South African Statistical Journal
Vol. 58, No. 1, pp. 57–79
https://doi.org/10.37920/sasj.2024.58.1.4

Multiscale decomposition of spatial lattice data for hotspot detection

René Stander, Inger Fabris-Rotelli and Ding-Geng Chen

Department of Statistics, University of Pretoria

Hotspot detection in spatial analysis identifies geographic areas with elevated event rates, facilitating more effective policy interventions aimed at reducing such incidents. In the current literature, several methods have been used to detect hotspots such as measures for local spatial association and spatial scan methods. However, the performance of these methods is limited for small-scale hotspots as well as spatial domains where the number of areas is small. In this work, we propose a new approach, making use of the Discrete Pulse Transform (DPT) to decompose spatial lattice data along with the multiscale Ht-index and the spatial scan statistic as a measure of saliency on the extracted pulses to detect significant hotspots. The proposed method outperforms the well-used local Getis-Ord statistic in a simulation study, especially on small-scale hotspots. The method is also illustrated on South African COVID-19 cases and South African crime data.

Keywords: COVID-19, Crime, Discrete Pulse Transform, Feature detection, Hotspot detection, Ht-index, Local Getis-Ord, Mutliscale decomposition, Multiscale Ht-index, Spatial lattice data, Spatial scan statistics, Spatial statistics.

1. Introduction

A spatial hotspot is defined as an area in the spatial domain where the occurrence of events or the rate of event occurrences is higher relative to the domain as well as the neighbouring areas (Wang et al., 2013; Chainey, 2014; Stresman et al., 2019; Zahran et al., 2020; Modiba et al., 2022). In the application section of this paper, this is discussed in a use-case scenario. In spatial lattice data, the spatial domain is divided into predefined subregions. For this reason, a hotspot in such data will be considered as an already defined subregion or a collection of these subregions in the spatial domain.

Detecting spatial hotspots is important to identify areas of concern and then to effectively and efficiently deploy preventative measures and resources (Chainey, 2014). Examples of applications include, but are not limited to, criminology (Chen et al., 2010), zoological sciences (Wu et al., 2013), public health (Mahara et al., 2016), disease mapping (Stresman et al., 2019), traffic accidents (Cheng et al., 2018) and forest fires (Said et al., 2017). Although hotspot detection is mostly used in the literature to identify areas of concern, it is important to note that hotspot detection methods can also be used to detect higher concentrations of positive outcomes.

Corresponding author: René Stander (rene.stander@up.ac.za)
MSC2020 subject classifications: 62M30, 62M40 62P10 62P25

To detect spatial hotspots, local indicators of spatial association (LISA) have been used in the literature (Anselin, 1995). These are the Getis-Ord statistic (Getis and Ord, 1992), Geary's C (Geary, 1954) and the Moran's I statistic (Moran, 1948). For all of these statistics, the value for the test statistic is calculated as a global measure of the overall degree of event clustering present in the entire spatial domain. All three have a counterpart to calculate the spatial association at each spatial unit, then called the local statistic as defined by Anselin (1995). In the case of spatial lattice data, each subregion is considered a spatial unit where measurements are observed. Then, the significant locations are detected according to the test statistic value of the local measure of association.

A commonly used LISA statistic, is the local Getis-Ord statistic (Getis and Ord, 1992). Since its development, it has been used in a number of applications (Lees, 2006; Songchitruksa and Zeng, 2010; Jana and Sar, 2016; Everett et al., 2021; Althaf et al., 2022). The test statistic values for the local Getis-Ord statistic approach normality as the number of areas increases. Therefore, when the number of areas is small, the normality assumption is violated and the results of the method are influenced negatively (Getis and Ord, 1992).

Another LISA statistic, also used in the literature, is the local Moran's I statistic (Anselin, 1995). It has been used in a variety of applications (Ruiz et al., 2004; Zhang et al., 2008; Bone et al., 2013; Tepanosyan et al., 2019). Similarly as for the local Getis-Ord statistic, Wang et al. (2023) also found that the results of the local Moran's I are jeopardised when the number of areas is small.

Another group of hotspot detection methods covered in the literature is the spatial scan method (Kulldorff, 1997). Being an effective hotspot detection method, especially in disease mapping, the spatial scan statistic uses a circular sliding window approach to scan the study area and effectively test whether the event occurrences are by chance using a likelihood ratio test. The exact properties, such as size and shape, of the hotspots present will always be unknown before-hand. Features of the sliding window determines the type and the size of the hotspots detected. The significance of the sliding window is also highlighted by Kulldorff (1997). Since the development of the spatial scan method, many different methods and sliding windows have been implemented. These methods are summarised in a review paper by French et al. (2022). Another limitation of the spatial scan statistic is that the performance of the method is not optimal when the hotspots to detect are small in size (Aamodt et al., 2006).

A hotspot can be seen as a salient feature, which is an important and distinctive part of the data (Kadir and Brady, 2001). A method applied to images, the Discrete Pulse Transform (DPT) algorithm, has been used to extract salient features (Fabris-Rotelli and Stein, 2018, 2020). It was first developed for one-dimensional arrays (Rohwer, 2006) and Anguelov and Fabris-Rotelli (2010) extended the algorithm to two-dimensional arrays. The DPT algorithm performs a hierarchical decomposition by recursively applying the LULU operators, L_n and U_n, $n = 1, \ldots, N$, where N is the size of the array. The L_n and U_n operators smooth local minimum and maximum sets of size n. As the DPT algorithm smooths the array using the LULU operators, the smoothed-out values are extracted as pulses (Fabris-Rotelli and Stein, 2018, 2020). See Fabris-Rotelli and Stein (2020) for an example). The size of the pulses is then called scales. By adding all the extracted pulses together, the original array is obtained and therefore reconstructed in full. Salient features can be extracted by reconstructing selective pulses, hence partial reconstruction (Fabris-Rotelli and Stein, 2020).

Stander et al. (2021) applied the DPT on irregular spatial lattice data for the first time. Previously, the DPT algorithm has only been applied to regular spatial lattice data (images). As far as the authors

are aware, this paper then presents the second implementation of the DPT on irregular spatial lattice data. In this paper, we propose a new hotspot detection method using the multiscale decomposition algorithm on irregular spatial lattice data to extract salient features. In this application, the salient features of interest will be hotspots in spatial lattice data.

Herein, the choice of which pulses to partially reconstruct is a two-step process. First, informative intervals are created with the multiscale Ht-index (Fabris-Rotelli and Stein, 2020). The multiscale Ht-index has been extended from the Ht-index developed by Jiang et al. (2013) which divided the scales of the pulses into intervals. Secondly, to know which intervals contain the salient features, we require some quantification method. We propose the use of the spatial scan statistic to quantify the saliency of the partially reconstructed pulses for the intervals of scales identified by the multiscale Ht-index.

The newly proposed hotspot detection method introduces an approach only requiring the counting model, the Poisson process. Therefore, the method is robust to the number of areas in the spatial domain considered. Also, quantifying the saliency for each interval of pulses separately adds to the robustness of the method for small-scale clusters. The proposed hotspot detection method will perform equally well on the number of events or the rate of events. Based on the application at hand, the user may determine the appropriate scenario.

The structure of the paper is as follows. In Section 2, we will cover some important theoretical concepts necessary for the proposed method and the proposed methodology. Section 3 is started with a proof of concept in the form of a simulation study, followed by two applications of the proposed method on South African crime and COVID-19 data in Section 4. The discussion of the results as well as the conclusion is in Section 5.

2. Methodology

2.1 Spatial lattice data

Spatial lattice data are observed over a collection of areas. This type of spatial data, observed over N areas, can be represented as

$$Z(D) = \{Z(A_1), Z(A_2), \ldots, Z(A_N)\},$$

where $D = \bigcup_{i=1}^{N} A_i$ and $i = 1, \ldots, N$ (Cressie, 1993). Here, $Z(\cdot)$ indicates the spatial process observed which can either be continuous or discrete. The observed spatial process at each area is typically the average or the sum of the measured attribute for A_i.

The areas in spatial lattice data are predefined and can either be regularly or irregularly shaped. When areas in spatial lattice data are regularly shaped, all the areas have the same shape and size. Spatial lattice data with regularly shaped areas are also known as image arrays, but some sources refer to them as gridded data (Li et al., 2015). These can also be referred to as *regular lattice data*. In the literature, spatial lattice data with irregularly shaped areas are referred to as either areal data (Banerjee et al., 2014) or spatial lattice data (Cressie, 1993). Examples of irregularly shaped areas include, but are not limited to ward boundaries, district boundaries and province boundaries[1]. To avoid confusion with the term *spatial lattice data*, we refer to such data as *irregular lattice data*. The

[1] Examples of three levels of administrative boundaries. This is the terminology used in South Africa.

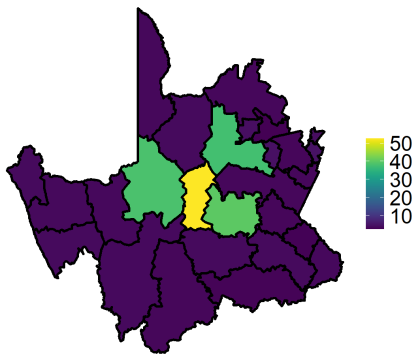


Figure 1. Illustration of irregular lattice data where each A_i represents a municipality in the Northern Cape province in South Africa. $Z(A_i)$ is simulated and indicated with a colour scale. The simulation of $Z(A_i)$ is done such that four neighbouring regions have higher values and consequently form a hotspot.

term *spatial lattice data* will then be used as a collective term to refer to both types of data, *irregular lattice data* and *regular lattice data*.

Figure 1 is an illustration of irregular lattice data which will be used as an illustrative example for the remainder of this section. The colour scale in Figure 1 represents the $Z(A_i)$ values of each A_i. Each A_i in the example is a municipality in the Northern Cape province in South Africa while $Z(A_i)$ are simulated such that four neighbouring regions have higher values and form a hotspot. The $Z(A_i)$ values are simulated by using spatial sampling to sample spatial random points throughout the entire spatial domain. The number of sampled points in A_i is then $Z(A_i)$. Then, in four neighbouring regions, additional spatial random points are sampled such that these areas have a higher $Z(\cdot)$ value and subsequently form a hotspot.

In the illustrative example for this section, we are considering the rate of event occurrences. The observed value at each subregion is standardised by dividing the number of events by the size of the region.

2.2 Connectivity

We define a morphological connection in Definition 1. The Discrete Pulse Transform works on connected sets, C_i, and is defined for a morphological connection.

Definition 1. Morphological connection (Serra, 1988). Let B be an arbitrary non-empty set. A family C of subsets of B is called a connected class or a connection on B if:
(1) $\emptyset \in C$;
(2) $\{x\} \in C$ for all $x \in B$; and
(3) For any family, $C_i \in C$ for $i \in I$,

$$\bigcap_{i \in I} C_i \neq \emptyset \Longrightarrow \bigcup_{i \in I} C_i \in C.$$

If a set C_i belongs to a connection C, then C_i is called a connected set.

Tobler's First Law of Geography states that *everything is related to everything else, but near things are more related than distant things* (Tobler, 1970). Therefore, when working with spatial data, neighbours are defined between the areas to account for the spatial dependence that may exist, following the idea of connectivity in Definition 1. For A_i, we define the set of neighbours as:

$$\mathcal{N}(A_i) = \bigcup_{k=1}^{q_i} \mathcal{N}^k(A_i),$$

where $\mathcal{N}^k(A_i)$, $k = 1, \ldots, q_i$, are the areas that satisfy the neighbourhood definition and are neighbours of A_i (Zhu et al., 2010). The number of areas, q_i, that satisfies the neighbourhood definition might differ for each A_i.

Different neighbourhood definitions exist in the literature, but for the purpose of this paper, we will only focus on the Rook contiguity neighbourhood definition (Bivand et al., 2013). Two areas, A_i and A_j, are defined to be Rook contiguity neighbours if they share a polygon edge[2]. In image analysis (regular lattice data), the Rook contiguity neighbours are also known as 4-connectivity[3]. According to Fabris-Rotelli (2013), the Rook contiguity neighbourhood definition satisfies Definition 1.

The neighbourhood object can be represented with an adjacency matrix. Some sources refer to the term *spatial weight matrix* (Zhu et al., 2010) or *proximity matrix* (Banerjee et al., 2014). The adjacency matrix is an $N \times N$ matrix indicating the neighbourhood relationship between spatial units,

$$\mathbf{W} = \left[w_{ij}\right]_{i,j=1}^{N},$$

where $w_{ii} = 0$, implying that $A_i \notin \mathcal{N}(A_i)$. For the Rook contiguity neighbourhood definition, the w_{ij} entries are either 0 or 1 (Bivand et al., 2013),

$$w_{ij} = \begin{cases} 1 & \text{if } A_j \in \mathcal{N}(A_i), \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The Rook contiguity neighbour definition is illustrated in Figure 2 using the illustrative example from Figure 1. The white star indicates, A_i, the area whose contiguity neighbours we are interested in. The red lines connect to $\mathcal{N}^1(A_i)$, $\mathcal{N}^2(A_i)$, and $\mathcal{N}^3(A_i)$, the areas with which A_i shares a boundary and satisfy the Rook contiguity neighbourhood definition.

2.3 LULU operators and the Discrete Pulse Transform

The LULU operators (Anguelov and Fabris-Rotelli, 2010) are non-linear and fully trend-preserving filters that operate on a signal $f \in \mathcal{A}(\mathbb{Z}^2)$, where $\mathcal{A}$ is a vector lattice with N data points. Traditionally, these operators have been applied on regular lattices with N areas and for the explanation of the theory, we will consider regular lattice data in two dimensions. The L_n operator smooths local maximum sets of size n and the U_n operator smooths local minimum sets of size n. These are defined as,

$$L_n(f)(x) = \max_{V \in \mathcal{N}_n(x)} \min_{y \in V} f(y), x \in \mathbb{Z}^2, \tag{2}$$

$$U_n(f)(x) = \min_{V \in \mathcal{N}_n(x)} \max_{y \in V} f(y), x \in \mathbb{Z}^2, \tag{3}$$

[2] Another contiguity neighbourhood definition is the Queen contiguity neighbours where A_i and A_j share either a polygon edge or a vertex.

[3] The Queen contiguity neighbours are also known as 8-connectivity.

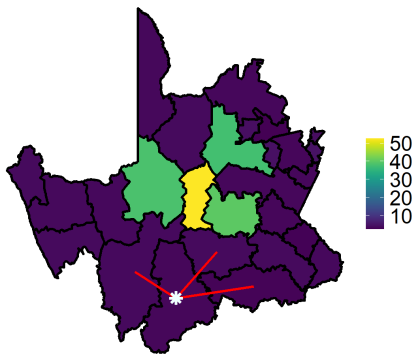


Figure 2. Illustration of the Rook contiguity neighbourhood definition. The white star indicates A_i. The red lines connect to $\mathcal{N}^1(A_i)$, $\mathcal{N}^2(A_i)$ and $\mathcal{N}^3(A_i)$.

where $\mathcal{N}_n(x)$ is the set of all connected sets of size $n+1$ containing point x, $\mathcal{N}_n = \{V \in C : x \in V,\ \text{card}(V) = n+1\}$, where C is a morphological connection and $\text{card}(V)$ denotes the number of elements in the set V. $x \in \mathbb{Z}^2$ is the position of the array cell in a regular grid, for example $x = (1, 1)$ is the first row, first column. The array cell itself is then an areal unit A_i, a regular polygon, i.e., a square. See Fabris-Rotelli and Stein (2018, 2020) and Stander et al. (2021) for details not repeated here. Local minimum and maximum sets are defined in Definition 3 but it is also necessary to define an adjacent set, $V \in C$ (see Definition 2).

Definition 2. Adjacent set (Anguelov and Fabris-Rotelli, 2010). Let $V \in C$. A point $x \notin V$ is adjacent to V if $V \cup \{x\} \in C$. The set of all points adjacent to V is

$$\text{adj}(V) = \{x \in \mathbb{Z}^2 : x \notin V,\ V \in C\}.$$

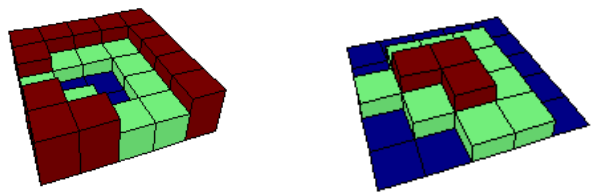
Definition 3. Local maximum set and local minimum set (Anguelov and Fabris-Rotelli, 2010). A connected subset V of $\mathbb{Z}^2$ is called a local maximum set of $f \in \mathcal{A}(\mathbb{Z}^2)$ if

$$\sup_{y \in adj(V)} f(y) < \inf_{x \in V} f(x).$$

Similarly V is a local minimum set if

$$\inf_{y \in adj(V)} f(y) > \sup_{x \in V} f(x).$$

Thus, a local minimum (maximum) set is a connected set with an adjacent set strictly more (less) than the value of the set, as illustrated in Figure 3. Figure 3(a) illustrates a local minimum set, set V, with the blue blocks, while the green blocks are the adjacent set, $adj(V)$. Figure 3(b) shows a local maximum set, set V, with the red blocks, and the adjacent set, $adj(V)$, with the green blocks. In Anguelov and Fabris-Rotelli (2010), it has been proven that the L_n and U_n operators can only



(a) (b)

Figure 3. Illustrations of (a) local minimum and (b) local maximum sets. In (a) the blue blocks are the local minimum set and the green blocks are the adjacent set, whereas in (b) the red blocks are the local maximum set and the green blocks are the adjacent set.

be applied to local minimum and local maximum sets. This implementation leads to a simpler computation than applying (2) and (3), i.e., from first principles.

The Discrete Pulse Transform (DPT) is a hierarchical decomposition algorithm that recursively applies the L_n and U_n operators for $n = 1, \ldots, N$, where N is the size of the array. At each iteration of the DPT algorithm, the smoothed signal $P_n(f)$ is subtracted from the identity operator. This results in signal f filtered by applying P_n (Anguelov and Fabris-Rotelli, 2010). The process starts by applying L_1 and U_1, hence smoothing local minimums and local maximums of size 1. The value of n is increased at each step from 2 up to N, obtaining

$$DPT(f) = (D_1(f), D_2(f), \ldots, D_N(f)),$$

where $D_1(f) = (I - P_1)(f)$, $D_n(f) = (I - P_n) \circ Q_{n-1}(f)$, $P_n = L_n \circ U_n$ or $P_n = U_n \circ L_n$, and $Q_n = P_1 \circ \cdots \circ P_n$, $n \in \mathbb{N}$, and where I is the identity operator. The filtered parts removed are called pulses, ϕ_{ns}, $s = 1, \ldots, \gamma(n)$, where $\gamma(n)$ is the number of pulses of scale n. For a connected set $V \subset \mathbb{Z}^2 \ni \text{card}(V) = n$, a pulse is defined as

$$\phi_{ns} = \begin{cases} c & \text{if } x \in V, \\ 0 & \text{otherwise,} \end{cases}$$

for some constant $c \in \mathbb{R}$. The original signal can be reconstructed in full as,

$$\sum_{n=1}^{N} D_n(f) = \sum_{n=1}^{N} \sum_{s=1}^{\gamma(n)} \phi_{ns}.$$

The implementation of the DPT algorithm is similar for irregular lattice data as for regular lattice data, as long as the neighbourhood structure satisfies the definition of a morphological connection. The only difference when applying the DPT algorithm to irregular lattice data compared to regular lattice data is the neighbourhood object being defined differently (Stander et al., 2021). According to Fabris-Rotelli (2013), the contiguity neighbours satisfy Definition 1. When applying the LULU operators and the DPT algorithm to irregular spatial lattice data, the subregions from the irregular lattice data, A_i, are used instead of $x \in \mathbb{Z}^2$.

A partially reconstructed signal can be attained by summing together pulses in a certain interval. For example, consider pulses with scales $[e_l; e_u]$, then the partial reconstruction is

$$f^p_{[e_l;e_u]}(D) = \sum_{n=\lceil e_l \rceil}^{\lfloor e_u \rfloor} D_n(f) = \sum_{s=1}^{\gamma(n)} \phi_{ns}, \tag{4}$$

and represented by the spatial process,

$$Z^p_{[e_l;e_u]}(D) = \left\{ Z^p_{[e_l;e_u]}(A_1), Z^p_{[e_l;e_u]}(A_2), \ldots, Z^p_{[e_l;e_u]}(A_N) \right\}. \tag{5}$$

When implementing the DPT algorithm, all possible connected sets are considered at each iteration. The inability to parallel process the DPT algorithm adds to the computational aspect. Laurie (2010) mentions that the algorithm has a complexity of $O(N^3)$. As the number of areas in the spatial lattice data increases, the complexity also increases rapidly. Graph-based algorithms, such as the Roadmaker's algorithm (Laurie, 2010), which was later on extended to the Roadmaker's Pavage algorithm (Stoltz, 2014), have been developed for implementation of the DPT algorithm. The implementation of the DPT decomposition for this paper is done with the Roadmaker's Pavage algorithm. This method is discussed in detail in Stander et al. (2021).

2.4 Multiscale Ht-index

Salient features can be extracted by partially reconstructing pulses with certain scales. The multiscale Ht-index (Fabris-Rotelli and Stein, 2020) is utilised to obtain informative intervals of the scales of the extracted pulses from the DPT algorithm.

The Ht-index utilises head/tail breaks in order to divide the data into sensible groups (Jiang et al., 2013). This method is applied to the scales of the pulses extracted using the DPT algorithm. For example, consider the pulses with the following scales 1, 1, 1, 1, 2, 2, 6, 9 respectively after decomposition. The first step is to calculate the mean of the scales, which is 2.875. This leads to 75% of the scales being less than the mean (1, 1, 1, 1, 2, 2). The process is repeated by only considering the scales larger than the mean in the previous step (6, 9). The mean of 6 and 9 is 7.5. Then 50% of the values are less than the mean and the process can stop. The calculation was performed twice, so the Ht-index is equal to 2. The intervals to consider for the extracted pulses are $[1; 2.875]$, $[6; 7.5]$, and $[9; 9]$.

With the multiscale version of the Ht-index, the calculations are repeated on the intervals calculated in each previous step. The multiscale version of the Ht-index allows for a visual representation in a tree-like structure as in Figure 4. The multiscale Ht-index calculated from the decomposition of the illustrative example in Figure 1 is shown in Figure 4. The initial intervals calculated with the Ht-index are $[1; 4.654]$, $[5; 10.222]$, $[12; 19.5]$, and $[27; 27]$. These are shown as the first (top) level. From here, the Ht-index calculation is repeated on each interval. For example, applying the Ht-index on the scales in the interval $[5; 10.222]$ results in intervals $[5; 7.571]$ and $[8; 10]$. These are shown in the second level below the interval used.

2.5 Saliency quantification

We propose the use of the spatial scan statistic as a measure to quantify the saliency in each interval resulting from the multiscale Ht-index.

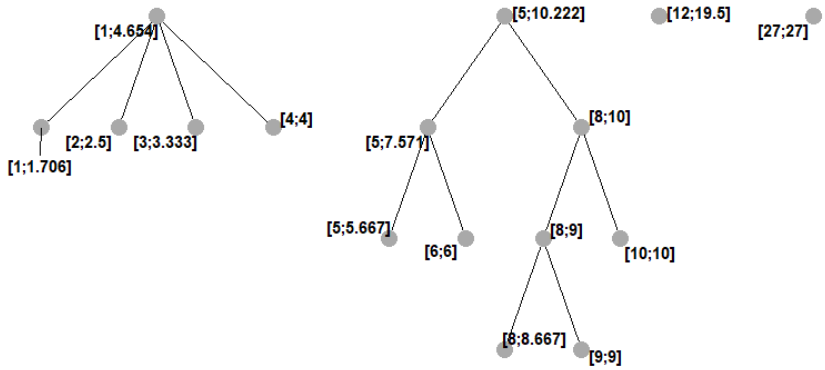


Figure 4. Visual representation of the multiscale Ht-index in a tree-like structure. The multiscale Ht-index applied to the decomposition of the illustrative example in Figure 1.

Let us consider a spatial scan statistic applied to spatial lattice data. The domain, D, consists of N areas with a total of n_D events occurring over the entire domain. We denote n_{A_i} to be the number of events occurring in area A_i. A sliding window approach is used in the process by iterating over each area.

The spatial scan statistic is defined for a Poisson model. The Poisson process is the logic counting process which is a popular choice in spatial statistics. The null hypothesis of homogeneity is tested against the alternative hypothesis $H_A : p > q$, $A_i \in D$, where p is the probability of a certain event occurring in area $A_i \subset D$ and q is the probability of a certain event occurring in area $\bar{A}_i \subset D \setminus A_i$.

The test statistic of the spatial scan statistic is obtained by using a likelihood ratio test. The test statistic is fully derived by Kulldorff (1997). The likelihood ratio is,

$$LR(A_i) = \begin{cases} \left(\frac{n_{A_i}}{\mu_{A_i}}\right)^{n_{A_i}} \left(\frac{n_D - n_{A_i}}{n_D - \mu_{A_i}}\right)^{n_D - n_{A_i}} & n_{A_i} > \mu_{A_i}, \\ 1 & \text{otherwise}, \end{cases}$$

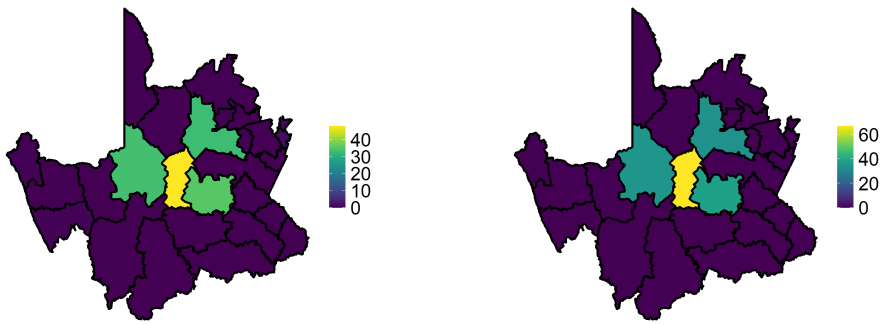
where μ_{A_i} is the expected number of events to occur in area A_i under the null hypothesis of homogeneity (Kulldorff, 1997; Duczmal et al., 2006; Ishioka et al., 2007). The test statistic is then,

$$\lambda = \max_{A_i \in D} LR(A_i). \tag{6}$$

The test statistic value from (6) is calculated for each interval from the multiscale Ht-index. As the scales of the pulses are integer values, we consider only the integer values within the interval in partial reconstruction. For example, if we consider the scales in the interval from the multiscale Ht-index, $[1; 4.654]$, then the partial reconstruction is obtained with (4) as,

$$f^p_{[1;4.654]}(D) = \sum_{n=1}^{4} D_n(f) = \sum_{s=1}^{\gamma(n)} \phi_{ns},$$

with the spatial process, $Z^p_{[1;4.654]}(D)$, as in (5). The partial reconstruction is shown in Figure 5(a). For $f^p_{[1;4.654]}(D)$, we calculate $LR(A_i), \forall A_i \in D$. The calculated non-zero values for $LR(A_i)$ are $\{38.491; 35.052; 67.117; 33.614\}$. These are shown in Figure 5(b). The test statistic value for the



(a) (b)

Figure 5. Considering the interval of scales $[1; 4.654]$ from the multiscale Ht-index from the illustrative example. (a) $f^{P}_{[1;4.654]}(D)$, the partial reconstruction of the scales. (b) $LR(A_i)$, the likelihood ratio, calculated for $f^{P}_{[1;4.654]}(D)$.

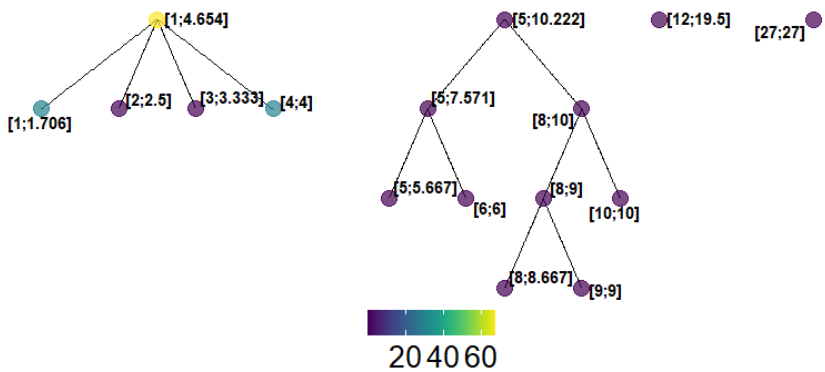


Figure 6. Test statistic value, λ, calculated for each interval from the multiscale Ht-index. The value of λ is indicated by colour.

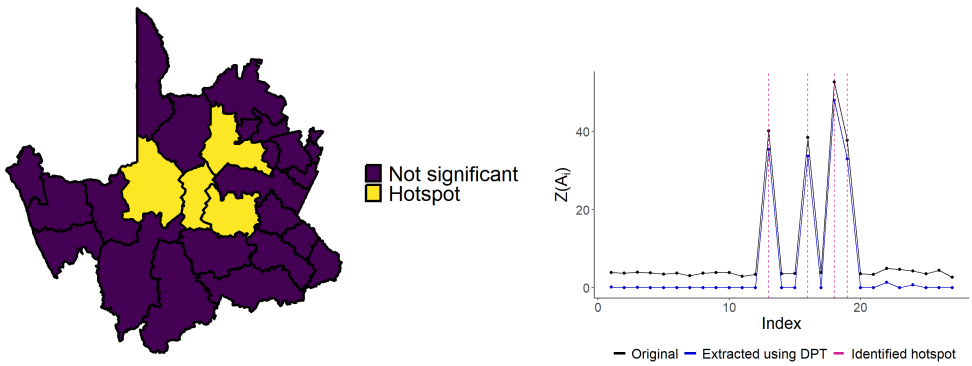
corresponding interval from the multiscale Ht-index is,

$$\lambda = \max_{A_i \in D} LR(A_i) = \max_{A_i \in D} \{38.491; 35.052; 67.117; 33.614\} = 67.117.$$

Figure 6 indicates the test statistic value for each interval by colour on the multiscale Ht-index visualisation from Figure 4. The interval with the maximum test statistic value is considered to contain the salient pulses. In this case, the interval with the maximum test statistic value is $[1; 4.654]$.

2.6 Hotspot identification

After a test statistic value is calculated for each interval from the multiscale Ht-index, the corresponding pulses whose scales are included in the interval are reconstructed according to (4). The hotspot areas are detected as the areas with standardised values of more than 1.96 at a significance level of 0.05. The values of the areas are standardised to have zero mean and a standard deviation of one. The observed values in spatial lattice data are usually the average of the events. Similar to the



(a) (b)

Figure 7. (a) Hotspot areas identified with the proposed algorithm. (b) Original values for each A_i are plotted against the index and partially reconstructed values of the salient pulses are plotted on the same axis.

Getis-Ord statistic, the limiting distribution of the averages is normal (Modiba et al., 2022).

In Figure 7(b), the original values for each A_i are plotted against the index of the area with the partially reconstructed values of the identified salient pulses on the same axis. This indicates the ability of the algorithm (before identifying the hotspots) to extract important features and information from data while smoothing noise. The identified hotspots are also shown in Figure 7(a). Recall from Figure 1 that the areas where the hotspots are identified, match where the hotspot was simulated.

In summary, the proposed hotspot detection method starts by first decomposing the irregular spatial lattice data, using the DPT algorithm, into pulses. The pulses from the decomposition can be partially reconstructed to detect salient features in the data. The multiscale Ht-index is used to divide the scales of the extracted pulses into intervals. If the scales included in an interval are $[e_l; e_u]$, then the partial reconstruction is, $f^p_{[e_l;e_u]}(D) = \sum_{n=\lceil e_l \rceil}^{\lfloor e_u \rfloor} D_n(f)$. For each interval from the multiscale Ht-index, the pulses with the relevant scales are partially reconstructed and (6) is applied. The interval resulting in the maximum test statistic value is considered to contain the salient pulses. A partial reconstruction is done again for these pulses whereafter the hotspot areas are detected as the areas with standardised values of more than 1.96.

3. Simulation study

To determine the accuracy and robustness of the proposed hotspot detection method, we perform a simulation study. In the simulation study, irregular spatial lattice data are simulated such that some of the areas within the spatial domain are known to be hotspot areas. The results for the proposed method are compared to the local Getis-Ord statistic as it is the most popular method in literature and the gold standard for hotspot detection in GIS software such as ArcGIS[4]. The local Getis-Ord

[4] https://pro.arcgis.com/en/pro-app/2.9/tool-reference/spatial-statistics/h-how-hot-spot-analysis-getis-ord-gi-spatial-stati.htm. Accessed on: 8 March 2023.

(Getis and Ord, 1992) statistic is defined as,

$$G_i^* = \frac{\sum_{j=1}^{N} w_{ij}x_j - \bar{x}\sum_{j=1}^{N} w_{ij}}{s\sqrt{\frac{N\sum_{j=1}^{N} w_{ij}^2 - \left(\sum_{j=1}^{N} w_{ij}\right)^2}{N-1}}},$$

where w_{ij} is the neighbourhood relationship between A_i and A_j as defined in (1), $x_j = Z(A_j)$, the value observed at A_j, and

$$s = \sqrt{\frac{1}{N}\sum_{j=1}^{N} x_j^2 - \bar{x}^2} \quad \text{and} \quad \bar{x} = \frac{1}{N}\sum_{j=1}^{N} x_j.$$

3.1 Simulation design

The simulations are conducted by simulating random points over the entire spatial domain from a Poisson Process with $N \times 50$ points, where N is the number of areas in the spatial domain. At each iteration, areas are selected and more points are simulated in those areas such that it can be considered a hotspot area. The simulated points are aggregated such that the total number of points in A_i is the value of the spatial process observed in the corresponding area, $Z(A_i)$. The proposed method as well as the local Getis-Ord statistic is applied to each of the simulations. The detected hotspots are then recorded and compared to where the hotspots are known to be simulated.

The simulation study is performed on two spatial domains, one with 20 areas and one with 52 areas. Variation in the simulation study is obtained by varying the size of the hotspot and the number of areas included in the hotspot. The hotspot is created by simulating either 5%, 30%, or 60% more than the total number of points to the randomly chosen hotspot areas. The choice of the number of hotspot areas to include in the simulation study is 1) one area to illustrate that the proposed method is able to detect small and single hotspot areas (to indicate the sensitivity of the method), 2) two neighbouring areas to illustrate the proposed method's ability to detect larger neighbouring areas, and 3) two separate areas for illustration on disjoint areas. To select the areas where more points are simulated to form a hotspot area, an area is selected completely at random. For the first choice, only the selected area is used. In the case of the second choice, an adjacent region to the selected area is also selected. Lastly, for the third choice, another region is selected that is not adjacent to the first selected region. For each combination of spatial domain, hotspot size, and number of hotspots, 500 simulations are performed.

3.2 Simulation results

Examples of some simulations on the spatial domain with 20 areas are shown in Figure 8. The $Z(A_i)$ are plotted on a map with red and green stars indicating where either the proposed method or local Getis-Ord detected hotspots. This can be seen in Figure 8(a)–(c). The original values, $Z(A_i)$, as well as the partially reconstructed values, $Z^p(A_i)$, are plotted against the index of the areas on the same axis. The partial reconstruction shown is for the pulses whose scales are included in the interval resulting in the maximum test statistic value from the multiscale Ht-index. Straight vertical lines show the areas where the hotspots are detected. This is done for both the proposed method (see Figure 8(d)–(f)) as well as local Getis-Ord (see Figure 8(g)–(i)).

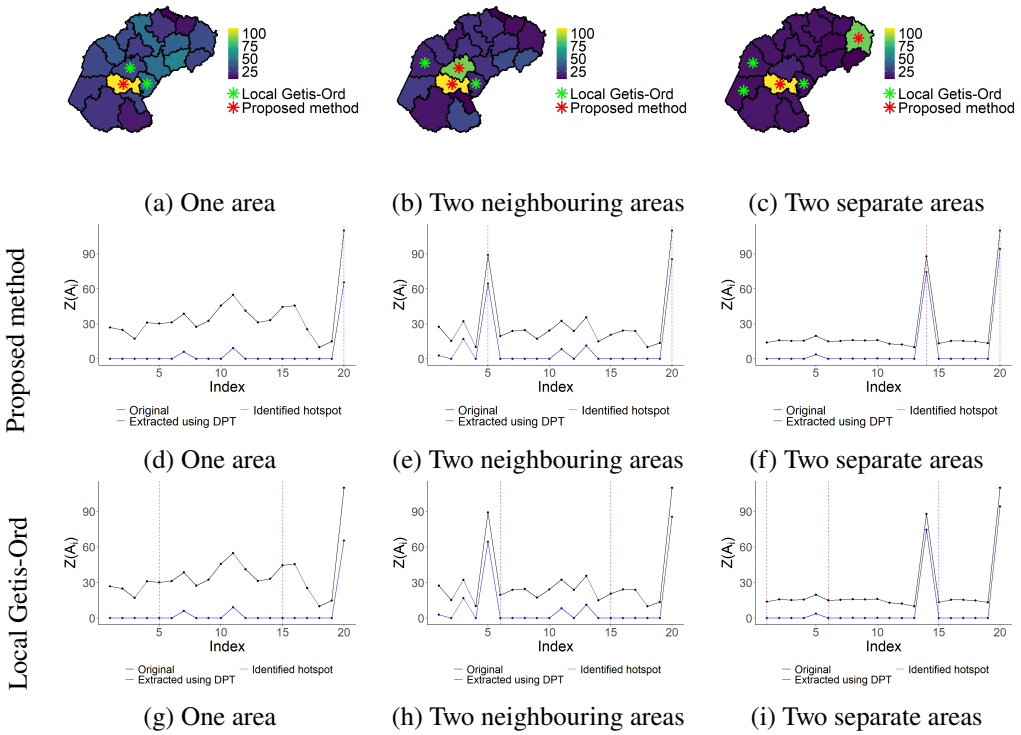


Figure 8. Examples of the simulations on the spatial domain with 20 areas with (a) one, (b) two neighbouring, and (c) two separate simulated hotspots. The stars on the maps indicate where either the proposed method or local Getis-Ord detects hotspots. The original values, $Z(\cdot)$, and the partially reconstructed values, $Z^P(\cdot)$, of the salient pulses are plotted against the index of the region. Identified hotspot areas are indicated with vertical lines for the proposed method ((d)–(f)) as well as local Getis-Ord ((g)–(i)).

The accuracy of the proposed method, especially on small-scale hotspots and spatial domains with a small number of areas, is illustrated in the example shown in Figure 8. The proposed method is able to accurately detect hotspots present in spatial lattice data while the local Getis-Ord statistic does not identify the hotspots correctly. A similar situation is presented in Figure 9 where simulations on the spatial domain with 52 areas are shown. Figure 9(a)–(c) shows the plotted $Z(A_i)$ values on a map with red and blue stars indicating where either the proposed method or local Getis-Ord detected hotspots. In Figure 9(d)–(f), the original values as well as the partially reconstructed values are plotted against the index of the area. Straight lines indicate where the proposed method detected the hotspots while Figure 9(g)–(i) show where local Getis-Ord detects the hotspots.

From Figure 8 and Figure 9, it can be seen that the proposed method has the ability to identify the hotspots in single areas while local Getis-Ord tends to indicate multiple neighbouring areas. The proposed method is more accurate and robust than the local Getis-Ord statistic, especially with small-scale hotspots.

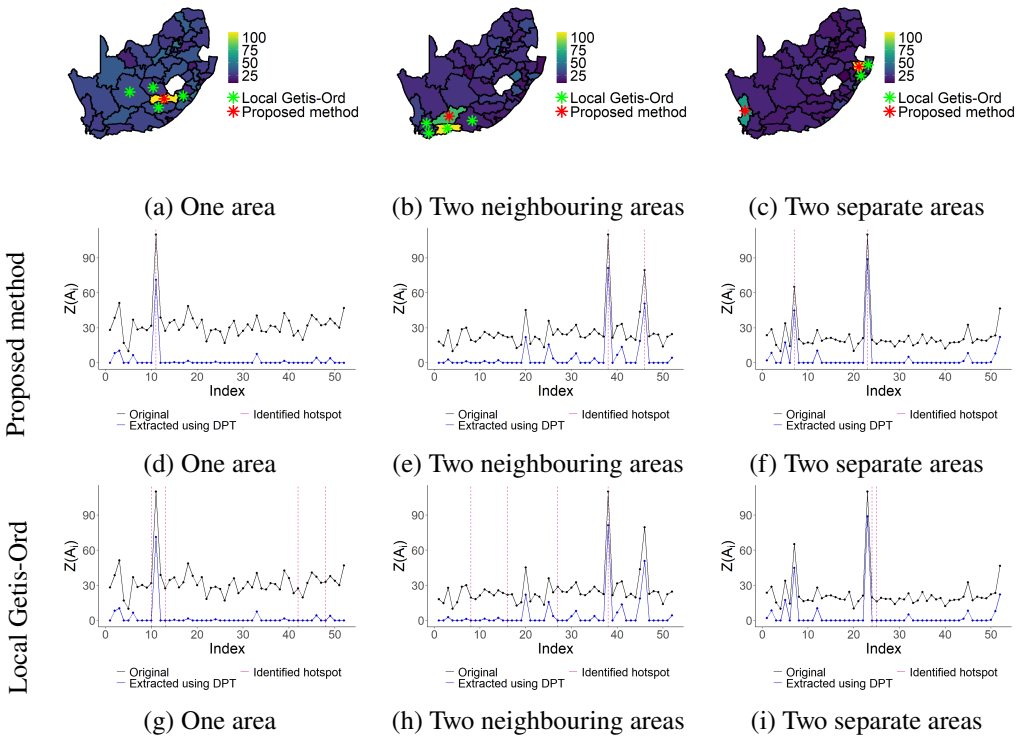


Figure 9. Examples of the simulations on the spatial domain with 52 areas with (a) one, (b) two neighbouring, and (c) two separate simulated hotspots. The stars on the maps indicate where either the proposed method or local Getis-Ord detects hotspots. The original values, $Z(\cdot)$, and the partially reconstructed values, $Z^p(\cdot)$, of the salient pulses are plotted against the index of the area. Identified hotspot areas are indicated with vertical lines for the proposed method ((d)–(f)) as well as local Getis-Ord ((g)–(i)).

The results for each combination of number of hotspots and size from the simulation study are shown in Table 1. We considered performance measures such as true positive rate[5] (TPR), true negative rate[6] (TNR), false positive rate[7] (FPR), false negative rate[8] (FNR), and accuracy[9]. It is clear from the table that the proposed method outperforms the local Getis-Ord statistic on the simulated data. Although the accuracy is consistent between the number of hotspots, the true positive rate is lower for when two hotspots are present than for when only one hotspot is present.

It is interesting to note that the TNR for local Getis-Ord is, although lower than the proposed method, still high. The reason for this is the small number of hotspot regions relative to the total regions in the domain which causes the inflation of the TNR.

[5] $TPR = TP/(TP+FN)$, where TP: True positive and FN: False negative.
[6] $TNR = TN/(TN+FP)$, where TN: True negative and FN: False positive.
[7] $FPR = FP/(FP+TN)$, where FP: False positive and TN: True negative.
[8] $FNR = FN/(FN+TP)$, where FN: False negative and TN: True positive.
[9] $Accuracy = (TN+TP)/(TN+TP+FN+FP)$.

Table 1. Results from the simulation study. The true positive rate (TPR), true negative rate (TNR), false positive rate (FPR), false negative rate (FNR), as well as the accuracy are shown for both the proposed method and the local Getis-Ord statistic. The results for the simulation study are shown for the different number of hotspots and the size of the hotspot separately. An overall performance metric is also calculated on the data combined.

		Proposed method			Local Getis-Ord statistic		
Number of hotspots	Size	TPR	TNR	Accuracy	TPR	TNR	Accuracy
		FPR	FNR		FPR	FNR	
One	5%	**0.999**	**0.998**	**0.998**	0.046	0.952	0.927
		0.002	**0.001**		0.048	0.954	
	30%	**0.993**	**0.998**	**0.998**	0.045	0.949	0.924
		0.002	**0.007**		0.051	0.955	
	60%	**0.997**	**0.998**	**0.998**	0.041	0.952	0.927
		0.002	**0.003**		0.048	0.959	
Two neighbouring	5%	**0.839**	**0.998**	**0.989**	0.376	0.939	0.909
		0.002	**0.161**		0.060	0.624	
	30%	**0.827**	**0.998**	**0.989**	0.394	0.939	0.909
		0.002	**0.173**		0.061	0.605	
	60%	**0.824**	**0.998**	**0.988**	0.375	0.939	0.908
		0.002	**0.176**		0.061	0.625	
Two separately	5%	**0.774**	**0.999**	**0.988**	0.004	0.954	0.901
		0.001	**0.226**		0.046	0.996	
	30%	**0.774**	**0.999**	**0.987**	0.001	0.955	0.902
		0.001	**0.226**		0.045	0.999	
	60%	**0.765**	**0.999**	**0.987**	0.002	0.954	0.901
		0.001	**0.235**		0.046	0.998	
Overall		**0.840**	**0.999**	**0.991**	0.162	0.948	0.912
		0.001	**0.159**		0.052	0.838	

4. Application

4.1 COVID-19 cases in South Africa

The COVID-19 pandemic hit the world hard in 2020. The first case of COVID-19 in South Africa was confirmed on 5 March 2020 and the entire country went into lockdown on 26 March 2020. Local governments and decision-makers around the world kept a close eye on the cases to be able to implement several lockdown measures to contain the spread of the virus. In such a case, the use of hotspot detection methods is useful to identify the areas at risk.

For this example, the total number of COVID-19 cases is considered at municipal level in South Africa for April – July 2020[10]. South Africa consists of nine provinces and 278 municipalities. The

[10] The right to use this data was approved by the NAS ethics committee NAS317/2022.

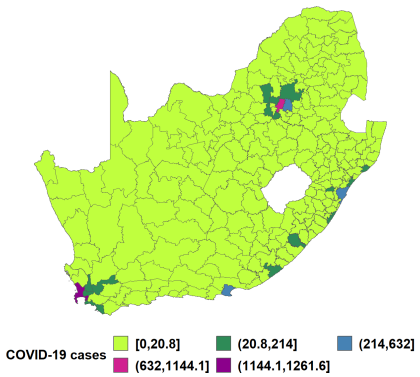
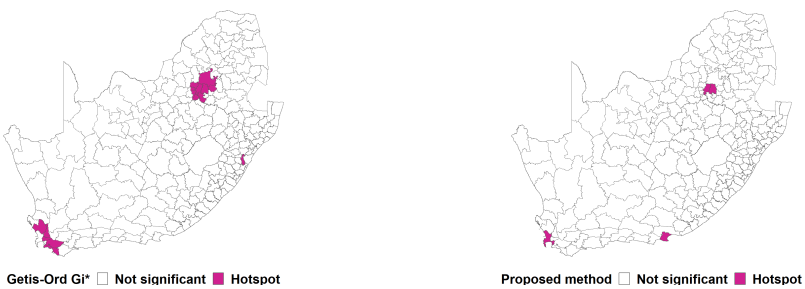


Figure 10. Total COVID-19 cases for April – July 2020 at municipal level, standardised by the area of the region.



(a) Local Getis-Ord (b) Proposed algorithm

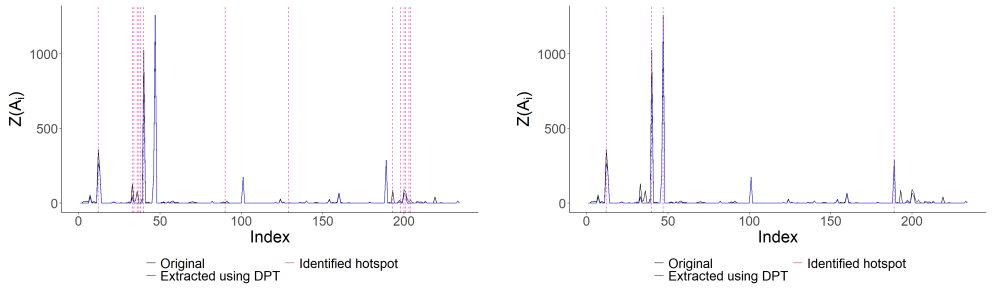
Figure 11. Detected hotspots shown on a map using (a) the local Getis-Ord statistic and (b) the proposed method.

observed values are standardised by dividing the observed counts by the area of the region. This is shown in Figure 10. The proposed method as well as the local Getis-Ord statistic is then applied to the standardised values.

The results for both the local Getis-Ord statistic and the proposed algorithm are shown in Figure 11(a) and Figure 11(b), respectively. It should be noted that the local Getis-Ord statistic detects larger areas as hotspots while the proposed method detects smaller, more precise, areas. This can be seen especially in the upper right hotspot areas (Gauteng area of South Africa). Also, take note that local Getis-Ord detects regions as hotspots that, by visually inspecting and comparing to Figure 10, are clearly not hotspots. The same applies to the areas that should clearly be detected as a hotspot but are missed by local Getis-Ord while detected with the proposed algorithm. Focussing some attention on the detected hotspot areas in the lower left of the map (Cape Town region), the area which has the clearest maximum value in the region is not detected by local Getis-Ord as a hotspot but is indeed identified by the proposed method.

The areas highlighted as hotspots in Figure 11(b) using the proposed method, are the metropolitan areas in South Africa. These areas are more likely to be hotspots of COVID-19 cases as the population is denser and the mobility of individuals is higher.

Another proof of concept is shown in Figure 12 where the original values as well as the partially



(a) (b)

Figure 12. The original values and the partially reconstructed values of the salient pulses are plotted against the index of the area. Identified hotspot areas are indicated with vertical lines for (a) local Getis-Ord as well as (b) the proposed method.

reconstructed values are plotted against the index of the area on the same axis. The vertical lines indicate the detected hotspots for the local Getis-Ord statistic (in Figure 12(a)) and the proposed algorithm (in Figure 12(b)). It can be seen that the proposed algorithm indicates the peaks in the data with higher accuracy while the local Getis-Ord statistic has many more false positives. The proposed method is detecting smaller scale hotspots than the local Getis-Ord statistic.

In the case of a pandemic such as COVID-19, the impact of stricter lockdown measures can have a severe impact on the already struggling economy of South Africa. Therefore, wrongly detecting an area as a hotspot should be avoided at all costs.

4.2 Crime in South Africa

South Africa has the third highest crime rate in the world[11]. Hotspot detection methods have been used by decision-makers around the world to deploy resources to contain criminal activity (Chainey, 2014). In this example, the crime cases of three different crime types (property, violent, and sexual crimes) are considered at police precinct level in South Africa[12]. The observed crime counts at each police precinct are standardised by dividing the observed counts by the area of the region.

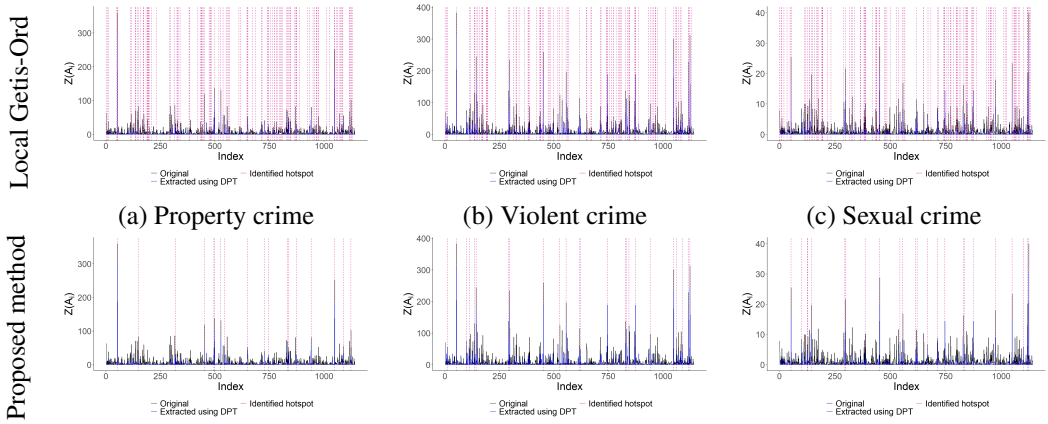
The proposed method as well as local Getis-Ord are applied to each of the crime types separately. Figure 13 displays the detected hotspots of property crimes (local Getis-Ord statistic in Figure 13(a) and proposed method in Figure 13(d)), violent crimes (local Getis-Ord statistic in Figure 13(b) and proposed method in Figure 13(e)) and sexual crimes (local Getis-Ord statistic in Figure 13(c) and proposed method in Figure 13(f)). Similarly to the previous example in Section 4.1, the local Getis-Ord statistic detects larger areas as hotspots while the proposed method detects smaller regions. A larger number of connected areas (a larger hotspot) can be seen to be identified as a hotspot by the proposed method in Figure 13(e) which is an indication of the proposed method's ability to also detect larger scale hotspots.

In Figure 14, the original values and the partially reconstructed values are plotted against the index of the area on the same axis. The vertical lines indicate the detected hotspots for the local

[11] https://worldpopulationreview.com/country-rankings/crime-rate-by-country. Accessed on: 22 February 2023

[12] Data used with permission from Prof Breetzke from the Department of Geography, Geoinformatics & Meteorology, University of Pretoria.

Figure 13. Detected hotspots shown on a map using (a)–(c) local Getis-Ord statistic and (d)–(f) the proposed method.



(a) Property crime (b) Violent crime (c) Sexual crime

(d) Property crime (e) Violent crime (f) Sexual crime

Figure 14. The original values and the partially reconstructed values of the salient pulses are plotted against the index of the area. Identified hotspot areas are indicated with vertical lines for (a)–(c) local Getis-Ord as well as (d)–(f) the proposed method.

Getis-Ord statistic (Figure 14(a–c)) and the proposed algorithm (Figure 14(d–f)). Similarly to the previous example, it can be seen that the proposed algorithm indicates the peaks in the data with higher accuracy while the local Getis-Ord statistic has many more false positives.

5. Discussion and conclusion

In this paper, a hotspot detection method that makes use of the Discrete Pulse Transform to decompose irregular spatial lattice data is proposed. The multiscale Ht-index together with the spatial scan statistic is then used as a measure to quantify the saliency on the extracted pulses to detect significant hotspots.

The proposed method was validated with a simulation study where it was compared to the local Getis-Ord statistic. It was concluded that the proposed method is more accurate and robust than the local Getis-Ord statistic. The proposed method has the ability to detect hotspots with a higher accuracy, precision, and robustness than the local Getis-Ord. One aspect that drew attention in Section 3 and 4 was that the proposed method is able to detect small-scale hotspots, while local Getis-Ord tends to identify multiple connected areas as hotspots.

Although the proposed method works well on the scenarios tested in this paper, the method has not been tested on other lattice structures. The type of examples shown here is what the size of the lattices should look like in applications, but the scalability has not been verified in the simulation study. More variety can be introduced in the simulation study by allowing the size and shape of the cluster to include more possibilities. In future applications, a sensitivity analysis can be performed to test the proposed method's sensitivity to the strength of the hotspot.

In future work, the proposed method can be extended to detect warm spots, cold spots, and cool spots. A warm spot is a hotspot with a lower significance, hence an area of concern but lower concern than a hotspot area. Cold spots are the opposite of hotspots and cool spots are the opposite

of cold spots[13]. Methods such as Monte Carlo simulation can be explored for the final hotspot detection step. The extension of this method to predict emerging hotspots can be implemented[14]. The prediction of hotspot areas has been investigated in recent publications in both criminology and disease mapping (Adepeju et al., 2016; Chen et al., 2020). In this paper, only the Rook contiguity neighbourhood definition has been considered. In future applications, the influence of different neighbourhood definitions can be investigated when implementing the DPT step of the proposed method. Other neighbourhood definitions that can be considered are distance-based and graph-based methods (Bivand et al., 2013).

Acknowledgements. This work is partially based upon research supported by the South Africa National Research Foundation (NRF) and South Africa Medical Research Council (SAMRC) (South Africa DST-NRF-SAMRC SARChI Research Chair in Biostatistics-Professor Ding-Geng Chen, Grant number 114613 and Grant number 137785). Opinions expressed and conclusions arrived at are those of the author and are not necessarily to be attributed to the NRF and SAMRC.

References

Aamodt, G., Samuelsen, S. O., and Skrondal, A. (2006). A simulation study of three methods for detecting disease clusters. *International Journal of Health Geographics*, **5**, 1–11.

Adepeju, M., Rosser, G., and Cheng, T. (2016). Novel evaluation metrics for sparse spatio-temporal point process hotspot predictions: A crime case study. *International Journal of Geographical Information Science*, **30**, 2133–2154.

Althaf, P., Kannemadugu, H. B. S., and Kumar, K. R. (2022). Hotspot analysis and long-term trends of absorbing aerosol index from dust emissions measured by the Ozone Monitoring Instrument at different urban locations in India during 2005–2018. *Atmospheric Environment*, **272**, 118933.

Anguelov, R. and Fabris-Rotelli, I. (2010). LULU operators and discrete pulse transform for multidimensional arrays. *IEEE Transactions on Image Processing*, **19**, 3012–3023.

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

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2014). *Hierarchical Modeling and Analysis for Spatial Data*. CRC Press, Boca Raton, FL.

Bivand, R. S., Pebesma, E. J., and Gomez-Rubio, V. (2013). *Applied Spatial Data Analysis with R*. Springer, New York, NY.

Bone, C., Wulder, M. A., White, J. C., Robertson, C., and Nelson, T. A. (2013). A GIS-based risk rating of forest insect outbreaks using aerial overview surveys and the local Moran's *I* statistic. *Applied Geography*, **40**, 161–170.

Chainey, S. (2014). *Examining the extent to which hotspot analysis can support spatial predictions of crime*. Ph.D. thesis, University College London.

[13] https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/hot-spot-analysis.htm Accessed on: 10 October 2023.

[14] https://pro.arcgis.com/en/pro-app/latest/tool-reference/space-time-pattern-mining/emerginghotspots.htm. Accessed on: 8 March 2023

Chen, P., Chen, T., and Yuan, H. (2010). GIS based crime risk analysis and management in cities. *In The 2nd International Conference on Information Science and Engineering*. IEEE, 3721–3724.

Chen, Z., Liu, F., Li, B., Peng, X., Fan, L., and Luo, A. (2020). Prediction of hot spot areas of hemorrhagic fever with renal syndrome in Hunan Province based on an information quantity model and logistical regression model. *PLoS Neglected Tropical Diseases*, **14**, e0008939.

Cheng, Z., Zu, Z., and Lu, J. (2018). Traffic crash evolution characteristic analysis and spatiotemporal hotspot identification of urban road intersections. *Sustainability*, **11**, 160.

Cressie, N. A. C. (1993). *Statistics for Spatial Data*. John Wiley & Sons, Hoboken, NJ.

Duczmal, L., Kulldorff, M., and Huang, L. (2006). Evaluation of spatial scan statistics for irregularly shaped clusters. *Journal of Computational and Graphical Statistics*, **15**, 428–442.

Everett, B. I., Fennessy, S. T., and van den Heever, N. (2021). Using hotspot analysis to track changes in the crustacean fishery off KwaZulu-Natal, South Africa. *Regional Studies in Marine Science*, **41**, 101553.

Fabris-Rotelli, I. (2013). *Discrete pulse transform of images and applications*. Ph.D. thesis, University of Pretoria.

Fabris-Rotelli, I. and Stein, A. (2018). Inhomogeneous spatial modelling of DPT pulses for marine images. *Spatial Statistics*, **28**, 257–270.

Fabris-Rotelli, I. and Stein, A. (2020). Use of fractals to measure anisotropy in point patterns extracted with the DPT of an image. *Spatial Statistics*, **42**, 100452.

French, J. P., Meysami, M., Hall, L. M., Weaver, N. E., Nguyen, M. C., and Panter, L. (2022). A comparison of spatial scan methods for cluster detection. *Journal of Statistical Computation and Simulation*, **92**, 3343–3372.

Geary, R. C. (1954). The contiguity ratio and statistical mapping. *The Incorporated Statistician*, **5**, 115–146.

Getis, A. and Ord, J. K. (1992). The analysis of spatial association by use of distance statistics. *Geographical Analysis*, **24**, 189–206.

Ishioka, F., Kurihara, K., Suito, H., Horikawa, Y., and Ono, Y. (2007). Detection of hotspots for three-dimensional spatial data and its application to environmental pollution data. *Journal of Environmental Science for Sustainable Society*, **1**, 15–24.

Jana, M. and Sar, N. (2016). Modeling of hotspot detection using cluster outlier analysis and Getis-Ord Gi* statistic of educational development in upper-primary level, India. *Modeling Earth Systems and Environment*, **2**, 1–10.

Jiang, B., Liu, X., and Jia, T. (2013). Scaling of geographic space as a universal rule for map generalization. *Annals of the Association of American Geographers*, **103**, 844–855.

Kadir, T. and Brady, M. (2001). Saliency, scale and image description. *International Journal of Computer Vision*, **45**, 83–105.

Kulldorff, M. (1997). A spatial scan statistic. *Communications in Statistics - Theory and Methods*, **26**, 1481–1496.

Laurie, D. P. (2010). The roadmaker's algorithm for the discrete pulse transform. *IEEE Transactions on Image Processing*, **20**, 361–371.

Lees, B. (2006). The spatial analysis of spectral data: Extracting the neglected data. *Applied GIS*, **2**, 14.

Li, P., Banerjee, S., Hanson, T. A., and McBean, A. M. (2015). Bayesian models for detecting difference boundaries in areal data. *Statistica Sinica*, **25**, 385–402.

Mahara, G., Wang, C., Huo, D., Xu, Q., Huang, F., Tao, L., Guo, J., Cao, K., Long, L., Chhetri, J. K., Gao, Q., Wang, W., Wang, Q., and Guo, X. (2016). Spatiotemporal pattern analysis of scarlet fever incidence in Beijing, China, 2005–2014. *International Journal of Environmental Research and Public Health*, **13**, 131.

Modiba, J., Fabris-Rotelli, I., Stein, A., and Breetzke, G. (2022). Linear hotspot detection for a point pattern in the vicinity of a linear network. *Spatial Statistics*, **51**, 100693.

Moran, P. A. P. (1948). The interpretation of statistical maps. *Journal of the Royal Statistical Society. Series B*, **10**, 243–251.

Rohwer, C. (2006). *Nonlinear Smoothing and Multiresolution Analysis*. Birkhäuser, Basel.

Ruiz, M. O., Tedesco, C., McTighe, T. J., Austin, C., and Kitron, U. (2004). Environmental and social determinants of human risk during a West Nile virus outbreak in the greater Chicago area, 2002. *International Journal of Health Geographics*, **3**, 1–11.

Said, S. N. B. M., Zahran, E. M. M., and Shams, S. (2017). Forest fire risk assessment using hotspot analysis in GIS. *The Open Civil Engineering Journal*, **11**, 786–801.

Serra, J. (1988). Mathematical morphology for Boolean lattices. *In Image Analysis and Mathematical Morphology, Volume II: Theoretical Advances*. Academic Press, New York, NY.

Songchitruksa, P. and Zeng, X. (2010). Getis–Ord spatial statistics to identify hot spots by using incident management data. *Transportation Research Record*, **2165**, 42–51.

Stander, R., Fabris-Rotelli, I., Chen, D. G., and Breetzke, G. (2021). Multiscale decomposition of spatial lattice data for feature detection. *In The 2nd South African Conference for Artificial Intelligence Research*. 123–138.

Stoltz, G. G. (2014). *Roadmakers pavage, pulse reformation framework and image segmentation in the discrete pulse transform*. Master's thesis, University of Pretoria.

Stresman, G., Bousema, T., and Cook, J. (2019). Malaria hotspots: Is there epidemiological evidence for fine-scale spatial targeting of interventions? *Trends in Parasitology*, **35**, 822–834.

Tepanosyan, G., Sahakyan, L., Zhang, C., and Saghatelyan, A. (2019). The application of local Moran's *I* to identify spatial clusters and hot spots of Pb, Mo and Ti in urban soils of Yerevan. *Applied Geochemistry*, **104**, 116–123.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, **46**, 234–240.

Wang, D., Ding, W., Lo, H., Stepinski, T., Salazar, J., and Morabito, M. (2013). Crime hotspot mapping using the crime related factors: A spatial data mining approach. *Applied Intelligence*, **39**, 772–781.

Wang, Y., Lv, W., Wang, M., Chen, X., and Li, Y. (2023). Application of improved Moran's *I* in the evaluation of urban spatial development. *Spatial Statistics*, **54**, 100736.

Wu, T., Walther, B. A., Chen, Y., Lin, R., and Lee, P. (2013). Hotspot analysis of Taiwanese

breeding birds to determine gaps in the protected area network. *Zoological Studies*, **52**, 29.

Zahran, E. M. M., Shams, S., Said, S. N. B. M., Zahran, E. S. M. M., Gadong, B., and Brunei-Muara, B. D. (2020). Validation of forest fire hotspot analysis in GIS using forest fire contributory factors. *Systematic Reviews in Pharmacy*, **11**, 249–255.

Zhang, C., Luo, L., Xu, W., and Ledwith, V. (2008). Use of local Moran's I and GIS to identify pollution hotspots of Pb in urban soils of Galway, Ireland. *Science of the Total Environment*, **398**, 212–221.

Zhu, J., Huang, H., and Reyes, P. E. (2010). On selection of spatial linear models for lattice data. *Journal of the Royal Statistical Society: Series B*, **72**, 389–402.

Manuscript received 2023-04-17, revised 2023-11-06, accepted 2023-12-04.