BOUNDARY ESTIMATION

A Dissertation
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agriculture and Applied Science

By

Yingfei Mu

In Partial Fulfillment of the Requirements
for the Degree of
DOCTOR OF PHILOSOPHY

Major Department:
Statistics

July 2015

Fargo, North Dakota

NORTH DAKOTA STATE UNIVERSITY

Graduate School

Title

BOUNDARY ESTIMATION

By

Yingfei Mu

The supervisory committee certifies that this dissertation complies with North Dakota State University's regulations and meets the accepted standards for the degree of

DOCTOR OF PHILOSOPHY

SUPERVISORY COMMITTEE:

Dr. Gang Shen
Chair

Dr. Rhonda Magel

Dr. Seung Won Hyun

Dr. Doğan Çömez

Approved:

08 July 2015
Date

Dr. Rhonda Magel
Department Chair

ABSTRACT

The existing statistical methods do not provide a satisfactory solution to determining the spatial pattern in spatially referenced data, which is often required by research in many areas including geology, agriculture, forestry, marine science and epidemiology for identifying the source of the unusual environmental factors associated with a certain phenomenon. This work provides a novel algorithm which can be used to delineate the boundary of an area of hot spots accurately and efficiently. Our algorithm, first of all, does not assume any pre-specified geometric shapes for the change-curve. Secondly, the computation complexity by our novel algorithm for change-curve detection is in the order of $O(n^2)$, which is much smaller than $2^{O(n^2)}$ required by the CUSP algorithm proposed in Müller&Song [8] and Carlstein's [2] estimators. Furthermore, our novel algorithm yields a consistent estimate of the change-curve as well as the underlying distribution mean of observations in the regions. We also study the hypothesis test of the existence of the change-curve in the presence of independence of the spatially referenced data. We then provide some simulation studies as well as a real case study to compare our algorithm with the popular boundary estimation method : Spatial scan statistic.

ACKNOWLEDGEMENTS

First and foremost, I must acknowledge my limitless thanks to my advisor, Dr. Gang Shen for his help. This work would have never become truth without his guidance.

I am grateful to the committee members, who worked with me till the completion of the present research. I highly appreciate the efforts expended by Dr. Rhonda Magel, Dr. Seung Won Hyun and Dr. Doğan Çömez.

I would like to take this opportunity to say warm thanks to all my beloved friends for being so supportive along the way of doing my thesis.

I also would like to express my wholehearted thanks to my family for their generous support they provided me throughout my entire life and particularly through the process of pursuing the master degree. Because of their unconditional love and prayers, I have the chance to complete this thesis.

Last but not least, deepest thanks go to all people who took part in making this thesis real.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

With the recent advance in remote censoring technology, which makes big spatially referenced data accessible to researchers in many areas: Environmental science, agriculture, forestry, marine science, and epidemiology among many others, the issue of identifying the spatial pattern eclipsed by the wealth of big data receive more and more attention in practice, as it is indispensable for researchers to locate the related influential environmental factors. For instance, Riggan et al. [11] reported the heavy concentration of cancer of trachea, bronchus and lung in United States southeastern region and attributed it to the region's cash crop and employment in the manufacturing industries which has extensive exposure to airborne fibers and dusts.

A spatial pattern could be characterized by a change-curve or boundary in statistics, which partitions the whole area of interest in such a way that the distribution of the observations made within each subareas has a distinct mean. Therefore, identifying the spatial pattern amounts to detecting the boundary of the subareas. The major statistical approaches in literature applied for the change-curve detection problem include: 1. the spatial clustering methods which classify health data into clusters (by distribution mean); 2. the change-curve detecting methods which search for the boundary (change-curve) between subareas of distinct distribution means. However, none of these existing methods provide a satisfactory solution, because they assume independence of data or restrictive shapes of the boundary and most are very computationally expensive. Like many other statistical studies, we start with a simplified scenario of lattice design: Observations are made at the $n \times n$ grid nodes formed by the equally spaced divisions along each coordinate axis on the $\mathbb{R}^2$ plane and the coordinates of all the nodes are then rescaled into the unit square $[0, 1]^2$; there might exist a change-curve that splits the unit square into two different regions, region of hot spots and cold spots, respectively. The two regions differ in mean of the spatially referenced observations, see e.g. Carlstein & Krishnamoorthy [2].

To be specific, let Y_s be the spatially referenced observations made at location s, $s \in (0, 1]^2$ (unit square in $\mathbb{R}^2$). We assume $Y_s = \mu_s + \epsilon_s$, where ϵ_s is strictly stationary noise with $E\epsilon_s = 0$ and $\varepsilon_s = \sigma^2$. In order not to interrupt introduction of the change-curve problem, we postpone our circumscription of the spatial structure of $\{\epsilon_s\}$, $s \in (0, 1]^2$ to §3.5. We assume that there may exist

$\theta_0 \subset (0,1]^2$ such that

$$\mu_s = \begin{cases} \mu_1, & s \in \theta_0 \\ \mu_2, & s \in \theta_0^c. \end{cases}$$

With $\mu_1 \neq \mu_2$, the boundary of θ_0, $\partial\theta_0$, is the change-curve that separates the the two regions θ_0 and θ_0^c in $(0,1]^2$. Figure 1.1 below illustrates the setting of the problem as discussed above.

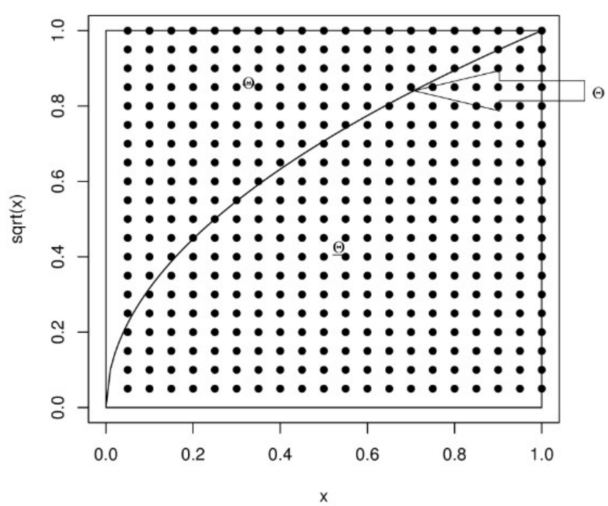


Figure 1.1. Grid with Boundary, the Two-Dimensional Case: n=20

Here we want to emphasize that the spatial location s at which observation Y_s is made is deterministic in this work. This is different from a boundary analysis or edge detection where the locations themselves may be random and Wombling methods [18] which are generally considered to detect regions of rapid change, typically lines or curves on the interpolated spatial surface.

2. LITERATURE REVIEW

This simplified dichotomous setting for the change-curve problem in §1, is actually considered by Müller&Song [8] among many others, e.g., Carlstein&Krishnamoorthy [2] studied the boundary between territories with relatively high and low cancer mortality rates in United States, though the hypothesized mortality rate actually varies from county to county and is not of dichotomous nature. Here we provide a brief review of the major existing statistical methods for the change-curve detection problem.

2.1. Spatial clustering methods

Spatial clustering methods determine whether spatial clusters exist and identify their shape and size via a global or local scale statistic, such as Moran's I statistic [7], Ripley's K function [12], or the scan statistic [6, 9, 14, 15]. The standard purely spatial scan statistic imposes circular window which is centered on each of several possible nodes positioned throughout the study region. For each node, the radius of the window varies continuously in size from zero to some pre-specified upper limit. In this way, the method creates a large number of distinct geographical circles flexible both in location and size which may serve as possible candidate clusters. Kulldorff et al. [6] further developed an elliptic version of the spatial scan statistic, using a scanning window of variable location, shape (eccentricity), angle and size. Jacquez et al. [5] proposed a new distribution free technique technique for cluster detection based on the b-statistic between the adjacent areas. Though their method relaxes the assumption of some unrealistic pre-specified cluster shapes that underpin almost all spatial clustering methods, it is essentially heuristic and the underlying theory is not well established.

2.2. Change-curve detecting methods

Carlstein&Krishnamoorthy [2] proposed a nonparametric method for finding out the change-curve from a class of candidates under some regularity conditions including smoothness of the change-curve. Their approach is exhaustively searching the change-curve within the candidate class that maximizes the criterion function and applicable to the linear or Lipschitz boundary. However, depending on the size of the class of candidate curves, this method could be very computationally expensive. Moreover, the consistency of this method requires the independence of the spatially referenced observations.

Müller&Song [8] suggested a cube splitting method (CUSP) which also searches the boundary by maximizing of a criterion function over the unions of cubes of aggregated pixels. It starts with an initial guess of the change-curve on a coarse level of approximation, then splits the peripheral cubes (near the boundary) into smaller cubes and verify whether the allocation of the cubes to the regions could increase the value of the criterion function. This cube splitting (refinement) step may then be iterated until the desired level of resolution is achieved. The central advantage of the CUSP method is its numerical feasibility and the number of cubes under consideraton for inclusion in the proposed regions may be kept small. However, the CUSP method is still computationally expensive and the related theory of this method is not well developed.

2.3. Kernel-based methods

Qiu [10] introduced a consistent estimator of the jump location curve for independent data based on difference of the two one-sided kernel smoothers. However, like other kernel-based methods which use a moving window for boundary detection and necessarily involve smoothing, it may wash out spatial heterogeneity by averaging observations within the chosen kernel and result in an inaccurate estimate.

3. BOUNDARY ESTIMATE

The change-curve detection method studied in this work is based on a criterion function similar to that in Carlstein&Krishnamoorthy [2] and Müller&Song [8]. Our estimate of the change-curve is essentially based on the projection of the spatially referenced data onto the unit square which utilizes a sort of monotonicity to search for the change-curve. It does not rely on the assumption of any pre-specified geometric shapes or independence of the spatially referenced data. Our change-curve estimator turns out to be computationally efficient and consistent.

3.1. Boundary candidates

We need some regularity conditions (RC) to circumscribe θ_0 within a suitable class of regions, Θ, which is large enough to cover the "regular" geometric shapes such as ellipsoid and polygons, and meanwhile small enough to allow a consistent estimate of the boundary from spatially referenced observations. We require $\forall\ \theta \in \Theta$,

RC1: $\theta \in \mathcal{B} \cap (0,1]^2$, where $\mathcal{B}$ is Borel σ-*algebra* in $\mathbb{R}^2$.

RC2: $\theta \in \mathcal{C} \cap (0,1]^2$, where $\mathcal{C}$ is a Vapnik-Čhervonenkis (VC) class of sets in in $\mathbb{R}^2$.

RC3: $\epsilon_0 < \lambda(\theta) < 1 - \epsilon_0$ for some $\epsilon_0 > 0$, where $\lambda(\cdot)$ is the Lebesgue measure on $\mathbb{R}^2$.

The regularity conditions above jointly define the suitable class of sets, Θ, in which we consider θ_0 resides.

Like many research works in the literature, our approach for the boundary detection also starts from the lattice design, which assumes that the spatially referenced observations Y_s are made at the $n \times n$ equally spaced nodes of $(0,1]^2$, i.e., $s \in J_n \triangleq \{(\frac{i}{n}, \frac{j}{n}) : i, j = 1, \ldots, n\}$. We shall emphasize here that later in this work when we talk about Y_s, it is tacitly understood that $s \in J_n$ and likewise, Y_s, $s \in \theta$, actually means $s \in \theta \cap J_n$. We drop off "$\cap J_n$" from "$\theta \cap J_n$" for ease of the notation.

Let A_s be the lower-left orthant of size $\frac{1}{n} \times \frac{1}{n}$ with respect to node $s \in J_n$, i.e., $A_s = (\frac{i-1}{n}, \frac{i}{n}] \times (\frac{j-1}{n}, \frac{j}{n}]$, and $\mathcal{H}_n$ be the *algebra* generated by all these lower-left orthants. Clearly, $\mathcal{H}_n$ is monotonously increasing as n increases. Let $\hat{\theta}_n$ be any estimate of θ_0 based on Y_s, $s \in J_n$.

Clearly, for any fixed n, it makes sense only if $\hat{\theta}_n \in \mathcal{H}_n$. Following Carlstein & Krishnamoorthy [2], throughout this work the error of estimation is defined to be $\lambda(\hat{\theta}_n \Delta \theta_0) \wedge \lambda(\hat{\theta}_n^c \Delta \theta_0)$, where Δ is the symmetric difference between the two sets, i.e.

$$A \Delta B \triangleq (A \cap B^c) \cup (A^c \cap B),$$

and $a \wedge b = \min(a, b)$. The following lemma indicates that the Lebesgue measure of the symmetric difference is a suitable metric for the estimation error.

Lemma 3.1.1. *Lebesgue measure of the symmetric difference, $\lambda(\Delta)$, is a pseudo-metric of the distance between two sets in Lebesgue σ-algebra.*

Proof. For any set $A, B, C \in \mathcal{B}$ (Borel), need to verify

1. non-negativity: $\lambda(A\Delta A) = \lambda(\emptyset) = 0$;
2. symmetry: $\lambda(A\Delta B) = \lambda((A \cap B^c) \cup (A^c \cap B)) = \lambda(B\Delta A)$;
3. subadditivity: $\lambda(A\Delta C) \le \lambda(A\Delta B) + \lambda(B\Delta C)$. This is so because $\lambda(A\Delta B) = \lambda(A \cap B^c) + \lambda(A^c \cap B)$ and

$$\lambda(A \cap B^c) = \lambda(A \cap B^c \cap C) + \lambda(A \cap B^c \cap C^c) \le \lambda(B^c \cap C) + \lambda(A \cap C^c),$$
$$\lambda(A^c \cap B) = \lambda(A^c \cap B \cap C) + \lambda(A^c \cap B \cap C^c) \le \lambda(A^c \cap C) + \lambda(B \cap C^c),$$

together with the fact $\lambda(A\Delta B) = \lambda(A\cap B^c)+\lambda(A^c\cap B)$ and $\lambda(B\Delta C) = \lambda(B\cap C^c)+\lambda(B^c\cap C)$.

□

To develop our new boundary detection algorithm, we introduce the following mappings $\bar{\mu}(\cdot)$ and $\bar{x}_n(\cdot)$, $\Theta \mapsto \mathbb{R}^1$, which associate the spatially referenced observations with their locations.

1. $\bar{\mu}(\theta) = \frac{1}{\lambda(\theta)} \int_\theta \mu_s \lambda(ds)$, where $\lambda(\cdot)$ is the Lebesgue measure on $\mathbb{R}^2$. For instance, with $\mu_1 = 1$ and $\mu_2 = 0$, $\bar{\mu}(\theta) = \frac{\lambda(\theta \cap \theta_0)}{\lambda(\theta)}$.
2. $\bar{x}_n(\theta) = \frac{1}{\sharp(\theta)} \sum_{s \in \theta} Y_s$, where $\sharp(\cdot)$ is the counting measure of the nodes contained in $\theta \in \mathbb{R}^2$ with mesh size $n^{-1} \times n^{-1}$. For instance, with $\Omega = (0, 1]^2$, $\sharp(\Omega) = n^2$.

To simplify our notations, we also introduce the empirical measure of $\theta \in \Theta$, namely, $\sharp_n(\theta) \triangleq \frac{1}{n^2}\sharp(\theta)$. Clearly, $\sharp_n(\theta) = \lambda(\theta)$ for $\forall\ n$, if $\theta \in \mathcal{H}_n$.

The regularity condition RC1 ensures that for $\forall\ \theta \in \Theta$ and $\epsilon > 0$, $\exists\ N_\theta$ and ψ_θ ($\psi_\theta \subseteq \theta$) such that $\psi_\theta \in \mathcal{H}_n$ for $\forall\ n \geq N_\theta$ and $0 \leq \lambda(\psi_\theta \Delta\theta) < \epsilon$. The regularity condition RC2 further guarantees such an approximation of θ via ψ_θ is uniform, i.e.,

Lemma 3.1.2. *for* $\forall\ \epsilon > 0$, $\exists\ N$ *such that* $\exists\ \psi_\theta \in \mathcal{H}_n$ *for* $\forall\ n \geq N$ *(*$\psi_\theta \subseteq \theta$*) and* $\sup_{\theta\in\Theta} \lambda(\psi_\theta \Delta\theta) < 2\epsilon$.

Proof. Θ is a VC class hence *totally bounded*, viz. Vaart & Wellner [17](§2), i.e., for $\forall\ \epsilon > 0$, $\exists$ $\theta_1, \ldots, \theta_M \in \Theta$ ($M < \infty$) such that $\min\limits_{1\leq k\leq M} \lambda(\theta\Delta\theta_k) < \epsilon$ for $\forall\ \theta \in \Theta$. Clearly, $\exists\ N < \infty$ and $\psi_k \subseteq \theta_k$ ($k = 1, \ldots, M$) such that $\psi_k \in \mathcal{H}_N$ for $\forall\ n \geq N$ and $\max\limits_{1\leq k\leq M} \lambda(\psi_k\Delta\theta_k) < \epsilon$. Note that both M and N above are independent of θ and $\lambda(\psi_k\Delta\theta) \leq \lambda(\theta\Delta\theta_k) + \lambda(\psi_k\Delta\theta_k)$, then the conclusion follows. □

3.2. Criterion function

Following Müller & Song [8] and Carlstein & Krishnamoorthy [2], we consider the following criterion function,

$$g(\theta) = \lambda(\theta)\lambda(\theta^c)|\bar{\mu}(\theta) - \bar{\mu}(\theta^c)|,\ \theta \in \mathcal{B} \cap (0, 1]^2.$$

Lemma 3.2.1. *Suppose there exists* $\theta_0 \in \Theta$ *and constants* μ_1, μ_2 *such that*

$$\mu_s = \begin{cases} \mu_1, & s \in \theta_0 \\ \mu_2, & s \in \theta_0^c, \end{cases}$$

then θ *maximizes* $g(\theta)$ *if and only if* $\lambda(\theta\Delta\theta_0) \wedge \lambda(\theta^c\Delta\theta_0) = 0$.

Proof. Observe that $|\mu_1 - \mu_2|$ can be rescaled to 1, so without loss of generality, assume $\mu_1 = 1$ and $\mu_2 = 0$. Clearly

$$\begin{aligned} g(\theta) &= \lambda(\theta)\lambda(\theta^c)|\bar{\mu}(\theta) - \bar{\mu}(\theta^c)| = \lambda(\theta)\lambda(\theta^c)|\frac{\lambda(\theta \cap \theta_0)}{\lambda(\theta)} - \frac{\lambda(\theta^c \cap \theta_0)}{\lambda(\theta^c)}| \\ &= |\lambda(\theta^c)\lambda(\theta_0) - \lambda(\theta^c)\lambda(\theta^c \cap \theta_0) - \lambda(\theta)\lambda(\theta^c \cap \theta_0)| \\ &= |\lambda(\theta^c)\lambda(\theta_0) - \lambda(\theta^c \cap \theta_0)| \\ &= |\lambda(\theta_0)[\lambda(\theta_0^c \cap \theta^c) + \lambda(\theta_0 \cap \theta^c)] - \lambda(\theta_0 \cap \theta^c)| \\ &= |\lambda(\theta_0)\lambda(\theta_0^c \cap \theta^c) - \lambda(\theta_0^c)\lambda(\theta_0 \cap \theta^c)| \\ &= \lambda(\theta_0)\lambda(\theta_0^c)|\frac{\lambda(\theta_0^c \cap \theta^c)}{\lambda(\theta_0^c)} - \frac{\lambda(\theta_0 \cap \theta^c)}{\lambda(\theta_0)}| \leq \lambda(\theta_0)\lambda(\theta_0^c) \end{aligned}$$

The last inequality is due to the fact $0 \leq \frac{\lambda(\theta_0^c \cap \theta^c)}{\lambda(\theta_0^c)}, \frac{\lambda(\theta_0 \cap \theta^c)}{\lambda(\theta_0)} \leq 1$. The equality holds if and only if $\lambda(\theta_0^c \cap \theta^c) = \lambda(\theta_0^c)$ and $\lambda(\theta_0 \cap \theta^c) = 0$ or $\lambda(\theta_0^c \cap \theta^c) = 0$ and $\lambda(\theta_0 \cap \theta^c) = \lambda(\theta_0)$, which implies $\lambda(\theta \Delta \theta_0) \wedge \lambda(\theta^c \Delta \theta_0) = 0$. □

Note that with $\mu_1 = 1$ and $\mu_2 = 0$, $g(\theta_0) = \lambda(\theta_0)\lambda(\theta_0^c)$. Lemma 3.2.1 simply implies that θ_0 is the unique global maxima of the criterion function $g(\theta)$ (unique in the sense of Lebesgue measure of symmetric difference is 0). This motivates the M-estimator of θ_0: maxima of the empirical version of $g(\theta)$, i.e., let

$$g_n(\theta) = \sharp_n(\theta)\sharp_n(\theta^c)|\bar{x}_n(\theta) - \bar{x}_n(\theta^c)|, \ \theta \in \mathcal{B} \cap (0, 1]^2,$$

then for any fixed n, based on the fact $\max_{\theta \in \mathcal{H}_n} g_n(\theta) = \max_{\theta \in \mathcal{B}} g_n(\theta)$, we may define

$$\tilde{\theta}_n \triangleq \arg\max_{\theta \in \mathcal{H}_n} g_n(\theta).$$

3.3. VC class

VC class is systematically discussed in Vaart & Wellner [17](§2.6). For self-containing purpose, we provide its definition and main properties here.

Let $\mathcal{C}$ be a class of subsets of a set $\mathcal{X}$. For an arbitrary set $\{x_1, \ldots, x_k\}$, we say that $\mathcal{C}$ *picks out* a certain subset of $\{x_1, \ldots, x_k\}$ if the subset can be expressed as $C \cap \{x_1, \ldots, x_k\}$ for some

$C \in \mathcal{C}$. The collection $\mathcal{C}$ is said to *shatter* $\{x_1, \ldots, x_k\}$ if each of its 2^k subsets can be picked out. The VC-index $V(\mathcal{C})$ of the class $\mathcal{C}$ is the smallest k for which no set of size k is shattered by $\mathcal{C}$. Clearly, the more refined $\mathcal{C}$ is, the larger its index is. The index $V(\mathcal{C})$ is formally defined as

$$V(\mathcal{C}) \quad = \quad \inf\{k : \max_{x_1,\ldots,x_k} \triangle_k(\mathcal{C}, x_1, \ldots, x_k) < 2^k\}$$

where $\triangle_k(\mathcal{C}, x_1, \ldots, x_k)$ is the number of subsets of $\{x_1, \ldots, x_k\}$ picked out by $\mathcal{C}$. A class of sets $\mathcal{C}$ is called VC if $V(\mathcal{C}) < \infty$.

As indicated in Vaart & Wellner [17], §2.6, let $\mathcal{C}$ and $\mathcal{D}$ be VC-classes of sets in a set $\mathcal{X}$ and $\phi : \mathcal{X} \mapsto \mathcal{Y}$ and $\psi : \mathcal{Z} \mapsto \mathcal{X}$ fixed functions, then

(1) $\mathcal{C}^c = \{C^c : C \in \mathcal{C}\}$ is VC;

(2) $\mathcal{C} \cap \mathcal{D} = \{C \cap D : C \in \mathcal{C}, D \in \mathcal{D}\}$ is VC;

(3) $\mathcal{C} \cup \mathcal{D} = \{C \cup D : C \in \mathcal{C}, D \in \mathcal{D}\}$ is VC;

(4) $\phi(\mathcal{C})$ is VC if ϕ is one-to-one;

(5) $\psi^{-1}(\mathcal{C})$ is VC;

(6) the sequential closure of $\mathcal{C}$ for pointwise convergence of indicator functions is VC; for VC-classes $\mathcal{C}$ and $\mathcal{D}$ in sets $\mathcal{X}$ and $\mathcal{Y}$,

(7) $\mathcal{C} \times \mathcal{D}$ is VC in $\mathcal{X} \times \mathcal{Y}$.

Moreover, by Theorem 2.6.4 in Vaart & Wellner [17], there exists a universal constant K such that for any VC class $\mathcal{C}$ of dimension ν and any $\epsilon \in (0, 1)$,

$$N(\epsilon, \mathcal{C}, \lambda(\Delta)) \leq K\nu(4e)^{\nu}\epsilon^{-(\nu-1)}$$

where $N(\epsilon, \mathcal{C}, \lambda(\Delta))$ is the *covering number* of the class $\mathcal{C}$ under radius ϵ (based on the pseudo-metric $\lambda(\Delta)$ as defined in §3.2). Clearly, an immediate consequence of this result is that VC class is totally bounded.

It turns out that a VC class may cover many commonly seen 2-dimension geometric shapes, including circles, ellipses and convex m-gons ($m \leq c_0 < \infty$). Unlike some other change-curve detection methods, we do not presume the change-curve to have specific geometric shapes. Instead, our regularity condition RC2 only assumes the change-curve belonging to a VC class. Note that the union and intersection of finite VC classes are still VC, the regularity condition RC2 actually enable us to consider the change-curve in a much larger class than the class of presumed geometric shapes, say, circles by scan statistics method.

3.4. Estimation algorithm

Note the cardinality of $\mathcal{H}_n$, $|\mathcal{H}_n| = 2^{n^2}$. A naive approach of obtaining M-estimator of the change-curve θ_0, $\tilde{\theta}_n$, is to search the global maxima of $g_n(\theta)$ across all the possible 2^{n^2} candidate sets in $\mathcal{H}_n$, which is necessarily computationally expansive. In this work, we introduce a new computationally efficient algorithm for detecting the change-curve θ_0 which utilizes a sort of monotonicity in mapping the change-curve. Its computational complexity turns out to be up to n^2 instead of 2^{n^2}.

To be specific, for fixed n, let $T_y \triangleq \bigcup\limits_{Y_s \geq y} A_s$, $T_y^c \triangleq \bigcup\limits_{Y_s < y} A_s$, where A_s is the lower-left orthant of size $\frac{1}{n} \times \frac{1}{n}$ with respect to the node $s \in J_n$. Then $T_y, T_y^c \in \mathcal{H}_n$ and $g_n(T_y) = \sharp_n(T_y)\sharp_n(T_y^c)|\bar{x}_n(T_y) - \bar{x}_n(T_y^c)|$. Let $y^* = \arg\max_{y \in \mathbb{R}} g_n(T_y)$, our estimator of θ_0 is simply $\hat{\theta}_n = T_{y^*}$. Clearly, for $\forall \ \{Y_s\}_{s \in J}$, where $J = \bigcup_{n=1}^{\infty} J_n$, $\{T_y\}_{y \in \mathbb{R}}$ is a VC-class of dimension 2 by definition and $g_n(T_y)$ has at most n^2 distinct values, each of which occurs at the value of the observed response. It turns out

Lemma 3.4.1. *For any fixed n, if all Y_s are distinct, then* $\max\limits_{\theta \in \mathcal{H}_n} g_n(\theta) = \max\limits_{y \in \mathbb{R}} g_n(T_y)$.

Proof. Clearly $\max\limits_{\theta \in \mathcal{H}_n} g_n(\theta) \geq \max\limits_{y \in \mathbb{R}} g_n(T_y)$, since $T_y \in \mathcal{H}_n$ for any real value y. Therefore, it only remains to show $\max\limits_{\theta \in \mathcal{H}_n} g_n(\theta) \leq \max\limits_{y \in \mathbb{R}} g_n(T_y)$. Observe that for $\forall \ \theta \in \mathcal{H}_n$ with $\sharp_n(\theta) = \sharp_n(T_y)$, $|\bar{x}_n(\theta) - \bar{x}_n(\theta^c)| \leq |\bar{x}_n(T_y) - \bar{x}_n(T_y^c)|$. Then

$$\sharp_n(\theta)\sharp_n(\theta^c)|\bar{x}_n(\theta) - \bar{x}_n(\theta^c)| \leq \sharp_n(T_y)\sharp_n(T_y^c)|\bar{x}_n(T_y) - \bar{x}_n(T_y^c)|$$

So with $\sharp_n(\theta) = \sharp_n(T_y)$, $g_n(\theta) \leq g_n(T_y)$. The conclusion follows by first taking $\sup_{y \in \mathbb{R}}$ on the righthand side and then taking $\sup_{\theta \in \mathcal{H}_n}$ on the lefthand side of the inequality above. □

Lemma 3.4.1 indicates though it's possible $\lambda(\hat{\theta}_n \Delta \tilde{\theta}_n) \wedge \lambda(\hat{\theta}_n \Delta \tilde{\theta}_n^c) \neq 0$, $g_n(\hat{\theta}_n) = g_n(\tilde{\theta}_n)$ always holds.

3.5. Spatial autocorrelation

As discussed in §3.1, the location of the spatially referenced data Y_s in the lattice design on $(0,1]^2$ with mesh size $n^{-1}\times n^{-1}$ is re-scaled to fit in the unit square. Therefore, the Euclidean distance between the any two arbitrary observations Y_v and Y_w, $v, w \in (0,1]^2$, is actually $n\|v-w\|$ rather than $\|v-w\|$, where $\|\cdot\|$ is the Euclidean distance. Now we circumscribe the spatial structure of the noise ϵ_s to the spatially referenced data Y_s in this lattice design. In addition to $\{\epsilon_s\}$, $s \in (0,1]^2$, being strict stationary with $E\epsilon_s = 0$ and $\operatorname{var}\epsilon_s = \sigma^2$, we further assume that $\{\epsilon_s\}$ has autocorrelation $\rho_n(d)$ satisfying $\rho_n(d) = o((nd)^{-2})$, where d is the scaled Euclidean distance between the location of two arbitrary observations in the unit square.

Let $\rho(\cdot)$ be the autocorrelation of $\{\epsilon_s\}$ at *unscaled* Euclidean distance. Clearly, $\rho_n(d) = \rho(nd)$. The polynomial rate decaying of $\rho_n(d)$ or equivalently $\rho(nd)$ guarantees that $\{\epsilon_s\}$ satisfies the ρ-mixing condition as discussed in Goldie & Greenwood [3] and Goldie & Greenwood [4] with the exponent of regularity being 2. This requirement on $\rho_n(d)$ concerns only the limiting situation of spatial correlation of $\rho(\cdot)$, hence it virtually accommodates many popular correlogram models in geostatistics, including

1. spherical correleogram: $\rho(d) = c\{1 - \frac{3}{2}(\frac{d}{r}) + \frac{1}{2}(\frac{d}{r})^3\}\mathbb{1}_{\{d\le r\}}$;
2. powered exponential correleogram: $\rho(d) = c\exp\{-(\frac{d}{r})^p\}$;
3. Matérn correleogram: $\rho(d) = \frac{c}{\Gamma(\nu)2^{\nu-1}}(\sqrt{2\nu}\frac{d}{r})^\nu K_\nu(\sqrt{2\nu}\frac{d}{r})$;

where r is the practical range of spatial dependency, $c \in (0,1]$, $p \in (0,2]$ and $\nu > 0$ are some universal constants, and $K_\nu(\cdot)$ is the 2^{nd} type modified Bessel function.

Lemma 3.5.1. *Let Y_s be as defined in §3.1. Suppose that $\{\epsilon_s\}$, $s \in (0,1]^2$, is strict stationary with $E\epsilon_s = 0$, $\operatorname{var}\epsilon_s = \sigma^2$, and $\rho(d) = o(d^{-2})$ (as $d \to \infty$), then there exists $\gamma > 1$ such that for any $\theta \in \Theta$,*

$$\lambda(\theta) \le \varliminf_{n\to\infty} \operatorname{var}(n^{-1}\sum_{s\in\theta} Y_s/\sigma) \le \varlimsup_{n\to\infty} \operatorname{var}(n^{-1}\sum_{s\in\theta} Y_s/\sigma) \le \gamma\lambda(\theta).$$

Proof. Observe that for any $\theta \in \Theta$,

$$\sigma^2 n^2 \sharp_n(\theta) \le \operatorname{var}(\sum_{s\in\theta} Y_s) \le \sigma^2 n^2 \sharp_n(\theta) + \sigma^2 n^2 \sharp_n(\theta) \sum_{k=1}^{\sqrt{2}n} 8k\rho_n(kn^{-1}),$$

so

$$\sharp_n(\theta) \le \mathrm{var}(n^{-1}\sum_{s\in\theta} Y_s/\sigma) \le \sharp_n(\theta)\{1+8\sum_{k=1}^{+\infty} k\rho(k)\}.$$

Note $\rho(k) = o(k^{-2})$, then $\sum_{k=1}^{+\infty} k\rho(k) < +\infty$. Let $\gamma \triangleq 1+8\sum_{k=1}^{+\infty} k\rho(k)$, the conclusion follows. □

Observe $\bar{x}_n(\theta) = \frac{1}{n^2\sharp_n(\theta)}\sum_{s\in\theta} Y_s$, an immediate consequence of this lemma is that for $\forall\ \theta \in \Theta$, $\mathrm{var}\,\bar{x}_n(\theta) = O(n^{-2})$.

3.6. Consistency of Estimator

Let $\hat{\theta}_n = \arg\max_{\theta\in\Theta} g_n(\theta)$, $\hat{\theta}_n$ turns out to be consistent in the sense of $\lambda(\hat{\theta}_n\Delta\theta_0) \wedge \lambda(\hat{\theta}_n^c\Delta\theta_0) \to 0$ as $n \to \infty$. To prove this result, we will show $|g_n(\theta) - g(\theta)| \xrightarrow{p} 0$ as $n \to \infty$ for $\forall\ \theta \in \Theta$. Thanks to Lemma 3.6.1 as below (cf. Vaart [16], §5, Theorem 5.7), the consistency of $\hat{\theta}_n$ then follows if uniformity of the convergence holds, i.e., $\sup_{\theta\in\Theta}|g_n(\theta) - g(\theta)| \xrightarrow{p} 0$.

Lemma 3.6.1. *Let $\theta_0 = \arg\max_{\theta\in\Theta} g(\theta)$ and $\delta(\theta_0,\epsilon)$ be the neighborhood of θ_0, i.e., $\lambda(\theta\Delta\theta_0) < \epsilon$, $\forall\ \theta \in \delta(\theta_0,\epsilon)$. Suppose $g(\theta_0) > g(\theta)$ if $\theta \notin \delta(\theta_0,\epsilon)$ for $\forall\ \epsilon > 0$ and $\sup_{\theta\in\Theta}|g_n(\theta) - g(\theta)| \xrightarrow{p} 0$, then $\lambda(\hat{\theta}_n\Delta\theta_0) \xrightarrow{p} 0$.*

Proof. Observe $g(\theta_0) \ge g(\hat{\theta}_n)$ and $g_n(\hat{\theta}_n) \ge g_n(\theta_0)$, then

$$\begin{aligned} 0 \le g(\theta_0) - g(\hat{\theta}_n) &= [g(\theta_0) - g_n(\theta_0)] + [g_n(\hat{\theta}_n) - g(\hat{\theta}_n)] + [g_n(\theta_0) - g_n(\hat{\theta}_n)] \\ &\le |g_n(\theta_0) - g(\theta_0)| + \sup_{\theta\in\Theta}|g_n(\theta) - g(\theta)|, \end{aligned}$$

so $0 \le g(\theta_0) - g(\hat{\theta}_n) \le o_p(1)$ by the virtue of $\sup_{\theta\in\Theta}|g_n(\theta) - g(\theta)| \xrightarrow{p} 0$. Appealing to the assumption $g(\theta_0) > g(\hat{\theta}_n)$ if $\hat{\theta}_n \notin \delta(\theta_0,\epsilon)$, the conclusion follows. □

Before we show the uniform convergence in probability of $g_n(\theta)$ to $g(\theta)$, we need one more important result.

Lemma 3.6.2. *Suppose that $\{\epsilon_s\}$, $s \in (0,1]^2$, is strict stationary with $E\epsilon_s = 0$, $\mathrm{var}\,\epsilon_s = \sigma^2$, and $\rho(d) = o(d^{-2})$ (as $d \to \infty$), then $\sup_{\theta\in\Theta}|\bar{x}_n(\theta) - \bar{\mu}(\theta)| \xrightarrow{p} 0$.*

Proof. Without loss of generality, let $\mu_1 = 1$, $\mu_2 = 0$. Then $\bar{\mu}(\theta) = \frac{\lambda(\theta\cap\theta_0)}{\lambda(\theta)}$ for $\forall\ \theta \in \Theta$. Hence

$$\begin{aligned} |\bar{x}_n(\theta) - \bar{\mu}(\theta)| &= |\frac{1}{\sharp(\theta)}\sum_{s\in\theta} Y_s - \frac{\lambda(\theta\cap\theta_0)}{\lambda(\theta)}| \\ &\leq |\frac{1}{\sharp(\theta)}\{\sum_{s\in\theta\cap\theta_0}(Y_s - 1) + \sum_{s\in\theta\cap\theta_0^c} Y_s\}| + |\frac{\sharp_n(\theta\cap\theta_0)}{\sharp_n(\theta)} - \frac{\lambda(\theta\cap\theta_0)}{\lambda(\theta)}|. \end{aligned}$$

First, observe $|\frac{\sharp_n(\theta\cap\theta_0)}{\sharp_n(\theta)} - \frac{\lambda(\theta\cap\theta_0)}{\lambda(\theta)}| \leq \frac{1}{\lambda(\theta)}\{|\sharp_n(\theta) - \lambda(\theta)| + |\sharp_n(\theta\cap\theta_0) - \lambda(\theta\cap\theta_0)|\}$, $\inf\limits_{\theta\in\Theta}\lambda(\theta) > \epsilon_0$, and $|\sharp_n(\theta) - \lambda(\theta)|, |\sharp_n(\theta\cap\theta_0) - \lambda(\theta\cap\theta_0)| \leq \sup_{\theta\in\Theta}\lambda(\psi_\theta\Delta\theta)$, then

$$\sup_{\theta\in\Theta}|\frac{\sharp_n(\theta\cap\theta_0)}{\sharp_n(\theta)} - \frac{\lambda(\theta\cap\theta_0)}{\lambda(\theta)}| = o(1)$$

follows, by taking $sup_{\theta\in\Theta}$ over the LHS of the last inequality above and appealing to Lemma 3.1.2.

Second, observe $Y_s - 1$, $\forall\ s \in \theta\cap\theta_0$ and Y_s, $\forall\ s \in \theta\cap\theta_0^c$ are identically distributed with mean 0. Appealing to Lemma 3.5.1, $\text{var}\{\frac{1}{\sharp(\theta)}[\sum_{s\in\theta\cap\theta_0}(Y_s - 1) + \sum_{s\in\theta\cap\theta_0^c} Y_s]\} = o(1)$. So, by Cauchy inequality, $|\frac{1}{\sharp(\theta)}\{\sum_{s\in\theta\cap\theta_0}(Y_s - 1) + \sum_{s\in\theta\cap\theta_0^c} Y_s\}| = o_p(1)$.

Note that Θ is a VC class hence a Gilivenko-Cantelli (GC) class (viz. Vaart & Wellner [17], §2.6), i.e., $|\frac{1}{\sharp(\theta)}\{\sum_{s\in\theta\cap\theta_0}(Y_s - 1) + \sum_{s\in\theta\cap\theta_0^c} Y_s\}| = o_p(1)$ implies

$$\sup_{\theta\in\Theta}|\frac{1}{\sharp(\theta)}\{\sum_{s\in\theta\cap\theta_0}(Y_s - 1) + \sum_{s\in\theta\cap\theta_0^c} Y_s\}| = o_p(1).$$

Combine the two equations above, the conclusion follows. □

Now we may proceed to to show the uniform convergence in probability of $g_n(\theta)$ to $g(\theta)$ for $\forall\ \theta \in \Theta$.

Theorem 3.6.1. *In the same setting as Lemma 3.6.2,* $\sup_{\theta\in\Theta}|g_n(\theta) - g(\theta)| \xrightarrow{p} 0$.

Proof. Without loss of generality, let $\mu_1 = 1$, $\mu_2 = 0$. Then $0 \leq \bar{\mu}(\theta) = \frac{\lambda(\theta\cap\theta_0)}{\lambda(\theta)} \leq 1$ for $\forall\ \theta \in \Theta$. Observe

$$
\begin{aligned}
&|g_n(\theta) - g(\theta)| = |\sharp_n(\theta)\sharp_n(\theta^c)|\bar{x}_n(\theta) - \bar{x}_n(\theta^c)| - \lambda(\theta)\lambda(\theta^c)|\bar{\mu}(\theta) - \bar{\mu}(\theta^c)|| \\
\leq\ & |\sharp_n(\theta)\sharp_n(\theta^c)[\bar{x}_n(\theta) - \bar{x}_n(\theta^c)] - \lambda(\theta)\lambda(\theta^c)[\bar{\mu}(\theta) - \bar{\mu}(\theta^c)]| \\
\leq\ & |\sharp_n(\theta)\sharp_n(\theta^c)\bar{x}_n(\theta) - \lambda(\theta)\lambda(\theta^c)\bar{\mu}(\theta)| + |\sharp_n(\theta)\sharp_n(\theta^c)\bar{x}_n(\theta^c) - \lambda(\theta)\lambda(\theta^c)\bar{\mu}(\theta^c)| \\
\leq\ & |\sharp_n(\theta)\sharp_n(\theta^c) - \lambda(\theta)\lambda(\theta^c)|\bar{\mu}(\theta) + \sharp_n(\theta)\sharp_n(\theta^c)|\bar{x}_n(\theta) - \bar{\mu}(\theta)| \\
& + |\sharp_n(\theta)\sharp_n(\theta^c) - \lambda(\theta)\lambda(\theta^c)|\bar{\mu}(\theta^c) + \sharp_n(\theta)\sharp_n(\theta^c)|\bar{x}_n(\theta^c) - \bar{\mu}(\theta^c)| \\
\leq\ & 4\sup_{\theta\in\Theta}|\sharp_n(\theta) - \lambda(\theta)| + 2\sup_{\theta\in\Theta}|\bar{x}_n(\theta^c) - \bar{\mu}(\theta^c)|
\end{aligned}
$$

Note $\theta^c \in \Theta$, $|\sharp_n(\theta) - \lambda(\theta)| \leq \sup_{\theta\in\Theta}\lambda(\phi_\theta\Delta\theta) = o(1)$ by the virtue of Lemma 3.1.2 and $\sup_{\theta\in\Theta}|\bar{x}_n(\theta) - \bar{\mu}(\theta)| = o(1)$ by Lemma 3.6.2, then take $\sup_{\theta\in\Theta}$ over the LHS of the inequality above the conclusion follows. □

With the help of Lemma 3.4.1, Lemma 3.6.1 and Theorem 3.6.1, the consistency of $\hat{\theta}_n$ or $\tilde{\theta}_n$ for θ_0 holds. The estimator for $\mu(\theta_0) = \mu_1$, $\mu(\theta_0^c) = \mu_2$ and σ^2 could be defined as follows:

$$
\begin{aligned}
&\hat{\mu}_1^{(n)} = \bar{x}_n(\hat{\theta}_n),\ \hat{\mu}_2^{(n)} = \bar{x}_n(\hat{\theta}_n^c), \\
&\hat{\sigma}_n^2 = \frac{1}{n^2}\Big\{\sum_{s\in\hat{\theta}_n}(Y_s - \bar{x}_n(\hat{\theta}_n))^2 + \sum_{s\in\hat{\theta}_n}(Y_s - \bar{x}_n(\hat{\theta}_n))^2\Big\}.
\end{aligned}
$$

Now we prove consistency of these estimates.

Theorem 3.6.2. *Suppose $\theta_0 \in \Theta$, Θ satisfies the regularity conditions RC1-RC3, and ϵ_s satisfies the assumptions in Lemma 3.6.2, then $\hat{\mu}_1^{(n)} \xrightarrow{p} \mu_1$ and $\hat{\mu}_2^{(n)} \xrightarrow{p} \mu_2$. Moreover, if $E\epsilon_s^4 < \infty$ and ϵ_s^2 has autocorrelation $\rho^\star(d) = d^{-2}$ (d unscaled distance), then $\hat{\sigma}_n^2 \xrightarrow{p} \sigma^2$.*

Proof. Without loss of generality, we assume $\mu_1 = 1$ and $\mu_2 = 0$ as before. Observe $|\mu(\hat{\theta}_n) - \mu(\theta_0)| = |\frac{1}{\lambda(\hat{\theta}_n)}\int_{\hat{\theta}_n\cap\theta_0} ds - 1| = \frac{\lambda(\hat{\theta}_n) - \lambda(\hat{\theta}_n\cap\theta_0)}{\lambda(\hat{\theta}_n)} = o_p(1)$ by Lemma 3.1.2 and $|\bar{x}_n(\hat{\theta}_n) - \mu(\hat{\theta}_n)| = o_p(1)$ by Lemma 3.6.2, then

$$
|\bar{x}_n(\hat{\theta}_n) - \mu(\theta_0)| \leq |\bar{x}_n(\hat{\theta}_n) - \mu(\hat{\theta}_n)| + |\mu(\hat{\theta}_n) - \mu(\theta_0)| = o_p(1),
$$

So $\hat{\mu}_1^{(n)} \xrightarrow{p} \mu_1$. Similarly, $\hat{\mu}_2^{(n)} \xrightarrow{p} \mu_2$. Note $\lambda(\hat{\theta}_n\Delta\theta_0) = o_p(1)$, $\bar{x}_n(\hat{\theta}_n) - 1 = o_p(1)$, and $\bar{x}_n(\hat{\theta}_n^c) = o_p(1)$, then

$$\begin{aligned}
\hat{\sigma}_n^2 &= \frac{1}{n^2}\Big\{\sum_{s\in\hat{\theta}_n}[Y_s-\bar{x}_n(\hat{\theta}_n)]^2+\sum_{s\in\hat{\theta}_n^c}[Y_s-\bar{x}_n(\hat{\theta}_n^c)]^2\Big\}\\
&= \frac{1}{n^2}\Big\{\sum_{s\in\hat{\theta}_n}(Y_s-1)^2+\sum_{s\in\hat{\theta}_n^c}Y_s^2\Big\}+o_p(1)\\
&= \frac{1}{n^2}\Big\{\sum_{s\in\hat{\theta}_n\cap\theta_0}\epsilon_s^2+\sum_{s\in\hat{\theta}_n^c\cap\theta_0^c}\epsilon_s^2\Big\}+\frac{1}{n^2}\Big\{\sum_{s\in\hat{\theta}_n\cap\theta_0^c}(Y_s-1)^2+\sum_{s\in\hat{\theta}_n^c\cap\theta_0}Y_s^2\Big\}+o_p(1)
\end{aligned}$$

Note ϵ_s^2, $s\in(0,1]^2$ is strictly stationary, and additionally $E\epsilon_s^4<\infty$, $\rho^\star(d)=d^{-2}$, then

$$\frac{1}{n^2}\Big\{\sum_{s\in\hat{\theta}_n\cap\theta_0}\epsilon_s^2+\sum_{s\in\hat{\theta}_n^c\cap\theta_0^c}\epsilon_s^2\Big\}=\sigma^2+o_p(1).$$

So $\hat{\sigma}_n^2=\sigma^2+o_p(1)$ follows. □

3.7. Existence of boundary

Now we develop a statistical hypothesis test to examine whether or not the change-curve exists. This is also an important issue to the boundary detection problem. In the setting of the boundary detection problem as described in the beginning of of this section, the hypothesis test of the existence of the change-curve could be simply formulated as follows:

$$H_0:\mu_1=\mu_2 \quad \text{vs} \quad H_a:\mu_1\neq\mu_2.$$

Before presenting our test statistic of this test, we shall study the asymptotic behavior of $n\sharp_n(\theta)[\bar{x}_n(\theta)-\bar{\mu}(\theta)]$, a stochastic process indexed by $\theta\in\Theta$. Without loss of generality, we assume $\mu_1=\mu_2=0$ under H_0, then $\bar{\mu}(\theta)\equiv 0$. Observe that

$$n\sharp_n(\theta)\bar{x}_n(\theta)=n^{-1}\sum_{s\in\theta}Y_s,$$

by Theorem 2.2 in Alexander & Pyke [1] or Theorem 2.5.1 in Vaart & Wellner [17], §2.5, one has

Lemma 3.7.1. *Suppose the noise ϵ_s, $s \in (0,1]^2$ are iid, then $n^{-1}\sum_{s\in\theta} Y_s/\sigma \xrightarrow{d} W(\theta)$, where $W(\theta)$ is a standard Wiener process indexed by θ in Θ with $EW(\theta_1) = 0$ and $\operatorname{cov}(W(\theta_1), W(\theta_2)) = \lambda(\theta_1 \cap \theta_2)$ for $\forall\ \theta_1, \theta_2 \in \Theta$, if Y_s are iid with $EY_s = 0$ and $\operatorname{var} Y_s = \sigma^2$, $s \in J_n$.*

Now we propose out test statistic $T_n \triangleq n \sup_{\theta\in\Theta} |g_n(\theta)|/\hat{\sigma}_n$, which is simply sup of the scaled criterion function $g_n(\theta) = \sharp_n(\theta)\sharp_n(\theta^c)|\bar{x}_n(\theta) - \bar{x}_n(\theta^c)|$, $\theta \in \Theta$, where $\hat{\sigma}_n$ is as defined in Theorem 3.6.2. We consider its null limiting distribution here.

Theorem 3.7.1. *Suppose the noise ϵ_s, $s \in (0,1]^2$ are iid, under H_0*

$$n \sup_{\theta\in\Theta} |g_n(\theta)|/\hat{\sigma}_n \xrightarrow{d} \sup_{\theta\in\Theta} |B(\theta)|$$

where $B(\theta)$ is the standard Brownian bridge stochastic process indexed by θ in Θ, i.e., $B(\theta) = W(\theta) - \lambda(\theta)W(\Omega)$, where $\Omega = (0,1]^2$.

Proof. Observe $\sharp_n(\theta)\bar{x}_n(\theta) + \sharp_n(\theta^c)\bar{x}_n(\theta^c) = \bar{x}_n(\Omega)$, then $\bar{x}_n(\theta) - \bar{x}_n(\theta^c) = \frac{\bar{x}_n(\theta) - \bar{x}_n(\Omega)}{1 - \sharp_n(\theta)}$. So $n|g_n(\theta)| = |n^{-1}\sum_{s\in\theta} Y_s - \sharp_n(\theta)n^{-1}\sum_{s\in\Omega} Y_s|$. Appeal to Theorem 1.1 in Vaart & Wellner [17] (Extended continuous mapping) and note that by Lemma 3.1.2 $\sup_{\theta\in\Theta} |\lambda(\theta) - \sharp_n(\theta)| \to 0$, $[n^{-1}\sum_{s\in\theta} Y_s - \sharp_n(\theta)n^{-1}\sum_{s\in\Omega} Y_s]/\sigma \xrightarrow{d} [W(\theta) - \lambda(\theta)W(\Omega)]$. Therefore with $\hat{\sigma}_n/\sigma \xrightarrow{p} 1$, the conclusion follows according to Slutsky's theorem [13]. □

By Lemma 3.2.1, $ng_n(\theta_0) = n\lambda(\theta_0)\lambda(\theta_0^c) \to \infty$ if θ_0 is the true change-curve under H_a. Therefore, given σ, a large test score of T_n is a support of H_a. So given $\sigma = 1$, at significance level α, H_a is concluded if $T_n > c_\alpha$, where c_α is the upper α^{th} percentile of $\sup_{\theta\in\Theta} |B(\theta)|$. Unfortunately, $\sup_{\theta\in\Theta} |B(\theta)|$ does not have an explicit distribution. We obtain $c_{0.05}$ via numeric simulation. Below is the histogram of the simulated values test score of T_n in $1,000$ simulations of $n = 20$ with $Y_s \overset{iid}{\sim} N(0,1)$. $c_{0.05}$ turns out to be 0.191 in this case.

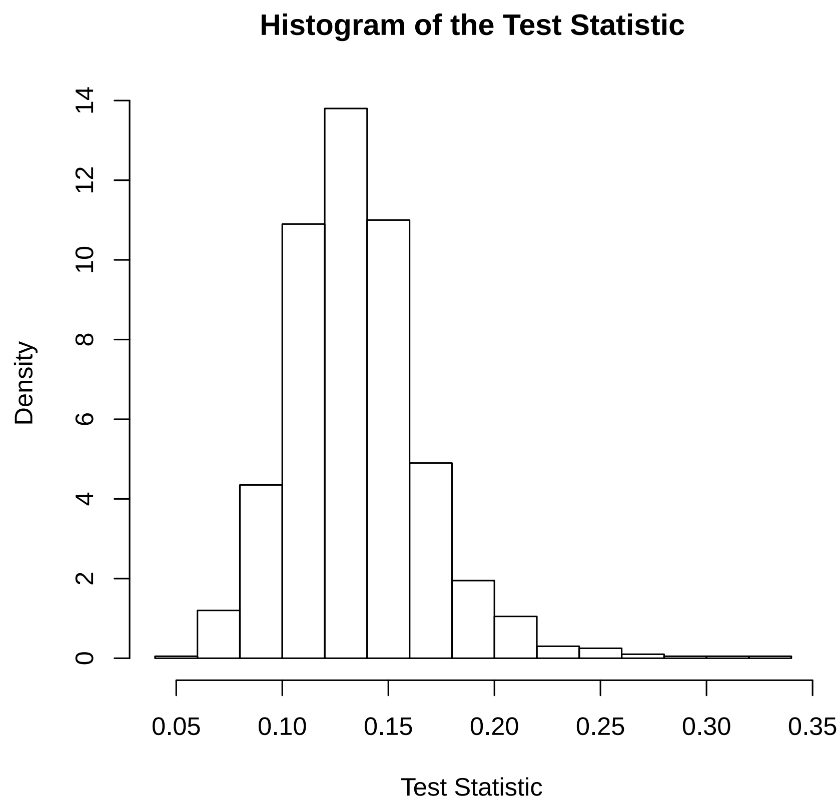


Figure 3.1. Histogram of T_n scores; n=20

As for the case of Y_s being spatially correlated, e.g. the noise ϵ_s satisfies the assumption in Lemma 3.5.1, though $\hat{\sigma}_n^2$ is consistent for σ^2 and $\operatorname{var} \bar{x}_n(\theta) = O(n^{-2})$, $\forall\ \theta \in \Theta$, but $\lim_{n\to\infty} \operatorname{var}(n^{-1} \sum_{s\in\theta} Y_s/\sigma)$ does not necessarily exist. Even if it exists, one has $n^{-1} \sum_{s\in\theta} Y_s/\sigma \xrightarrow{d} \xi(\theta)$, where $\xi(\theta)$ is some Gaussian stochastic process indexed by θ in Θ, not the standard Weiner process $W(\theta)$. This is so because $\operatorname{var}(n^{-1} \sum_{s\in\theta} Y_s/\sigma)$ depends on the shape of θ, not purely $\lambda(\theta)$.

4. NUMERIC STUDY

4.1. Simulation data

In this section, we conduct numeric study to examine the performance of our new algorithm in detection of the change-curve and compare it with that of the popular spatial clustering method: Scan statistics. To be compliant with the setting of the change-curve detection problem as discussed in §1, our simulation study is solely confined to the $n \times n$ lattice design in the unit square $(0,1]^2$ with $n = 30$, which has a change-curve splitting the unit square $(0,1]^2$ into two regions: θ_0 and θ_0^c. We consider two cases of the change-curve as described below:

$$C_1 : \{(x,y) \in (0,1]^2 \mid y = \sqrt{x}\},$$

$$C_2 : \{(x,y) \in (0,1]^2 \mid (x-0.5)^2 + (y-0.5)^2 = 0.4^2\},$$

We set the noise ϵ_s having a marginal distribution $N(0,1)$ for $\forall\ s \in (0,1]^2$. We also consider the following two cases of its spatial autocorrelation

$$D_1 : \rho(d) = 0,$$

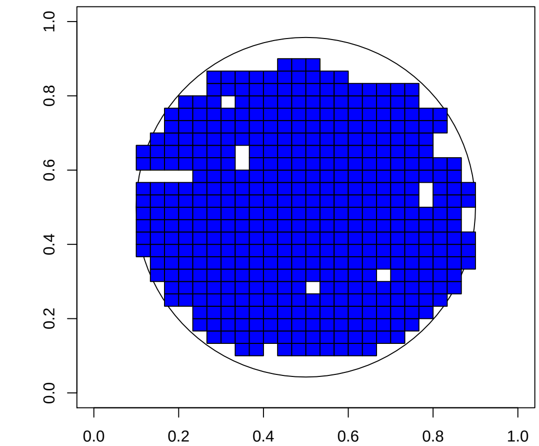
$$D_2 : \rho(d) = \exp\{-20d\},$$

where d is the *scaled* Euclidean distance.

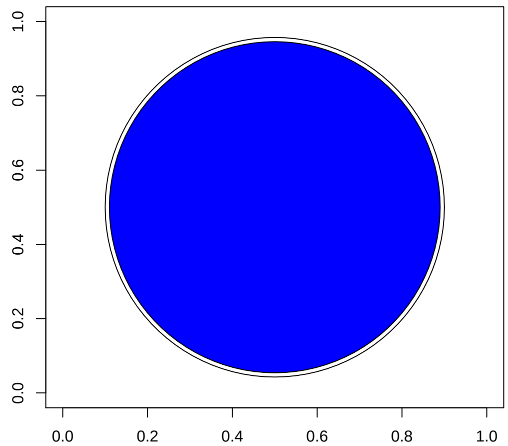
For each combination of the change-curve C_1, C_2 and spatial autocorrelation D_1, D_2, we first simulate the observations $Y_s = \mu_s + \epsilon_s$ at all the nodes with $\mu_s = 100$ if $s \in \theta_0$ and $\mu_s = 0$ if $s \in \theta_0^c$, then apply our algorithm and the scan statistics to determine the change-curve in the unit square, respectively. Computation of the change-curve via scan statistics method is implemented in the SaTScan software. The two resulting estimates of the change-curve in each case are compared for the estimation error $\lambda(\hat{\theta}_n \Delta \theta_0) \wedge \lambda(\hat{\theta}_n \Delta \theta_0^c)$.

4.1.1. Independent data

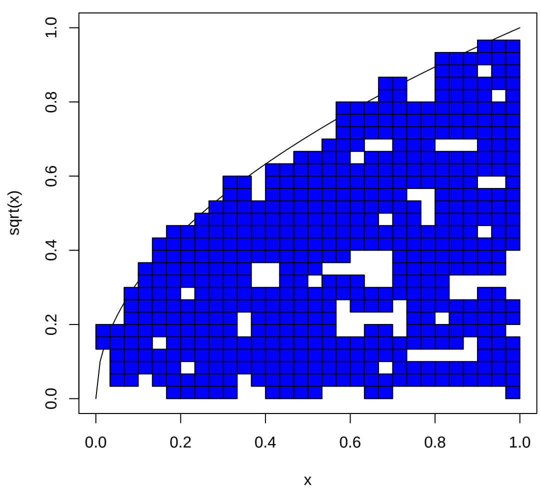
The following four plots illustrate how our method and spatial scan statistic works for the independent case, D_1. When the true boundary is a circle, the spatial scan statistic works great. However, when the boundary has a hyperbolic shape, the spatial scan statistic does not perform as well as our method.



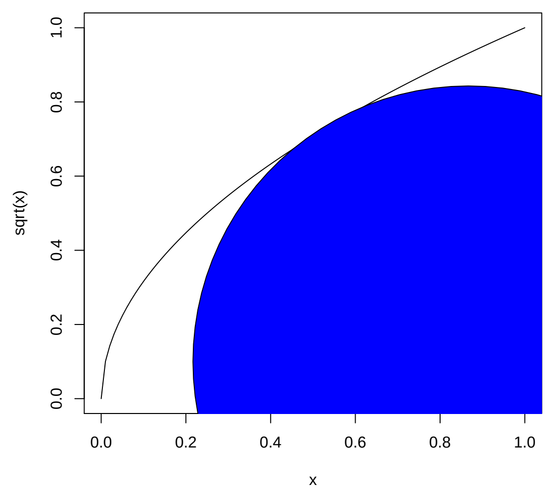
(a) $C_2\&D_1$, Our Method



(b) $C_2\&D_1$, Spatial Scan Statistic



(c) $C_1\&D_1$, Our Method



(d) $C_1\&D_1$, Spatial Scan Statistic

Figure 4.1. Comparison between Our Method and Spatial Scan Statistic; $n = 30$

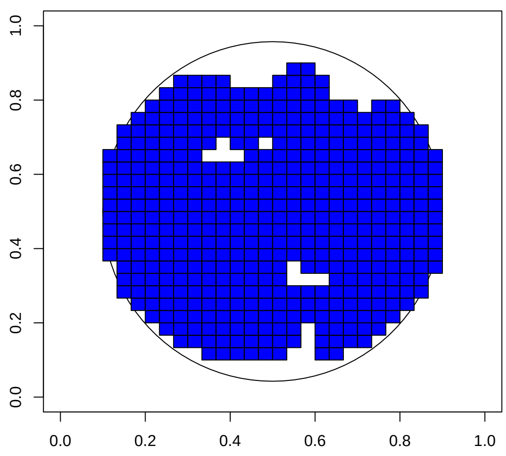
This can also be demonstrated by calculating the symmetric difference between the true boundary and the boundary estimate given by our method and spatial scan statistics, which is listed in the following table.

Table 4.1. Symmetric Difference between the True Boundary and the Boundary Estimate: Independent Data

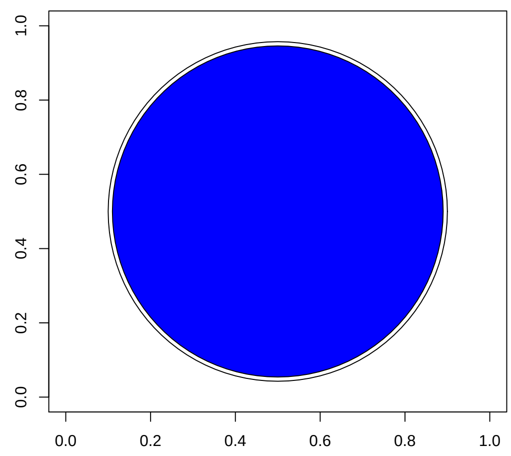
	Circular	Hyperbolic
Our Method	0.0833	0.1110
Spatial Scan Statistic	0.0079	0.1493

4.1.2. Correlated data

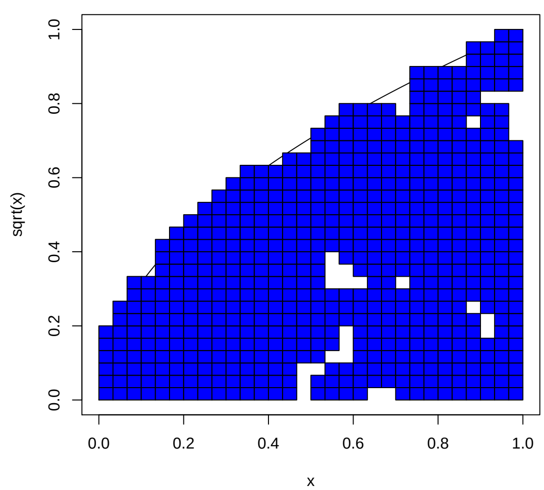
The following four plots shows the boundary estimate given by our method if the data are exponentially correlated with the autocorrelation function given in D_2. Similar with the independent case, we can see that the boundary estimate is close to the true boundary when $n = 30$. We also compare the boundary estimate given by our method and spatial scan statistic for both the circular case and the hyperbolic case. Obviously, the spatial scan statistic performs better than our method when the boundary is a circular shape yet our method precedes the spatial scan statistic when the boundary is a hyperbolic shape.



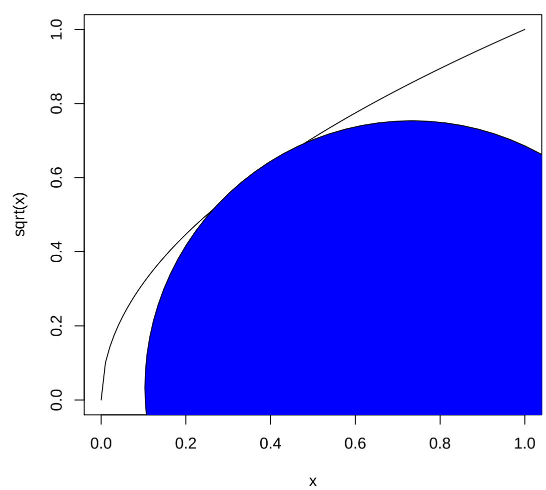
(a) $C_2 \& D_2$, Our Method



(b) $C_2 \& D_2$, Spatial Scan Statistic



(c) $C_1 \& D_2$, Our Method



(d) $C_1 \& D_2$, Spatial Scan Statistic

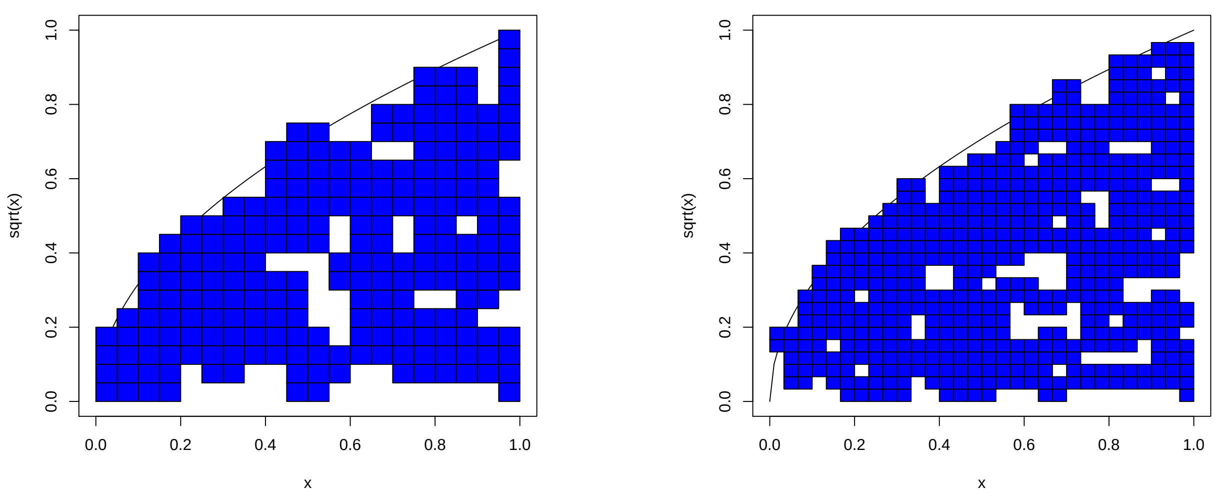
Figure 4.2. Comparison between Our Method and Spatial Scan Statistic; $n = 30$

Similar to the independent case, we provide the symmetric difference of the true boundary and the boundary estimate given by our method and spatial scan statistic for the correlated case in the following table, which can also be used to compare the performance of our method and spatial scan statistic.

Table 4.2. Symmetric Difference between the True Boundary and the Boundary Estimate: Correlated Data

	Circular	Hyperbolic
Our Method	0.0600	0.0267
Spatial Scan Statistic	0.0079	0.1378

The following two figures show how our method works when $y = \sqrt{x}$ is the true change curve with the points from the upper part form a standard normal distribution and these of the lower part form a normal distribution with the same standard error and 100 as their mean respectively with the unit square equally spaced into a 20×20 and 30×30 grid. It is clear that the larger the n is, the greater the method works, which might suggest that our estimate is consistent. Besides, the symmetric difference between the regions constructed by the true boundary and our boundary estimate for $n = 20$ is about 0.135 while that for $n = 30$ is about 0.111, which also suggest that our boundary estimate is consistent when n goes to infinity.

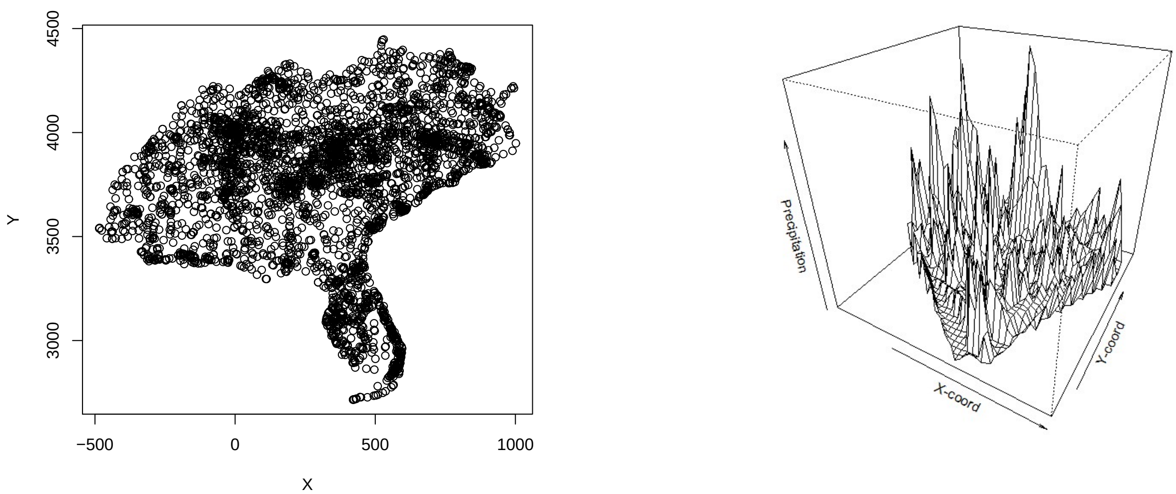


(a) n=20 (b) n=30

Figure 4.3. Consistency of Our Method

4.2. Live data

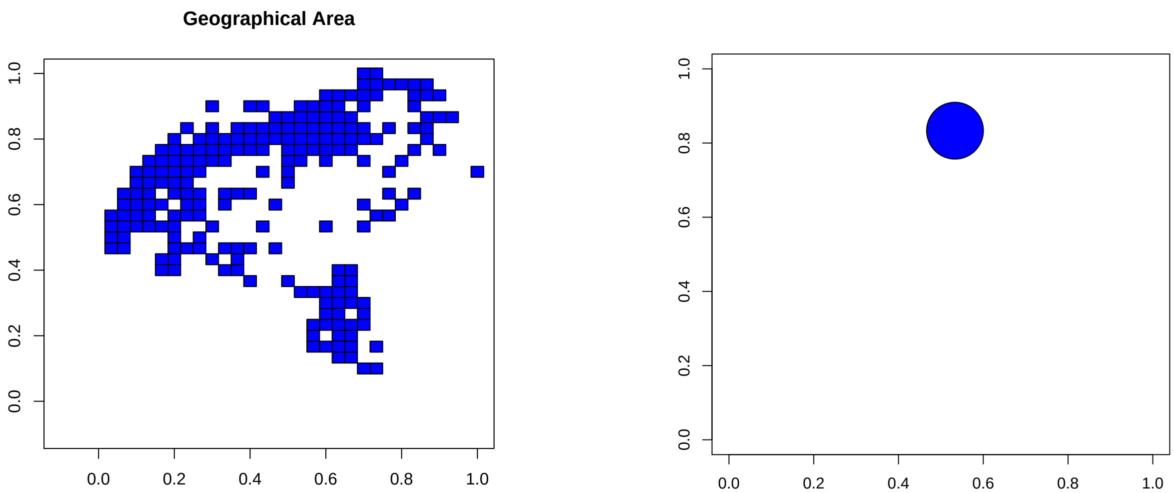
This data set we pick includes the daily precipitation on July 15th, 2014 from 10 southeastern states based on the definition of the Association of American Geographers: Alabama, Florida, Georgia, Kentucky, Mississippi, North Carolina, South Carolina, Tennessee, Virginia, and West Virginia. It also has the information of the min and max of daily temperature which is useless here. The useful variables we will use are longitude, latitude, elevation and precipitation (PRCP). We also translated the longitude and latitude into X and Y coordinates which is a usual practice when we deal with a relatively small geographic region. Originally, this data set has about 4000 observations, each of which is a specific weather station. We removed those with missing values on the variables of interest and we end up with 3840 observations. First of all, we created two plots as following, one shows the shape of the area and the other one shows the distribution of precipitation over this region.



(a) Shape of the Area (b) Distribution of the Precipitation over the Area

Figure 4.4. Live Data

To be compliant with the setting as discussed in §1, we first rescale the region into a unit square $(0, 1]^2$ and then equally space it into a 30×30 grid. Within each small cube, we define the observation to be the mean of the precipitation of those points that fall in the small cube. and the value of the observation is simply set to 0 for those small cubes which do not have any point. Then we conduct our boundary estimate analysis using both our method and spatial scan statistics. The first plot of the following two plots show the boundary estimate obtained by our method while the second one show the boundary estimate obtained by spatial scan statistic.



(a) Our Method

(b) Spatial Scan Statistic

Figure 4.5. Application of Our Method and Spatial Scan Statistic to Live Data

5. CONCLUDING REMARKS

In conclusion, we develop an algorithm in this work to detect the change-curve of spatially referenced data, by profiling the spatially referenced data and using the criterion function as proposed by Müller&Song [8]. Our estimator of the change-curve is simply the maxima of the criterion function. Our algorithm is robust because it does not assume any pre-specified geometric shapes for the change-curve and is computationally efficient because it uses monotonicity in search of the boundary. In the lattice design with mesh size $n^{-1} \times n^{-1}$, the computation complexity by our novel algorithm for change-curve detection is in the order of $O(n^2)$, much smaller than $2^{O(N^2)}$ required by the CUSP algorithm proposed in Müller&Song [8] and Carlstein's [2] estimators. Our novel algorithm yields a consistent estimate of the change-curve as well as the underlying distribution mean of observations in the regions. We also study the hypothesis test of the existence of the change-curve in the presence of independence of the spatially referenced data. We apply our algorithm and spatial scan statistic to simulated data for both the independent case and the correlated case with the true boundary is circular shaped and hyperbolic shaped. We conclude that although our method does not perform as well as spatial scan statistic works for the circular shaped boundaries, it seems outperform the scan statistic method when the boundary is hyperbolic shaped.

We will apply our algorithm to some other shaped boundaries and compare the results with some other methods in the future. Besides, though our work is on a simplified dichotomous lattice design like many other interesting studies, it can be extended to a more complicated scenario.

REFERENCES

[1] K. Alexander and R. Pyke. A uniform central limit theoren for set-indexed partial-sum processes with finite variance. *The Annals of Probability*, 14(2):582–597, 1986.

[2] E. Carlstein and C. Krishnamoorthy. Boundary estimation. *Journal of the American Statistical Association*, 87:430–438, 1992.

[3] C. Goldie and P. Greenwood. Characterisations of set-indexed brownian motion and assiciated conditions for finite dimensional convergence. *The Annals of Probability*, 11(3):802–816, 1986.

[4] C. Goldie and P. Greenwood. Variance of set-indexed sums of mixing random variables and weak convergence of set-indexed processes. *The Annals of Probability*, 14(3):817–839, 1986.

[5] G. M. Jacquez, A. Kaufmann, and P. Goovaerts. Boundaries, links and clusters: a new paradigm in spatial analysis? *Environmental and Ecological Statistics*, 15(4):403–419, 2008.

[6] M. Kulldorff, L. Huang, L. Pickle, and L. Duczmal. An elliptic spatial scan statistic. *Stat Med*, 25(22):3929–3943, 2006.

[7] P.A.P. Moran. Notes on continuous stochastic phenomena. *Biometrika*, 37(1-2):17–23, 1950.

[8] H. Muller and K. Cube Song. Splitting in multidimensional edge estimation. *Change-point Problems, IMS Lecture Notes - Monograph Series*, 23:210–223, 1994.

[9] J.I. Naus. Approximations for distributions of scan statistics. *Journal of the American Statistical Association*, 77(377):177–183, 1982.

[10] P. Qiu. Nonparametric estimation of jump surface. *Sankhya (Series A)*, 59(2):268–294, 1997.

[11] W. Riggan, J. Creason, W. Nelson, M. Manton, Woodbury K., E. Stallard, J. Pellom, and A. Beaubier. *U.S. Cancer mortality rates and trends, 1950-1979(vol.IV:maps).* U.S. Government Printing Office, Washington, D.C., 1987.

[12] B.D. Ripley. The second-order analysis of stationary point processes. *Journal of Applied Probability*, 13:255–266, 1976.

[13] E. Slutsky. *Über stochastische asymptoten und grenzwerte.* *Metron (in German)*, 5(3):3–89, 1925.

[14] T. Tango and K. Takahashi. A flexibly shaped spatial scan statistic for detiecting clusters. *International Journal of Health Geographics*, 4:4–11, 2005.

[15] T. Tango, K. Takahashi, and K. Kohriyama. A space-time scan statistic for detecting emerging outbreaks. *Biometrics*, 67(1):106–115, 2011.

[16] V. Vaart. *Asymptotic statistics*. Cambridge University Press, New York, USA, 1998.

[17] V. Vaart and J. Wellner. *Weak convergence and empirical processes*. Springer, New York, USA, 1996.

[18] W. Womb. Differential systematics. *Science*, 114:315–322, 1951.