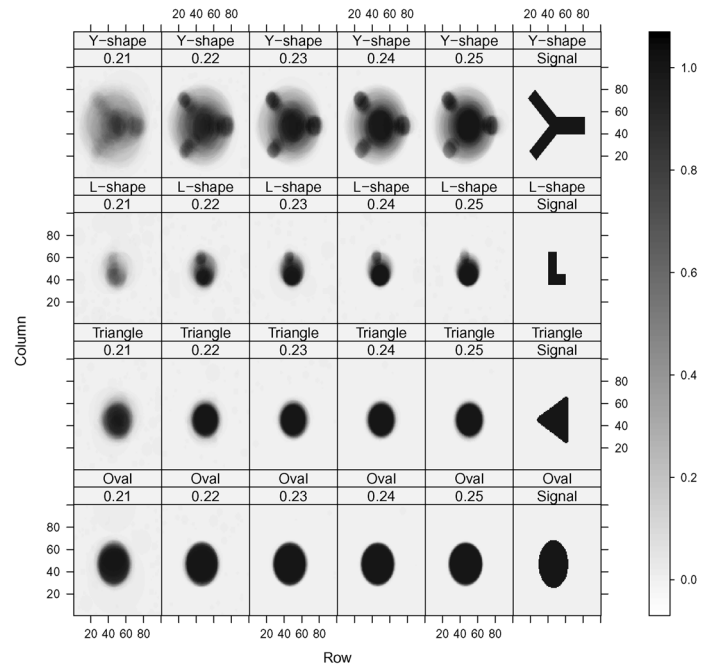


Figure 6. The probability map of identification from the spatial scan method by different alternatives.

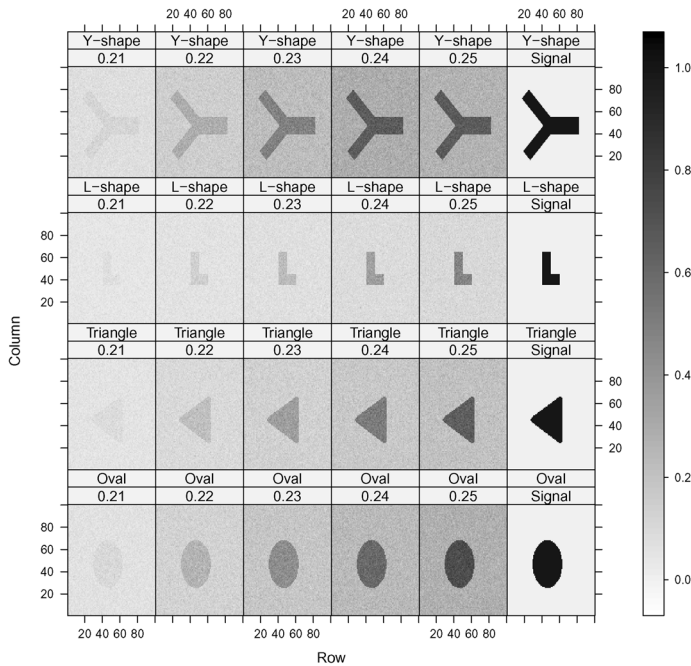


Figure 5. The probability map of identification from the individual tests adjusted by false discovery rate by different alternatives.

simulation setting. All figures shows that when the signal is weak, the chance of the signal region being detected is still much higher than the noise region.

Figure 4 provides the probability map by the MCD method. It shows that the actually shapes are almost correctly identified. There is a slight dilution effect by showing oval shapes on the corners. But overall the complicated shapes such as *L* and *Y* are correctly identified.

Figure 5 provides the probability map of the pixels being detected by the single scale method after adjusting for the false discovery rate. On average, the shapes are correctly identified. But this figure shows that the probability of correctly identified is relatively small. Even with a strong signal of 0.25, the probability of identification is about 60%.

Figure 6 provides a similar probability map of the pixels being detected by the Spatial Scan algorithm. We use the default circles as the scan window. It shows that for the oval shape signal, even when the signal strength is low, the Spatial Scan algorithm provides excellent performance. For other shapes, the performance is not bad in terms of the sensitivity. But it clearly misidentifies the shapes for triangle, *L* and *Y*-shapes.

### 3.3 Comparisons between different numbers of scales

Two-scale MCD method has been used in the previous subsection. In this subsection, we compare it with a five-scale MCD method, where the maximum scanning windows of these two methods are the same. In addition, we compare with another single scale method: the original scale tests using the local variability measure adjusted threshold. Here we only consider a single setting with the *L*-shape, with the

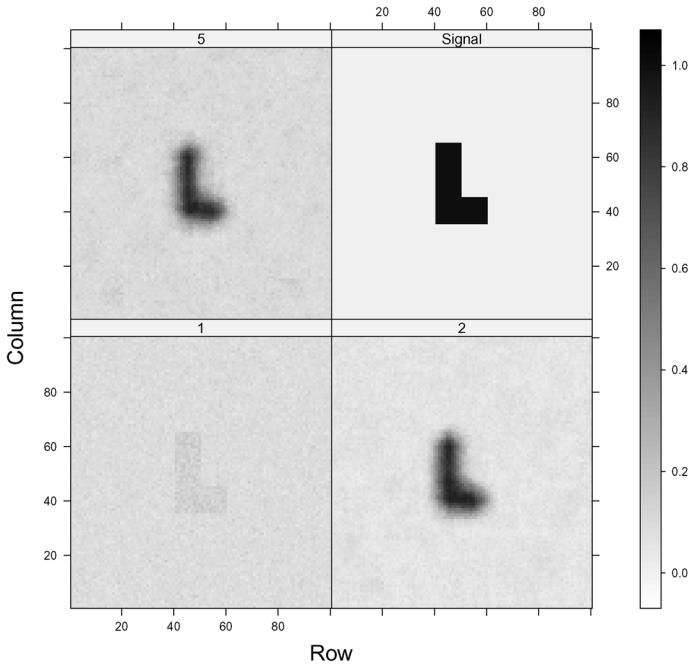


Figure 7. Comparison of the detection probability for single scale, two-scale, and five-scale MCD methods, under the setting of  $\text{Bin}(100, .2)$  vs.  $\text{Bin}(100, .22)$ .

alternative probability of success as 0.22. Figure 7 shows the detection probability at each pixels.

It shows that the signal scale method plus the local variability measure adjustment is not powerful. In the signal region, it only has 20% chance being detected. The two-scale MCD method with maximum scan window as 5 increases the detection probability to 80%. The five-scale MCD method with the same maximum scan window is very similar to the two-scale one. We also compared the ROC curves between 2-scale and 5-scale. It turns out that on average the 5-scale ROC has larger area under the curve than the 2-scale method, but they are very close to each other. See one example ROC curve comparisons in Figure 8. We note that the maximum scanning window is more important than the number of scales.

#### 4. EXAMPLES: FUNCTIONAL MRI DATA

In this section we apply the three aforementioned methods to a real fMRI data to illustrate their differences in real data application. The fMRI data was firstly analyzed in Maitra (2009, 2010) and Maitra et al. (2002).

The top left panel in Figure 9 shows 9 slices of the fMRI images in a total of 22 slices. Each individual image has  $128 \times 128$  pixels. All 9 images show activities in different regions, by showing dark colors for positive values. The response is a transformation of  $p$ -values from a previous study, and we are only interested in detection of the positive regions. By visual inspection of the raw data, it appears that

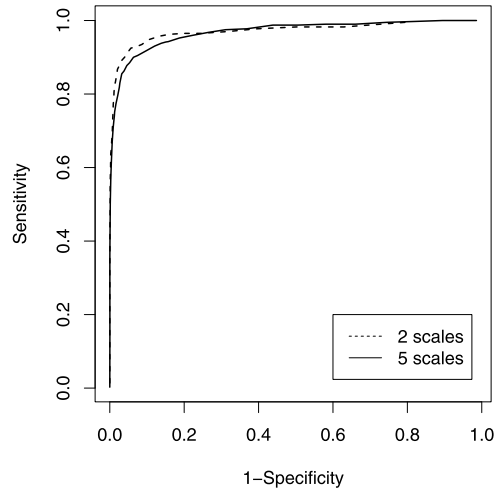


Figure 8. The ROC curves for one specific simulation example. The dashed one is for a two-scale MCD method, and the solid one is for a five-scale MCD method.

there are some positive activities at the following regions: central right (slice 1), central left and central right (slice 5), top left (slice 11), bottom central (slice 15), top central (slice 18), central middle or top central (slice 20), and central part (slice 22). The shapes of these regions seem to be very irregular.

The bottom left panel shows the single scale detection method with controlling false discovery rate (FDR) (20% false discovery rate). They highlight the regions that have been mentioned in the previous paragraph, but there are also many more outside spots, which is likely to be false positives. In addition, these candidate regions do not naturally form clusters in the figures, which makes it hard to interpret.

The top right panel shows the analysis results by the spatial Scan method. The method shows several important clusters in the plot, such as the clusters mentioned in previous paragraph. However, these regions are identified as a round shape as suggested by the spatial scan method, and do not reveal any useful shape information which is important for the analysis of fMRI data.

The bottom right panel shows the results generated by the MCD method. Within each slice, we identified several clusters, which are irregularly shaped. These regions form natural clusters, and also validate the initial visual impression. These results show that our method is suitable for detection of irregular shaped spatial signals.

#### 5. DISCUSSION

In this paper, we proposed a multiresolution cluster detection method, which is shown to have better performance empirically compared to single scale testing methods controlling for FDR and spatial scan statistics for detecting spatial clusters with irregular shape. We conjecture that the



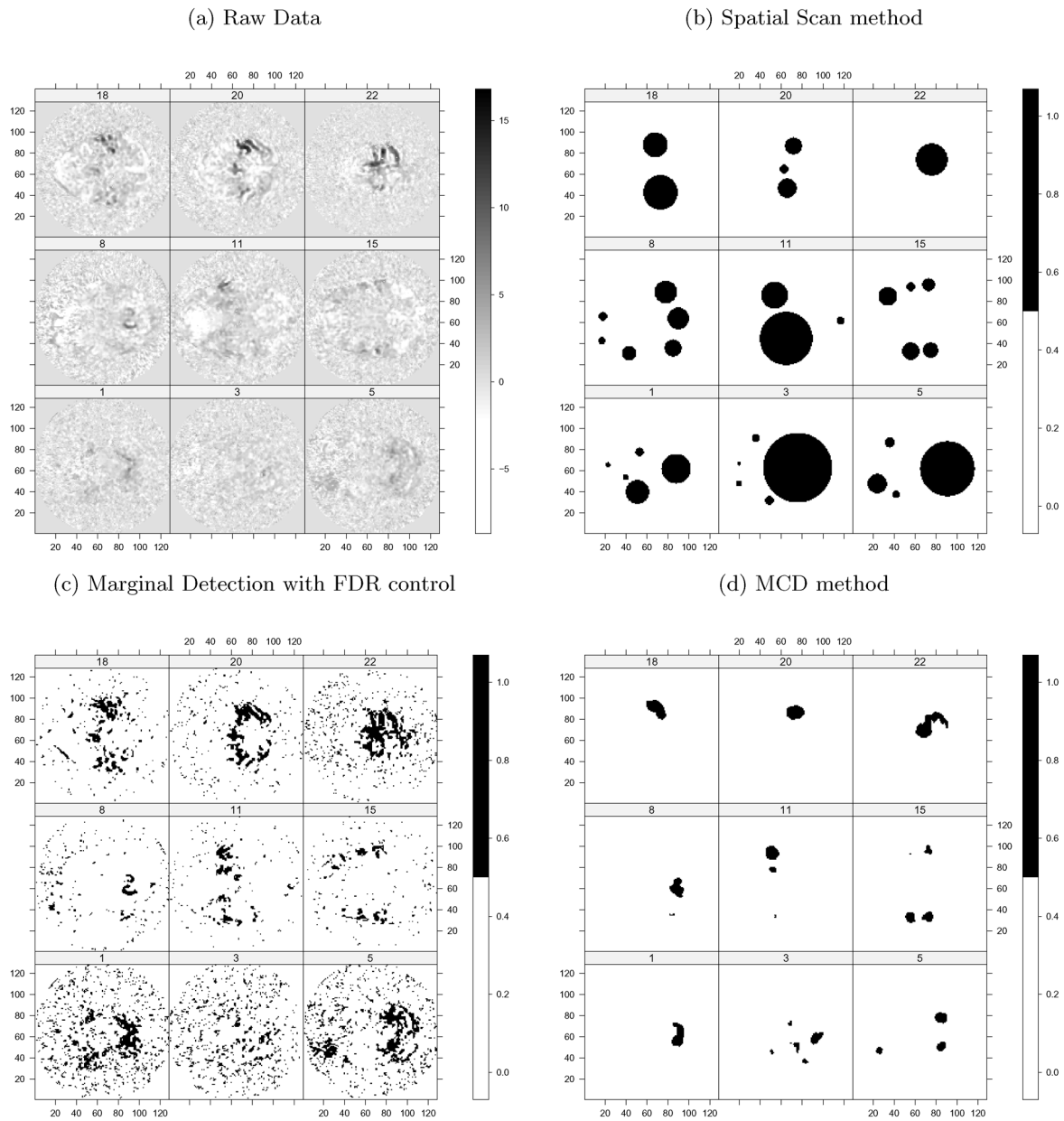


Figure 9. The detection result for an fMRI data set. The selected 9 slices of image are shown in panel (a). Panel (b) shows the detection result by the spatial scan method. It identify some interesting regions, but their shapes are regular. Besides, there are some tiny regions outside the big dots, which may correspond to false discovery. Panel (c) shows marginal detection method with FRD control level at 0.20. It also highlights some regions in each plot, but these identified spots do not naturally form (continuous) region. Panel (d) provided the detection result by the MCD method. It highlights several irregular shaped regions in each image.

power of MCD method is larger than methods based on single scale (e.g. the individual test combined with multiple comparison), the theoretical proof of which will be a future research topic. An early study of multiresolution method in one dimensional situation (Zhang et al., 2007) provides a foundation for this type of research. Another issue that is worth investigating is the selection of number of scales and

the largest scale for the MCD method, which we did not fully address in this paper.

## APPENDIX: DERIVATION OF THE PROPOSITIONS

The test statistics can be derived relatively straightforward as the follows:

## Derivation of the test statistic under Binomial assumption

The likelihood function can be calculate by using the following relation:

$$(2) \quad \begin{aligned} P(X^1 = x_1, X^2 = x_2, \dots, X^k = x_k) \\ = P(X^1 = x_1, X^2 - X^1 = x_2 - x_1, \dots, \\ X^k - X^{(k-1)} = x_k - x_{k-1}). \end{aligned}$$

Note that  $\{X^i - X^{(i-1)}\}$  are independent to each other.

Under the null hypothesis,  $X^i - X^{(i-1)} \sim \text{Bin}(N^i - N^{(i-1)}, p_0)$ . The distribution of the alternative is more complicated.

Let  $\hat{p}_{ij} = \frac{X_{ij}+1}{N_{ij}+2}$ . Then we have

$$\hat{p}_0 = \text{Median}(\hat{p}_{ij}),$$

which gives a robust estimate of  $\hat{p}_0$ . Thus the likelihood under the null is

$$\begin{aligned} P(X^1 = x_1, X^2 = x_2, \dots, X^k = x_k) \\ = P(X^1 = x_1, X^2 - X^1 = x_2 - x_1, \dots, \\ X^k - X^{(k-1)} = x_k - x_{k-1}) \\ = P(X^1 = x_1)P(X^2 - X^1 = x_2 - x_1) \dots \\ \times P(X^k - X^{k-1} = x_k - x_{k-1}) \\ = \binom{N^1}{x_1} \hat{p}_0^{x_1} (1 - \hat{p}_0)^{(N^1 - x_1)} \binom{N^2 - N^1}{x_2 - x_1} \hat{p}_0^{x_2 - x_1} \\ \times (1 - \hat{p}_0)^{((N^2 - N^1) - (x_2 - x_1))} \dots \binom{N^k - N^{k-1}}{x_k - x_{k-1}} \hat{p}_0^{x_k - x_{k-1}} \\ \times (1 - \hat{p}_0)^{((N^k - N^{k-1}) - (x_k - x_{k-1}))}. \end{aligned}$$

The maximum likelihood for the entire parameter space is quite challenging to be calculated. We may further introduce the following assumptions to simplify the calculation: For each region among the following  $k$ :  $\mathcal{D}_1, \mathcal{D}_2 - \mathcal{D}_1, \dots$ , and  $\mathcal{D}_k - \mathcal{D}_{k-1}$ ; we assume that the probability of successes are the same. Thus, we could have the following estimate of the probability of successes:

$$\hat{p}_1 = \max_{\mathcal{D}_1}(\text{Median}(\hat{p}_{ij}), \hat{p}_0),$$

$$\hat{p}_r = \max_{\mathcal{D}_r - \mathcal{D}_{r-1}}(\text{Median}(\hat{p}_{ij}), \hat{p}_0), \quad \forall r \geq 2.$$

Based on this, the likelihood for the entire parameter space is

$$\begin{aligned} P(X^1 = x_1, X^2 = x_2, \dots, X^k = x_k) \\ = P(X^1 = x_1)P(X^2 - X^1 = x_2 - x_1) \dots \\ \times P(X^k - X^{k-1} = x_k - x_{k-1}) \\ = \binom{N^1}{x_1} \hat{p}_1^{x_1} (1 - \hat{p}_1)^{(N^1 - x_1)} \binom{N^2 - N^1}{x_2 - x_1} \hat{p}_2^{x_2 - x_1} \end{aligned}$$

$$\begin{aligned} \times (1 - \hat{p}_2)^{((N^2 - N^1) - (x_2 - x_1))} \dots \binom{N^k - N^{k-1}}{x_k - x_{k-1}} \hat{p}_k^{x_k - x_{k-1}} \\ \times (1 - \hat{p}_k)^{((N^k - N^{k-1}) - (x_k - x_{k-1}))}. \end{aligned}$$

Thus, the likelihood ratio will be

$$\begin{aligned} \Lambda &= \frac{\sup_{\Theta_0} L(\theta; x)}{\sup_{\Theta} L(\theta; x)} \\ &= \left(\frac{\hat{p}_0}{\hat{p}_1}\right)^{x_1} \left(\frac{1 - \hat{p}_0}{1 - \hat{p}_1}\right)^{(N^1 - x_1)} \left(\frac{\hat{p}_0}{\hat{p}_2}\right)^{x_2 - x_1} \\ &\times \left(\frac{1 - \hat{p}_0}{1 - \hat{p}_2}\right)^{((N^2 - N^1) - (x_2 - x_1))} \dots \left(\frac{\hat{p}_0}{\hat{p}_k}\right)^{x_k - x_{k-1}} \\ &\times \left(\frac{1 - \hat{p}_0}{1 - \hat{p}_k}\right)^{((N^k - N^{k-1}) - (x_k - x_{k-1}))}. \end{aligned}$$

Thus, the typical likelihood ratio statistic  $-2 \log \Lambda$  can be easily derived.

## Derivation of the test statistic under Poisson assumption

We will use the same calculation formula in (2) For Poisson case, under the null hypothesis,  $X^i - X^{i-1} \sim P(|\mathcal{D}_i| - |\mathcal{D}_{i-1}| \lambda_0)$  for all  $i \geq 2$ . We can define  $\hat{\lambda}_0 = \text{Median}_{\mathcal{D}}(X_{ij})$ . Under the alternative or the entire parameter space, the distribution may be the follows,  $X^i - X^{i-1} \sim P(|\mathcal{D}_i| - |\mathcal{D}_{i-1}| \lambda_i)$ , and

$$\hat{\lambda}_i = \max\left(\frac{x_i - x_{i-1}}{|\mathcal{D}_i| - |\mathcal{D}_{i-1}|}, \hat{\lambda}_0\right).$$

Let  $m_k^* = \mathcal{D}_k - \mathcal{D}_{k-1}$ , we have, under  $H_0$ ,

$$\begin{aligned} P(X^1 = x_1, X^2 = x_2, \dots, X^k = x_k) \\ = P(X^1 = x_1, X^2 - X^1 = x_2 - x_1, \dots, \\ X^k - X^{k-1} = x_k - x_{k-1}) \\ = \frac{\hat{\lambda}_0^{x_1}}{x_1!} e^{-\hat{\lambda}_0} \frac{(m_2 \hat{\lambda}_0)^{x_2 - x_1}}{(x_2 - x_1)!} e^{-m_2 \hat{\lambda}_0} \dots \\ \times \frac{(m_k \hat{\lambda}_0)^{x_k - x_{k-1}}}{(x_k - x_{k-1})!} e^{-m_k \hat{\lambda}_0}, \end{aligned}$$

and under  $H_1$ ,

$$\begin{aligned} P(X^1 = x_1, X^2 = x_2, \dots, X^k = x_k) \\ = P(X^1 = x_1, X^2 - X^1 = x_2 - x_1, \dots, \\ X^k - X^{k-1} = x_k - x_{k-1}) \\ = \frac{\hat{\lambda}_1^{x_1}}{x_1!} e^{-\hat{\lambda}_1} \frac{(m_2 \hat{\lambda}_2)^{x_2 - x_1}}{(x_2 - x_1)!} e^{-m_2 \hat{\lambda}_2} \dots \\ \times \frac{(m_k \hat{\lambda}_k)^{x_k - x_{k-1}}}{(x_k - x_{k-1})!} e^{-m_k \hat{\lambda}_k}. \end{aligned}$$

The likelihood ratio leads to

$$\begin{aligned}\Lambda &= \frac{\frac{\hat{\lambda}_0^{x_1}}{x_1!} e^{-\hat{\lambda}_0} \frac{(m_2 \hat{\lambda}_0)^{x_2 - x_1}}{(x_2 - x_1)!} e^{-m_2 \hat{\lambda}_0} \dots \frac{(m_k \hat{\lambda}_0)^{x_k - x_{k-1}}}{(x_k - x_{k-1})!} e^{-m_k \hat{\lambda}_0}}{\frac{\hat{\lambda}_1^{x_1}}{x_1!} e^{-\hat{\lambda}_1} \frac{(m_2 \hat{\lambda}_2)^{x_2 - x_1}}{(x_2 - x_1)!} e^{-m_2 \hat{\lambda}_2} \dots \frac{(m_k \hat{\lambda}_k)^{x_k - x_{k-1}}}{(x_k - x_{k-1})!} e^{-m_k \hat{\lambda}_k}} \\ &= \left(\frac{\hat{\lambda}_0}{\hat{\lambda}_1}\right)^{x_1} \exp(\hat{\lambda}_1 - \hat{\lambda}_0) \left(\frac{\hat{\lambda}_0}{\hat{\lambda}_2}\right)^{x_2 - x_1} \exp(m_2(\hat{\lambda}_2 - \hat{\lambda}_0)) \\ &\quad \dots \left(\frac{\hat{\lambda}_0}{\hat{\lambda}_k}\right)^{x_k - x_{k-1}} \exp(m_k(\hat{\lambda}_k - \hat{\lambda}_0)).\end{aligned}$$

This leads to the proposition

### Derivation of the test statistic under normal assumption

Under the null hypothesis, we know that  $X^i - X^{i-1} \sim N(m_i \mu_0, m_i \sigma^2)$  for all  $i \geq 2$ . We can define  $\hat{\mu}_0 = \text{Median}_{\mathcal{D}}(X_{ij})$ .

Under the alternative or the entire parameter space, the distribution may be the follows,  $X^i - X^{i-1} \sim N(m_i \mu_i, m_i \sigma^2)$ , and

$$\hat{\mu}_i = \max\left(\frac{x_i - x_{i-1}}{|\mathcal{D}_i| - |\mathcal{D}_{i-1}|}, \hat{\mu}_0\right).$$

This leads to, under  $H_0$ ,

$$\begin{aligned}P(X^1 = x_1, X^2 = x_2, \dots, X^k = x_k) &= P(X^1 = x_1, X^2 - X^1 = x_2 - x_1, \dots, \\ &\quad X^k - X^{k-1} = x_k - x_{k-1}) \\ &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x_1 - \hat{\mu}_0)^2}{2\sigma^2}\right\} \frac{1}{\sqrt{2\pi m_2}\sigma} \\ &\quad \times \exp\left\{-\frac{((x_2 - x_1) - m_2 \hat{\mu}_0)^2}{2m_2 \sigma^2}\right\} \\ &\quad \dots \frac{1}{\sqrt{2\pi m_k}\sigma} \exp\left\{-\frac{((x_k - x_{k-1}) - m_k \hat{\mu}_0)^2}{2m_k \sigma^2}\right\};\end{aligned}$$

and, under  $H_1$ ,

$$\begin{aligned}P(X^1 = x_1, X^2 = x_2, \dots, X^k = x_k) &= P(X^1 = x_1, X^2 - X^1 = x_2 - x_1, \dots, \\ &\quad X^k - X^{k-1} = x_k - x_{k-1}) \\ &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x_1 - \hat{\mu}_1)^2}{2\sigma^2}\right\} \frac{1}{\sqrt{2\pi m_2}\sigma} \\ &\quad \times \exp\left\{-\frac{((x_2 - x_1) - m_2 \hat{\mu}_2)^2}{2m_2 \sigma^2}\right\} \\ &\quad \dots \frac{1}{\sqrt{2\pi m_k}\sigma} \exp\left\{-\frac{((x_k - x_{k-1}) - m_k \hat{\mu}_k)^2}{2m_k \sigma^2}\right\}.\end{aligned}$$

This leads to

$$\begin{aligned}\Lambda &= \exp\left\{\frac{1}{2\sigma^2}[(x_1 - \hat{\mu}_1)^2 - (x_1 - \hat{\mu}_0)^2] \right. \\ &\quad + \frac{1}{2m_2\sigma^2}[((x_2 - x_1) - m_2 \hat{\mu}_2)^2 - ((x_2 - x_1) - m_2 \hat{\mu}_0)^2] \\ &\quad + \dots + \frac{1}{2m_k\sigma^2}[((x_k - x_{k-1}) - m_k \hat{\mu}_k)^2 \\ &\quad \left. - ((x_k - x_{k-1}) - m_k \hat{\mu}_0)^2\right\} \\ &= \exp\left\{\frac{1}{2\sigma^2}[2x_1(\hat{\mu}_1 - \hat{\mu}_0) + \hat{\mu}_1^2 - \hat{\mu}_0^2] \right. \\ &\quad + \frac{1}{2m_2\sigma^2}[2m_2(x_2 - x_1)(\hat{\mu}_2 - \hat{\mu}_0) + m_2^2(\hat{\mu}_2^2 - \hat{\mu}_0^2)] \\ &\quad + \dots + \frac{1}{2m_k\sigma^2}[2m_k(x_k - x_{k-1})(\hat{\mu}_k - \hat{\mu}_0) \\ &\quad \left. + m_k^2(\hat{\mu}_k^2 - \hat{\mu}_0^2)\right\}.\end{aligned}$$

This leads to the proposition

### Derivation of Theorem 2.1

Let  $Y_0$  be the measurement at the specific location,  $Y_1, Y_2, Y_3$ , and  $Y_4$  are the four measurements collected in  $\mathcal{D}_1$ . Note that  $Y_i \sim N(\mu, 1)$ . If  $Y_i$  is in the noise region,  $Y_i \sim N(0, 1)$ , and if  $Y_i$  is in the signal region,  $Y_i \sim N(\delta, 1)$ .

Use the same notation, we have  $X_1 = Y_0$ , and  $X_2 = \sum_{i=0}^4 Y_i$ . The test statistic is defined as

$$\begin{aligned}-2 \log \Lambda &= 2X_1(\hat{\mu}_1 - \hat{\mu}_0) + (\hat{\mu}_0^2 - \hat{\mu}_1^2) \\ &\quad + 2(X_2 - X_1)(\hat{\mu}_2 - \hat{\mu}_0) + m_2(\hat{\mu}_0^2 - \hat{\mu}_2^2).\end{aligned}$$

It is straightforward to have

$$\hat{\mu}_0 = \frac{1}{N} \sum_{i=0}^{N-1} Y_i, \quad \hat{\mu}_1 = Y_0, \quad \hat{\mu}_2 = \frac{1}{4} \sum_{i=1}^4 Y_i,$$

and  $m_2 = 4$ .

Note that  $\hat{\mu}_0 \sim N(n_s \delta / N, 1/N)$ . Similarly, we can derive the distribution of  $\hat{\mu}_1$  and  $\hat{\mu}_2$  as follows.

1. If all  $Y_0, Y_1, \dots, Y_4$  are in the noise region,  $\hat{\mu}_1 \sim N(0, 1)$ , and  $\hat{\mu}_2 \sim N(0, 1/4)$ .
2. If  $Y_0$  is a noise point, but at the boundary as shown in Figure 2 (a) and (b), then  $\hat{\mu}_1 \sim N(0, 1)$ , and  $\hat{\mu}_2 \sim N(2\delta/4, 1/4)$  (case (a)) or  $\hat{\mu}_2 \sim N(\delta/4, 1/4)$  (case (b)).
3. If  $Y_0$  is a signal point, but at the boundary as shown in Figure 2 (c) and (d), then  $\hat{\mu}_1 \sim N(\delta, 1)$ , and  $\hat{\mu}_2 \sim N(2\delta/4, 1/4)$  (case (c)) or  $\hat{\mu}_2 \sim N(3\delta/4, 1/4)$  (case (d)).
4. If  $Y_0, Y_1, \dots, Y_4$  are all in the signal region,  $\hat{\mu}_1 \sim N(\delta, 1)$ , and  $\hat{\mu}_2 \sim N(\delta, 1/4)$ .

Note that the test statistic,  $T$  can be written as

$$T = 2Y_0 \left( Y_0 - \frac{1}{N} \sum_{i=0}^{N-1} Y_i \right) + \left( \left( \frac{1}{N} \sum_{i=0}^{N-1} Y_i \right)^2 - Y_0^2 \right) \\ + 2 \sum_{i=1}^4 Y_i \left( \frac{1}{4} \sum_{i=1}^4 Y_i - \frac{1}{N} \sum_{i=0}^{N-1} Y_i \right) \\ + 4 \left( \left( \frac{1}{N} \sum_{i=0}^{N-1} Y_i \right)^2 - \left( \frac{1}{4} \sum_{i=1}^4 Y_i \right)^2 \right).$$

For each individual  $T$ , it can be shown that when  $N \rightarrow \infty$ , we have

$$T \xrightarrow{d} \hat{\mu}_1^2 + 4\hat{\mu}_2^2.$$

By using the above consideration on whether  $Y_0$  is in the noise region, boundary, and signal region, we have the following results

1. When all  $Y_0, Y_1, \dots, Y_4$  are in the noise region,  $\hat{\mu}_1^2 \sim \chi_1^2(0)$ ,  $(2\hat{\mu}_2)^2 \sim \chi_1^2(0)$ , and thus  $T \xrightarrow{d} \chi_2^2(0)$ .
2. When  $Y_0$  is in noise region, and as shown in Figure 2(a),  $\hat{\mu}_1^2 \sim \chi_1^2(0)$ ,  $2\hat{\mu}_2 \sim N(\delta, 1)$ , and thus  $(2\hat{\mu}_2)^2 \sim \chi_1^2(\delta^2)$ . This leads to  $T \xrightarrow{d} \chi_2^2(\delta^2)$ .
3. When  $Y_0$  is in noise region, and as shown in Figure 2(b),  $\hat{\mu}_1^2 \sim \chi_1^2(0)$ ,  $2\hat{\mu}_2 \sim N(\delta/2, 1)$ , and thus  $(2\hat{\mu}_2)^2 \sim \chi_1^2(\delta^2/4)$ . This leads to  $T \xrightarrow{d} \chi_2^2(\delta^2/4)$ .
4. When  $Y_0$  is in signal region, and as shown in Figure 2(c),  $\hat{\mu}_1^2 \sim \chi_1^2(\delta^2)$ ,  $2\hat{\mu}_2 \sim N(\delta, 1)$ , and thus  $(2\hat{\mu}_2)^2 \sim \chi_1^2(\delta^2)$ . This leads to  $T \xrightarrow{d} \chi_2^2(2\delta^2)$ .
5. When  $Y_0$  is in signal region, and as shown in Figure 2(d),  $\hat{\mu}_1^2 \sim \chi_1^2(\delta^2)$ ,  $2\hat{\mu}_2 \sim N(3\delta/2, 1)$ , and thus  $(2\hat{\mu}_2)^2 \sim \chi_1^2(9\delta^2/4)$ . This leads to  $T \xrightarrow{d} \chi_2^2(13\delta^2/4)$ .
6. When  $Y_0, Y_1, \dots, Y_4$  are in the signal region,  $\hat{\mu}_1^2 \sim \chi_1^2(\delta^2)$ ,  $2\hat{\mu}_2 \sim N(2\delta, 1)$ , and thus  $(2\hat{\mu}_2)^2 \sim \chi_1^2(4\delta^2)$ . This leads to  $T \xrightarrow{d} \chi_2^2(5\delta^2)$ .

Note that  $\hat{\mu}_1$  and  $\hat{\mu}_2$  are independent to each other. Thus, we will have following relations:

$$E(T) = \begin{cases} 2 & Y_0 \text{ is an inner point in the noise region,} \\ f(\delta) & Y_0 \text{ is an boundary point,} \\ g(\delta) & Y_0 \text{ is an inner point in the signal region,} \end{cases}$$

where  $2 + \delta^2/4 \leq f(\delta) \leq 2 + 13\delta^2/4$ , and  $g(\delta) = 2 + 5\delta^2$ . It is straightforward that, if  $\delta > 0$ ,

$$2 < f(\delta) < g(\delta).$$

Similarly, we have

$$V(T) = \begin{cases} 4 & Y_0 \text{ is an inner point in the noise region,} \\ v_1(\delta) & Y_0 \text{ is an boundary point,} \\ v_2(\delta) & Y_0 \text{ is an inner point in the signal region,} \end{cases}$$

where  $2(2 + \delta^2/2) \leq v_1(\delta) \leq 2(2 + 13\delta^2/2)$ , and  $v_2(\delta) = 2(2 + 10\delta^2)$ .

Let us assume that we have  $n_n$  inner points in the noise region,  $n_b$  boundary points, and  $n_s$  inner points in the signal region. The average  $T(y_0)$  will have

$$\frac{1}{n_n} \sum_{Y_0 \in \mathcal{N}} T(Y_0) \xrightarrow{d} N \left( 2, \frac{4}{n_n} \right), \\ \frac{1}{n_b} \sum_{Y_0 \in \mathcal{B}} T(Y_0) \xrightarrow{d} N \left( f(\delta), \frac{v_1(\delta)}{n_b} \right), \\ \frac{1}{n_s} \sum_{Y_0 \in \mathcal{S}} T(Y_0) \xrightarrow{d} N \left( g(\delta), \frac{v_2(\delta)}{n_s} \right).$$

Thus, if  $n_n, n_b$ , and  $n_s \rightarrow \infty$ , we will have stochastically

$$P \left( \frac{1}{n_n} \sum_{Y_0 \in \mathcal{N}} T(Y_0) < \frac{1}{n_b} \sum_{Y_0 \in \mathcal{B}} T(Y_0) < \frac{1}{n_s} \sum_{Y_0 \in \mathcal{S}} T(Y_0) \right) \rightarrow 1.$$

## Derivation of Theorem 2.2

Let  $\mathcal{B}$  be the boundary point, which is defined as at least one of its neighbor 4 points is from another group.  $\mathcal{B}^c$  be non-boundary point. The local variability is defined as  $\frac{1}{4} \sum_{i=0}^4 (Y_i - \bar{Y})^2$ . Let us focus on  $\tilde{V}(Y) = \sum_{i=0}^4 (Y_i - \bar{Y})^2$ . It is straightforward to show that if  $Y \in \mathcal{B}^c$ ,  $\tilde{V}(Y) \sim \chi_4^2(0)$ . Similarly as derivation in the former section, let  $Y_i \sim N(\mu_i, 1)$ . If  $Y_0 \in \mathcal{B}$ , we have  $\sum Y_i^2 \sim \chi_5^2(\sum \mu_i^2)$ . It can be shown that  $\mu_i^2 = k\delta^2$ , where  $k$  is the number of signal points in the local region (note that,  $k = 1, 2, 3, 4$ ). Let  $\bar{Y} = \frac{1}{5} \sum_{i=0}^4 Y_i$ , It is straightforward that  $\bar{Y} \sim N(k\delta/5, 1/5)$ , or  $\sqrt{5}\bar{Y} \sim N(k\delta/\sqrt{5}, 1)$ . And thus

$$\tilde{V}(Y) \sim \chi_4^2(C),$$

where

$$C = k\delta^2 - k^2\delta^2/5, \quad k = 1, 2, 3, 4.$$

Then,  $E(\tilde{V}(Y)) = 4 + C$ , and  $\text{Var}(\tilde{V}(Y)) = 8 + 4C$ , if  $Y \in \mathcal{B}$ . Similarly,  $E(\tilde{V}(Y)) = 4$ , and  $\text{Var}(\tilde{V}(Y)) = 8$ , when  $Y \in \mathcal{B}^c$ . By the central limit theorem, we have

$$\text{Ave}_{Y \in \mathcal{B}} \tilde{V}(Y) \xrightarrow{d} N \left( 4 + C, \frac{8 + 4C}{n_b} \right), \\ \text{Ave}_{Y \in \mathcal{B}^c} \tilde{V}(Y) \xrightarrow{d} N \left( 4, \frac{8}{N - n_b} \right).$$

And thus

$$\text{Ave}_{Y \in \mathcal{B}} \tilde{V}(Y) - \text{Ave}_{Y \in \mathcal{B}^c} \tilde{V}(Y) \xrightarrow{d} N \left( C, \frac{8 + 4C}{n_b} + \frac{8}{N - n_b} \right).$$

If we want to  $P(\text{Ave}_{Y \in \mathcal{B}} \tilde{V}(Y) - \text{Ave}_{Y \in \mathcal{B}^c} \tilde{V}(Y) > 0) \rightarrow 1$ , we need to have

$$\Phi(a) \rightarrow 1,$$

where  $a = \sqrt{\frac{CNp_B(1-p_B)}{8+4(1-p_B)}}$ .

Here  $p_B$  be the proportion of boundary point in the image. So if  $\frac{\delta^2 N p_B (1-p_B)}{8+4(1-p_B)} \rightarrow \infty$ , then  $a \rightarrow \infty$ . This leads to  $P(\text{Ave}_{Y \in \mathcal{B}} \tilde{V}(Y) - \text{Ave}_{Y \in \mathcal{B}^c} \tilde{V}(Y) > 0) \rightarrow 1$ . And then  $P(\text{Ave}_{Y \in \mathcal{B}} V(Y) - \text{Ave}_{Y \in \mathcal{B}^c} V(Y) > 0) \rightarrow 1$ .

Received 4 January 2012

## REFERENCES

- BAKER, R. D. (2004). A modified knox test of space-time clustering. *Journal of Applied Statistics* **31** 457–463. [MR2080594](#)
- BENJAMINI, Y. and HOCHBERG, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57** 289–300. [MR1325392](#)
- CHAUDHURI, P. and MARRON, J. (1999). SiZer for exploration of structures in curves. *Journal of the American Statistical Association* **94** 807–823. [MR1723347](#)
- CHAUDHURI, P. and MARRON, J. (2000). Scale space view of curve estimation. *The Annals of Statistics* **28** 408–428. [MR1790003](#)
- COSTA, M. and KULLDORFF, M. (2009). Applications of spatial scan statistics: A review. *Scan Statistics* 129–152. [MR2796293](#)
- DIGGLE, P., ROWLINGSON, B. and SU, T.-L. (2005). Point process methodology for on-line spatio-temporal disease surveillance. *Environmetrics* **16** 423–434. [MR2147534](#)
- GLAZ, J. and BALAKRISHNAN, N. (1999). *Scan Statistics and Applications*, Birkhäuser, Boston. [MR1697781](#)
- GLAZ, J., NAUS, J. and WALLENSTEIN, S. (2001). *Scan Statistics*, Springer Verlag. [MR1869112](#)
- GLAZ, J., POZDNYAKOV, V. and WALLENSTEIN, S. (2009). *Scan Statistics: Methods and Applications*, Birkhauser. [MR2761094](#)
- KULLDORFF, M. (1999). Spatial scan statistics: Models, calculations, and applications. *Scan Statistics and Applications* 303–322. [MR1697758](#)
- KULLDORFF, M. (2010). *SaTScan-Software for the Spatial, Temporal, and Space-Time Scan Statistics*. Harvard Medical School and Harvard Pilgrim Health Care, Boston.
- KULLDORFF, M., HUANG, L., PICKLE, L. and DUCZMAL, L. (2006). An elliptic spatial scan statistic. *Statistics in Medicine* **25** 3929–3943. [MR2297401](#)
- KULLDORFF, M., TANGO, T. and PARK, P. J. (2003). Power comparisons for disease clustering tests. *Computational Statistics & Data Analysis* **42** 665–684. [MR1977177](#)
- LIANG, S., BANERJEE, S. and CARLIN, B. P. (2009). Bayesian wombling for spatial point processes. *Biometrics* **65** 1243–1253. [MR2756512](#)
- LINDBERG, T. (1993). *Scale-Space Theory in Computer Vision*, Springer.
- LINDBERG, T. (1994). Scale-space theory: A basic tool for analyzing structures at different scales. *Journal of Applied Statistics* **21** 225–270.
- LOH, J. M. (2011). K-scan for anomaly detection in disease surveillance. *Environmetrics* **22** 179–191. [MR2843350](#)
- LOH, J. M., LINDQUIST, M. A. and WAGER, T. D. (2008). Residual analysis for detecting mis-modeling in fMRI. *Statistica Sinica* **18** 1421–1448. [MR2468275](#)
- MAITRA, R. (2009). Assessing certainty of activation or inactivation in test-retest fMRI studies. *Neuroimage* **47** 88–97. doi: 10.1016/j.neuroimage.2009.03.073.
- MAITRA, R. (2010). A re-defined and generalized percent-overlap-of-activation measure for studies of fMRI reproducibility and its use in identifying outlier activation maps. *Neuroimage* **50** 124–135. doi: 10.1016/j.neuroimage.2009.11.070.
- MAITRA, R., ROYS, S. R. and GULLAPALLI, R. P. (2002). Test-retest reliability estimation of fMRI data. *Magnetic Resonance in Medicine* **48** 62–70.
- NAUS, J. (1965a). Clustering of random points in two dimensions. *Biometrika* **52** 263–267. [MR0211433](#)
- NAUS, J. (1965b). The distribution of the size of the maximum cluster of points on a line. *Journal of the American Statistical Association* 532–538. [MR0183041](#)
- STOREY, J. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64** 479–498. [MR1924302](#)
- WHEELER, D. C. (2007). A comparison of spatial clustering and cluster detection techniques for childhood leukemia incidence in Ohio, 1996–2003. *International Journal of Health Geographics* **6**.
- ZHANG, C., FAN, J. and YU, T. (2011). Multiple testing via  $FDR_l$  for large-scale imaging data. *Annals of Statistics* **39** 613–642. [MR2797858](#)
- ZHANG, L., ZHU, Z., JEFFAY, K., MARRON, J. and SMITH, F. (2008). Multi-resolution anomaly detection for the internet. In: *Proc. INFOCOM Workshops 2008, IEEE, IEEE*, pp. 1–6.
- ZHANG, L., ZHU, Z. and MARRON, J. S. (2007). Multiresolution anomaly detection method for long range dependent time series. Technical Report, UNC/STOR/07/12.
- ZHANG, T. and LIN, G. (2009). Spatial scan statistics in loglinear models. *Computational Statistics and Data Analysis* **53** 2851–2858. [MR2667595](#)

Lingsong Zhang  
Department of Statistics  
Purdue University  
150 N. University St.  
West Lafayette, IN 47906  
USA  
E-mail address: [lingsong@purdue.edu](mailto:lingsong@purdue.edu)

Zhengyuan Zhu  
Department of Statistics & Statistical Laboratory  
Iowa State University  
Snedecor Hall  
Ames, IA 50011-1210  
USA  
E-mail address: [zhuz@iastate.edu](mailto:zhuz@iastate.edu)