

Spatial CUSUM for Signal Region Detection

Abstract

Detecting weak clustered signal in spatial data is a useful but challenging task in applications such as medical image and epidemiology. A more efficient detection algorithm can provide more precise early warning, and effectively reduce the decision risk and cost. To date, many methods have been developed to detect signals with spatial structure. However, most of the existing methods are either too conservative or computationally too intensive. In this paper, we consider a novel method, named Spatial CUSUM (SCUSUM), which combines the idea of CUSUM procedure and false discovery rate controlling. We develop theoretical properties of the method which indicates that asymptotically SCUSUM can reach high classification accuracy. In the simulation study we demonstrate that our SCUSUM method is more sensitive and better at detecting weak spatial signals compared to some existing method. This new method is applied to a real fMRI dataset as illustration, and more irregular weak spatial signals are detected in the images.

Keywords: Spatial signal detection, monitoring-warning system, weak signal detection, CUSUM, FDR, weak dependence, fMRI.

1. Introduction

Spatial signal detection is an important topic in many fields, including astrophysics ([Abazajian and Kaplinghat \(2012\)](#); [Gladders and Yee \(2000\)](#)), brain imaging analysis ([Craddock et al. \(2012\)](#); [Zhang et al. \(2011\)](#); [Blumensath et al. \(2013\)](#); [Shen et al. \(2013\)](#)), epidemiology ([Kulldorff and Nagarwalla \(1995\)](#); [Tango \(2000\)](#); [Wheeler \(2007\)](#)), meteorology ([Sun et al. \(2015\)](#)) etc. Typically, given a spatial domain \mathcal{D} , e.g. a brain image or a geographical map, if there is no spatial signal, all the observations could be regarded to follow the same distribution. While with the existing of spatial signals, the responses within a unknown sub-region are from a different distribution. Locating signal region with low signal-noise ratio is meaningful in the early detection and warning systems: in the early stages of abnormality, the spatial signal is very weak compared with the measurement noise; however, an accurate early warning could effectively reduce the decision risk and avoid unnecessary but lethal cost. This kind of warning systems has been studied and applied in many practical situations, e.g. disease monitoring ([Thomson and Connor \(2001\)](#); [Grover-Kopec et al. \(2005\)](#)), weather monitoring ([Breed \(2011\)](#)) etc. Therefore, there will be a huge breakthrough if weak spatial signals could be efficiently identified.

So far, many methods and algorithms have been developed for spatial signal detection. One class of methods to identify spatial clusters is the spatial scan statistics ([Glaz et al. \(2009\)](#); [Glaz et al. \(2001\)](#); [Priebe et al. \(2005\)](#); [Glaz and Balakrishnan \(2012\)](#) etc.) Scan statistics, also known as window statistics, was first proposed in [Naus \(1965\)](#). The idea is to perform likelihood ratio tests on all the scan windows can windows of different sizes and locations and identify the significant windows as clusters. This method was designed to find unusual clusters of randomly positioned points. [Naus \(1982\)](#) developed the asymptotic distribution for the scan statistics and proposed the method to find the maximum cluster of points on a line or circle, the length of the longest success run in

Bernoulli trials, and the generalized birth-day problem. [Kulldorff \(1999\)](#) extended the framework of the conventional scan statistics to multidimensional scenario, including two-dimensional scan statistics on the plane or on a sphere, three-dimensional scan statistics in space or in spacetime. [Glaz and Zhang \(2004\)](#) derived a variant of scan statistics, called multiple scan statistics of variable window sizes, for independent and identically distributed Bernoulli trials. They showed that using multiple scan statistics could improve the effectiveness in detecting signals. However, one of the main problems for scan statistics is that: if the shape of the true cluster is not circle or ellipsoid, the power of the traditional scan statistics will significantly reduce. Also, without p -value adjustment, the detection result from scan statistics might be too aggressive ([Zhang et al. \(2010\)](#).)

Another branch of them is based on multiple testing and false discovery rate (FDR) controlling ([Benjamini and Hochberg \(1995\)](#); [Benjamini and Yekutieli \(2001\)](#); [Genovese et al. \(2002\)](#); [Miller et al. \(2001\)](#); [Zhang et al. \(2011\)](#); [Tango \(2000\)](#); [Sun et al. \(2015\)](#).) Multiple hypothesis testing is concerned with testing several statistical hypotheses simultaneously, and false discovery rate is a criterion designed to control the expected proportion of rejected null hypotheses that are incorrect rejections:

$$\text{FDR} = \mathbb{E}\left[\frac{\#\text{incorrect rejections}}{\#\text{rejected null hypotheses}}\right]. \quad (1)$$

In spatial signal detection, the statistical hypotheses are about whether locations belongs to signal region or not. [Genovese et al. \(2002\)](#) applied multiple testing to functional neuroimaging data and used FDR to find a threshold for signal classification. Their experiments showed that FDR worked more conservative when the correlations between hypotheses are high. [Miller et al. \(2001\)](#) performed FDR to astrophysical data, and showed that FDR had a similar rate of correct detections but with significantly fewer false detections compared with some standard testing procedures. In [Tango \(2000\)](#), multiple testing are used on disease map data in the Tokyo Metropolitan Area to detecting spatial disease clusters. To consider the spatial information as well as improve detecting effectiveness, [Zhang et al. \(2011\)](#) proposed a testing procedure named FDR_L . Via aggregating the p -value nearby, FDR_L could avoid the lack of identification phenomenon and improve the detection sensitivity. [Sun et al. \(2015\)](#) developed an oracle procedures, which optimally control the false discovery rate, false discovery exceedance and false cluster rate, for multiple testing of spatial signals. And the tropospheric ozone data in eastern USA were analyzed with their method to show the detection effectiveness. Although FDR and its variant methods have good statistical interpretation and easy implementation, they require the priori knowledge about the null distribution and also they might miss many signals when signal-noise ratio is small.

In this paper, we will introduce a novel detecting method, named Spatial CUSUM (SCUSUM), to identify spatial signal region, which is noise-distribution free method. In this method, we only assume the mean of signal region is different (usually higher) from that of indifference region, and we do not care much about the distribution of noise process as long as it's zero-mean and independence. SCUSUM consists of two parts: first applying moving window and CUSUM cut-off to estimate signal weight for each location, then determining a threshold with FDR controlling. Moving window method has been broadly used on analyzing time series data and spatial data ([Páez et al. \(2008\)](#); [Haas \(1990\)](#)). The idea is to utilize the data in the neighbor region to capture local feature. In our work, we use this idea to project the spatial domain into a sequence, which could be analyzed by CUSUM procedure. The CUSUM procedure (or cumulative summation) is a well-known method to locate changepoints in time series ([Horváth and Hušková \(2012\)](#); [Cho et al. \(2016\)](#); [Gromenko et al. \(2017\)](#); [Wang and Samworth \(2018\)](#); [Aue et al. \(2009\)](#).) Through the CUSUM transformation,

tion, the testing statistics could be compared with standard Brownian Bridge, to test whether the changepoint exists no not. If it exists, then the location of changepoint is the point reaching the maximum of CUSUMs. In our work, via repeating CUSUM cut-off with moving window, the frequency of being detected could be calculated for each location. We define this frequency as signal weight. Then, with bounded density estimation method, we could approximate the null density and alternative density. A proper threshold could be found based on FDR to identify the singal region. Our theoretical results show that SCUSUM could asymptotically reduce the misclassification rate to zero with probability 1. The experiment parts provide the evidence that our easily-implementing method could detect more weak spatial signals, compared with some existing methods.

The rest of the paper is organized as follows. The problem is formulated mathematically in Section 2. We will describe our proposed method and its theoretical properties in Section 3. Section 4 present simulation comparisons between SCUSUM and FDR_L under different signal strengths and noise dependence structures. An application of four methods (SCUSUM, scan statistics, FDR and FDR_L), to a real fMRI data is given in Section 5. And conclusions are given in Section 6. The proofs are shown in Appendix.

2. Problem formulation

Let \mathcal{D} be the entire spatial domain, s present the location belonging to \mathcal{D} and $x(s)$ be the observed data at location s . Consider $\mathcal{D}_{\mathcal{A}}$ be the signal region in \mathcal{D} , and its complement set $\mathcal{D}_{\mathcal{A}}^c$ be the indifference region. We assume that under H_0 , there is no signal region (i.e. $\mathcal{D}_{\mathcal{A}} = \emptyset$) and $x(s)$ has the same mean μ ; while under H_1 , $x(s)$ has mean μ_1 if $s \in \mathcal{D}_{\mathcal{A}}$ and μ_0 if $s \in \mathcal{D}_{\mathcal{A}}^c$.

Hence, the following additive model for the random variables $X = \{x(s), s \in \mathcal{D}\}$ is considered:

$$x(s) = \mu_0 \mathbb{I}(s \in \mathcal{D}_{\mathcal{A}}^c) + \mu_1 \mathbb{I}(s \in \mathcal{D}_{\mathcal{A}}) + \epsilon(s), \quad (2)$$

where both μ_0 and μ_1 are the unknown mean (w.l.o.g, assume $\mu_1 \geq \mu_0$), $\epsilon(s)$ is the independent noise and $\mathbb{I}(\cdot)$ is the indicator function. Note here, we don't assign any distribution model to the noise. The only requirement for noise is that it has zero mean and i.i.d. Our goal is to identify the signal region $\mathcal{D}_{\mathcal{A}}$.

3. Proposed method

In this section, we will give the details of our proposed method named Spatial CUSUM (SCUSUM), which has two steps: 1) For each location, we first estimate the signal weight, which is expected to be large in the signal region $\mathcal{D}_{\mathcal{A}}$, while small in the indifference region $\mathcal{D}_{\mathcal{A}}^c$. 2) Given a significant level α , a threshold is determined based on FDR idea. In section 3.1, we describe the way of using the moving window idea to project a spatial domain into a sequence and then estimating the signal weight for each location with the CUSUM cut-off. In section 3.2, we introduce how to estimate the null distribution f_{H_0} and alternative distribution f_{H_1} with estimated signal weights, following by the step to determine the detection threshold. In section 3.3, we briefly discuss neighbor size selection for moving window in the first step.

3.1. The first step: signal weight estimation

Our signal weight estimation method is inspired by the CUSUM procedure for changepoint detection in time series. However, for spatial signal detection, the conventional CUSUM is impractical,

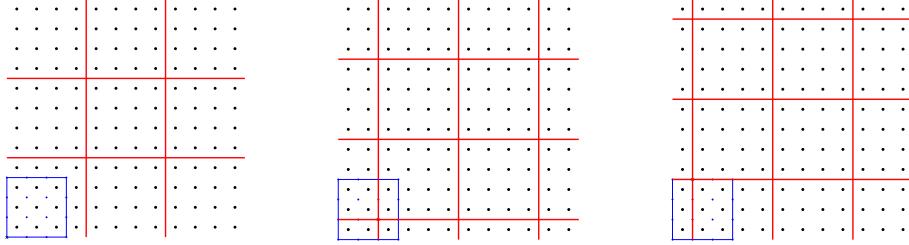


Figure 1: Moving window to divide spatial domain: The blue points inside blue square is $\{(p + \frac{1}{2}, q + \frac{1}{2}), p = 1 : k, q = 1 : k\}$, we select initial grid point from the set; The black points present the observed locations; red grids are boundary lines for blocks.

mainly due to lack of natural order for spatial observations, which are located in \mathbb{R}^2 or \mathbb{R}^3 (in our work, we focus on \mathbb{R}^2 .) Hence, we consider to use the moving window technique to construct proper sequences.

Given a square spatial domain \mathcal{D} and a neighbor size k , we can select one point (x, y) from $\{(p + \frac{1}{2}, q + \frac{1}{2}), p = 1 : k, q = 1 : k\}$. Then we can divide \mathcal{D} into b non-overlapping blocks $\{B_i\}_{i=1}^b$ of size $k \times k$, with (x, y) as the initial grid point (see Figure 1.) So we have $n = \sum_{i=1}^b n_i$, where n is the total number of observations and n_i is the observations in the block B_i .

According to model 2, if block B_i is inside the signal region, i.e. $B_i \in \mathcal{D}_s$, then all the observations in it follow $x(s) = \mu_1 + \epsilon(s)$, $\forall x \in B_i$; if B_i is inside the indifference region, i.e. $B_i \in \mathcal{D}_d^c$, then $x(s) = \mu_0 + \epsilon(s)$, $\forall x \in B_i$; if B_i is at the boundary of \mathcal{D}_s and \mathcal{D}_d^c , then the observations $x(s)$ follows a mixture model: $x(s) = \mu_1 + \epsilon(s)$ with probability p_i and $x(s) = \mu_0 + \epsilon(s)$ with $1 - p_i$, where p_i is the ratio of signal points inside B_i (see Figure 2.)

Next, based on the above division, we construct two sequences to capture the feature of these spatial observations. The first sequence is sample sequence: A random sample, denoted as γ_i , is drawn from block B_i . It could be regarded as the 'representative' for this block. Meanwhile, we construct the second sequence by computing the block mean without the 'representative', $\tilde{\mu}_i = \frac{\sum_{x \in B_i} x - \gamma_i}{\sum_{x \in \mathcal{D}} \mathbb{I}(x \in B_i) - 1}$. As the number of observations n_i in B_i increases, the pseudo block mean gets closer to the true block mean, i.e. $\tilde{\mu}_i \rightarrow \frac{\sum_{x \in B_i} x}{\sum_{x \in \mathcal{D}} \mathbb{I}(x \in B_i)}$. Hence, the pseudo block mean $\tilde{\mu}_i$ could present the local block mean. Based on the analysis in the last paragraph, we

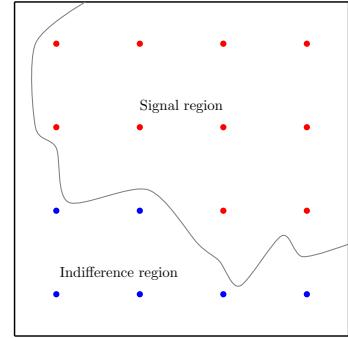


Figure 2: The block at the boundary of signal region and indifference region. In this case, $p_i = 0.625$.

could easily derive the following results for $\{\gamma_i\}_{i=1}^b$ and $\{\tilde{\mu}_i\}_{i=1}^b$:

$$\gamma_i = \begin{cases} \mu_1 + \epsilon, & \text{if } B_i \in \mathcal{D}_{\mathcal{A}} \\ \mu_0 + \epsilon, & \text{if } B_i \in \mathcal{D}_{\mathcal{A}}^c \\ \mu_1 z + \mu_0(1 - z) + \epsilon, & \text{if } B_i \text{ at boundary} \end{cases} \quad (3)$$

$$\mathbb{E}[\tilde{\mu}_i] = \begin{cases} = \mu_1, & \text{if } B_i \in \mathcal{D}_{\mathcal{A}} \\ = \mu_0, & \text{if } B_i \in \mathcal{D}_{\mathcal{A}}^c \\ \approx p_i \mu_1 + (1 - p_i) \mu_0, & \text{if } B_i \text{ at boundary} \end{cases} \quad (4)$$

where $z \sim Ber(p_i)$. (3) and (4) show that even though $\{\gamma_i\}_{i=1}^b$ and $\{\tilde{\mu}_i\}_{i=1}^b$ are independent (see Lemma 1), they have similar patterns: the closer B_i is to $\mathcal{D}_{\mathcal{A}}$, the more likely it has large γ_i and $\tilde{\mu}_i$, and vice versa. Hence, we could consider to rearrange $\{\gamma_i\}_{i=1}^b$ according to the decreasing order of $\{\tilde{\mu}_i\}_{i=1}^b$, denoted as $\{\gamma_i^*\}_{i=1}^b$. Intuitively, if there is no signal region, then $\{\gamma_i^*\}$ should be around μ_0 ; otherwise, $\{\gamma_i^*\}_{i=1}^b$ should have three parts: the first part presenting signal blocks is around μ_1 , the second part is the interim from μ_1 to μ_0 and the third part is indifference blocks around μ_0 (see Figure 3.)

Lemma 1 *Based on model 2, $\{\gamma_i\}$ and $\{\tilde{\mu}_i\}$ are independent. As the number of observations in each block n_i goes to infinity, i.e. $\min n_i \rightarrow \infty$, we have the following: under the null hypothesis H_0 : there is no signal region, then $\{\gamma_i^*\}$ is an i.i.d sequence; under the alternative hypothesis H_1 : signal region exists, then there exists l_1 and l_2 with $0 \leq l_1 < l_2 \leq b$,*

$$\mathbb{E}[\gamma_i^*] = \begin{cases} \mu_1, & \text{if } 0 \leq i < l_1 \\ \in (\mu_0, \mu_1) & \text{if } l_1 \leq i < l_2 \\ \mu_0, & \text{if } l_2 \leq i \leq n, \end{cases} \quad (5)$$

and l_1 is the number of blocks inside $\mathcal{D}_{\mathcal{A}}$, and $(b - l_2)$ is the number of blocks inside $\mathcal{D}_{\mathcal{A}}$.

Lemma 1 shows that under H_1 , the projected sequence $\{\gamma_i^*\}_{i=1}^b$ has a changepoint in $[l_1, l_2]$ and the conventional CUSUM could help locate a cut-off index near or inside the interval. First, we compute the CUSUM statistics for $\{\gamma_i^*\}_{i=1}^b$ at each location:

$$\tilde{\gamma}_r = \left| \sum_{i=1}^r \gamma_i^* - \frac{r}{b} \sum_{i=1}^b \gamma_i^* \right|. \quad (6)$$

Then the cut-off index is $t = \arg \max_i \tilde{\gamma}_i$. The following theorem guarantees the accuracy of the cut-off index.

Theorem 1 *Under the alternative hypothesis: if the signal region exists, as the number of block b and the number of observations in each block n_i go to infinity, then the cut-off index t based on CUSUM procedure will fall into the interval $[l_1, l_2]$ with probability 1, i.e. $\mathbb{P}(l_1 \leq t \leq l_2) = 1$ as $b \rightarrow \infty$ and $\min n_i \rightarrow \infty$.*

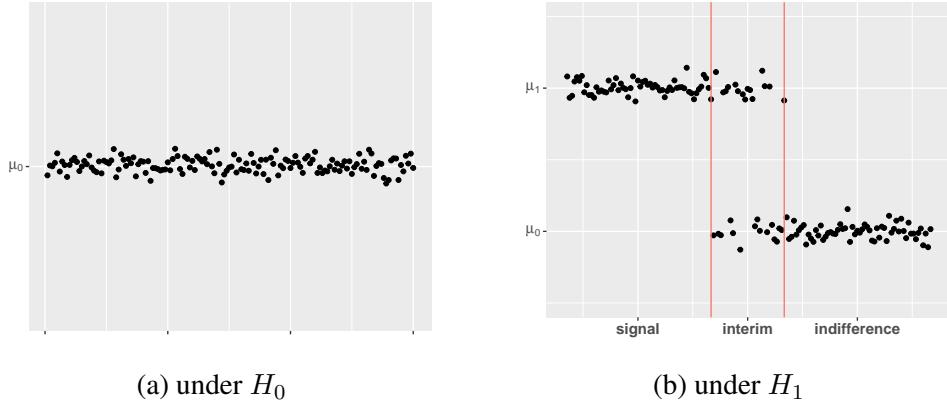


Figure 3: The possible patterns of $\{\gamma_i^*\}_{i=1}^b$: (a) presents the scenario without signal region and $\{\gamma_{(i)}^*\}_{i=1}^b$ are around μ_0 ; (b) shows the pattern with signal region and there are three parts: signal, interim and indifference.

Theorem 1 ensures that the cut-off procedure could asymptotically separate the signal region and indifference (see Figure 3 (b)). Also, with the given spatial domain, as the block size becomes finer (equivalent to $b \rightarrow \infty$), the number of the blocks at the boundary is decreasing. Hence we have $(l_2 - l_1)/b \rightarrow 0$. Combining the results from Theorem 1, the number of misclassified locations goes to zero.

Above theoretical results require $b \rightarrow \infty$ and $\min n_i \rightarrow \infty$. In practice, with limited observations, the detected result might be affected by the initial point selection, especially when the signal region is not regular. Thus, we could eliminate the effect of initial point by going through all the possible initial points (see Figure 1.) We summarize our method in Algorithm 1. Also, we could sufficiently extract the local information and eliminate the effect of randomly sampled 'representatives' by repeated Algorithm 1 more than once. With these steps, we could estimate the signal weights $\{w(s)\}$ (or $\{\tilde{w}(s)\}$) by computing the detected frequency of each location. The larger signal weight means the location is more likely to belong to the signal region.

3.2. The second step: Threshold estimation with FDR

With the estimated signal weight and a given significant level α , we could identify the signal region with FDR. Firstly, the weights are in $[0, 1]$, and could be considered as the possibilities that the locations have signals. Thus we could use density estimation with the boundary correction method to estimate the distribution of the signal weights, $f(x)$. Many density estimations have been studied in previous works (Chen (1999); Cowling and Hall (1996); Jones and Foster (1996); Cattaneo et al. (2017).) In our work, we use the local polynomial density estimation method from Cattaneo et al. (2017).

In the following, we analyze the characteristic of $f(x)$, which could help us estimate the threshold. Under the null hypothesis H_0 , with Lemma 1, we know that $\{\gamma_{(i)}^*\}_{i=1}^b$ are i.i.d. and the corresponding blocks are random indexed. Hence with CUSUM cut-off procedure, the distribution for signal weight $f(x)$ is symmetric and has lower value with $x = 0$ and $x = 1$ (see Figure 4 (a).)

Algorithm 1 Moving Window detecting method for signal weight

Require: observed data $\{x(s)\}$ on grid $\{(p, q), p = 1 : n, q = 1 : n\}$, neighbor size k , repeat times m ;

Ensure: corresponding signal weights $\{w(s)\}$ or $\{\tilde{w}(s)\}$;

- 1: **for** (x, y) in $\{(p + \frac{1}{2}, q + \frac{1}{2}), p = 1 : k, q = 1 : k\}$ **do**
- 2: Divide \mathcal{D} into blocks $\{B_i\}_{i=1}^b$ of size $k \times k$ based on (x, y) ;
- 3: Sample one observation from each block γ_i ;
- 4: Estimate the block mean $\tilde{\mu}_i$;
- 5: Reorder $\{\gamma_i\}_{i=1}^b$ according to $\{\tilde{\mu}_i\}_{i=1}^b$ decreasingly as $\{\gamma_i^*\}_{i=1}^b$;
- 6: Conduct CUSUM transformation on $\{\gamma_i^*\}_{i=1}^b$ as $\{\tilde{\gamma}_{(i)}^*\}_{i=1}^b$;
- 7: Find the location t where $\{\tilde{\gamma}_{(i)}^*\}_{i=1}^b$ reaches maximum;
- 8: Define the blocks corresponding to the first t elements in $\{\gamma_i\}_{i=1}^b$ as signal block, and so do the observations in these blocks;
- 9: **end for**
- 10: Compute corresponding signal weight $w(s) = \frac{\text{detected times for } x(s)}{k^2}$;
(Option)
- 11: Repeat above produce m times and obtain $\{w^i(s)\}_{i=1}^m$;
- 12: Compute the average signal weights at each location $\{\tilde{w}^i(s)\} : \tilde{w}^i(s) = \sum_{i=1}^m w^i(s)/m$;

Lemma 2 Under the null hypothesis H_0 , as the number of block b and the number of observations in each block n_i go to infinity, then the density for signal weights $f(x)$ is symmetric.

Under the alternative hypothesis H_1 , the distribution $f(x)$ should be composed by the null part $f_{H_0}(x)$ and alternative part $f_{H_1}(x) : f(x) = f_{H_0}(x) + f_{H_1}(x)$. The observations inside the signal region are more likely to be detected, i.e. the corresponding signal weight gets close to 1 and vice versa for the observations inside the indifference region. With finer block division, the fraction of the observations in the blocks at the boundary goes to 0. Therefore, $f_{H_1}(x)$ has the 'peak' near 1 and $f_{H_0}(x)$ has the 'peak' around 0, which implies that $f(x)$ has two 'peaks' near the boundaries separately and a 'valley' in the middle of $[0, 1]$. Then we can use the line search to locate the 'valley', say $(t^*, f(t^*))$, and conduct linear interpolation between the two points $(t^*, f(t^*))$ and $(1, 0)$. Obviously, the null density $f_{H_0}(x)$ is controlled by

$$\tilde{f}_{H_0}(x) = \begin{cases} f(x), & \text{if } 0 \leq x \leq t^* \\ f(t^*)(1 - \frac{x - t^*}{1 - t^*}), & \text{if } t^* < x \leq 1 \end{cases} \quad (7)$$

which could be used as estimated null density (see Figure 4 (b).) Recall the definition of the marginal false discovery rate (mFDR) (Genovese and Wasserman (2002); Sun et al. (2015)):

$$\text{mFDR} = \frac{\mathbb{E}[\#\text{false positive}]}{\mathbb{E}[\#\text{rejected}]} \quad (8)$$

Thus, we could control mFDR with given significant level α by finding a threshold c so that

$$c = \arg \min_x \left(\frac{\tilde{f}_{H_0}(x)}{f(x)} \leq \alpha \right). \quad (9)$$

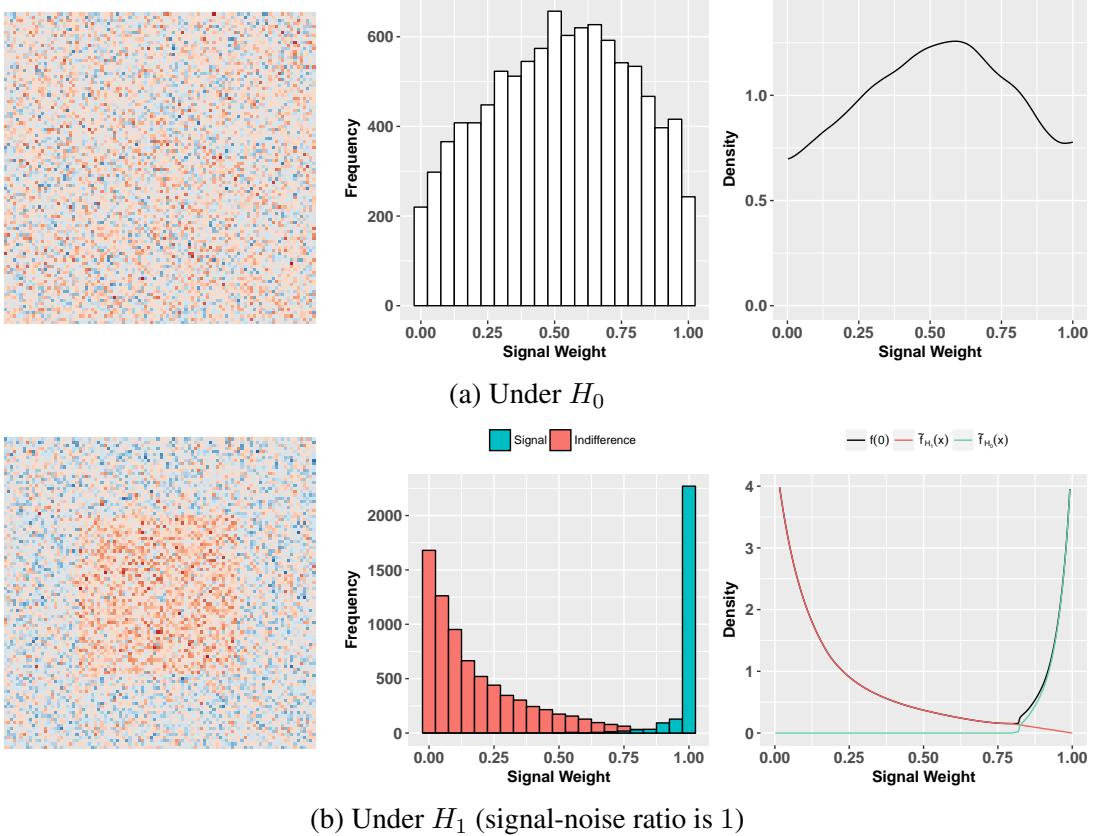


Figure 4: The idea for the second step: (a) Under H_0 , there is no signal (first column.) The histogram of signal weights are shown in the second column and it's symmetric with 'peak' around 0.5. The estimated density is shown in the third column; (b) Under H_1 , the histogram (the second column) is composed with two parts and has higher values around boundaries 0 and 1, lower values at the middle part. The estimated densities are shownd in third column: the black curve is $f(x)$, green one is estimated null density $\tilde{f}_{H_0}(x)$ and red one is estimated alternative density $\tilde{f}_{H_1}(x) = f(x) - \tilde{f}_{H_0}(x)$.

Algorithm 2 The signal region detection method

Require: Signal weights $\{w(s)\}$ or $\{\tilde{w}^i(s)\}$, significant level α ;

Ensure: corresponding detected result;

- 1: Estimate the density curve $f(x)$ based on signal weights, $x \in [0, 1]$;
 - 2: Estimate null density $f_{H_0}(x)$ and alternative density $f_{H_1}(x)$ with (7);
 - 3: Compute mFDR and find the threshold c with (9);
 - 4: Obtain the detected result with the threshold c ;
-

The observations with signal weight larger than c are the detected signals. This step is summarized in Algorithm 2.

3.3. Neighbor size selection

In this part, we consider the selection of neighbor size k , and this mainly affects the accuracy in Section 3.1. Intuitively, the larger k means the larger block and tends to over-smooth; while the smaller k implying the small block might lose spatial information.

In order to make the "right" cut-off, we need to ensure two points: 1) the variance of the pseudo block mean $\{\tilde{\mu}_i\}$ should be as small as possible, so that we could reasonably rearrange the 'representatives' $\{\gamma_i\}$; 2) the length of the rearranged 'representative' sequence $\{\gamma_i^*\}$ should be as long as possible, which could ensure the cut-off location t fall into $[l_1, l_2]$ with probability 1. For the first point, we need to make the number of observations in each block n_i go to infinity; for the second point, the length of $\{\gamma_i^*\}$ is the number of blocks b . And the relationship between b , k and n_i could be approximated as:

$$\begin{cases} \min n_i \approx k^2 \\ b \approx n/k^2 \end{cases} \quad (10)$$

Hence, we could get the following trade-off problem:

$$k_{opt} = \arg \min_k k^2 + C_1 \frac{n}{k^2} = \sqrt[4]{C_1 n}, \quad (11)$$

where C_1 is a given weight to reflect which part we want to emphasize and n is the total number of observations.

Of course, the above analysis is based on the theoretical result. In practice, the neighbor size selection depends on the specific problem and application. Related discussions on neighbor size selection could be found in existing works (Wang et al. (2006); Hall et al. (1995); Sun et al. (2015).)

4. Simulation Study

In this section, we will use simulation to show the effectiveness of our proposed method. We compare SCUSUM with FDR_L , because the two methods are designed to detect irregular signals with false discovery rate controlling. All the examples are simulated in the image with 100×100 pixels. Although in model 2 we didn't specify the distribution for noise process, we consider independent standard normal distribution $N(0, 1)$ for noise term and generate the data according to the model:

$$x(i, j) = \mu(i, j) + \epsilon(i, j), \quad i, j = 1, \dots, 100, \quad (12)$$

where $\mu(i, j) = 0$ for $(i, j) \in \mathcal{D}_{\mathcal{A}}^c$, and $\mu(i, j) \neq 0$ for $(i, j) \in \mathcal{D}_{\mathcal{A}}$, the 'L' shape and 'H' shape shown in Figure 5: the black region is signal region $\mathcal{D}_{\mathcal{A}}$ and white region is the indifference part.. The total number of signal pixels is 1288. Here we mainly concern about the accuracy of classification, both false positive and false negative. We set the repeated time m in Algorithm 1 as 50. Additional, for FDR_L , we do a standard normal test on each pixels and then apply the algorithm on the corresponding p -values.

In Table 1, the simulation results for SCUSUM and FDR_L are shown. Under each setting, we repeat simulation 100 times. For the two algorithms, we preset the significant level α as 0.05. In the signal region, μ ranges from 0.8 to 2. Also to show the effect of neighbor size selection, we choose k from $\{3, 5, 10\}$. We can see that when our method could control FDR under the presetted significant level $\alpha = 0.05$ with small neighbor size $k = 3, 5$; while FDR_L would allow FDR a little

Table 1: Detected accuracy comparision between SCUSUM and FDR_L : the signal-noise ratio is changed from 0.8 to 2 and neighbor size is chose from $\{3, 5, 10\}$. In the two algorithms, significant level α is 0.05.

neighbor size	Signal μ	SCUSUM			FDR_L		
		false negative	false positive	FDR	false negative	false positive	FDR
k=3	0.8	0.6338	0.0010	0.0186	0.9831	0.0003	0.0917
	1	0.4286	0.0010	0.0115	0.9380	0.0006	0.0484
	1.5	0.1953	0.0005	0.0044	0.5158	0.0039	0.0508
	2	0.1271	0.0006	0.0042	0.1478	0.0072	0.0535
k=5	0.8	0.3750	0.0009	0.0094	0.8034	0.0020	0.0586
	1	0.2750	0.0009	0.0082	0.5351	0.0047	0.0612
	1.5	0.1599	0.0019	0.0147	0.0971	0.0112	0.0770
	2	0.1014	0.0027	0.0194	0.0159	0.0150	0.0928
k=10	0.8	0.3240	0.0066	0.0588	0.1433	0.0286	0.1793
	1	0.2526	0.0084	0.0692	0.0569	0.0394	0.2163
	1.5	0.1642	0.0133	0.0968	0.0076	0.0651	0.3051
	2	0.1288	0.0149	0.1034	0.0027	0.0751	0.3358

bit higher than 0.05. When the neighbor size k is small ($k = 3$), SCUSUM outperforms both in false negative and false positive: though the false negative for SCUSUM is 0.6338 when $k = 3$ and $\mu = 0.8$, the corresponding false positive is 0.001, a very small proportion and the false negative for FDR_L under the same setting is 0.9831, almost 1. In addition, when we allow the neighbor size k to be 5, imply a little larger block, the false negative for SCUSUM would decrease to be less than 0.50 while the false positive is controlled less than 0.003. Compared with FDR_L , with neighbor size as 5, the false negative is over 0.50 with low signal strength 0.8 and 1.0, while the false positive is almost higher than 0.003. These experiment data implies that our proposed method, SCUSUM, performs better than FDR_L .

In Figure 6, we show the probability maps of the pixels being detected by the two methods with give $\alpha = 0.05$. Here we set the neighbor size $k = 5$, which we think has performance from Table 1 (relatively lower false positive and false negative). We range the signal strength from 0.5 to 2. In the probability maps, the darker the color is, the higher probability the corresponding point is signal. Comparing the probability maps with the ground truth (see Figure 5), it could be easily see that with hige signal strength $\mu \geq 1.5$, the two methods have almost the same performance; while the signal is much too low (e.g. $\mu = 0.5$), SCUSUM has a better detected result than FDR_L . Also we can see that in the results of FDR_L there are some shadows outside of 'L' and 'H' signal region, which are false positive; while for SCUSUM, the detections for indifference region are more 'white' (no shadows.) To some degree, these probability maps are consistent with the experiment data in Tabel 1.

Though in model 2 we assume the noise to be independence, now we try to apply SCUSUM to dependence spatial observations, and compare with FDR_L . Here we consider to use the Exponential

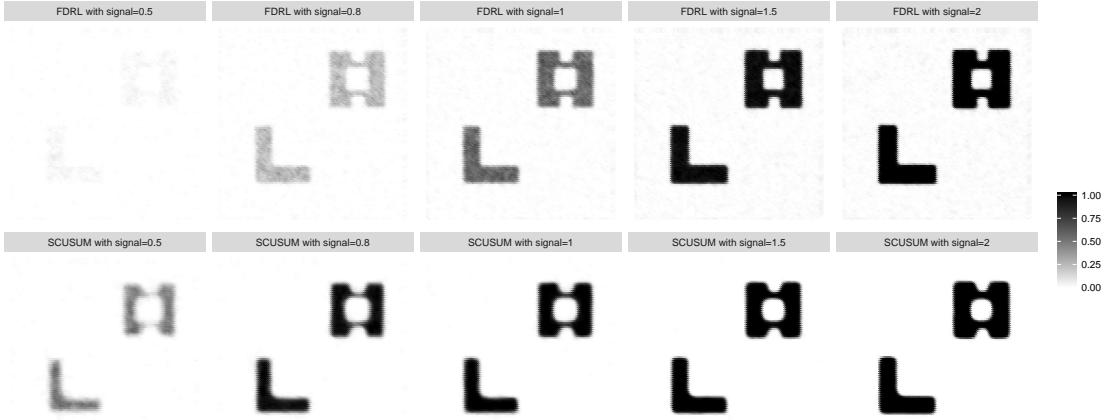


Figure 6: Comparision of the detection probability for SCUSUM and FDR_L under different signal strength: the darker the color is, the higher probability the corresponding point is signal.

Covariance Model ([Gelfand et al. \(2010\)](#)) to generate dependence data. The covariance matrix is

$$C(s_i, s_j) = \exp\left(-\frac{\|s_i - s_j\|}{r}\right), \quad (13)$$

where s_i and s_j are two locations, $\|s_i - s_j\|$ is the distance between the two location, and r is the dependence scale. The larger r means the stronger dependence. Then the data are generate from multivariate normal distribution with above covariance matrix and correponding mean, μ_0 for $\mathcal{D}_{\mathcal{A}^c}$ and μ_1 for $\mathcal{D}_{\mathcal{A}}$. We range scale r from $\{0.1, 0.3, 0.5\}$, and the corresponding covariances for unit distance are $\{0.00004, 0.03567, 0.13533\}$. We set the neighbor size $k = 5$ for both two algorithms and $\alpha = 0.05$ to control marginal FDR. The results are shown in Figure 7 and Tabel 2. It could be seen that with the weak dependence, SCUSUM could still detect the signal region efficiently while control the false posive. However for FDR_L , the false negative increases significantly with stronger dependence (see Table 2). Also Table 2 shows that larger dependence scale leads to larger false positive, false negative and FDR for SCUSUM. This gives us a hint that for larger scale dependence noise we need to choose relatively larger blocks. Nevertheless, we can see that SCUSUM could recognize the signal region with higher probability than FDR_L , when the noise dependence is weak.

5. Real Data Experiment

In this section, we apply four methods, SCUSUM, scan statistics, the conventional FDR and FDR_L , to a real fMRI data to illustrate their differences in real data application. The fMRI data has been analyzed in some previous works ([Maitra \(2009\)](#); [Zhang and Zhu \(2012\)](#) etc.)

The first column in Figure 8 shows six slices of the fMRI images in a total of 22 slices. Each individual image has 128×128 pixels. All these images show activities in different regions by heat maps. The pixels' values are the transformations of p -values from a previous study, which should follow a standard normal distribution, and we only care about detection of the regions with positive values.

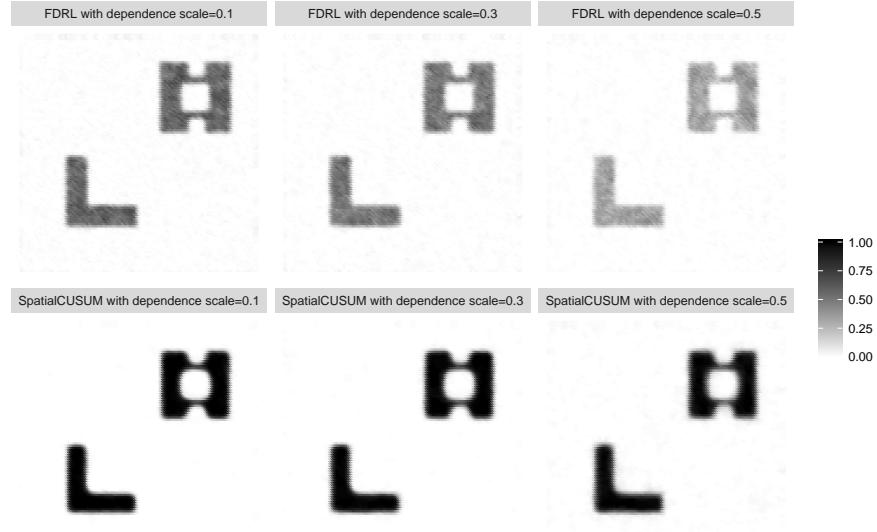
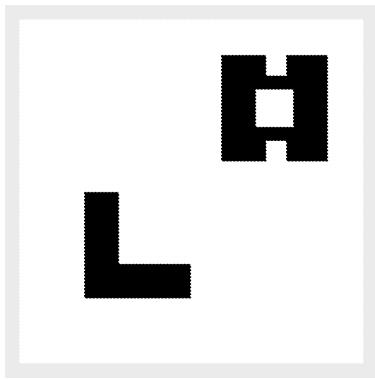


Figure 7: Comparision of the detection probability for SCUSUM and FDR_L on data with different dependence scales: the darker the color is, the higher probability the corresponding point is signal. Here we set the neighbor size $k = 5$ and signal $\mu = 1$.

Table 2: Summary for detection probabilities on dependence data. The dependence scale is changed from 0.1 to 0.5, neighbor size is chose as 5 and signal strength is 1.

Dependence Scale		$r = 0.1$	$r = 0.3$	$r = 0.5$
false negative	SCUSUM	0.2721	0.2820	0.3108
	FDR_L	0.5213	0.5879	0.7317
false positive	SCUSUM	0.00095	0.00154	0.00418
	FDR_L	0.00499	0.00411	0.00280
FDR	SCUSUM	0.0086	0.0141	0.0386
	FDR_L	0.0622	0.0604	0.0590

As for the conventional FDR, the lack of identification phenomenon happened, e.g in the fifth slice all the signals are missed. Also without considering spatial correlation, many detected 'signals' are scattered around, especially in the fourth slice, which means that some of them might be false positive. Although FDR_L could make full use of the neighbor information of spatially structured data and improve the detection ef-



fectiveness (more signals are detected in the second column,) many weak signals are missed, e.g. in fifth slice and sixth slice many red parts are not detected. For Scan statistics, though almost all the 'hot' pixels are detected, the 'signal' regions are too large, which is doubtable.

Similar with the conclusion in Section 4, our method is more likely to detect some weak signals compared with FDR methods and successfully control the false positive. In all the six slices, the detected regions are larger and cover the regions detected by FDR methods. Meanwhile, within each slice, SCUSUM could identify several irregularly shaped clusters. We can see that the detected regions form natural clusters, and they are the spatially grouped 'hot' parts in the raw images. These results show that our proposed method might be more suitable for detection of irregular shaped and weak spatial signals.

6. Conclusion

In this work, we proposed a spatial signal detection method, SCUSUM, which could accommodate the local spatial information. SCUSUM consists of two steps: firstly signal weights are estimated by moving window projecting and CUSUM cut-off; then a threshold is determined with given significant level α to control marginal false discovery rate. Our simulation study shows that our method have better performance compared to the conventional FDR procedure and FDR_L method. Empirically, SCUSUM tends to detect spatially grouped and weak signals, which are missed by the other two methods. Finally, our method is applied to a real fMRI data to illustrate its detection effectiveness.

In model 2, though our method doesn't need to specify the distribution for noise process, the noise ϵ is assumed to be independent, and this is a strong assumption in spatial statistics. In the future work, it could be possible to consider a dependent noise process. In our simulation, the result shows that with weak spatial dependent noise process, SCUSUM could still maintain its effectiveness. Another key issue is to consider how to combine multi-source image data. So far, we only consider the observations are scalar and use moving window idea to project spatial observations to a sequence. However, this projecting method could be problematic when the observation of a spatial location is high-dimensional data or functional data.

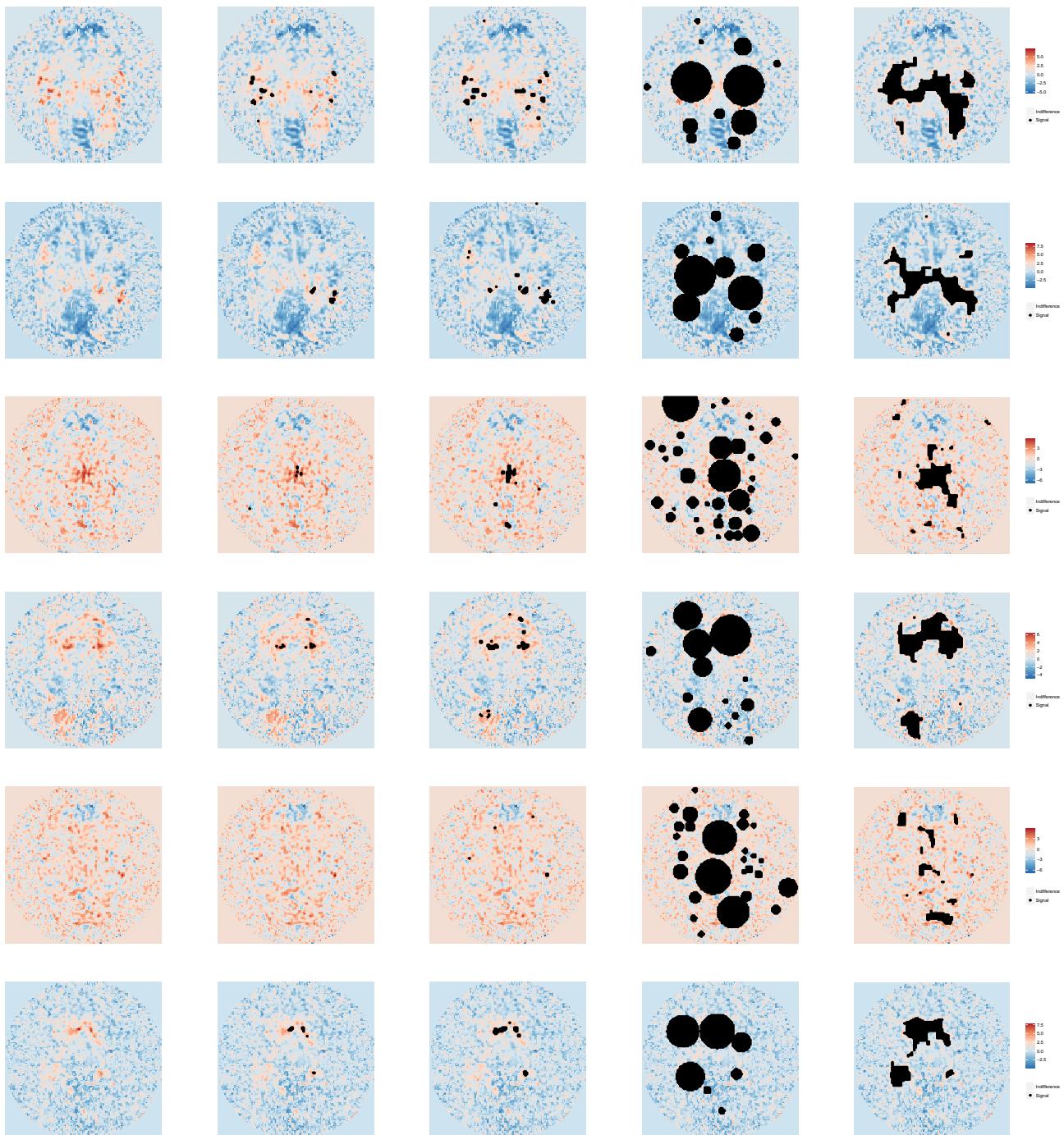


Figure 8: Detected signal regions for the real fMRI dataset. The first column is the raw data. Conventional FDR approach (second column), FDR_L approach (third column), Scan statistics (fourth column) and our proposed SCUSUM method (fifth column) are shown. The black part is the detected signal region. Here, significant level is $\alpha = 0.0001$.

References

- Kevork N Abazajian and Manoj Kaplinghat. Detection of a gamma-ray source in the galactic center consistent with extended emission from dark matter annihilation and concentrated astrophysical emission. *Physical Review D*, 86(8):083511, 2012.
- Alexander Aue, Robertas Gabrys, Lajos Horváth, and Piotr Kokoszka. Estimation of a change-point in the mean function of functional data. *Journal of Multivariate Analysis*, 100(10):2254–2269, 2009.
- Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, pages 289–300, 1995.
- Yoav Benjamini and Daniel Yekutieli. The control of the false discovery rate in multiple testing under dependency. *Annals of statistics*, pages 1165–1188, 2001.
- Thomas Blumensath, Saad Jbabdi, Matthew F Glasser, David C Van Essen, Kamil Ugurbil, Timothy EJ Behrens, and Stephen M Smith. Spatially constrained hierarchical parcellation of the brain with resting-state fmri. *Neuroimage*, 76:313–324, 2013.
- David S Breed. Weather monitoring techniques, November 15 2011. US Patent 8,060,308.
- Matias D Cattaneo, Michael Jansson, and Xinwei Ma. Ipdensity: Local polynomial density estimation and inference. 2017.
- Song Xi Chen. Beta kernel estimators for density functions. *Computational Statistics & Data Analysis*, 31(2):131–145, 1999.
- Haeran Cho et al. Change-point detection in panel data via double cusum statistic. *Electronic Journal of Statistics*, 10(2):2000–2038, 2016.
- Ann Cowling and Peter Hall. On pseudodata methods for removing boundary effects in kernel density estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 551–563, 1996.
- R Cameron Craddock, G Andrew James, Paul E Holtzheimer, Xiaoping P Hu, and Helen S Mayberg. A whole brain fmri atlas generated via spatially constrained spectral clustering. *Human brain mapping*, 33(8):1914–1928, 2012.
- Alan E Gelfand, Peter Diggle, Peter Guttorp, and Montserrat Fuentes. *Handbook of spatial statistics*. CRC press, 2010.
- Christopher Genovese and Larry Wasserman. Operating characteristics and extensions of the false discovery rate procedure. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(3):499–517, 2002.
- Christopher R Genovese, Nicole A Lazar, and Thomas Nichols. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage*, 15(4):870–878, 2002.

- Michael D Gladders and HKC Yee. A new method for galaxy cluster detection. i. the algorithm. *The Astronomical Journal*, 120(4):2148, 2000.
- Joseph Glaz and Narayanaswamy Balakrishnan. *Scan statistics and applications*. Springer Science & Business Media, 2012.
- Joseph Glaz and Zhenkui Zhang. Multiple window discrete scan statistics. *Journal of Applied Statistics*, 31(8):967–980, 2004.
- Joseph Glaz, Joseph I Naus, Sylvan Wallenstein, Sylvan Wallenstein, and Joseph I Naus. *Scan statistics*. Springer, 2001.
- Joseph Glaz, Vladimir Pozdnyakov, and Sylvan Wallenstein. *Scan statistics: methods and applications*. Springer Science & Business Media, 2009.
- Oleksandr Gromenko, Piotr Kokoszka, and Matthew Reimherr. Detection of change in the spatiotemporal mean function. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(1):29–50, 2017.
- Emily Grover-Kopec, Mika Kawano, Robert W Klaver, Benno Blumenthal, Pietro Ceccato, and Stephen J Connor. An online operational rainfall-monitoring resource for epidemic malaria early warning systems in africa. *Malaria Journal*, 4(1):6, 2005.
- Timothy C Haas. Kriging and automated variogram modeling within a moving window. *Atmospheric Environment. Part A. General Topics*, 24(7):1759–1769, 1990.
- Peter Hall, Joel L Horowitz, and Bing-Yi Jing. On blocking rules for the bootstrap with dependent data. *Biometrika*, 82(3):561–574, 1995.
- Lajos Horváth and Marie Hušková. Change-point detection in panel data. *Journal of Time Series Analysis*, 33(4):631–648, 2012.
- MC Jones and PJ Foster. A simple nonnegative boundary correction method for kernel density estimation. *Statistica Sinica*, pages 1005–1013, 1996.
- Martin Kulldorff. Spatial scan statistics: models, calculations, and applications. In *Scan statistics and applications*, pages 303–322. Springer, 1999.
- Martin Kulldorff and Neville Nagarwalla. Spatial disease clusters: detection and inference. *Statistics in medicine*, 14(8):799–810, 1995.
- Ranjan Maitra. Assessing certainty of activation or inactivation in test-retest fmri studies. *Neuroimage*, 47(1):88–97, 2009.
- Christopher J Miller, Christopher Genovese, Robert C Nichol, Larry Wasserman, Andrew Connolly, Daniel Reichart, Andrew Hopkins, Jeff Schneider, and Andrew Moore. Controlling the false-discovery rate in astrophysical data analysis. *The Astronomical Journal*, 122(6):3492, 2001.
- JL Naus. Clustering of random points in two dimensions. *Biometrika*, 52(1-2):263–266, 1965.

- Joseph I Naus. Approximations for distributions of scan statistics. *Journal of the American Statistical Association*, 77(377):177–183, 1982.
- Antonio Páez, Fei Long, and Steven Farber. Moving window approaches for hedonic price estimation: an empirical comparison of modelling techniques. *Urban Studies*, 45(8):1565–1581, 2008.
- Carey E Priebe, John M Conroy, David J Marchette, and Youngser Park. Scan statistics on enron graphs. *Computational & Mathematical Organization Theory*, 11(3):229–247, 2005.
- Xilin Shen, Fuyuze Tokoglu, Xenios Papademetris, and R Todd Constable. Groupwise whole-brain parcellation from resting-state fmri data for network node identification. *Neuroimage*, 82:403–415, 2013.
- Wenguang Sun, Brian J Reich, T Tony Cai, Michele Guindani, and Armin Schwartzman. False discovery control in large-scale spatial multiple testing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(1):59–83, 2015.
- Toshiro Tango. A test for spatial disease clustering adjusted for multiple testing. *Statistics in medicine*, 19(2):191–204, 2000.
- Madeleine C Thomson and Stephen J Connor. The development of malaria early warning systems for africa. *Trends in parasitology*, 17(9):438–445, 2001.
- Jigang Wang, Predrag Neskovic, and Leon N Cooper. Neighborhood size selection in the k-nearest-neighbor rule using statistical confidence. *Pattern Recognition*, 39(3):417–423, 2006.
- Tengyao Wang and Richard J Samworth. High dimensional change point estimation via sparse projection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(1):57–83, 2018.
- David C Wheeler. A comparison of spatial clustering and cluster detection techniques for childhood leukemia incidence in ohio, 1996–2003. *International Journal of Health Geographics*, 6(1):13, 2007.
- Chunming Zhang, Jianqing Fan, and Tao Yu. Multiple testing via fdrl for large scale imaging data. *Annals of statistics*, 39(1):613, 2011.
- Lingsong Zhang and Zhengyuan Zhu. Spatial multiresolution cluster detection method. *arXiv preprint arXiv:1205.2106*, 2012.
- Zhenkui Zhang, Renato Assunção, and Martin Kulldorff. Spatial scan statistics adjusted for multiple clusters. *Journal of Probability and Statistics*, 2010, 2010.

Appendix A. Proof for Lemma 1

Proof In assumed model 2, the noise processes are independent. Thus, in the i th block, the sampled representative γ_i is independent with the rest observations, which implies γ_i and $\tilde{\mu}_i$ are independent, i.e. $\gamma_i \perp \tilde{\mu}_i$. Also the representatives and pseudo block means between blocks are independent. These lead that $\{\gamma_i\}$ and $\{\tilde{\mu}_i\}$ are independent.

Under the null hypothesis H_0 , i.e. there is no signal and $\mathbb{E}[\tilde{\mu}_i] = \mu_0$, pseudo block mean sequence $\{\tilde{\mu}_i\}$ are i.i.d, as well as representative sequence $\{\gamma_i\}$. For the distribution of $\{\gamma_i^*\}$, we have following:

$$\begin{aligned} f(\gamma_1^* \leq y_1, \dots, \gamma_b^* \leq y_b) &= \sum_{([1], \dots, [b]) \in S_b} f(\gamma_{[1]} \leq y_1, \dots, \gamma_{[b]} \leq y_b | \tilde{\mu}_{[1]} \geq \dots \geq \tilde{\mu}_{[b]}) f(\tilde{\mu}_{[1]} > \dots > \tilde{\mu}_{[b]}) \\ &= \sum_{([1], \dots, [b]) \in S_b} f(\gamma_{[1]} \leq y_1, \dots, \gamma_{[b]} \leq y_b | \tilde{\mu}_{[1]} \geq \dots \geq \tilde{\mu}_{[b]}) \frac{1}{b!} \\ &= \sum_{([1], \dots, [b]) \in S_b} f(\gamma_{[1]} \leq y_1, \dots, \gamma_{[b]} \leq y_b) \frac{1}{b!} \\ &= \sum_{([1], \dots, [b]) \in S_b} f(\gamma_1 \leq y_1, \dots, \gamma_b \leq y_b) \frac{1}{b!} \\ &= f(\gamma_1 \leq y_1, \dots, \gamma_b \leq y_b). \end{aligned}$$

Here $([1], \dots, [b])$ is the one possible decreasing order for $\{\tilde{\mu}_i\}$, S_b presents the set of all the possible orders. The first equation is according to bayesian formula; the second equation is because of $f(\tilde{\mu}_{[1]} > \dots > \tilde{\mu}_{[b]}) = \int_{\tilde{\mu}_{[1]} > \dots > \tilde{\mu}_{[b]}} dm(\tilde{\mu}_{[1]}, \dots, \tilde{\mu}_{[b]}) = 1/n!$ under the independence of $\{\tilde{\mu}_i\}$; the third equation is due to independence between $\{\gamma_i\}$ and $\{\tilde{\mu}_i\}$; the fourth equation is because $\{\gamma_i\}$ are independent; the fifth equation is due to the cardinality of S_b is $1/b!$. From above result, we reach that $\{\gamma_i^*\}$ are also i.i.d, having the same distribution with $\{\gamma_i\}$.

Under the alternative hypothesis H_1 , as the number of observations in each block n_i goes to infinity, the weak law of larger number supports that $\tilde{\mu}_i \xrightarrow{p} \mathbb{E}[\tilde{\mu}_i]$. Thus, with (3) and (4), (5) holds. And l_1 is the number of block inside the signal region, $(b - l_2)$ is the number of block inside the indifference region and $(l_2 - l_1)$ is the number of block at the boundary. ■

Appendix B. Proof for Theorem 1

Proof The proof idea is similar with Aue et al. (2009). W.l.o.g, here we consider the variance of noise processes is 1, $\mu_0 = 0$ and $\mu_1 = \mu_0 + \Delta = \Delta$. Define the ratio of signal region to the entire spatial domain is θ , hence, as blocks become finer and finer ($b \rightarrow \infty$), the ratio of the blocks with signal representative to the total blocks is getting closer to θ .

Under H_1 , define the following events:

$$\begin{cases} A_1 := \{\mu_i \leq \mu_j, B_i \in \mathcal{D}_{\mathcal{A}}, B_j \in \mathcal{D}_{\mathcal{A}^c}, \forall i, j\}, \\ A_2 := \{\mu_k \leq \mu_j, B_i \in \mathcal{D}_{\mathcal{A}}, B_k \text{ at boundary}, \forall i, k\}, \\ A_3 := \{\mu_j \leq \mu_k, B_j \in \mathcal{D}_{\mathcal{A}^c}, B_k \text{ at boundary}, \forall j, k\}. \end{cases} \quad (14)$$

Event $(A_1 \cup A_2 \cup A_3)^c$ presents the scenario that the order of $\{\gamma_i^*\}$ is from signal blocks $[1, l_1]$ to interim (boundary) blocks $[l_1, l_2]$ and then to indifferent blocks $(l_2, b]$ (see Figure 3 (b).) And with Lemma 1, $\mathbb{P}((A_1 \cup A_2 \cup A_3)^c) = 1$, as $\min n_i \rightarrow \infty$.

With event $(A_1 \cup A_2 \cup A_3)^c$, following we show that the probability of cut-off location $t = \arg \max_i \tilde{\gamma}_i$ falling into $[l_1, l_2]$ would converge to 1. To proof that, we define statistics

$$Q(r) = \tilde{\gamma}_r^2 = \left(\sum_{i=1}^r \gamma_i^* - \frac{r}{b} \sum_{i=1}^b \gamma_i^* \right)^2. \quad (15)$$

First consider the probability of event $B_1(N) = \{t \geq (l_2 + N)\}$, with $(l_2 + N) \leq b$, N is a fixed constant. Define $R(r; l_2) = Q(r) - Q(l_2)$, and note that $Q(l_2)$ is a constant.

$$\begin{aligned} R(r; l_2) &= Q(r) - Q(l_2) \\ &= \left(\sum_{i=1}^r \gamma_i^* - \frac{r}{b} \sum_{i=1}^b \gamma_i^* \right)^2 - \left(\sum_{i=1}^{l_2} \gamma_i^* - \frac{l_2}{b} \sum_{i=1}^b \gamma_i^* \right)^2 \\ &= \underbrace{\left[\sum_{l_2+1}^r \gamma_i^* - (r - l_2)\bar{\gamma}^* \right]}_{(I)} \underbrace{\left[\sum_{i=1}^{l_2} \gamma_i^* + \sum_{j=1}^{l_2} \gamma_j^* - (r + l_2)\bar{\gamma}^* \right]}_{(II)} \end{aligned}$$

where $\bar{\gamma}^* = \frac{1}{b} \sum_{i=1}^b \gamma_i^*$. And with equation (5), following equations hold:

$$(I) = \sum_{i=l_2+1}^r \epsilon_i - (r - l_2) \frac{1}{b} \sum_{i=1}^b \epsilon_i - (r - l_2)\theta\Delta, \quad (16)$$

$$\begin{aligned} (II) &= \sum_{i=1}^r \epsilon_i + \theta b\Delta + \sum_{j=1}^{l_2} \epsilon_j + \theta b\Delta - (r + l_2) \frac{1}{b} \sum_{i=1}^b \epsilon_i - (r + l_2)\theta\Delta \\ &= \sum_{i=1}^r \epsilon_i + \sum_{j=1}^{l_2} \epsilon_j - (r + l_2) \frac{1}{b} \sum_{i=1}^b \epsilon_i + (2\theta b - (r + l_2)\theta)\Delta, \end{aligned} \quad (17)$$

Define the following statistics:

$$\begin{cases} E^1(r; l_2) := \sum_{i=l_2+1}^r \epsilon_i - (r - l_2) \frac{1}{b} \sum_{i=1}^b \epsilon_i, \\ E^2(r; l_2) := \sum_{i=1}^r \epsilon_i + \sum_{j=1}^{l_2} \epsilon_j - (r + l_2) \frac{1}{b} \sum_{i=1}^b \epsilon_i, \\ D^1(r; l_2) := -(r - l_2)\theta\Delta, \\ D^2(r; l_2) := (2\theta b - (r + l_2)\theta)\Delta. \end{cases} \quad (18)$$

So $(I) = E^1(r; l_2) + D^1(r; l_2)$ and $(II) = E^2(r; l_2) + D^2(r; l_2)$.

As $b \rightarrow \infty$, we have

$$\begin{aligned} \max_{(l_2+N) \leq r \leq b} D^1(r; l_2) D^2(r; l_2) &= \max_{(l_2+N) \leq r \leq b} [-(r - l_2)\theta\Delta][(2\theta b - (r + l_2)\theta)\Delta] \\ &= \max_{(l_2+N) \leq r \leq b} -\theta^2 \Delta^2 (r - l_2) (2 - \frac{r + l_2}{b}) b \\ &= -\theta^2 \Delta^2 N (2 - \frac{2l_2 + N}{b}) b, \end{aligned}$$

the last equation is due to the $D^1(r; l_2)D^2(r; l_2)$ reaches the maximum with $r = (l_2 + N)$. Also $\forall \epsilon \geq 0$, we have

$$\begin{aligned} & \lim_{b \rightarrow \infty} \sup \mathbb{P} \left(\max_{(l_2+N) \leq r \leq b} D^1(r; l_2)D^2(r; l_2) > -\epsilon \right) \\ &= \lim_{b \rightarrow \infty} \sup \mathbb{P} \left(-\theta^2 \Delta^2 N \left(2 - \frac{2l_2 + N}{b} \right) b > -\epsilon \right) \\ &= \lim_{b \rightarrow \infty} \sup \mathbb{P} \left(\theta^2 \Delta^2 N \left(2 - \frac{2l_2 + N}{b} \right) b \leq \epsilon \right) = 0 \end{aligned}$$

If we could prove $D^1D^2(r; l_2)$ is the leading term in $R(r; l_2)$, then

$$\begin{aligned} \lim_{b \rightarrow \infty} \sup \mathbb{P}(B_1(N)) &= \lim_{b \rightarrow \infty} \sup \mathbb{P}(t \geq l_2 + N) \\ &= \lim_{b \rightarrow \infty} \sup \mathbb{P} \left(\max_{(l_2+N) \leq r \leq b} R(r; l_2) > 0 \right) \\ &= \lim_{b \rightarrow \infty} \sup \mathbb{P} \left(\theta^2 \Delta^2 N \left(2 - \frac{2l_2 + N}{b} \right) b + O(1) \leq \epsilon \right) = 0 \end{aligned}$$

Hence $\mathbb{P}(t \geq l_2) = \cup_{N=0}^{b-l_2} \mathbb{P}(B_1(N)) = 0$.

Following lemmas support that $D^1(r; l_2)D^2(r; l_2)$ is the leading term in $R(r; l_2)$, with $r \in [l_2 + N, b]$.

Lemma 3 *With the assumptions of Theorem 1, given $N, \forall \epsilon > 0$,*

$$\lim_{b \rightarrow \infty} \sup \mathbb{P} \left(\max_{(l_2+N) \leq r \leq b} \frac{|E^1(r; l_2)E^2(r; l_2)|}{|D^1(r; l_2)D^2(r; l_2)|} \geq \epsilon \right) = 0. \quad (19)$$

Proof With the brief derivation, we have,

$$\begin{aligned} & \max_{(l_2+N) \leq r \leq b} \frac{|E^1(r; l_2)E^2(r; l_2)|}{|D^1(r; l_2)D^2(r; l_2)|} \\ & \leq \max_{(l_2+N) \leq r \leq b} \frac{|E^1(r; l_2)||E^2(r; l_2)|}{\theta \Delta^2 (r - l_2)(2b\theta - (l_2 + r))} \\ &= O(1) \max_{(l_2+N) \leq r \leq b} \frac{|E^1(r; l_2)||E^2(r; l_2)|}{(r - l_2)(l_2 + r)} \\ &= O(1) \underbrace{\max_{(l_2+N) \leq r \leq b} \frac{|E^1(r; l_2)|}{(r - l_2)}}_{(III)} \underbrace{\frac{|E^2(r; l_2)|}{(l_2 + r)}}_{(IV)} \end{aligned}$$

the last equation is because of $l_2 > \theta\Delta$.

For (III),

$$\begin{aligned} \frac{|E^1(r; l_2)|}{(r - l_2)} &\leq \frac{|\sum_{i=l_2+1}^r \epsilon_i| + |(r - l_2)\frac{1}{b} \sum_{i=1}^b \epsilon_i|}{(r - l_2)} \\ &= \frac{|\sum_{i=l_2+1}^r \epsilon_i|}{(r - l_2)} + |\frac{1}{b} \sum_{i=1}^b \epsilon_i| \xrightarrow{p} 0, \end{aligned}$$

the last equation is due to the law of iterated logarithm and the weak law of large number. Similarly, for (IV),

$$\frac{|E^2(r; l_2)|}{(l_2 + r)} \leq \frac{|\sum_{i=1}^r \epsilon_i + \sum_{j=1}^{l_2} \epsilon_j| + |(r + l_2) \frac{1}{b} \sum_{i=1}^b \epsilon_i|}{(l_2 + r)} \xrightarrow{p} 0.$$

Hence, with continuous mapping theorem, we have the result. \blacksquare

Lemma 4 *With the assumptions of Theorem 1, given $N, \forall \epsilon > 0$,*

$$\lim_{b \rightarrow \infty} \sup \mathbb{P}(\max_{(l_2+N) \leq r \leq b} \frac{|E^1(r; l_2)D^2(r; l_2)|}{|D^1(r; l_2)D^2(r; l_2)|} \geq \epsilon) = 0. \quad (20)$$

Proof Similarly, we have

$$\begin{aligned} & \max_{(l_2+N) \leq r \leq b} \frac{|E^1(r; l_2)D^2(r; l_2)|}{|D^1(r; l_2)D^2(r; l_2)|} \\ & \leq O(1) \max_{(l_2+N) \leq r \leq b} \frac{|E^1(r; l_2)|}{(r - l_2)} \frac{|D^2(r; l_2)|}{(r + l_2)}. \end{aligned}$$

From Lemma 3, we know that $\frac{|E^1(r; l_2)|}{(r - l_2)} \xrightarrow{p} 0$. With directly derivation, $\frac{|D^2(r; l_2)|}{(r + l_2)} = O(1)$. Hence, $\max_{(l_2+N) \leq r \leq b} \frac{|E^1(r; l_2)D^2(r; l_2)|}{|D^1(r; l_2)D^2(r; l_2)|} \xrightarrow{p} 0$. \blacksquare

Lemma 5 *With the assumptions of Theorem 1, given $N, \forall \epsilon > 0$,*

$$\lim_{b \rightarrow \infty} \sup \mathbb{P}(\max_{(l_2+N) \leq r \leq b} \frac{|D^1(r; l_2)E^2(r; l_2)|}{|D^1(r; l_2)D^2(r; l_2)|} \geq \epsilon) = 0. \quad (21)$$

Proof The idea is the same with Lemma 4:

$$\max_{(l_2+N) \leq r \leq b} \frac{|D^1(r; l_2)E^2(r; l_2)|}{|D^1(r; l_2)D^2(r; l_2)|} \leq O(1) \max_{(l_2+N) \leq r \leq b} \frac{|D^1(r; l_2)|}{(r - l_2)} \frac{|E^2(r; l_2)|}{(r + l_2)} \xrightarrow{p} 0, \quad (22)$$

with $\frac{|D^1(r; l_2)|}{(r - l_2)} = O(1)$ and $\frac{|E^2(r; l_2)|}{(r + l_2)} \xrightarrow{p} 0$. \blacksquare

The above three lemmas support that $D^1(r; l_2)D^2(r; l_2)$ is the leading term in $R(r; l_2)$, with $r \in [l_2 + N, b]$. Hence, the probability of event $\{t > l_2\} \cup (A_1 \cup A_2 \cup A_3) \rightarrow 0$, as $b \rightarrow \infty$ and $\min n_i \rightarrow \infty$.

For the other side, consider event $B_2(N) = \{t \leq (l_1 - N)\}$, with a given N and $l_1 \geq N$. Similarly define $R(r; l_1) = Q(r) - Q(l_1)$, and following we will show $R(r; l_1) < 0, \forall r < l_1$

asymptotically with probability 1.

$$\begin{aligned}
R(r; l_1) &= Q(r) - Q(l_1) \\
&= \left(\sum_{i=1}^r \gamma_i^* - \frac{r}{b} \sum_{i=1}^b \gamma_i^* \right)^2 - \left(\sum_{i=1}^{l_1} \gamma_i^* - \frac{l_1}{b} \sum_{i=1}^b \gamma_i^* \right)^2 \\
&= \underbrace{\left[- \sum_{r+1}^{l_1} \gamma_i^* - (r - l_1) \bar{\gamma}^* \right]}_{(V)} \underbrace{\left[\sum_{i=1}^r \gamma_i + \sum_{j=1}^{l_1} \gamma_j - (r + l_1) \bar{\gamma}^* \right]}_{(VI)}
\end{aligned}$$

Define the following statistics:

$$\begin{cases} E^1(r; l_1) := - \sum_{i=r+1}^{l_1} \epsilon_i - (r - l_1) \bar{\gamma}^*, \\ E^2(r; l_1) := \sum_{i=1}^r \epsilon_i + \sum_{j=1}^{l_1} \epsilon_j - (r + l_1) \bar{\gamma}^*, \\ D^1(r; l_1) := -(l_1 - r)(1 - \theta)\Delta, \\ D^2(r; l_1) := (r + l_1)(1 - \theta)\Delta. \end{cases} \quad (23)$$

Similarly, we have $(V) = E^1(r; l_1) + D^1(r; l_1)$, $(VI) = E^2(r; l_1) + D^2(r; l_1)$. Also, as $b \rightarrow \infty$, $l_1 = \theta b \rightarrow \infty$, so we have

$$\max_{1 \leq r \leq (l_1 - N)} D^1(r; l_1) D^2(r; l_1) = [-N(r + l_1)(1 - \theta)^2 \Delta^2]$$

and

$$\begin{aligned}
&\limsup_{b \rightarrow \infty} \mathbb{P}(\max_{1 \leq r \leq (l_1 - N)} D^1(r; l_1) D^2(r; l_1) > -\epsilon) \\
&= \limsup_{b \rightarrow \infty} \mathbb{P}(-N(r + l_1)(1 - \theta)^2 \Delta^2 > -\epsilon) \\
&= \limsup_{b \rightarrow \infty} \mathbb{P}(N(r + \theta b)(1 - \theta)^2 \Delta^2 \leq \epsilon) = 0
\end{aligned}$$

Following lemma shows that $D^1(r; l_1) D^2(r; l_1)$ is the leading term in $R(r; l_1)$, with $r \in [1, l_1 - N]$.

Lemma 6 *With the assumptions of Theorem 1, given $N, \forall \epsilon > 0$, we have*

$$\begin{cases} \limsup_{b \rightarrow \infty} \mathbb{P}(\max_{1 \leq r \leq (l_1 - N)} \frac{|E^1(r; l_1) E^2(r; l_1)|}{|D^1(r; l_1) D^2(r; l_1)|} \geq \epsilon) = 0, \\ \limsup_{b \rightarrow \infty} \mathbb{P}(\max_{1 \leq r \leq (l_1 - N)} \frac{|D^1(r; l_1) E^2(r; l_1)|}{|D^1(r; l_1) D^2(r; l_1)|} \geq \epsilon) = 0, \\ \limsup_{b \rightarrow \infty} \mathbb{P}(\max_{1 \leq r \leq (l_1 - N)} \frac{|E^1(r; l_1) D^2(r; l_1)|}{|D^1(r; l_1) D^2(r; l_1)|} \geq \epsilon) = 0. \end{cases} \quad (24)$$

Proof Similarly with Lemma 3-5, we have following:

$$\begin{cases} \frac{|E^1(r; l_1)|}{(l_1 - r)} \leq \frac{|\sum_{i=r+1}^{l_1} \gamma_i^*| + |\frac{1}{b} \sum_{i=1}^b \gamma_i^*|}{l_1 - r} \xrightarrow{p} 0; \\ \frac{|E^2(r; l_1)|}{(l_1 + r)} \leq \frac{|\sum_{i=1}^r \gamma_i^* + \sum_{i=1}^{l_1} \gamma_i^*| + |\frac{1}{b} \sum_{i=1}^b \gamma_i^*|}{l_1 + r} \xrightarrow{p} 0; \\ \frac{|D^1(r; l_1)|}{(l_1 - r)} = \frac{|D^2(r; l_1)|}{(l_1 + r)} = O(1). \end{cases} \quad (25)$$

Hence, we have the results

$$\begin{cases} \max_{1 \leq r \leq (l_1 - N)} \frac{|E^1(r; l_1)E^2(r; l_1)|}{|D^1(r; l_1)D^2(r; l_1)|} \leq O(1) \max_{1 \leq r \leq (l_1 - N)} \frac{|E^1(r; l_1)||E^2(r; l_1)|}{(l_1 - r)(l_1 + r)} \xrightarrow{p} 0; \\ \max_{1 \leq r \leq (l_1 - N)} \frac{|D^1(r; l_1)E^2(r; l_1)|}{|D^1(r; l_1)D^2(r; l_1)|} \leq O(1) \max_{1 \leq r \leq (l_1 - N)} \frac{|D^1(r; l_1)||E^2(r; l_1)|}{(l_1 - r)(l_1 + r)} \xrightarrow{p} 0; \\ \max_{1 \leq r \leq (l_1 - N)} \frac{|E^1(r; l_1)D^2(r; l_1)|}{|D^1(r; l_1)D^2(r; l_1)|} \leq O(1) \max_{1 \leq r \leq (l_1 - N)} \frac{|E^1(r; l_1)||D^2(r; l_1)|}{(l_1 - r)(l_1 + r)} \xrightarrow{p} 0. \end{cases} \quad (26)$$

■

With above conclusion, we have

$$\begin{aligned} \lim_{b \rightarrow \infty} \sup \mathbb{P}(B_2(N)) &= \lim_{b \rightarrow \infty} \sup \mathbb{P}(t \leq l_1 - N) \\ &= \lim_{b \rightarrow \infty} \sup \mathbb{P}\left(\max_{1 \leq r \leq (l_1 - N)} R(r; l_1) > 0\right) \\ &= \lim_{b \rightarrow \infty} \sup \mathbb{P}(N(r + \theta b)(1 - \theta)^2 \Delta^2 + O(1) \leq \epsilon) = 0 \end{aligned}$$

which implies $\mathbb{P}(t \leq l_1) = \cup_{N=1}^{l_1} \mathbb{P}(B_2(N)) = 0$.

With these results, we have

$$\begin{aligned} &\lim_{b \rightarrow \infty} \mathbb{P}(\{t \in [l_1, l_2]\}) \\ &\leq 1 - \lim_{b \rightarrow \infty} \sup \mathbb{P}(\{t < l_1\} \cup \{t > l_2\} \cup (A_1 \cup A_2 \cup A_3)) \\ &\rightarrow 1, \end{aligned}$$

as $b \rightarrow \infty$ and $\min n_i \rightarrow \infty$.

■

Appendix C. Proof for Lemma 2

Proof This lemma is easy to prove: From Lemma 1, as $\min n_i \rightarrow \infty$, $\{\gamma_i^*\}$ are i.i.d. Consider sequence $\{y_1, \dots, y_b\}$, then $\mathbb{P}(\gamma_1^* = y_1, \dots, \gamma_b^* = y_b) = \mathbb{P}(\gamma_1^* = y_b, \dots, \gamma_b^* = y_1)$. Note that in the left side $\{\gamma_i^*\}$ equal to the reversed sequence $\{y_b, \dots, y_1\}$. Via applying CUSUM cut-off on $\{y_1, \dots, y_b\}$

and $\{y_b, \dots, y_1\}$, the detection result is opposite. Hence the distribution for signal weights under H_0 is symmetric. ■