Statistica Sinica **32** (2022), 1961-1981
doi:https://doi.org/10.5705/ss.202020.0342

INFERENCE FOR STRUCTURAL BREAKS IN SPATIAL MODELS

Ngai Hang Chan[1], Rongmao Zhang[2,3] and Chun Yip Yau[1]

[1] *The Chinese University of Hong Kong,*
[2] *Zhejiang University and* [3] *Zhejiang University City College*

Abstract: Testing for structural changes in spatial trends constitutes an important issue in many biomedical and geophysical applications. In this paper, a novel statistic based on a discrepancy measure over small blocks is proposed. This measure can be used not only to construct tests for structural breaks, but also to identify the change-boundaries of the breaks. The asymptotic properties and limit distributions of the proposed tests are also established. To derive the asymptotics, the notion of spatial physical dependence is adopted to account for the spatial dependence structure. A bootstrap procedure is applied to the proposed statistic to handle the asymptotic variance of the limit distribution. The method is illustrated by means of simulations and a data analysis.

Key words and phrases: Change-boundaries, discrepancy measure, inference, nonstationary processes, spatial trends.

1. Introduction

Inference for second-order stationary spatial statistical models with a constant mean and a stationary covariance structure has been a topic of active study; see, for example, the seminal text of Cressie (1993) for a comprehensive introduction. When data are affected by topographical structures, sudden events, abrupt policy changes, and other local issues, the second-order stationarity assumption becomes questionable. A misspecified model can often result in inefficient inference and inaccurate predictions. Thus, it is important to test for stationarity, and when nonstationarity is detected, to identify a change boundary. The main objectives of this study are to deal with these two tasks in a structural break context. Specifically, consider the two-dimensional spatial trend model:

$$Y_{\boldsymbol{i}} = \mu\left(\frac{\boldsymbol{i}}{\boldsymbol{n}}\right) + \varepsilon_{\boldsymbol{i}}, \quad \boldsymbol{i} = (i_1, i_2) \in [\mathbf{1}, \boldsymbol{n}] \cap \mathbb{Z}^2, \tag{1.1}$$

Corresponding author: Ngai Hang Chan, Department of Statistics, The Chinese University of Hong Kong, Shatin, Hong Kong, China. E-mail: nhchan@sta.cuhk.edu.hk.

where $\boldsymbol{n} = (n_1, n_2)$, $\boldsymbol{i}/\boldsymbol{n} = (i_1/n_1, i_2/n_2)$, $[\mathbf{1}, \boldsymbol{n}] = [1, n_1] \times [1, n_2]$, μ is an unknown measurable function, and the noise $\{\varepsilon_{\boldsymbol{i}}\}$ is a mean-zero spatial process with expression $\varepsilon_{\boldsymbol{i}} = g(\eta_{\boldsymbol{i}-\boldsymbol{j}}, \boldsymbol{j} \in \mathbb{Z}^2)$, for some measurable function g and independent and identically distributed (i.i.d.) spatial random variables $\{\eta_{\boldsymbol{i}}\}$. It is important to test whether there are abrupt changes in μ. We consider an abrupt change as the Hölder discontinuity in the function μ. Specifically, a function f is said to be Hölder continuous with an exponent $\alpha > 0$ on $I = [0, 1] \times [0, 1] \subset \mathbb{R}^2$ if there exists a constant C such that, for any $\boldsymbol{s}, \boldsymbol{t} \in I$,

$$\sup_{\boldsymbol{s},\boldsymbol{t}} |f(\boldsymbol{s}) - f(\boldsymbol{t})| \leq C||\boldsymbol{t} - \boldsymbol{s}||^{\alpha},$$

where $|| \cdot ||$ denotes the Euclidean norm. When $\alpha = 1$, f is called a Lipschitz continuous function. With different values of α, we can measure changes in the degrees of smoothness in μ. Therefore, considering Hölder continuity is more general than merely testing for changes in the constant mean levels in μ.

Let $C^m(I)$ be the collection of functions having up to mth-order derivatives on I, $H^{\alpha}(I)$ be the set of Hölder continuous functions with exponent α on I, and $PH^{\alpha}(I)$ be the set of piecewise Hölder continuous functions with exponent α on I; that is, there exist a partition $\{B_1, \ldots, B_p\}$ of I (i.e., $\bigcup_{i=1}^{p} B_i = I$, $B_i \cap B_j = \emptyset$) and p different Hölder continuous functions $\{f_j\}$ such that

$$\mu(\boldsymbol{s}) = \sum_{j=1}^{p} f_j(\boldsymbol{s}) I(\boldsymbol{s} \in B_j)\,, \tag{1.2}$$

and $\mu(\cdot)$ satisfies $\min_{\mathbf{s}\in\partial B_i(\delta)\cap B_i,\, \mathbf{t}\in\partial B_i(\delta)\cap B_i^c} |\mu(\mathbf{s}) - \mu(\mathbf{t})| \geq \vartheta_0 > 0$, where ∂B_i is the boundary of B_i and $\partial B_i(\delta) = \{\mathbf{x} : \min_{\mathbf{s}\in\partial B_i} \|\mathbf{x} - \mathbf{s}\| \leq \delta\}$. For model (1.1), two related issues are considered:

1. We test the hypothesis that the trend $\mu(\cdot)$ contains no structural break. In the context of Hölder continuity, the null hypothesis is $H_0 : \mu(\cdot) \in H^{\alpha}(I)$. The alternative hypothesis is $H_1 : \mu(\cdot) \in PH^{\alpha}(I)$.

2. We detect possible change-boundaries (∂B_j), which partition I into p subregions, such that $\mu(\cdot)$ is Hölder continuous within each subregion, but not Hölder continuous on any two adjacent subregions, that is, $\mu(\cdot) \in PH^{\alpha}(I)$, as defined in (1.2).

Issue 1 tests for the stationarity assumption commonly made in spatial statistics. When $\mu(\boldsymbol{s}) = c_1 I(\boldsymbol{s} \in B) + c_2 I(\boldsymbol{s} \in B^c)$ and the noise $\{\varepsilon_{\boldsymbol{i}}\}$ is independent, issue 2 is equivalent to an edge estimation in image processing or the detection

of disease outbreaks in public health, which have been actively pursued in the literature. See, for example, Song et al. (2011); Muller and Song (1994), and Otto and Schmid (2016) for edge estimation; see Kulldorff (2001); Huang, Kulldorff and Gregorio (2007), and Neill (2012) for detection of disease outbreaks.

Issues 1 and 2 have been studied extensively in a time series context; for examples, see Bick and Rosenblatt (1973); Bai and Perron (1998, 2003); Davis, Lee and Rodriguez-Yam (2006); Wu and Zhao (2007); Harchaoui and Lévy-Leduc (2010); Chen and Hong (2012); Chan, Yau and Zhang (2014); Aue, Rice and Sönmez (2018). Structural break phenomena of spatial data may also be found in many practical problems: the detection of tumors in computed tomography scans, described in Otto and Schmid (2016); the estimation of the change boundary for the support of a multivariate probability density, as given in Hall, Peng and Rau (2001); and the estimation of edges in image processing, studied in Tsybakov (1994). Some procedures for detecting change-boundaries have also been developed. For example, De Martino et al. (2008) and Song et al. (2011) classify functional magnetic resonance imaging (fMRI) spatial patterns by combining feature selection and support-vector machines (SVMs). The SVM training algorithm is applied recursively to eliminate irrelevant voxels (features) and estimate informative spatial patterns; see also Mwangi, Tian and Soares (2014). We adopt a different approach to detect edges based on spatial change-locations, which is similar to change-point detection in time series. For example, Muller and Song (1994) assume that the edges divide the region into two subregions, and estimate the change curve over candidate plateau sets based on a cumulative sum (CUSUM)-type test statistic. Otto and Schmid (2016) assume that the means of the observations are different between the interior and the exterior of a circle of radius δ, and then estimate the circular change boundary based on the maximum likelihood function under the Gaussian assumption. These methods fail when changes occur in more than two regions. Furthermore, no spatial dependence structure is considered.

In this study we pursue a different approach, based on detecting discontinuous points of μ, to depict the change boundary. To elucidate the main idea, we take a two-dimensional random field as an example. In such a field, changes occur along curves between two adjacent regions. To detect such changes, one must identify the boundary curve. Such a curve may be approximated by line segments connecting a number of grid points, the discontinuous vertices (change points) of μ. We can therefore detect a possible spatial change region by identifying the discontinuous vertices, and then connecting them with the segments to construct an approximate change-boundary curve. Even though edge estimation in image

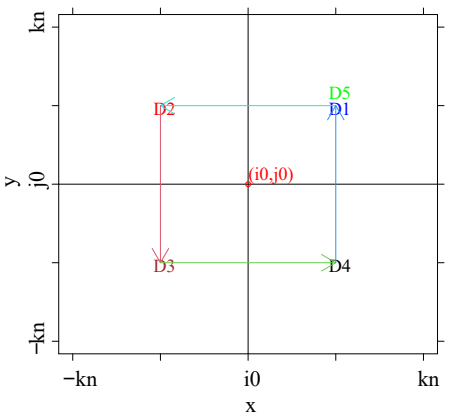


Figure 1. Four blocks D_i, for $i = 1, 2, \ldots, 4$, with side length k_n, used to detect changes at a spatial point $t_0 = (i_0/n_1, j_0/n_2)$. An arrow from D_i to D_{i+1} indicates a difference in the total discrepancy $\mu(\boldsymbol{s})$ between D_i and D_{i+1} in (1.3).

processing (also known as the multivariate change-point problem) is an area of active research, for example, see Carlstein, Müller and Siegmund (1994), it seems that no formal definition has ever been given for a spatial change-point based on the distances between different quadrants. We now give such a definition for the two-dimensional case, which can be extended easily to higher dimensions.

Given a spatial point $\boldsymbol{t}_0 = (t_1^0, t_2^0) \in \mathbb{R}^2$, define $i_0 = [n_1 t_1^0]$, $j_0 = [n_2 t_2^0]$, $D_1 = [i_0, i_0 + k_n] \times [j_0, j_0 + k_n]$, $D_2 = [i_0 - k_n, i_0] \times [j_0, j_0 + k_n]$, $D_3 = [i_0 - k_n, i_0] \times [j_0 - k_n, j_0]$, $D_4 = [i_0, i_0 + k_n] \times [j_0 - k_n, j_0]$, and $D_5 = D_1$, where k_n is the number of lattice points on each side of a block, also known as the block length. This setup is shown in Figure 1. We say a point $\boldsymbol{t}_0$ is a change point of the trend μ if the total discrepancy of $\mu(\mathbf{s})$ on D_i, for $i = 1, 2, 3, 4$, that is,

$$\mathrm{Dist}(\boldsymbol{t}_0) := \sum_{i=1}^{4} \left[\int_{D_i} \frac{\mu(\boldsymbol{s})\, d\boldsymbol{s}}{|D_i|} - \int_{D_{i+1}} \frac{\mu(\boldsymbol{s})\, d\boldsymbol{s}}{|D_{i+1}|} \right]^2, \tag{1.3}$$

is large, where $|D_i|$ denotes the area of the set D_i.

Define

$$S_1([\boldsymbol{n} \cdot \boldsymbol{t}_0]) = \frac{1}{k_n^2} \sum_{i=i_0}^{i_0+k_n} \sum_{j=j_0}^{j_0+k_n} Y_{ij}, \qquad S_2([\boldsymbol{n} \cdot \boldsymbol{t}_0]) = \frac{1}{k_n^2} \sum_{i=i_0-k_n}^{i_0} \sum_{j=j_0}^{j_0+k_n} Y_{ij},$$

$$S_3([\boldsymbol{n} \cdot \boldsymbol{t}_0]) = \frac{1}{k_n^2} \sum_{i=i_0-k_n}^{i_0} \sum_{j=j_0-k_n}^{j_0} Y_{ij}, \quad S_4([\boldsymbol{n} \cdot \boldsymbol{t}_0]) = \frac{1}{k_n^2} \sum_{i=i_0}^{i_0+k_n} \sum_{j=j_0-k_n}^{j_0} Y_{ij},$$

and let $S_5([\boldsymbol{n}\cdot\boldsymbol{t}_0]) = S_1([\boldsymbol{n}\cdot\boldsymbol{t}_0])$, where $[\boldsymbol{n}\cdot\boldsymbol{t}_0] = ([n_1t_1^0],[n_2t_2^0]) = (i_0,j_0)$.

To test whether a point $\boldsymbol{t}_0 \in \mathbb{R}^2$ is a change point of μ, we propose a statistic T_n based on a local discrepancy measure on $\boldsymbol{t}_0$, which is an estimator of (1.3) given by

$$T_n := T([\boldsymbol{n}\cdot\boldsymbol{t}_0]) = \sum_{i=1}^{4}(S_i([\boldsymbol{n}\cdot\boldsymbol{t}_0]) - S_{i+1}([\boldsymbol{n}\cdot\boldsymbol{t}_0]))^2 =: \sum_{i=1}^{4} M_{n,i}^2. \tag{1.4}$$

The reason for using a local instead of a global discrepancy measure is that nearby regions usually contain more similar information than those that are far apart. As a result, a large T_n value indicates that discontinuities exist between the four quadrants. It is shown that the proposed statistic has an asymptotic χ^2-type distribution under the null hypothesis (i.e., no change point), and diverges to infinity under a spatial change point, and therefore provides consistent detection.

The proposed test statistic T_n also enables us to identify change boundaries. This procedure involves splitting the region into smaller blocks, and identifying possible discontinuous points in these blocks. The change-boundary may then be recovered by connecting these vertices. Because any subset of $\mathbb{R}^2$ can be approximated by smaller squares, this idea is similar to the approximation of integrals by Riemann sums.

Compared with the subset-scan detection scheme of Kulldorff (2001); Neill (2012); Huang, Kulldorff and Gregorio (2007), the proposed method is easier to implement with less of a computational burden. Our method scans the region at most n^2 times, while subset scanning detects change areas by scanning all possible subsets, which may require up to 2^n scans.

The remainder of the paper is organized as follows. In Section 2, we establish the asymptotic properties of the proposed tests. Section 3 identifies change boundaries using an illustrative example. Section 4 considers a practical issue related to the limit distribution, namely spatial bootstrapping. Simulations and two real examples are given in Section 5. Section 6 concludes the paper. All proofs are presented in the Supplementary Material.

2. Test for Stationarity

In this section, we consider the test statistic (1.4) for the existence of structural breaks in μ. Let $Y_{ij} = 0$, for $i \le 0$ or $i > n_1$ or $j \le 0$ or $j > n_2$, and $n = n_1n_2$. Define an integrated discrepancy G_n by

$$G_n = \frac{k_n^2}{n} \sum_{i=0}^{[n_1/k_n]} \sum_{j=0}^{[n_2/k_n]} T(ik_n, jk_n)\,. \tag{2.1}$$

Such a statistic has an intuitive meaning. If μ in model (1.1) is a $H^\alpha(I)$ continuous function, then G_n is likely to be small because all of the $\{T(ik_n, jk_n)\}$ are small. Thus, we use G_n to solve issue 1.

To establish the asymptotic distributions of T_n and G_n, we first introduce the dependence structure for the noise. In spatial statistics, the notion of dependence is often modeled using a strong mixing condition (e.g., α-mixing) on the underlying sigma fields; more information about strong mixing can be found in Doukhan (1994); Lin and Lu (1996). However, as Wu (2005) pointed out, it is not easy to determine whether a stationary random field satisfies the mixing condition. In this study, we adopt the notion of the physical dependence measure introduced by Wu (2005). In particular, let $\{\eta'_{\boldsymbol{i}}\}$ be an i.i.d. copy of $\{\eta_{\boldsymbol{i}}\}$, and let $\widetilde{\varepsilon}_{\boldsymbol{i}}$ be the coupled version of $\varepsilon_{\boldsymbol{i}}$ given by

$$\widetilde{\varepsilon}_{\boldsymbol{i}} = g(\widetilde{\eta}_{\boldsymbol{i}-\boldsymbol{j}}, \boldsymbol{j} \in \mathbb{Z}^2),$$

where $\widetilde{\eta}_{\boldsymbol{i}} = \eta_{\boldsymbol{i}}$ if $\boldsymbol{i} \neq 0$, and $\widetilde{\eta}_{\boldsymbol{i}} = \eta'_i$ if $\boldsymbol{i} = 0$. Let $\delta_{\boldsymbol{i},p} = E(|\varepsilon_{\boldsymbol{i}} - \widetilde{\varepsilon}_{\boldsymbol{i}}|^p)^{1/p} =: ||\varepsilon_{\boldsymbol{i}} - \widetilde{\varepsilon}_{\boldsymbol{i}}||_p$ be the physical dependence measure of $\{\varepsilon_{\boldsymbol{i}}\}$, which measures the effect of a single noise at the origin on an observation at $\boldsymbol{i}$. In addition, let $\Delta_p = \sum_{\boldsymbol{i}\in\mathbb{Z}^2} \delta_{\boldsymbol{i},p}$ be the stability measure as given in Machkouri, Volný and Wu (2013), with $\Delta_p(m) = \sum_{\boldsymbol{i}\in\mathbb{Z}^d,\, ||\boldsymbol{i}||>m} \delta_{\boldsymbol{i},p}$. It can be shown that when $\Delta_4 < \infty$, the fourth moment exists and the process is short-range dependent. Two typical examples of the physical dependence are the following:

1. Linear random field: $\varepsilon_{\boldsymbol{i}} = \sum_{\boldsymbol{j}\in\mathbb{Z}^d} \alpha(\boldsymbol{i}-\boldsymbol{j})\eta_{\boldsymbol{j}}$ for certain real numbers $\{\alpha(\boldsymbol{i})\}$ and i.i.d. sequence $\{\eta_{\boldsymbol{t}}\}$; see Lahiri and Robinson (2016).

2. Spatial autoregressive models:

$$\boldsymbol{y}_n = \lambda_0 \boldsymbol{W}_n \boldsymbol{y}_n + \boldsymbol{u}_n,$$

where $\boldsymbol{y}_n = (y_{\boldsymbol{1}}, \ldots, y_{\boldsymbol{n}})^T$, $\boldsymbol{W}_n$ is an $n \times n$ matrix in which the diagonal elements are all zero, and $\boldsymbol{u} = (\eta_{\boldsymbol{1}}, \ldots, \eta_{\boldsymbol{n}})^T$, $\eta_{\boldsymbol{t}}$ are i.i.d. random errors.

Let

$$\Omega = \begin{pmatrix} 2 & -1 & 0 & -1 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ -1 & 0 & -1 & 2 \end{pmatrix}.$$

Then, we have the following asymptotic results for T_n and G_n:

Theorem 1. *Suppose that model* (1.1) *holds with noise satisfying* $\Delta_4 < \infty$, $\mu(\boldsymbol{s}) \in PH^{\alpha}[I]$, *and the block length* k_n *satisfies* $1/k_n + k_n^{1+\alpha}(n_1+n_2)^{\alpha}/n^{\alpha} \to 0$. *Then,*

(i) For T_n, *we have*

$$k_n^2\left[T_n - 2\sum_{i=1}^{4}(\mu_i - \mu_{i+1})(M_{n,i} - EM_{n,i}) - \sum_{i=1}^{4}(\mu_i - \mu_{i+1})^2\right] \xrightarrow{d} \sigma^2 \sum_{i=1}^{4} Z_i^2, \tag{2.2}$$

where $\sigma^2 = \sum_{\boldsymbol{j}\in\mathbb{Z}^2} \mathrm{E}(\varepsilon_{\boldsymbol{0}}\varepsilon_{\boldsymbol{j}})$, $(Z_1, \ldots, Z_4)$ *is a four-dimensional normally distributed random vector with mean vector zero and covariance matrix* Ω, $\mu_i = \lim_{n\to\infty} \mathrm{E}(S_i([\boldsymbol{n}\cdot\boldsymbol{t}_0]))$, *for* $i = 1, 2, 3, 4$, *and* $\mu_1 = \mu_5$.

(ii) For G_n, *we have*

$$(\sqrt{n}k_n)(G_n - \mathrm{E}G_n) \xrightarrow{d} N(0, \sigma_0^2), \ \textit{for some } \sigma_0^2 > 0. \tag{2.3}$$

Furthermore, under $H_0 : \mu \in H^{\alpha}(I)$,

$$(\sqrt{n}k_n)(G_n - 8\sigma^2 k_n^{-2}) \xrightarrow{d} N(0, \sigma_0^2)\,. \tag{2.4}$$

Because k_n is the block size and $2(n_1+n_2)$ and $n = n_1 n_2$ are the perimeter and area, respectively, of the spatial region, the condition $1/k_n + k_n^{(1+\alpha)}(n_1 + n_2)^{\alpha}/n^{\alpha} \to 0$ in Theorem 1 asserts that the block size grows to infinity at a rate slower than the $\alpha/(1+\alpha)$ power of the ratio of the area to the perimeter of a spatial region. To give more insight into Theorem 1, we provide the following remarks:

Remark 1. Theorem 1 indicates that if $\boldsymbol{t}_0$ is not a break point of $\mu(\cdot)$, then $k_n^2 T_n$ converges to a generalized chi-squared distribution, as the other two terms disappear in this case. In contrast, if $\boldsymbol{t}_0$ is a break point, then $k_n^2 T_n = O_p(k_n^2)$ because $k_n \sum_{i=1}^{4}(\mu_i - \mu_{i+1})(M_{n,i} - EM_{n,i}) \xrightarrow{d} \sum_{i=1}^{4}(\mu_i - \mu_{i+1})Z_i$, by Lemma 1 and the continuous mapping theorem (see the proof of Theorem 1), which means the second term on the left-hand side of (2.2) is $O_p(k_n)$. When $\boldsymbol{t}_0$ is a break point, $(\mu_i - \mu_{i+1})^2 > 0$ for at least one of $i = 1, \ldots, 4$, which results in the dominance of the third term in (2.2) with order $O(k_n^2)$.

Remark 2. When μ is Hölder continuous and $k_n^{1+\alpha}(n_1+n_2)^{\alpha}/n^{\alpha} \to 0$, then $\mathrm{E}G_n$ converges to a quantity that does not depend on μ. This is because when μ is Hölder continuous, for any given $\boldsymbol{s} = (s_1, s_2) \in [(i-1)k_n/n_1, (i+1)k_n/n_1] \times$

$[(j-1)k_n/n_2, (j+1)k_n/n_2]$,

$$\left|\mu(\boldsymbol{s}) - \mu\left(\frac{ik_n}{n_1}, \frac{jk_n}{n_2}\right)\right| \le C\left(\left(\frac{k_n}{n_1}\right)^{\alpha} + \left(\frac{k_n}{n_2}\right)^{\alpha}\right), \tag{2.5}$$

which implies that for any $\boldsymbol{t}_0 = (t_1^0, t_2^0)$,

$$\begin{aligned}
&\mathrm{E}\left[\sum_{i=1}^{4}(S_i([\boldsymbol{n}\cdot\boldsymbol{t}_0]) - S_{i+1}([\boldsymbol{n}\cdot\boldsymbol{t}_0]))^2\right] \\
&= \sum_{i=1}^{4}(S_i^{\mu}([\boldsymbol{n}\cdot\boldsymbol{t}_0]) - S_{i+1}^{\mu}([\boldsymbol{n}\cdot\boldsymbol{t}_0])^2 + \sum_{i=1}^{4}\mathrm{E}\Big[\bar{S}_i([\boldsymbol{n}\cdot\boldsymbol{t}_0]) - \bar{S}_{i+1}([\boldsymbol{n}\cdot\boldsymbol{t}_0])\Big]^2 \\
&\le C\left(\left|\frac{k_n}{n_1}\right|^{\alpha} + \left|\frac{k_n}{n_2}\right|^{\alpha}\right)^2 + \sum_{i=1}^{4}\mathrm{E}\Big[\bar{S}_i([\boldsymbol{n}\cdot\boldsymbol{t}_0]) - \bar{S}_{i+1}([\boldsymbol{n}\cdot\boldsymbol{t}_0])\Big]^2 \\
&= \sum_{i=1}^{4}\mathrm{E}\Big[\bar{S}_i([\boldsymbol{n}\cdot\boldsymbol{t}_0]) - \bar{S}_{i+1}([\boldsymbol{n}\cdot\boldsymbol{t}_0])\Big]^2(1+o(1)),
\end{aligned}$$

as $k_n^{1+\alpha}(n_1+n_2)^{\alpha}/n^{\alpha} \to 0$ and $\mathrm{E}[\bar{S}_i([\boldsymbol{n}\cdot\boldsymbol{t}_0]) - \bar{S}_{i+1}([\boldsymbol{n}\cdot\boldsymbol{t}_0])]^2 = O(1/k_n^2)$, where $S_i^{\mu}([\boldsymbol{n}\cdot\boldsymbol{t}_0])$ and $\bar{S}_i([\boldsymbol{n}\cdot\boldsymbol{t}_0])$ are defined analogously to $S_i([\boldsymbol{n}\cdot\boldsymbol{t}_0])$, with Y_{ij} replaced with $\mu_{ij} = \mu(i/n_1, j/n_2)$ and ε_{ij}, respectively.

Remark 3. If $\mu \in H^{\alpha}(I)$ and $k_n^{1+\alpha}(n_1+n_2)^{\alpha}/n^{\alpha} \to 0$, then by (2.5), we have $\mathrm{E}G_n = 8\sigma^2 k_n^{-2}(1+o(1))$. However, when $\mu \in PH^{\alpha}(I)$, $\mathrm{E}G_n \ge Ck_n(1/n_1 + 1/n_2) + 8\sigma^2 k_n^{-2}$. This implies that if $k_n^3(1/n_1 + 1/n_2) \to \infty$, then

$$\begin{aligned}
\sqrt{n}k_n(G_n - 8\sigma^2 k_n^{-2}) &= \sqrt{n}k_n(\mathrm{E}G_n - \mathrm{E}G_n + \mathrm{E}G_n - 8\sigma^2 k_n^{-2}) \\
&= \sqrt{n}k_n(\mathrm{E}G_n - \mathrm{E}G_n) + \sqrt{n}k_n(\mathrm{E}G_n - 8\sigma^2 k_n^{-2}) \\
&= \sqrt{n}k_n(\mathrm{E}G_n - \mathrm{E}G_n) + \sqrt{n}k_n^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)(1+o(1)) \\
&\to \infty.
\end{aligned}$$

Thus, one can test whether changes happen in different regions using the statistics $g_n = \sqrt{n}k_n(G_n - 8\sigma^2 k_n^{-2})$. One might also consider applying the statistics $\widetilde{G}_n = \max_{1\le i\le[n_1/k_n], 1\le j_1\le[n_2/k_n]} T(ik_n, jk_n)$ instead. It would be interesting to establish the asymptotic distribution of $\widetilde{G}_n$ and compare its performance with G_n.

3. Boundary-Change Detection

Boundary-change detection is a very important problem in nonstationary spatial statistics. To pursue a solution, we propose a threshold point evaluation

(TPE) algorithm for boundary detection, which evaluates T_n at each location, and deletes those values smaller than a certain threshold. In particular, we estimate the boundary using a three-step procedure:

1. Calculate T_n at each lattice point $\boldsymbol{s}_i$ with spacing k_n.
2. Select $\boldsymbol{s}_i$, such that $T_n(\boldsymbol{s}_i) > \lambda$ for some prescribed threshold λ, as possible change points $\boldsymbol{s}_1^*, \boldsymbol{s}_2^*, \ldots, \boldsymbol{s}_{K_{\max}}^*$.
3. Connect the adjacent lattice points among $\boldsymbol{s}_1^*, \boldsymbol{s}_2^*, \ldots, \boldsymbol{s}_{K_{\max}}^*$ with line segments to visualize a change boundary.

Remark 4. To implement the testing and TPE algorithm, we consider the following choices of parameters k_n and λ. First, taking $k_n = cn^{1/5} = c(n_1 n_2)^{1/5}$, for $c > 0$, satisfies the theoretical requirements $1/k_n + k_n^{1+\alpha}(n_1 + n_2)^\alpha / n^\alpha \to 0$, for $\alpha = 1$ and $k_n^3(1/n_1 + 1/n_2) \to \infty$. The case $\alpha = 1$ corresponds to Hölder continuity of the spatial trend μ with order equal to one, which is convenient and weak enough for many practical situations. On the other hand, when considering smoother functions, such as piecewise constant functions, a larger k_n of order $(n_1 n_2)^{1/4}$ can be employed. Our simulation experiments suggest that $k_n = 2(n_1 n_2)^{1/5}$ gives good performance in most cases.

For the selection of λ, when μ is Hölder continuous and satisfies the conditions of Theorem 1, then for $x > 0$,

$$\begin{aligned} P\Big\{ \max_i T_n(\boldsymbol{s}_i) > x \Big\} &\le \sum_i P\{k_n^2 T_n(\boldsymbol{s}_i) > k_n^2 x\} \\ &\le n k_n^{-2} \sum_{i=1}^{4} P\left\{ Z_i^2 > \frac{k_n^2 x}{4\sigma^2} \right\} (1 + o(1)) \\ &= 4 n k_n^{-2} P\left\{ |Z_1| > \frac{k_n \sqrt{x}}{2\sigma} \right\} (1 + o(1)) \\ &\le 16 \sigma n k_n^{-3} x^{-1} \exp\left\{ \frac{-x k_n^2}{8\sigma^2} \right\} (1 + o(1)), \end{aligned}$$

which implies that if $\lambda = 16\sigma^2 \log n / k_n^2$, then $P\{\max_i T_n(\boldsymbol{s}_i) > \lambda\} \to 0$. However, if $\boldsymbol{s}_0$ is a change point, then $T_n(\boldsymbol{s}_0)$ has the same order as k_n^2. As a result, we can choose a threshold such as $\lambda = 16\sigma^2 \log n / k_n^2$, $\log n$, or $O(\log n / k_n)$ in the TPE algorithm. Based on our simulation experience, we suggest $\lambda = 16 \log n / k_n$ as a rule-of-thumb choice.

We now give an example to illustrate the above algorithm. Consider model (1.1) with i.i.d. standard normal white noise and the trend $\mu(\boldsymbol{s}) = 0$ ($\boldsymbol{s} = (s_1, s_2)$)

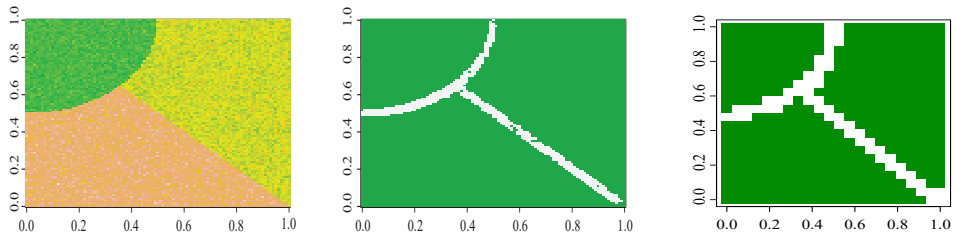


Figure 2. The left panel plots the true boundaries for three different regions, where the green area corresponds to $\mu(\boldsymbol{s}) = 0$, the yellow to $\mu(\boldsymbol{s}) = 5$, and the gray to $\mu(\boldsymbol{s}) = 10$. The estimated boundaries and regions with $k_n = n^{1/4}/4$ and $k_n = 2n^{1/5}$ are plotted in the middle and right panels, respectively.

for $s_1^2+(s_2-1)^2 < 1/4$, $\mu(\boldsymbol{s}) = 5$ for $\boldsymbol{s}$ for $s_1^2+(s_2-1)^2 \geq 1/4$, and $s_1+s_2 > 1$ and $\mu(\boldsymbol{s}) = 10$ otherwise. The corresponding areas are plotted in green, yellow, and gray in the left panel of Figure 2. In the simulation, we set $k_n = n^{1/4}/4$ and $2n^{1/5}$, where $n_1 = n_2 = 100$ and $n = n_1 n_2$, and choose a cutoff with $\lambda = 16 \log n/k_n$. All blocks containing change points (i.e., $T_n(\boldsymbol{s}_i) > \lambda$) are used to estimate the change boundaries. The estimated change blocks are depicted in white; see the right panel of Figure 2, which estimates the true boundaries very well.

4. Bootstrapping

From equations (2.3) and (2.4), the proposed test statistic G_n has a normal limit distribution, which involves two nuisance parameters σ^2 and σ_0^2. Although σ^2 and σ_0^2 can be estimated using the asymptotic variance estimators of the observations $y(\mathbf{s}_i)$ and the quantities G_n respectively (e.g., Sherman (2010)), the estimators can sometimes be unstable. As an alternative, we use a spatial block bootstrap procedure, called the grid-based block bootstrap (GBBB), introduced by Lahiri and Zhu (2006), to directly approximate the asymptotic distribution without estimating σ^2 and σ_0^2. Specifically, let b_n be the block size, and let

$$\mathcal{J}_n = \{\boldsymbol{j} \in \mathbb{Z}^2 : \{\boldsymbol{j}b_n + [0,1)^2 b_n\} \cap [\mathbf{1}, \boldsymbol{n}] \neq \emptyset\}$$

be the minimal set of indices $\boldsymbol{j} \in \mathbb{Z}^2$, such that the disjoint hyper-cubes $\{\boldsymbol{j}b_n + [0,1)^2 b_n, \boldsymbol{j} \in \mathcal{J}_n\}$ with side lengths b_n constitute a covering of $[\mathbf{1}, \boldsymbol{n}] = [1, n_1] \times [1, n_2]$. Let

$$\mathcal{I}_n = \{\boldsymbol{i} \in \mathbb{Z}^2 : \{\boldsymbol{i} + [0,1)^2 b_n\} \subseteq [\mathbf{1}, \boldsymbol{n}]\}$$

be the index set of all overlapping hyper-cubes contained in $[\mathbf{1}, \boldsymbol{n}]$. Consider a

$|\mathcal{J}_n|$ out of $|\mathcal{I}_n|$ version of the GBBB algorithm, as follows:

1. Let $\boldsymbol{I_j}$, $\boldsymbol{j} \in \mathcal{J}_n$ be i.i.d. with a uniform distribution on the set $\mathcal{I}_n$:

$$P^*(\boldsymbol{I_j} = \boldsymbol{i}) = \frac{1}{|\mathcal{I}_n|}, \boldsymbol{i} \in \mathcal{I}_n.$$

Let $A_n(j) = \boldsymbol{j}b_n + [0,1)^2 b_n$ and $B_n(\boldsymbol{I_j}) = \boldsymbol{I_j} + [0,1)^2 b_n$, $\boldsymbol{j} \in \mathcal{J}_n$. Define $\{Y^*_{\boldsymbol{t}}\}$, the block bootstrap series of $\{Y_{\boldsymbol{t}}\}$, as follows:

$$\{Y^*_{\boldsymbol{t}},\ \boldsymbol{t} \in A_n(\boldsymbol{j})\} := Y^*(A_n(\boldsymbol{j})) = Y(B_n(\boldsymbol{I_j})) =: \{Y_{\boldsymbol{t}}, \boldsymbol{t} \in B_n(\boldsymbol{I_j})\}, \boldsymbol{j} \in \mathcal{J}_n.$$

2. We calculate the bootstrap version of G_n by

$$G^*_n = \frac{1}{|\mathcal{J}_n|} \sum_{j \in \mathcal{J}_n} T^*_n(A_n(\boldsymbol{j})), \tag{4.1}$$

where

$$T^*_n(A_n(\boldsymbol{j})) = \frac{k_n^2}{b_n^2} \sum_{\boldsymbol{i}:[(\boldsymbol{i}-\boldsymbol{1})k_n,(\boldsymbol{i}+\boldsymbol{1})k_n] \subseteq A_n(\boldsymbol{j})} T^*_n(\boldsymbol{i}k_n)\,,$$

and $T^*_n(\boldsymbol{i}k_n)$ is defined analogously to $T_n(\boldsymbol{i}k_n)$ in Section 2, replacing $Y_{\boldsymbol{t}}$ with $Y^*_{\boldsymbol{t}}$.

Theorem 2. *Suppose that the conditions of Theorem* 1 *hold,* $k_n = o(b_n)$*, and* $b_n = o(\min\{n_1, n_2\})$*; then, for any* $x \in \mathbb{R}$*,*

$$P^*\{\sqrt{|\mathcal{J}_n|}b_n k_n[G^*_n - \mathrm{E}^*(G^*_n)] \le x\} = P\{\sqrt{n}k_n(G_n - \mathrm{E}G_n) \le x\} + o_p(1),$$

where P^* *and* E^* *denote the probability and expectation value, respectively, of the bootstrapped sample.*

From the simulation experiments in Section 5, it can be seen that various choices of block sizes that satisfy $b_n = o(\min\{n_1, n_2\})$ give excellent performance. In practice, a convenient choice is $b_n = \min\{n_1, n_2\}^{2/3}$ or $n^{1/3}$ when $n_1 = n_2$.

5. Simulations and Applications

5.1. Simulations

We illustrate the proposed method using simulated examples and two data sets. Let $\varepsilon_{\mathbf{s}}$ be a noise process generated by

$$\varepsilon_{ij} = a_1\varepsilon_{i-1,j} + a_2\varepsilon_{i,j-1} + a_3\varepsilon_{i-1,j-1} + \eta_{ij}, \tag{5.1}$$

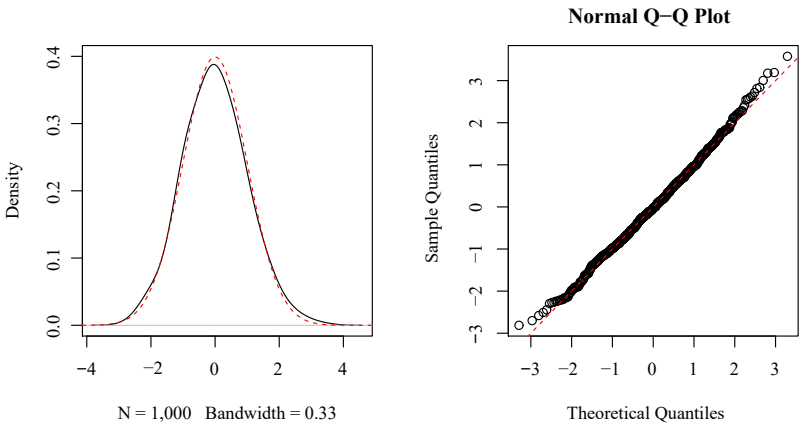


Figure 3. The density and the QQ plot of $(\sqrt{n}k_n)(G_n - EG_n)$ for $\mu(\boldsymbol{s}) = 0$, with noise drawn from (5.1) and $(a_1, a_2, a_3) = (0.3, 0.2, 0.3)$, given in the left and right panels, respectively. The dotted red curves are the density (left panel) and the QQ plot (right panel)of a standard normal distribution.

where $\{\eta_{ij}\}$ is a sequence of i.i.d. standard normal random variables.

First, we simulate the distribution of $g_n = (\sqrt{n}k_n)(G_n - \mathrm{E}(G_n))$. The data are generated from model (1.1) with $\mu(\boldsymbol{s}) = 0$, and the noise is generated from (5.1) with $(a_1, a_2, a_3) = (0.3, 0.2, 0.3)$ and $(n_1, n_2) = (250, 250)$. We set $k_n = 2(n_1 n_2)^{1/5}$ and draw 1,000 samples. The corresponding distribution of $g_n/sd(g_n)$ is plotted in Figure 3, where the left panel gives the density and the right panel gives the QQ plot, and $sd(g_n)$ denotes the standard deviation of g_n. Here, G_n approximates a normal distribution well, as shown in Figure 3.

Second, we consider the power and size of G_n in testing $H_0 : \mu(\boldsymbol{s}) = 0$ vs. $H_1 : \mu(\boldsymbol{s}) = \sigma \cdot I(1/3 \le s, t \le 2/3)$ for different values of σ and (n_1, n_2). We set $k_n = 2(n_1 n_2)^{1/5}$ and use the noise given in (5.1) with $(a_1, a_2, a_3) = (0.3, 0.2, 0.3)$ in all of the simulations. The power of G_n over 500 replications, with significance level $\alpha = 0.05$, is reported in Table 1. In addition, the size of G_n over 500 replications, with $(n_1, n_2) = (100, 100), (200, 200)$, and $(250, 250)$ and with significance levels $\alpha = 0.05$ and 0.10, is reported in Table 2. Table 1 shows that G_n works reasonably well in detecting spatial change points, especially when $n = n_1 n_2$ and σ are large. Furthermore, the power increases as σ and n increase. This makes sense, because when σ increases, changes become bigger and easier to detect, and as n increases, Theorem 1 shows that the asymptotic power tends to one. Similarly, Table 2 shows that the sizes of G_n approximate well to the significance level. This result confirms the conclusion of Theorem 1.

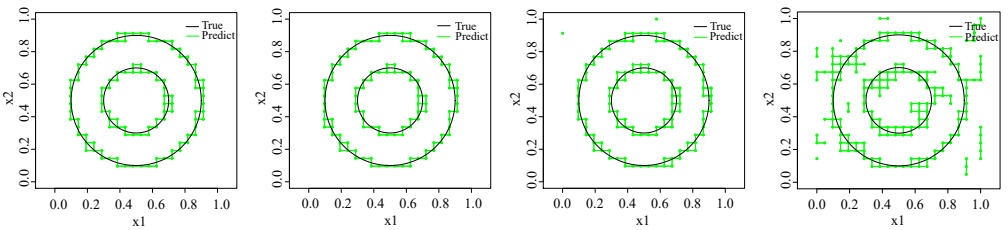
Third, we consider the performance of the TPE algorithm in detecting the

Table 1. The power of G_n in testing $H_0 : \mu(\boldsymbol{s}) = 0$ vs. $H_1 : \mu(\boldsymbol{s}) = \sigma \cdot I(1/3 \le s_1, s_2 \le 2/3)$.

σ	(n_1, n_2)=(50, 50)	(100, 100)	(150, 150)	(200, 200)	(250, 250)	(500, 500)
0.5	0.132	0.132	0.161	0.189	0.178	0.458
1	0.41	0.441	0.703	0.881	0.872	1.000
2	0.985	1.000	1.000	1.000	1.000	1.000

Table 2. The size of G_n in testing $H_0 : \mu(\boldsymbol{s}) = 0$ vs. $H_1 : \mu(s) \neq 0$.

α=0.05			α=0.10		
$(n_1, n_2) = (100, 100)$	(200, 200)	(250, 250)	(100, 100)	(200, 200)	(250, 250)
0.045	0.047	0.038	0.090	0.097	0.095

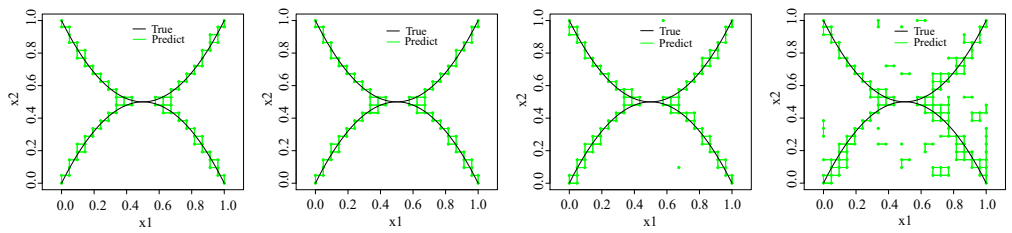


(a) (b) (c) (d)

Figure 4. The true circles with center $(0.5, 0.5)$ and radii 0.2 and 0.4 and their estimates from the TPE algorithm with threshold $\lambda = 16 \log n/k_n$, where the data in the three subregions from inside to outside are generated from model (1.1) with mean functions $\mu(\boldsymbol{s}) =$ 0, 5, and 10, respectively, and the noise is generated from (5.1). The sub-figures (a)–(d) are plotted for noise with $(a_1, a_2, a_3) = (0, 0, 0)$, $(0.3, 0.2, 0.3)$, $(0.6, 0.8, -0.48)$, and $(0.6, 0.9, -0.54)$ respectively.

change boundary, with noise given by (5.1) and $(a_1, a_2, a_3) = (0, 0, 0), (0.3, 0.2, 0.3), (0.6, 0.8, -0.48)$, and $(0.6, 0.9, -0.54)$. The first example consists of two concentric circles, both with center $(0.5, 0.5)$ and with respective radii 0.2 and 0.4. These circles split the region into three subregions. The data within the circle with radius 0.2 have a zero mean function, those between the two circles have a mean function $\mu(\boldsymbol{s}) = 5$, and those outside the circle with radius 0.4 have a mean function $\mu(\boldsymbol{s}) = 10$. In this example, we take $n_1 = n_2 = 500$ and $k_n = 2(n_1 n_2)^{1/5}$, and compute T_n on all of the lattices $(ik_n, jk_n), i, j = 1, 2, \ldots, [500/k_n]$. The threshold λ in the TPE algorithm is taken as $\lambda = 16 \log n/k_n$, $n = n_1 n_2 = 500^2$, and all lattices with $T_n > \lambda$ are taken as possible change points. The true circles and their change-point estimates (dots) based on the TPE are plotted in Figure 4.

The change boundary in the second example consists of two parabolas, the first given by $y = 2(x - 0.5)^2 + 0.5$, and the second by $y = -2(x - 0.5)^2 + 0.5$,



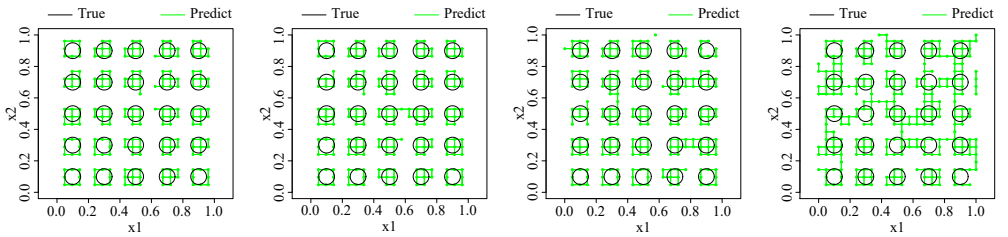
(a) (b) (c) (d)

Figure 5. The true parabolas $y = 2(x-0.5)^2+0.5$ and $y = -2(x-0.5)^2+0.5$ and their estimates based on the TPE algorithm with threshold $\lambda = 16\log n/k_n$, where sub-figures (a)–(d) are plotted for noise with $(a_1, a_2, a_3) = (0, 0, 0)$, $(0.3, 0.2, 0.3)$, $(0.6, 0.8, -0.48)$, and $(0.6, 0.9, -0.54)$, respectively.

which partition the region into four subregions. The mean function of the data in the upper region with lower boundary $y = 2(x-0.5)^2+0.5$ and the data in the lower region with upper boundary $y = -2(x-0.5)^2+0.5$ is taken as $\mu(\boldsymbol{s}) = 5$, while the mean function of the data in the other two subregions is taken as $\mu(\boldsymbol{s}) = 0$; the noise is generated just as in the first example. We choose sample size $n_1 = n_2 = 500$, block length $k_n = 2(n_1 n_2)^{1/5}$, and threshold $\lambda = 16\log n/k_n$. The true parabolas and their TPE estimates are plotted in Figure 5.

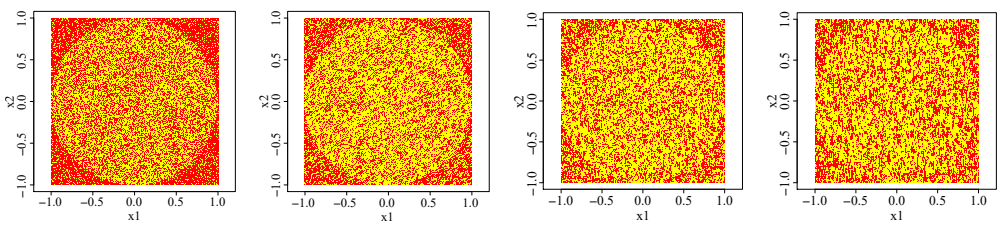
We also consider estimating the change boundary under a large number of dispersed change regions. Twenty-five circular regions centered at (c_x, c_y), for $c_x, c_y \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$, each with radius 0.05, are assigned with mean function $\mu(\boldsymbol{s}) = 5$, while the mean function of the data outside the circular regions is taken as $\mu(\boldsymbol{s}) = 0$. The noise is generated as in the first example. We choose sample size $n_1 = n_2 = 500$, block length $k_n = 2(n_1 n_2)^{1/5}$, and threshold $\lambda = 16\log n/k_n$. The true parabolas and their TPE estimates are plotted in Figure 6. Figures 4 to 6 show that the proposed procedure works reasonably well, but that noise dependence affects the performance; that is, the weaker the spatial autocorrelation between the noise, the better is the performance.

Finally, we consider the power of the bootstrap procedure. To this end, data inside and outside the circle with radius 1 are generated from model (1.1) with mean $\mu(\boldsymbol{s}) = 1$ and $\mu(\boldsymbol{s}) = 0$, respectively, and noise (5.1) with $(a_1, a_2, a_3) = (0, 0, 0)$, $(0.3, 0.2, 0.3)$, $(0.6, 0.7, -0.42)$, and $(0.6, 0.8, -0.48)$, as shown in Figure 7. We set $n_1 = n_2 = 500 = \sqrt{n}$ and repeat each simulation 1,000 times. Based on these 1,000 samples, the power of G_n^* in detecting structural breaks for different (k_n, b_n) with significance level $\alpha = 0.05$ is reported in Table 3. From Table 3, the bootstrap procedure is demonstrated to work well, except in the case $b_n = n^{1/2}$, where $|\mathcal{J}_n| = 1$ and the condition $b_n = o(\min\{n_1, n_2\})$ in Theorem 2



(a) (b) (c) (d)

Figure 6. The true circular regions $(x - c_x)^2 + (y - c_y)^2 = 0.05^2$, where $c_x, c_y \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$, and their estimates based on the TPE algorithm with threshold $\lambda = 16 \log n / k_n$, where sub-figures (a)–(d) are plotted for noise with $(a_1, a_2, a_3) = (0, 0, 0)$, $(0.3, 0.2, 0.3)$, $(0.6, 0.8, -0.48)$, and $(0.6, 0.9, -0.54)$, respectively.



(a) (b) (c) (d)

Figure 7. Data with mean $\mu(\boldsymbol{s}) = 1$ inside the circle with radius 1, $\mu(\boldsymbol{s}) = 0$ outside the circle, and noise (5.1) with $(a_1, a_2, a_3) = (0, 0, 0)$, $(0.3, 0.2, 0.3)$, $(0.6, 0.7, -0.42)$, and $(0.6, 0.8, -0.48)$ are plotted in (a)–(d) respectively.

Table 3. The power of G_n^* under H_1 for circular-boundary data.

(a_1, a_2, a_3)	(k_n, b_n)						
	$(n^{1/3}, n^{1/2})$	$(n^{1/4}, n^{1/2})$	$(n^{1/3}, n^{9/20})$	$(n^{1/4}, n^{9/20})$	$(n^{1/3}, n^{2/5})$	$(n^{1/4}, n^{2/5})$	$(n^{1/4}, n^{1/3})$
(0, 0, 0)	0.527	0.822	0.992	1.000	1.000	1.000	1.000
(0.3,0.2,0.3)	0.508	0.754	0.999	1.000	1.000	1.000	1.000
(0.6,0.7,-0.42)	0.455	0.969	0.939	1.000	1.000	1.000	1.000
(0.6,0.8,-0.48)	0.322	0.995	1.000	0.998	1.000	1.000	1.000

does not hold.

5.2. Data analysis

In this subsection, we apply G_n to test the stationarity of two data sets. The first is the precipitation data set from the National Climatic Data Center for the years 1895 to 1997, available at `www.image.ucar.edu/GSP/Data/US.monthly.met`. This data set has been analyzed extensively, including by

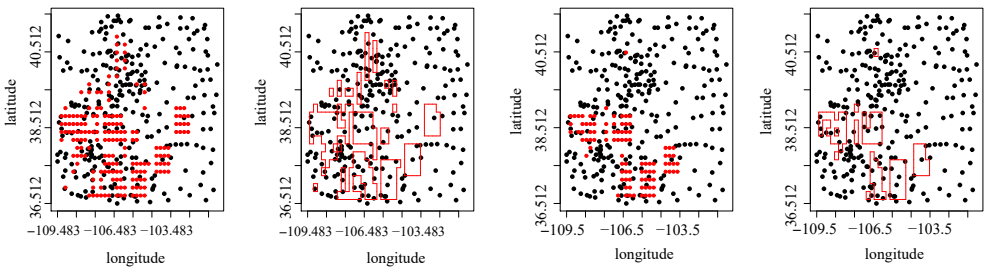
Table 4. The values of G_n and $\widetilde{G}_n$ for the monthly total precipitation in Colorado in December of 1986, with length $k_n = [(n_1 n_2)^{1/\delta}]$.

n_1	δ	4	4.5	5	5.5	6	6.5	7	7.5	8
200	k_n	14	11	8	7	6	5	5	4	4
	G_n	478.94	547.45	526.98	483.58	433.47	341.85	341.85	242.63	242.63
	$\widetilde{G}_n$	1916.06	1452.74	923.36	691.26	522.78	372.60	372.60	212.49	212.49
500	k_n	22	16	12	10	8	7	6	5	5
	G_n	537.93	465.53	304.41	242.79	165.98	133.70	98.79	70.83	70.83
	$\widetilde{G}_n$	6282.35	3844.14	2259.48	1699.52	1037.24	748.68	478.72	214.89	214.89

Johns et al. (2003); Furrer, Genton and Nychka (2006); Kaufman, Schervish and Nychk (2008). Assuming the data to be stationary, Liang et al. (2013) used this data set to analyze anomalies in monthly total precipitation, which are defined as monthly totals, standardized by the long-term mean and standard deviation for each station. It is informative to test for the stationarity assumption before pursuing further analysis. Consider the monthly total precipitation in Colorado in December of 1986. One reason for examining this month is that it includes a large data set from 283 stations. In our analysis, we split the range of longitudes ($l_1 = 8.463°$) and latitudes ($l_2 = 4.955°$) of the 283 stations into rectangles of length and width $(l_1/500, l_2/500)$ (i.e., $n_1 = n_2 = 500$) and $(l_1/200, l_2/200)$ (i.e., $n_1 = n_2 = 200$), and take the block length k_n to be $[(n_1 n_2)^{1/\delta}]$ with different δ in computing T_n. The corresponding values of G_n and $\widetilde{G}_n = (\sqrt{n} k_n \widehat{\sigma}_0)(G_n - 8\widehat{\sigma}^2 k_n^{-2})$ are reported in Table 4, where $\widehat{\sigma}_0^2$ and $\widehat{\sigma}^2$ are the asymptotic variance estimators for σ_0^2 and σ^2, respectively. By (2.4) and the stationarity assumption, $\widetilde{G}_n$ converges to a standard normal distribution, which means that all of the values of $\widetilde{G}_n$ in Table 4 yield very small p-values; this casts doubt on the stationarity assumption.

We also use $\widetilde{G}_n$ to test the December data for other years, and find that all of the p-values are very small, which indicates again that the stationarity assumption is questionable. The nonstationarity of this data set is confirmed by the bootstrap procedure with a significance level of 0.05. The locations of the largest values over $\lambda = 8 \log n/k_n$ and $16 \log n/k_n$ of T_n, and the change boundary based on these change points for $n_1 = n_2 = 200$ and $k_n = [2(n_1 n_2)^{1/5}]$, are plotted in (a) and (b) of Figure 8, respectively.

Although different λ may lead to different quantities of change points and different change boundaries, Figure 8 shows that changes in monthly total precipitation occur in the west of Colorado, especially the southwestern region, but not in the east. Because plains cover most of eastern Colorado, precipitation varies little. However, most of western Colorado is made up of mountains, foothills,



(a) $\lambda = 8\sigma^2 \log n/k_n$ (b) $\lambda = 16 \log n/k_n$

Figure 8. The locations of those points with T_n over λ, and the corresponding change boundaries based on the detected change points for monthly total precipitation in Colorado, USA, are marked in red for (a) $\lambda = 8\sigma^2 \log n/k_n$ and (b) $\lambda = 16 \log n/k_n$.

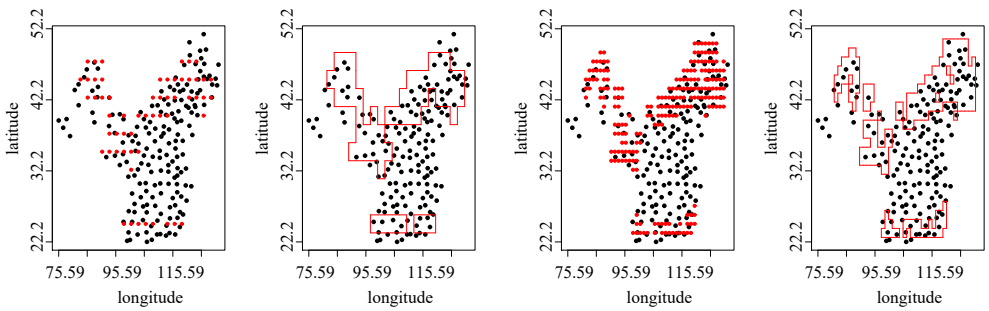
Table 5. The values of G_n and $\widetilde{G}_n$ with length $k_n = [(n_1 n_2)^{1/\delta}]$ for the average monthly temperature in China, in January.

n_1	δ	4	4.5	5	5.5	6	6.5	7	7.5	8
	k_n	10	8	6	5	5	4	4	3	3
100	G_n	143.98	117.22	125.05	143.34	143.34	146.55	146.55	112.30	112.30
	$\widetilde{G}_n$	499.05	345.63	227.63	181.96	181.96	125.24	125.24	17.57	17.57
	k_n	14	11	8	7	6	5	5	4	4
200	G_n	121.61	137.77	146.55	143.70	112.30	92.34	92.34	66.12	66.12
	$\widetilde{G}_n$	1259.64	1031.56	724.03	557.01	405.98	238.08	238.08	47.35	47.35

high plains, and desert lands. Our result confirms the finding of Doesken, Pielke, Sr. and Bliss (2003) that mountains and surrounding valleys greatly affect local climate.

The second data set consists of monthly temperatures (in degrees Celsius) recorded at 176 monitoring stations in China from January 1970 to December 2000. All series have length 372. In our analysis, we consider the average monthly temperature for January over the entire 31 years. In this data set, the range of longitudes amongst all stations is 56.01°, and the range of latitudes is 29.24°. As in the previous example, we split the ranges of longitudes and latitudes into 200 subintervals of length $l_1/200$ and width $l_2/200$ (i.e., $n_1 = n_2 = 200$), and 100 subintervals of length $l_1/100$ and width $l_2/100$ (i.e., $n_1 = n_2 = 100$); the value of k_n is taken to be $(n_1 n_2)^{1/\delta}$, as in the first example. We then calculate the corresponding G_n and $\widetilde{G}_n$, as shown in Table 5.

Because $\widetilde{G}_n$ converges to a standard normal distribution under the stationarity assumption, Table 5 shows that at a significance level of 0.05, the average monthly temperature varies between locations, and thus the stationarity assump-



(a) $\lambda = 16 \log n/k_n, n = 100^2$ (b) $\lambda = 16 \log n/k_n, n = 200^2$

Figure 9. The change locations and corresponding change boundaries for monthly temperature in China in January for the period $1970 - 2000$ based on $T_n > 16 \log n/k_n$ and $k_n = [2(n_1 n_2)^{1/5}]$, for (a) $n_1 = n_2 = 100$ and (b) $n_1 = n_2 = 200$.

tion is questionable. To determine possible change points, we consider those T_n with values larger than $\lambda = 16 \log n/k_n$, where $k_n = [2(n_1 n_2)^{1/5}]$, $n = n_1 n_2$, and $n_1 = n_2$ is chosen to be 100 and 200. The corresponding change points and their boundaries are plotted in Figure 9.

Figure 9 shows that different λ can lead to some changes on the boundaries, but that major change regions located in Xinjiang province and on the northern border remain similar. Little change occurs in the southern regions. These phenomena can be explained intuitively as follows. Because Xinjiang province is surrounded by mountains, plains, and deserts, its temperature varies more than in the other regions. Similarly, climate patterns in the northern part of China are more severely affected by the winter monsoon. As a result, the observed temperature changes are most pronounced in Xinjiang province and the northern part of China, as indicated in Figure 9.

6. Conclusion

This paper proposes a discrepancy measure for the identification of changes in spatial domains. The underlying idea is similar to the approximation of integrals by Riemann sums in local regions, in which in-fill asymptotics can be implemented. Also called fixed-domain asymptotics, these are based on observations that become increasingly dense in some fixed, bounded region as the sample size increases, as seen in Cressie (1993). Using this measure, we present novel statistical tests for structural changes and the specific forms of the underlying trends. We also establish the asymptotic properties of these tests and limit distributions. Furthermore, an algorithm to identify change-boundaries is provided.

The proposed procedures are illustrated using simulations and real data.

Although this study mainly focuses on the lattice case, it is plausible (although nontrivial) to extend the results to the nonlattice case by applying kernel smoothing to T_n. In addition, the approach presented here can be applied to test structural covariances by considering the product process $\{Y(\boldsymbol{s})Y(\boldsymbol{t})\}$. These challenging issues are left to future research.

Supplementary Material

Proofs of the main theorems are presented in Supplementary Materials.

Acknowledgments

We would like to thank the co-editors, Professor Hans-Georg Müller and Professor Xiaotong Shen, for their patience and helpful advice during the editorial process. We would also like to thank the associate editor and two anonymous referees for their helpful comments and thoughtful suggestions. This research was supported, in part, by grants from HKSAR-RGC-GRF Nos 14308218, 14307921 and 14325216 and HKSR-RGC-TRF No. T32-101/15-R (Chan), NSFC (No.11771390/1217427), the Zhejiang Provincial Natural Science Foundation (No. LZ21A010002), and the Fundamental Research Funds for the Central Universities (Zhang), and HKSAR-RGC-GRF Nos 14302719, 14305517, 14304221 (Yau).

References

Aue, A., Rice, G. and Sönmez, O. (2018). Detecting and dating structural breaks in functional data without dimension reduction. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **80**, 509–529.

Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica* **66**, 47–78.

Bai, J. and Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics* **18**, 1–22.

Bickel, P. J. and Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates. *The Annals of Statistics* **1**, 1071–1095.

Carlstein, E., Müller, H. G. and Siegmund, D. (1994). *Change-Point Problems*. Institute of Mathematical Statistics, Hayward.

Chan, N. H., Yau, C. Y and Zhang, R. M. (2014). Group Lasso for structural break time series. *Journal of the American Statistical Association* **109**, 590–599.

Chen, B. and Hong, Y. (2012). Testing for smooth structural changes in time series models via nonparametric regression. *Econometrica* **80**, 1157–1183.

Cressie, N. A. C. (1993). *Statistics for Spatial Data*. Wiley, New York.

Davis, R. A., Lee, T. C. M. and Rodriguez-Yam, G. A. (2006). Structure break estimation for nonstationary time series models. *Journal of the American Statistical Association* **101**, 223–239.

De Martino, F., Valente, G., Staeren, N., Ashburner, J., Goebel, R. and Formisano, E. (2008). Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns. *NeuroImage* **43**, 44–58.

Doesken, N. J., Pielke, Sr., R. A. and Bliss, O. A. P. (2003). *Climate of Colorado, Climatography of the United States No.60*. Colorado Climate Center, Colorado State University.

Doukhan, P. (2003). *Mixing: Properties and Examples*. Lecture Notes in Statistics 85, Springer, New York.

Furrer, R., Genton M. G. and Nychka, D. (2006). Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics* **15**, 502–523.

Hall, P., Peng, L and Rau, C. (2001). Local likelihood tracking of fault lines and boundaries. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **63**, 569–582.

Harchaoui, Z. and Lévy-Leduc, C. (2010). Multiple change-point estimation with a total variation penalty. *Journal of the American Statistical Association* **105**, 1480–1493.

Huang, L., Kulldorff, M. and Gregorio, D. (2007). A spatial scan statistic for survival data. *Biometrics* **63**, 109–118.

Johns, C. J., Nychka, D., Kittel, T. G. F. and Daly, C. (2003). Infilling sparse records of spatial fields. *Journal of the American Statistical Association* **98**, 796–806.

Kaufman, C. G., Schervish, M. J. and Nychka, D. W. (2008). Covariance tapering for likelihood-based estimation in large spatial data sets. *Journal of the American Statistical Association* **103**, 1545–1555.

Kulldorff, M. (2001). Prospective time periodic geographical disease surveillance using a scan statistic. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **164**, 61–72.

Lahiri, S. N. and Robinson, P. M. (2016). Central limit theorems for long-range dependent spatial linear processes. *Bernoulli* **22**, 345–375.

Lahiri, S. N. and Zhu, J. (2006). Resampling methods for spatial regression models under a class of stochastic designs. *The Annals of Statistics* **34**, 1774–1813.

Liang, F., Cheng, Y., Song, Q., Park, J. and Yang, P. (2013). A resampling-based stochastic approximation method for analysis of large geostatistical data. *Journal of the American Statistical Association* **108**, 325–339.

Lin, Z. and Lu, C. (1996). *Limit Theory on Mixing Dependent Random Variables*. Kluwer Academic Publishers, Dordrecht.

Machkouri, M., Volný, D. and Wu, W. B. (2013). A central limit theorem for stationary random fields. *Stochastic Processes and their Applications* **123**, 1–14.

Müller, H. G. and Song, K. (1994). Cube splitting in multidimensional edge estimation. In *Change-Point Problems* (Edited by E. Carlstein, H. G. Müller and D. Siegmund), 317–329. Institute of Mathematical Statistics, Hayward.

Mwangi, B., Tian, T. S. and Soares, J. C. (2014). A review of feature reduction techniques in neuroimaging. *Neuroinformatics* **12**, 229–244.

Neill, D. B. (2012). Fast subset scan for spatial pattern detection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **74**, 337–360.

Otto, P. and Schmid, W. (2016). Detection of spatial change points in the mean and covariances of multivariate simultaneous autoregressive models. *Biometrical Journal* **58**, 1113–1137.

Sherman, M. (2010). *Spatial Statistics and Spatio-Temporal Data: Covariance Functions and Directional Properties*. Wiley, New York.

Song, S. T., Zhan, Z. C., Long, Z. Y., Zhang, J. C. and Yao, L. (2011). Comparative study of SVM methods combined with voxel selection for object category classification on fMRI data. *PLoS One* **6**, e17191.

Tsybakov, A. B. (1994). Multidimensional change-point problems and boundary estimation. In *Change-Point Problems* (Edited by E. Carlstein, H. G. Müller and D. Siegmund), 317–329. Institute of Mathematical Statistics, Hayward.

Wu, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 14150–14154.

Wu, W. B. and Zhao, Z. (2007). Inference of trends in time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **69**, 391–410.

Ngai Hang Chan

Department of Statistics, The Chinese University of Hong Kong, Shatin, Hong Kong, China.

E-mail: nhchan@sta.cuhk.edu.hk

Rongmao Zhang

School of Mathematical Science, Zhejiang University, Hangzhou, Zhejiang, China.

E-mail: rmzhang@zju.edu.cn

Chun Yip Yau

Department of Statistics, The Chinese University of Hong Kong, Shatin, Hong Kong, China.

E-mail: cyyau@sta.cuhk.edu.hk

(Received April 2019; accepted March 2021)