Assessing Spatial Stationarity and Segmenting Spatial Processes into Stationary Components

ShengLi Tzeng\textsuperscript{a}, Bo-Yu Chen\textsuperscript{b}, Hsin-Cheng Huang\textsuperscript{c,*}

\textsuperscript{a}Department of Applied Mathematics, National Sun Yat-Sen University, Taiwan.  
\textsuperscript{b}Department of Statistics, Purdue University, USA.  
\textsuperscript{c}Institute of Statistical Science, Academia Sinica, Taiwan.

Abstract

In this research, we propose a novel technique for visualizing nonstationarity in geostatistics, particularly when confronted with a single realization of data at irregularly spaced locations. Our method hinges on formulating a statistic that tracks a stable microergodic parameter of the exponential covariance function, allowing us to address the intricate challenges of nonstationary processes that lack repeated measurements. We implement the fused lasso technique to elucidate nonstationary patterns at various resolutions. For prediction purposes, we segment the spatial domain into stationary sub-regions via Voronoi tessellations. Additionally, we devise a robust test for stationarity based on contrasting the sample means of our proposed statistics between two selected Voronoi subregions. The effectiveness of our method is demonstrated through simulation studies and its application to a precipitation dataset in Colorado.

Keywords: Fused lasso, Geostatistics, irregularly spaced data, microergodic parameter, nonstationary spatial process, spatial clustering, spatial visualization, stationarity test, Voronoi tessellation

*Corresponding author

Email address: hchuang@stat.sinica.edu.tw (Hsin-Cheng Huang)
1. Introduction

Consider a spatial process \( \{y(s) : s \in D\} \) of interest defined on a region \( D \subset \mathbb{R}^2 \). Suppose that we observe data \( z \equiv (z(s_1), \ldots, z(s_n))' \) at \( n \) spatial locations, which may be irregularly spaced, according to the measurement equation:

\[
z(s_i) = y(s_i) + e(s_i); \quad i = 1, \ldots, n,
\]

where \( e(s_1), \ldots, e(s_n) \sim N(0, \tau^2) \) are white-noise variables, representing measurement errors. A major problem in geostatistics, called kriging, is to predict \( y(s_0) \) at any location \( s_0 \in D \) based on \( z \). For simplicity, we assume that the mean function of the process \( y(\cdot) \) is known and, without loss of generality, zero. Then for a given covariance function of \( y(\cdot) \), the ordinary-kriging predictor of \( y(s_0) \) is

\[
\hat{y}(s_0) = \left( c + \frac{1 - c' \Sigma^{-1} 1}{1' \Sigma^{-1} 1} \right)' \Sigma^{-1} Z,
\]

where \( c \equiv \text{cov}(z, y(s_0)), \Sigma \equiv \text{var}(z), \) and \( 1 = (1, \ldots, 1)' \).

Given a realization noisy data \( z \) at \( n \) locations, it is typical to assume that the covariance function of \( y(\cdot) \) is stationary. A commonly used stationary covariance model is the isotropic Matérn family (Matérn, 1986) given by

\[
\text{cov}(y(s), y(s + u)) = \frac{\sigma^2}{2^{\nu-1} \Gamma(\nu)} \left( \frac{\sqrt{2\nu}}{\alpha} \| u \| \right)^\nu K_\nu \left( \frac{\sqrt{2\nu}}{\alpha} \| u \| \right); \quad s, s + u \in \mathbb{R}^2,
\]

where \( K_\nu(\cdot) \) is the modified Bessel function of the second kind of order \( \nu \), \( \sigma^2 \) is a variance parameter, and \( \theta \equiv (\alpha, \nu)' \) consists of a scale parameter \( \alpha \) and a smoothness parameter \( \nu \). It’s important to emphasize that spatial covariance functions don’t always exhibit stationarity. Sometimes, they can be markedly influenced by local conditions and topographical variations, leading to substantial deviations from stationarity. Visualizing the nonstationary attributes from a single dataset presents a challenge. For illustration, Figure 1(a1) depicts a zero-mean stationary process with a Matérn covariance function. In contrast, Figure 1(a2) presents a zero-mean piecewise stationary process characterized by two distinct Matérn covariance functions. While the nonstationarity in Figure 1(a2) is apparent, determining which
one among Figures 1(b1) and 1(b2) (consisting of 400 random samples from the processes in Figures 1(a1) and 1(a2) respectively) exhibits nonstationarity is not straightforward.

Several approaches have been proposed for testing spatial stationarity. Fuentes (2005) pioneered a frequency domain test for spatial samples on a regular grid. Jun and Genton (2012) proposed a test that partitions the spatial domain into two non-intersecting fields for irregularly spaced data. More recently, Bandyopadhyay and Rao (2017) unveiled a test leveraging the Fourier transform in the frequency domain, catering to irregularly spaced data. In addition, local indicators of spatial autocorrelation (LISA) have been proposed by Anselin (1995) for lattice data. To our knowledge, there seems to be an absence of spatial dependence indices crafted explicitly for irregularly sampled data within geostatistics.

In this study, we introduce a local statistic designed to highlight nonstationary characteristics within geostatistical datasets. This is achieved by employing a robust local estimation of a microergodic parameter inherent to the exponential covariance model. We leverage the fused lasso methodology to illuminate nonstationary patterns across varying resolutions. To enhance the accuracy of spatial predictions using stationary models, we segment the spatial domain into homogenous sub-regions utilizing Voronoi tessellations. A rigorous test for spatial stationarity is established by comparing the sample means of the estimated microergodic parameters between two Voronoi subregions. If the stationarity is violated, we further partition $D$ into $K$ components $\{D_1, \ldots, D_K\}$ such that each process $\{y(s) : s \in D_k\}$ is stationary, for $k = 1, \ldots, K$. It’s worth mentioning that both Guinness and Fuentes (2015) and Muyskens et al. (2022) have crafted techniques to segregate domain $D$ into stationary subregions. However, the approach by Guinness and Fuentes requires data on a regular grid, while the method by Muyskens et al. (2022) necessitates a regular grid shape for their base partition due to the algorithm’s reliance on circulant embedding. It’s important to note that the choice of grid resolution influences segmentation and increases computational demands when selecting among different resolutions.

The rest of this paper is organized as follows. In Section 2, we develop a statistic to
Figure 1: (a1) A zero-mean stationary process; (a2) A zero-mean nonstationary process; (b1) Data sampled from the process in (a1) at 400 locations using simple random sampling; (b2) Data sampled from the process in (a2) at 400 locations using simple random sampling; (c1) Local spatial indices based on the data in (b1); (c2) Local spatial indices based on the data in (b2).
monitor spatial heterogeneity at multiple resolutions. The statistic is designed to estimate a microergodic parameter of the exponential variogram with data that may be irregularly spaced. We then provide an approach to partition the domain $D$ into $K$ homogeneous subregions using Voronoi tessellations (Voronoi, 1908). Section 3 gives the proposed test for stationarity using a $t$-type statistic based on the Voronoi subregions obtained from Section 2 with $K = 2$. Some simulation results are given in Section 4. An application to a precipitation dataset in Colorado is provided in Section 5. Finally, Section 6 concludes with a summary.

2. Segmenting spatial processes into stationary components

2.1. A statistic for local spatial dependence

First, we construct a statistic to monitor the local spatial dependence of $y(\cdot)$ around $s_i$, for $i = 1, \ldots, n$. To find local structure around $s_i$, we consider a neighborhood set of $s_i$:

$$N_i \equiv \{ j : \| s_j - s_i \| \leq r, j \neq i \}; \quad i = 1, \ldots, n,$$

where $\| \cdot \|$ is the Euclidean distance and $r > 0$ is an appropriate radius. If $y(\cdot)$ is an isotropic stationary process around $s_i$, then its variogram at distance $h$ is

$$2\gamma_{y,i}(h) \equiv E(y(s) - y(s_i))^2, \text{ for } h = \| s - s_i \| \leq r; \quad i = 1, \ldots, n.$$

It follows that for $j \in N_i$,

$$2\gamma_{z,i}(\| s_j - s_i \|) \equiv E\left((z(s_j) - z(s_i))^2\right) = 2\gamma_y(\| s_j - s_i \|) + 2\delta_{ij}\tau^2; \quad i = 1, \ldots, n,$$

where $\delta_{ij} \equiv 1 \text{ if } i = j; \ 0 \text{ otherwise.}$ To reflect the local behavior, it is desirable to consider $N_i$'s with a small $r$. To obtain a statistic that is robust to outliers, we utilize a squared-root transform and apply the following approximation formula (Cressie and Hawkins, 1980):

$$\frac{|z(s_j) - z(s_i)|^{1/2}}{\{2\gamma_{z,i}(\| s_j - s_i \|)\}^{1/4}} \approx N\left(2^{1/4}\pi^{-1/2}\Gamma(3/4), 2^{1/2}\{\pi^{-1/2} - \pi^{-1}\Gamma(3/4)^2\}\right); \quad s_j \in N_i. \quad (5)$$

We first assume that $\tau^2$ is known, and consider a local exponential semi-viariogram:

$$\gamma_{y,i}(h) = \sigma_i^2(1 - \exp(h/\alpha_i)); \quad 0 \geq h, \ i = 1, \ldots, n,$$
parametrized by variance $\sigma_i^2$ and range parameter $\alpha_i$; $i = 1, \ldots, n$. However, it is well known that both $\sigma_i^2$ and $\alpha_i$ are unidentifiable under the infill asymptotic framework (Zhang, 2004). Instead, we focus on their ratio, $\sigma_i^2/\alpha_i$, a microergodic parameter that can be consistently estimated. Applying a Taylor expansion to $(2\gamma_{z,i}(h))^{1/4}$ at $h = 0$, we obtain

$$
\gamma_{z,i}(h)^{1/4} = \left\{ \sigma_i^2(1 - \exp(-h/\alpha_i)) + \tau^2 \right\}^{1/4} = \tau^{1/2} + \tau^{-3/2}\sigma_i^2 h/(4\alpha_i) + O(h^2); \ i = 1, \ldots, n.
$$

Substituting $\tau^{1/2} + \tau^{-3/2}\sigma_i^2 ||s_j - s_i||/(4\alpha_i)$ above for $\gamma_{z,i}(||s_j - s_i||)^{1/4}$ in (5) leads to

$$
E\left(\frac{|z(s_i) - z(s_j)|^{1/2} - C_1}{C_2||s_i - s_j||}\right) \approx \sigma_i^2/\alpha_i; \ i = 1, \ldots, n,
$$

(6)

where $C_1 \equiv 2^{1/2}\pi^{-1/2}\Gamma(3/4)(\tau^2)^{1/4}$ and $C_2 \equiv 2^{-3/2}\pi^{-1/2}\Gamma(3/4)(\tau^2)^{-3/4}$. Note that the left-hand side of (6) depends on $||s_j - s_i||$, but the right-hand side does not. In addition, from (5), for small $||s_j - s_i||$,

$$
\text{var}\left(\frac{|z(s_i) - z(s_j)|^{1/2} - C_1}{C_2||s_i - s_j||}\right) \approx \frac{C_3}{||s_j - s_i||^2},
$$

where $C_3 \equiv 2(\pi^{-1/2} - \pi^{-1}\Gamma(3/4)^2)/C_2$. This motivates us to use the following weighted average as our local spatial indices to monitor the heterogeneity of spatial dependence:

$$
\xi_i \equiv \frac{1}{|N_i|} \sum_{j \in N_i} \omega_{ij} \sum_{j \in N_i} \left\{ \omega_{ij} \frac{|z(s_j) - z(s_i)|^{1/2} - C_1}{||s_j - s_i||} \right\}; \ i \in \mathcal{I},
$$

(7)

where $\omega_{ij} \equiv ||s_j - s_i||^2$, $\mathcal{I} \equiv \{ i : |N_i| > 0, i = 1, \ldots, n \}$, and $|N_i|$ denotes the number of elements in $N_i$. In practice, we recommend choosing $r = \{5|D|/(n\pi)\}^{1/2}$ in (4), so that $|N_i| \approx 5$ on average, for $i \in \mathcal{I}$. Figures 1(c1) and 1(c2) show the proposed local spatial indices of (7) based on the data in Figures 1(b1) and 1(b2), respectively.

When $\tau^2$ is unknown, we estimate it based on a linear extrapolation to the zero ordinate of $\gamma_{z,i}(\cdot)$ at two small lags, determined by $\mathcal{P}_k \equiv \{(i,j) : d_{k-1}^* < |s_i - s_j| \leq d_k^*, i < j\}; k = 1, 2$, where $0 = d_0^* < d_1^* < d_2^*$. Specifically, we compute the robust semivariogram estimates of Cressie and Hawkins (1980) based on pairs in $\mathcal{P}_k$:

$$
\hat{\gamma}_k = \left( \frac{\sum_{(i,j) \in \mathcal{P}_k} |z(s_i) - z(s_j)|^{1/2}/m_k}{2(0.457 + 0.494/m_k + 0.045/m_k^2)} \right)^4; \ k = 1, 2,
$$

(8)
where $m_k$ is the number of pairs on $P_k$; $k = 1, 2$. Applying linear extrapolation while imposing constraints for a nonnegative slope and intercept, we obtain

$$\hat{\gamma}^2 = \max \left(0, \hat{\gamma}_1 - d_1 \max \left(0, \frac{\hat{\gamma}_2 - \hat{\gamma}_1}{d_2 - d_1} \right) \right),$$

(9)

where $d_k = \sum_{(i,j) \in P_k} |s_i - s_j|/m_k$ is the average distance among pairs in $P_k$, for $k = 1, 2$.

Two distinct subregions are discernible from Figure 2. In contrast, there is no clear pattern from Figure 3.

2.2. Multiresolution spatial visualization

We note from (5) that $\xi_i$ is approximately Gaussian with $E(\xi_i) \approx C_2 \sigma_i^2 / \alpha_i$, for $i \in I$. Let $n^* \equiv |I|$, and without loss of generality, assume that $I = \{1, \ldots, n^*\}$. If $y(\cdot)$ is globally stationary, we have $\sigma_1^2 = \cdots = \sigma_n^2$ and $\alpha_1 = \cdots = \alpha_n$, and hence $E(\xi_1) \approx \cdots \approx E(\xi_n^*)$. We can apply a spatial-clustering approach to segment $D$ into stationary components based on $\xi \equiv (\xi_1, \ldots, \xi_{n^*})'$.

To explore the spatial nonstationarity evident in the data presented in Figure 1(b2) at various resolutions, we decompose $D$ into disjoint Voronoi cells, denoted as $\{B_1, \ldots, B_{n^*}\}$, corresponding to $\{s_1, \ldots, s_{n^*}\}$. Each point in the Voronoi cell $B_k$ is closer to $s_k$ than any other point in the set $\{s_1, \ldots, s_{n^*}\}$. Then we cluster $\{B_1, \ldots, B_{n^*}\}$ into homogeneous components using the fused lasso (Tibshirani et al., 2005):

$$\sum_{i=1}^{n^*} (\xi_i - \beta_i)^2 + \rho \sum_{(j,k) \in E} |\beta_j - \beta_k|,$$

where $E$ is obtained by linking between any two cells that share a boundary and $\rho \geq 0$ is a regularization parameter. The resulting images at multiple resolutions with different tuning parameter values of $\rho$ are shown in Figure 2. The corresponding images for the data in Figure 1(b1) from a stationary process are shown in Figure 3. Comparing the two collections of images, it is evident that Figure 2 showcases at least two major components with distinct values across different resolutions. In contrast, Figure 3 consistently exhibits a single prominent homogeneous component throughout all resolutions.
Figure 2: Multiresolution spatial visualization for a piecewise stationary process using the proposed fused lasso method.

Figure 3: Multiresolution spatial visualization for a stationary process using the proposed fused lasso method.
2.3. Voronoi tessellations into stationary components

The fused lasso approach given in the previous section is used for visualization. We aim to partition the region $D$ into stationary components for subsequent spatial prediction. To achieve this, we utilize the Voronoi subregions constructed from $K$ seeds denoted as $\{p_1, \ldots, p_K\} \subset \mathbb{R}^2$. Given these seeds, we derive the corresponding Voronoi tessellation, subdividing $D$ into $K$ distinct components $D_1, \ldots, D_K$. Define $n_k \equiv \sum_{i \in I} I(s_i \in D_k); k = 1, \ldots, K$. Let $S_K$ be the set of all possible $K$ seeds. To identify the optimal set of $K$ seeds $\{p_1, \ldots, p_K\}$ from $S_k$, we apply the following objective function grounded in independent normal likelihood:

$$f_K(p_1, \ldots, p_K; \xi) = \sum_{k=1}^{K} \sum_{i \in I; s_i \in D_k} \left\{ \log(\hat{v}_k(D_k)) - \log \phi \left( \frac{\xi_i - \hat{\mu}_k(D_k)}{\hat{v}_k(D_k)} \right) \right\},$$

(10)

where

$$\hat{\mu}_k(D_k) \equiv \frac{1}{n_k} \sum_{i \in I; s_i \in D_k} \xi_i,$$

and

$$\hat{v}_k^2(D_k) \equiv \frac{1}{n_k} \sum_{i \in I; s_i \in D_k} (\xi_i - \hat{\mu}_k(D_k))^2,$$

are the maximum likelihood (ML) estimators of the mean and the variance of $\{\xi_i : s_i \in D_k, i \in I\}$, for $k = 1, \ldots, K$, and $\phi(\cdot)$ is the probability density function of the standard normal distribution. The proposed segmentation of $D$ into $K \geq 2$ components is determined by

$$\{p_1^{(K)}, \ldots, p_K^{(K)}\} \equiv \arg\min_{\{p_1, \ldots, p_K\} \in S_K} f_K(p_1, \ldots, p_K; \xi)$$

(11)

with the corresponding Voronoi tessellation $\hat{D}_1^{(K)}, \ldots, \hat{D}_K^{(K)}$.

We propose a simple algorithm to find the solution of (11). Its pseudo-code is outlined in Algorithm [1].

3. The proposed test for stationarity

We can utilize the proposed segmentation with $K = 2$ to test for spatial stationarity. We consider the following hypothesis test:

$$H_0 : y(\cdot) \text{ is stationary} \; \text{versus} \; H_1 : y(\cdot) \text{ is not stationary}.$$
Algorithm 1 Find the solution of (11) with a given $K$ based on $\xi$.

Require:
\{\(p_1, \ldots, p_K\)\}: the initial seeds obtained from a deterministic $K$-means algorithm of Nidheesh et al. (2017);
\{\(D_1, \ldots, D_K\)\}: the Voronoi tessellations corresponding to \(\{p_1, \ldots, p_K\}\).

Ensure:
\begin{align*}
\text{repeat} & \\
& \text{for } k \leftarrow 1 \text{ to } K \text{ do} \\
& \quad \text{update } p_k \text{ by replacing it from } \{s_i \in D_k : i = 1, \ldots, n^*\} \text{ such that } f_K(p_1, \ldots, p_K; \xi) \text{ is minimized;} \\
& \quad \text{update } \{D_1, \ldots, D_K\} \text{ corresponding to the current seeds } \{p_1, \ldots, p_K\}; \\
& \text{end for} \\
\text{until} & \text{ no further reduction of } f_K(p_1, \ldots, p_K; \xi) \text{ is possible.}
\end{align*}

Based on the two subregions $\hat{D}_1^{(2)}$ and $\hat{D}_2^{(2)}$ selected by (11) with $K = 2$, we propose the following two-sample $t$ statistic:
\begin{equation}
T = \frac{|\bar{\mu}_1(\hat{D}_1^{(2)}) - \bar{\mu}_2(\hat{D}_2^{(2)})|}{\sqrt{\hat{v}_1^2(\hat{D}_1^{(2)})/(\hat{n}_1 - 1) + \hat{v}_2^2(\hat{D}_2^{(2)})/(\hat{n}_2 - 1)}}, \tag{12}
\end{equation}
where $\hat{n}_k \equiv \sum_{i=1}^{n^*} I(s_i \in \hat{D}_k^{(2)}); k = 1, 2$. The distribution of $T$ is complicated because there is a selection process involved in obtaining $\hat{D}_1^{(2)}$ and $\hat{D}_2^{(2)}$. So, we apply a Monte Carlo (MC) method to find the null distribution of $T$. Specifically, we assume that under $H_0$, $y(\cdot)$ is a Gaussian process with the isotropic Matérn covariance model of (3). We estimate $\theta$ and $\sigma^2$ in (3) by ML. The ML estimator of $\theta = (\alpha, \nu)'$ can be obtained by minimizing the negative log profile likelihood:
\begin{equation}
\hat{\theta} \equiv (\hat{\alpha}, \hat{\nu})' \equiv \arg \min_{\theta} \left\{ \frac{1}{2} \log |\Omega(\theta)| + \frac{K}{2} \log \{z'\Omega(\theta)^{-1}z\} + \text{constant} \right\}, \tag{13}
\end{equation}
where $\Omega(\theta)$ is an $n \times n$ correlation matrix whose $(i, j)$-th entry is $\text{cov}(y(s_i), y(s_j)) + \tau^2 \delta_{ij}/\sigma^2$. Then the ML estimator of $\sigma^2$ is given by:
\begin{equation}
\hat{\sigma}^2 \equiv \frac{1}{n} z'\Omega(\hat{\theta})^{-1}z. \tag{14}
\end{equation}

To implement the proposed MC method, first, we simulate data $z^{(m)}$, for $m = 1, \ldots, M$, based on (1) and (3) with $\theta$ and $\sigma^2$ replaced by $\hat{\theta}$ in (13) and $\hat{\sigma}^2$ in (14). Next, we compute
$T_m$ in (12) based on $z^{(m)}$. Then the MC $p$-value of the proposed test is

$$
\hat{p} = \frac{1}{M+1} \sum_{m=1}^{M} I(T_m > T). \quad (15)
$$

Although we introduce the segmentation method before hypothesis testing, in practice, we first perform the stationarity test and obtain the $p$-value $\hat{p}$ of (15). We use a stationary model for subsequent analysis if $\hat{p} \geq 0.05$. Otherwise, we apply the proposed spatial segmentation method to partition $D$ into $K$ stationary subregions with $K$ selected by minimizing Bayesian information criterion (BIC) (Schwarz, 1978):

$$
\text{BIC}(K) = f_K(\hat{p}_1^{(K)}, \ldots, \hat{p}_K^{(K)}) + 4K \log(n^*). \quad (16)
$$

4. Simulation studies

4.1. Testing stationarity

We examined the size of the proposed stationarity test under $H_0$ by performing the same simulation experiment as in Section 7.1.1 of Bandyopadhyay and Rao (2017). We considered a zero-mean spatial process $\{y(s) : s \in D\}$ on a region $D = [-5/2, 5/2] \times [-5/2, 5/2]$ with a Matérn covariance function of (3). We generated data according to (1) with $\sigma^2 = 1$, $\nu = 1/2$, $\alpha \in \{1/3, 2/3, 1, 4/3, 2\}$, and $\tau^2 \in \{0, 0.01\}$. In addition, we considered various sample sizes $n \in \{50, 100, 500, 1000, 2000\}$ and two distributions for sampling locations, including a uniform distribution and a clustered distribution with two clusters (see details in Bandyopadhyay and Rao, 2017), resulting in a total of $5 \times 2 \times 5 \times 2 = 100$ combinations.

We compared our method with BR’s (Bandyopadhyay and Rao, 2017). The empirical Type-I error rates under various settings for the uniform and the clustered distributions are shown in Table 1 and Table 2 respectively. Although our method shows a few elevated Type-I error rates when spatial dependence is strong, overall, the Type-I error rates are close to the nominal level. On the other hand, the Type-I error rates for the BR’s method tend to be too large for a few cases under the uniform design and too small for many instances under the clustered design. The distributions of p-values for various scenarios under $H_0$ are
Table 1: Empirical Type-I errors for our method and BR’s method (Bandyopadhyay and Rao, 2017) under various scenarios with a uniform sampling design based on 500 simulated replicates.

| n   | Method | $\tau^2 = 0$ | $\tau^2 = 0.01$ |
|-----|--------|--------------|-----------------|
|     |        | $\alpha = \frac{1}{3}$ | $\alpha = \frac{2}{3}$ | $\alpha = 1$ | $\alpha = \frac{1}{3}$ | $\alpha = \frac{2}{3}$ | $\alpha = 1$ | $\alpha = \frac{1}{3}$ | $\alpha = \frac{2}{3}$ | $\alpha = 1$ |
| 50  | Ours   | 0.068 | 0.054 | 0.052 | 0.058 | 0.060 | 0.062 | 0.060 | 0.056 | 0.058 | 0.054 |
| 50  | BR     | 0.030 | 0.020 | 0.040 | 0.050 | 0.090 | 0.030 | 0.020 | 0.050 | 0.050 | 0.080 |
| 100 | Ours   | 0.034 | 0.042 | 0.044 | 0.054 | 0.058 | 0.034 | 0.046 | 0.044 | 0.056 | 0.050 |
| 100 | BR     | 0.030 | 0.030 | 0.030 | 0.040 | 0.040 | 0.030 | 0.050 | 0.040 | 0.050 | 0.040 |
| 500 | Ours   | 0.046 | 0.054 | 0.056 | 0.052 | 0.050 | 0.040 | 0.050 | 0.048 | 0.054 | 0.048 |
| 500 | BR     | 0.020 | 0.030 | 0.020 | 0.030 | 0.030 | 0.040 | 0.080 | 0.070 | 0.130 | 0.120 |
| 1000| Ours   | 0.072 | 0.072 | 0.068 | 0.070 | 0.054 | 0.040 | 0.080 | 0.070 | 0.070 | 0.068 |
| 1000| BR     | 0.050 | 0.050 | 0.060 | 0.060 | 0.080 | 0.080 | 0.100 | 0.100 | 0.140 | 0.130 |
| 2000| Ours   | 0.078 | 0.066 | 0.064 | 0.064 | 0.072 | 0.074 | 0.074 | 0.068 | 0.076 | 0.080 |
| 2000| BR     | 0.070 | 0.060 | 0.090 | 0.090 | 0.080 | 0.070 | 0.090 | 0.090 | 0.180 | 0.180 |

Table 2: Empirical Type-I errors for our method and BR’s method (Bandyopadhyay and Rao, 2017) under various scenarios with a clustered sampling design based on 500 simulated replicates.

| n   | Method | $\tau^2 = 0$ | $\tau^2 = 0.01$ |
|-----|--------|--------------|-----------------|
|     |        | $\alpha = \frac{1}{3}$ | $\alpha = \frac{2}{3}$ | $\alpha = 1$ | $\alpha = \frac{1}{3}$ | $\alpha = \frac{2}{3}$ | $\alpha = 1$ | $\alpha = \frac{1}{3}$ | $\alpha = \frac{2}{3}$ | $\alpha = 1$ |
| 50  | ours   | 0.064 | 0.060 | 0.074 | 0.072 | 0.078 | 0.058 | 0.060 | 0.070 | 0.076 | 0.072 |
| 50  | BR     | 0.020 | 0.020 | 0.010 | 0.010 | 0.020 | 0.020 | 0.020 | 0.020 | 0.030 | 0.020 |
| 100 | ours   | 0.082 | 0.074 | 0.072 | 0.062 | 0.054 | 0.080 | 0.078 | 0.064 | 0.074 | 0.068 |
| 100 | BR     | 0.020 | 0.020 | 0.020 | 0.020 | 0.020 | 0.020 | 0.030 | 0.020 | 0.030 | 0.030 |
| 500 | ours   | 0.060 | 0.052 | 0.058 | 0.060 | 0.056 | 0.064 | 0.052 | 0.052 | 0.060 | 0.060 |
| 500 | BR     | 0.020 | 0.020 | 0.010 | 0.010 | 0.010 | 0.010 | 0.020 | 0.020 | 0.010 | 0.010 |
| 1000| ours   | 0.056 | 0.058 | 0.064 | 0.066 | 0.060 | 0.060 | 0.060 | 0.060 | 0.062 | 0.062 |
| 1000| BR     | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.020 |
| 2000| ours   | 0.052 | 0.062 | 0.074 | 0.076 | 0.076 | 0.058 | 0.058 | 0.072 | 0.074 | 0.076 |
| 2000| BR     | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 |

displayed in Figures A.1-A.4. They are all very close to the uniform distribution on (0, 1) as we anticipate.

Next, we investigated the power of the proposed test following the same setups in Bandyopadhyay and Rao (2017). We considered three scenarios. In the first two scenarios, we replaced the stationary Matérn covariance function of (3) by a nonstationary Matérn covariance function with $\lambda = 20$ and 40, respectively:

$$
cov(y(s_1), y(s_2)) = |\Sigma_\lambda(s_1)|^{1/4}|\Sigma_\lambda(s_2)|^{1/4}\{\Sigma_\lambda(s_1) + \Sigma_\lambda(s_2)\}/2|^{-1/2}\exp\left(-|Q_\lambda(s_1, s_2)|\right),
$$

(17)
Table 3: Empirical powers for our and BR’s methods (Bandyopadhyay and Rao, 2017) under various scenarios based on 500 simulated replicates.

| n    | Method | $\tau^2 = 0$ | $\tau^2 = 0.01$ |
|------|--------|--------------|-----------------|
|      |        | $\lambda = 20$ | $\lambda = 20$ |
|      |        | $\lambda = 40$ | $\lambda = 40$ |
|      |        | 4 blocks      | 4 blocks        |
| 50   | Our    | 0.050         | 0.050           |
|      |        | 0.058         | 0.058           |
|      |        | 0.074         | 0.074           |
| 50   | BR     | 0.020         | 0.020           |
|      |        | 0.030         | 0.030           |
|      |        | 0.050         | 0.050           |
| 100  | Our    | 0.060         | 0.060           |
|      |        | 0.044         | 0.044           |
|      |        | 0.106         | 0.106           |
| 100  | BR     | 0.050         | 0.050           |
|      |        | 0.040         | 0.040           |
|      |        | 0.040         | 0.040           |
| 500  | Our    | 0.340         | 0.340           |
|      |        | 0.136         | 0.136           |
|      |        | 0.570         | 0.570           |
| 500  | BR     | 0.190         | 0.190           |
|      |        | 0.100         | 0.100           |
|      |        | 0.110         | 0.110           |
| 1000 | Our    | 0.760         | 0.760           |
|      |        | 0.266         | 0.266           |
|      |        | 0.926         | 0.926           |
| 1000 | BR     | 0.470         | 0.470           |
|      |        | 0.350         | 0.350           |
|      |        | 0.240         | 0.240           |
| 2000 | Our    | 0.990         | 0.990           |
|      |        | 0.570         | 0.570           |
|      |        | 1.000         | 1.000           |
| 2000 | BR     | 0.700         | 0.700           |
|      |        | 0.850         | 0.850           |
|      |        | 0.360         | 0.360           |

where

$$Q_\lambda(s_1, s_2) \equiv 2(s_1 - s_2)' \left\{ \Sigma_\lambda(s_1) + \Sigma_\lambda(s_2) \right\}^{-1}(s_1 - s_2),$$

$$\Sigma_\lambda(s) \equiv \begin{pmatrix} \log \left( \frac{s_x}{\lambda} + \frac{3}{4} \right) & -\frac{\|s\|^2}{\lambda^2} \\ \frac{\|s\|^2}{\lambda^2} & \log \left( \frac{s_x}{\lambda} + \frac{3}{4} \right) \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0.5 \end{pmatrix} \begin{pmatrix} \log \left( \frac{s_x}{\lambda} + \frac{3}{4} \right) & -\frac{\|s\|^2}{\lambda^2} \\ \frac{\|s\|^2}{\lambda^2} & \log \left( \frac{s_x}{\lambda} + \frac{3}{4} \right) \end{pmatrix},$$

and $s = (s_x, s_y)'$. For the third scenario, we considered a zero-mean piecewise stationary process $\{y(s) : s \in D\}$ by dividing $D = [-5/2, 5/2] \times [-5/2, 5/2]$ into $2 \times 2$ blocks of equal sizes. The processes on four blocks are mutually independent and have the Matérn covariance functions of (3), with $\sigma^2 = 1$, $\nu = 1/2$, and four different values of $\alpha \in \{1, 1/3, 1/2, 2/3\}$ for the four blocks. For each scenario, we considered the uniform sampling design and generated data according to (1) with $\tau^2 \in \{0, 0.01\}$ and $n \in \{50, 100, 500, 1000, 2000\}$, resulting in 10 different combinations. The empirical powers are displayed in Table 3 based on 500 simulated replicates. Except for a few cases in Scenario 2 with $\gamma = 40$ and $n \geq 1000$, our method is more powerful than the BR’s method in detecting spatial nonstationarity.

4.2. Spatial segmentation

We investigated the cluster recovery ability of the proposed method in spatial segmentation. We considered a region $D = [0, 1]^2$ and decomposed it into $D_1 \cup D_2$ as shown in Figure
We generated a zero-mean spatial process \( \{ y(s) : s \in D \} \) on \( D \) based on

\[
y(s) = w_1(s; a) \eta_1(s) + w_2(s; a) \eta_2(s); \quad s \in D,
\]

where

\[
w_k(s; a) \equiv \frac{\exp(-d(s, D_k)/a)}{\exp(-d(s, D_1)/a) + \exp(-d(s, D_2)/a)}; \quad s \in D, \quad k \in \{1, 2\},
\]

are weight functions with \( a > 0 \) controlling the degree of smoothness for process \( y(\cdot) \) around the boundary between \( D_1 \) and \( D_2 \), \( d(s, D_k) \equiv \min_{s^* \in D_k} \|s - s^*\| \), and \( \eta(s) \equiv (\eta_1(s), \eta_2(s))' \) is a zero-mean bivariate spatial process with a bivariate exponential covariance function:

\[
\text{cov}(\eta_k(s), \eta_{k'}(s^*)) = \left( \frac{2\alpha_k \alpha_{k'}}{\alpha_k^2 + \alpha_{k'}^2} \right)^{1/2} \exp\left(-\frac{(\alpha_k^2 + \alpha_{k'}^2)^{1/2}}{2^{1/2} \alpha_k \alpha_{k'}} \|s - s^*\| \right); \quad k, k' \in \{1, 2\}.
\]

We generated data according to (11) and (18) with \( \tau^2 = 0 \), \( \alpha_1 = 0.1 \), and \( \alpha_2 \in \{0.1, 0.2, 0.3, 0.4, 0.5\} \). Additionally, we considered \( n \in \{100, 500\} \) and \( a \in \{0.01, 0.1\} \), resulting in a total of 20 combinations. Note that \( \alpha_2 \) controls the degree of nonstationarity. When \( \alpha_1 = \alpha_2 \), we obtain \( y(\cdot) \) to be a stationary process with \( \text{cov}(y(s), y(s^*)) = \exp(-\|s - s^*\|/\alpha_1) \) regardless of the value of \( a \). By contrast, a larger departure of \( \alpha_2 \) from \( \alpha_1 \) indicates a higher degree of nonstationarity. These features can be seen in Figure 5, which shows realizations of \( y(\cdot) \) with \( \alpha_2 \in \{0.1, 0.2, 0.3, 0.4, 0.5\} \).

We applied the proposed optimization of (11) to segment \( D \) into Voronoi subregions \( \hat{D}_1, \ldots, \hat{D}_K \). We selected the final \( K \) according to BIC of (16). The performance of an estimated clustering \( \tilde{D} = \{\hat{D}_1, \ldots, \hat{D}_K\} \) is evaluated using the Rand index (Rand, 1971).
Figure 5: Realizations of $y(\cdot)$ from models with various $\alpha_2$ values, where a larger $\alpha_2$ value corresponds to a higher degree of nonstationarity, and $\alpha_2 = 0.1$ corresponds to a stationary process.

Table 4: Proportions of selecting the correct number (i.e., $K = 2$) of clusters under various situations based on 500 simulated replicates.

| $\alpha_2$ | $a = 0.01$ | $a = 0.1$ |
|------------|------------|------------|
|            | $n = 100$  | $n = 500$  |
|            | $n = 100$  | $n = 500$  |
| 0.1        | 0.326      | 0.296      |
| 0.2        | 0.446      | 0.736      |
| 0.3        | 0.560      | 0.800      |
| 0.4        | 0.642      | 0.794      |
| 0.5        | 0.694      | 0.834      |

based on $\{s_1, \ldots, s_n\}$:

$$R(D, \tilde{D}) \equiv \frac{n_{00} + n_{11}}{n_{00} + n_{01} + n_{10} + n_{11}},$$

where $D = \{D_1, D_2\}$ is the true clustering,

$n_{00}$ is the number of point pairs that are in different clusters under both $D$ and $\tilde{D}$,

$n_{01}$ is the number of point pairs that are in the same cluster under $D$ but in different clusters under $\tilde{D}$,

$n_{10}$ is the number of point pairs that are in different clusters under $D$ but in the same cluster under $\tilde{D}$,

$n_{11}$ is the number of point pairs that are in the same cluster under both $D$ and $\tilde{D}$.

Tables 4 and 5 show the proportions of selecting the correct number of clusters and the average Rank Index values based on our method under various situations. As expected, our method performs better for a smaller $a$ and a larger $n$. 

15
Table 5: Average Rand index values based on the number of clusters selected by BIC under various situations based on 500 simulated replicates.

| $\alpha_2$ | $a = 0.01$ | $a = 0.1$ |
|------------|------------|------------|
| $n = 100$  | $n = 500$  | $n = 100$  | $n = 500$  |
| 0.1        | 0.563      | 0.547      | 0.556      | 0.544      |
| 0.2        | 0.603      | 0.746      | 0.582      | 0.663      |
| 0.3        | 0.650      | 0.845      | 0.609      | 0.748      |
| 0.4        | 0.689      | 0.884      | 0.660      | 0.778      |
| 0.5        | 0.735      | 0.905      | 0.686      | 0.793      |

5. An application to precipitation data in Colorado

In this section, we applied our method to a precipitation dataset in Colorado. The dataset can be obtained from the Geophysical Statistics Project at the National Center for Atmospheric Research (http://www.image.ucar.edu/GSP/Data/US.monthly.met/CO.html), which has been analyzed previously by Paciorek and Schervish (2006) and Qadir et al. (2021). It consists of monthly total precipitation (in mm) recorded at 367 weather stations across Colorado from 1895 to 1997. It is well known that Western Colorado is mountainous with more significant topographical variability than Eastern Colorado.

Following Qadir et al. (2021), we considered the cumulative precipitations in the year 1992 and analyzed the data observed at 254 stations with no missing observations after applying the log transformation. Figure 6(a) shows the precipitation data we analyzed.

We estimated $\tau^2$ based on (8) and (9) by selecting a small $d_1^*$ and $d_2^*$ so that $m_1 = m_2 = 250$. Applying the proposed test of (12) described in Section 3, we obtained a p-value smaller than 0.01 for testing spatial stationarity, suggesting that the underlying process is likely nonstationary. We then segmented the process into stationary processes based on subregions by applying the proposed spatial segmentation based on (11) introduced in Section 2.3. From (16), we obtained the BIC values 359.75, 220.31, 213.82, and 228.96, for $K = 1, \ldots, 4$, respectively, where $K = 1$ corresponds to the stationary exponential model. The smallest BIC value is achieved at $K = 3$. Figure 6(b)-(d) shows the segmentation results based on $K = 2, 3, 4$. Even though we did not utilize any additional information (such as
Figure 6: (a) Precipitation amounts (mm in log scale) at 254 stations in Colorado in 1992; (b) Two subregions obtained by the proposed methods; (c) Three subregions obtained by the proposed methods; (d) Four subregions obtained by the proposed methods.

elevation) other than precipitations, Colorado Eastern Plains, which tend to have a different climate pattern from the rest, are automatically segmented as a subregion for $K \in \{2, 3, 4\}$, demonstrating that the proposed spatial segmentation method is effective.

We also investigated whether the proposed segmentation enhances spatial prediction. We randomly split the data into training data $\{z(s_i) : i \in I_{\text{train}}\}$ (consisting of 204 observations) and test data $\{z(s_i) : i \in I_{\text{test}}\}$ (with 50 observations). Using the training data, we applied the proposed spatial segmentation method (11) introduced in Section 2.3 with $K = 1, \ldots, 4$. Upon identifying the $K$ subregions through our methodology, we conducted spatial prediction
by fitting an exponential covariance model to each subregion independently, operating under the assumption that the data were generated from $[1]$. Our approach considered $y(\cdot)$ as a piecewise stationary process, in line with the decomposition. For every subregion, the model parameters were estimated using Maximum Likelihood (ML). Subsequently, we harnessed ordinary kriging from equation $[2]$ to derive the predictive surface for each subregion. To gauge the performance of our predictors, we utilized the root mean squared prediction error (RMSPE) criterion:

$$\text{RMSPE} = \left\{ \frac{1}{50} \sum_{i \in I_{\text{test}}} \left( \hat{y}(s_i) - z(s_i) \right)^2 \right\}^{1/2}.$$  

We also evaluated the performance of probabilistic forecast using the continuous ranked probability score (CRPS, Geniting and Raftery, 2007):

$$\text{crps}(F, z) = \int_{-\infty}^{\infty} (F(t) - I(t \geq z))^2 \, dt,$$

where $F(\cdot)$ is the predictive cumulative distribution function, $z \in \mathbb{R}$ is an observation, and $I(\cdot)$ is an indicator function. We computed the CRPS based on test data:

$$\text{CRPS} = \frac{1}{50} \sum_{i \in I_{\text{test}}} \text{crps}(\hat{F}(s_i), z(s_i)),$$

where for $i \in I_{\text{test}}, \hat{F}(s_i)$ is a generic predictive cumulative distribution function of $z(s_i)$.

We randomly split the data into training and test data 200 times and obtained 200 predicted values and prediction standard deviations at each location. Figure 7 shows boxplots of the RMSPE and CRPS values for $K = 1, \ldots, 4$. Our method performs better than the stationary model in terms of RMSPE and CRPS regardless of $K = 2, 3, 4$.

6. Summary

We develop a statistic to track nonstationarity by focusing on a microergodic parameter. This innovation enables us to simultaneously detect changes in both spatial variances and spatial ranges, from which we can segment the region into stationary components using Voronoi tesselations. The proposed method is designed for data observed at irregularly spaced locations without repeated measurements.
Additionally, we introduce a novel test to detect the nonstationarity of a spatial process. Our test is not only computationally efficient, but it also properly controls the Type-I error rate, proving to be more powerful than existing methods. Compared to the test by Bandyopadhyay and Rao (1997), which tends to underperform with an irregular sampling design, our test remains largely unaffected by the irregularity of data locations.

The proposed stationarity test offers another advantage: it can point out where the nonstationarity occurs once rejected. As a result, we can perform kriging by applying a stationary model to each component separately. It is also conceivable to take this further by establishing a divide-and-conquer strategy to combine the results. These avenues present promising research directions, especially when dealing with massive spatial data. Further investigations along these lines, including the construction of nonstationary models based on locally stationary processes and the development of scalable methods for kriging, are of significant interest but fall beyond the scope of this paper. We intend to explore these areas in future work.

Figure 7: Prediction performances of the precipitation data in Colorado based on 200 pairs of randomly split training and test data: (a) Boxplots of RMSPEs; (b) Boxplots of CRPSs.
Figure A.1: The distributions of p-values under $H_0$ for various scenarios with $\tau^2 = 0$ under a uniform sampling design.

Appendix A.

In this section, we display the distributions of p-values for various scenarios under $H_0$ in Section 4.1.

References

Anselin, L. (1995). Local indicators of spatial association–LISA, *Geographical Analysis*, 27, 93–115.

Bandyopadhyay, S. and Rao, S. S. (2017). A test for stationarity for irregularly spaced spatial data, *Journal of Royal Statistical Society, Series B*, 79, 95–123.

Cressie, N. (1993). Statistics for Spatial Data, rev. edn, Wiley, New York, NY.
Figure A.2: The distributions of p-values under $H_0$ for various scenarios with $\tau^2 = 0.01$ under a uniform sampling design.
Figure A.3: The distributions of p-values under $H_0$ for various scenarios with $\tau^2 = 0$ under a clustered sampling design.
Figure A.4: The distributions of p-values under $H_0$ for various scenarios with $\tau^2 = 0.01$ under a clustered sampling design.
Cressie, N. and Hawkins, D. M. (1980). Robust estimation of the variogram: I, *Mathematical Geology*, 12, 115–125.

Fuentes, M. (2005). A formal test for non-stationarity of spatial stochastic processes. *Journal of Multivariate Analysis*, 96, 30–54.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation, *Journal of the American Statistical Association*, 102, 359–378.

Guinness, J., and Fuentes, M. (2015). Likelihood approximations for big nonstationary spatial temporal lattice data, *Statistica Sinica*, 25, 329–349.

Jun, M. and Genton, M. (2012) A test for stationarity of spatio-temporal random fields on planar and spherical domains, *Statistica Sinica*, 22, 1737–1764.

Matérn, B. (1986). *Spatial Variation*, 2nd ed., Springer-Verlag, Berlin.

Muyskens, A., Guinness, J., and Fuentes, M. (2022). Partition-based nonstationary covariance estimation using the stochastic score approximation, *Journal of Computational and Graphical Statistics*, 31, 1025–1036.

Nidheesh, N., Nazeer, K. A., and Ameer, P. M. (2017). An enhanced deterministic K-Means clustering algorithm for cancer subtype prediction from gene expression data. *Computers in biology and medicine*, 91, 213-221.

Paciorek, C. J. and Schervish, M. J. (2006). Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics*, 17, 483-506.

Qadir, G. A., Sun, Y. and Kurtek, S. (2021). Estimation of spatial deformation for nonstationary processes via variogram alignment. *Technometrics*, 63, 548–561.

Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods, *Journal of the American Statistical Association*, 66, 846–850.
Schwarz, G. (1978). Estimating the dimension of a model, *The Annals of Statistics*, 6, 461–464.

Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., and Knight, K. (2005). Sparsity and smoothness via the fused lasso, *Journal of the Royal Statistical Society, Series B*, 67, 91–108.

Voronoi, G. (1908). Nouvelles applications des paramètres continus à la théorie des formes quadratiques. Premier mémoire. Sur quelques propriétés des formes quadratiques positives parfaites, *Journal für die reine und angewandte Mathematik (Crelles Journal)*, 1908, 97–102.

Zhang, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics, *Journal of the American Statistical Association*, 99, 250–261.