7

Spatiotemporal Point Pattern Analysis Using Ripley's K Function

Alexander Hohl, Minrui Zheng, Wenwu Tang, Eric Delmelle, and Irene Casas

CONTENTS

7.1 Introduction

Many geospatial phenomena are involved with movement across space, for instance, movement of humans and animals, the dispersal of plants, and the diffusion of infectious disease (Gould 1969; Demšar et al. 2015). Our ability to collect fine-resolution spatiotemporal data with high accuracy has

substantially improved, given the rapid development of geospatial technologies: location-aware smartphones track their owner's daily movements, sensor networks record biophysical variables in real time, and volunteered geographic information is easily accessible online (Kwan and Neutens 2014). While the study of geospatial movement or diffusion phenomena has received increased attention, the analysis of the spatiotemporal data associated with these phenomena remains challenging (Diggle 2013; Goodchild 2013; An et al. 2015). In this study, we aim to investigate the use of Ripley's K function for the analysis of spatiotemporal point patterns to gain insight into this challenge.

The Ripley's K function (Ripley 1976) is a quantitative approach that falls within the domain of spatial and spatiotemporal point pattern analysis. Spatial point pattern analysis is concerned with quantifying the distribution of point events in 2D geographic space (Illian et al. 2008). It has widespread applications, such as in plant ecology (Wiegand and Moloney 2004; Perry, Miller, and Enright, 2006), epidemiology (Gatrell et al. 1996), and criminology (Anselin et al. 2000). Ripley's K function characterizes a given set of points and distinguishes between random, clustered, and regular patterns. However, the use of point pattern analysis for evaluating spatiotemporally explicit phenomena lags behind in the availability of spatiotemporal datasets. It is important to note that spatiotemporal does not equal 3D, due to the orthogonal relationship between space and time (Nakaya and Yano 2010) and due to the peculiarity of the temporal dimension that clearly distinguishes it from the 2D spatial dimensions (Aigner et al. 2007). This further complicates the analysis of spatiotemporal point patterns.

The objective of this work is to investigate the capability of Ripley's K function-based point pattern analysis for the study of dynamic geospatial phenomena. Specifically, we focus on the combined use of global and local forms of Ripley's K function and present a case study of dengue fever in the city of Cali, Colombia to illustrate the benefits of this methodology.

The remainder of the chapter is organized as follows: Section 7.2 provides the background about global Ripley's K function, followed by Section 7.3, which discusses the temporal extension of the K function and its local variant, as well as specific details about our own implementation of the local Ripley's K function for spatiotemporal point pattern analysis. In Section 7.4, we present a case study where we apply Ripley's K function, followed by results (Section 7.5) and conclusions (Section 7.6).

7.2 Background: Global Ripley's K Function for Spatial Point Pattern Analysis

Different approaches exist to evaluate the level of spatial clustering among point events (Bailey and Gatrell 1995; Delmelle 2009). For instance, the

quadrant analysis essentially counts the number of events within each cell (quadrant) of a grid imposed on a study area. The results are compared with the expected frequency of occurrence if the mechanism generating those events was a homogeneous Poisson process (corresponding to point patterns that exhibit complete spatial randomness—i.e., CSR). Despite its ease of implementation, quadrant analysis is an area-based approach that aggregates original point events into quadrant counts, which makes this approach sensitive to the design of quadrants (e.g., shape and size). The nearest neighbor statistic (Diggle, Besag, and Gleaves 1976) alleviates this problem by testing whether point events are closer together (or farther apart) than expected under CSR based on nearest-neighbor distance. This distance-based approach suffers from some limitations, most notably that clustering is typically detected at a relatively small scale, and that distance among events is the only parameter governing the statistic. While the nearest-neighbor approach uses distances only to the closest events and hence only considers the smallest scales of patterns, Ripley's K Function provides a superior alternative in that it evaluates point patterns at different scales (Ripley 1976). By comparing the spatial pattern of the observed data points to simulated data, the K function can indicate for each of the scales evaluated whether the observed point pattern follows a random, clustered, or regular configuration.

Ripley's K function is a statistical approach computed on a set of point events distributed in n-dimensional space, and estimates the second-order property (variance) exhibited by the data. It takes into account (1) the number and (2) distance between the point events, and allows for quantifying how much the observed pattern deviates from randomness at multiple spatial scales (Bailey and Gatrell 1995; Dixon 2013). The theoretical K function, given a set of point-events S, is calculated by dividing E, the number of events that are expected to fall within distance d, by the intensity λ of S (first-order property) as in:

$$K(d) = E(d)/\lambda \tag{7.1}$$

Equation 7.1 is computed by centering a circle of radius d on each sampling point and counting the number of neighboring events that fall inside it. In this case, the number and locations of the sampling points coincide with the event locations. Dividing the total number of events n by the area of the circle πd^2 results in estimated intensity λ. Ripley's K function is the cumulative distribution of observed point events S with increasing distance. It is expected that $K(d) = \pi d^2$ if the point distribution conforms to CSR, $K(d) > \pi d^2$ if the points cluster within distance d, and $K(d) < \pi d^2$ if the data exhibit a regular pattern. The K function is a second-order analysis of point patterns usually in a two-dimensional space (Haase 1995; Dixon 2013). Second-order effects are caused by the spatial dependence in the process. In essence, Ripley's K function approach uses a circular search window (h) around each event (i) and counts how many other events are observed in that window. The window

then moves to the next event until all events (n events) in the study area are visited. The process is repeated at different spatial scales. Specifically,

$$K(h) = (A/n^2) * \sum_{i}^{n} \sum_{j}^{n} (I_h(d_{ij})/w_{ij}) \tag{7.2}$$

Equation 7.2 evaluates the structural characteristics of a given set of events, where d_{ij} is the distance between events i and j, and A is the size of the study region. The term w_{ij} is a factor to correct for edge effects. The K function is potentially biased as edge effects arise when circles intersect the boundary of the study region. There are different methods to deal with edge effects of the K function, which have been thoroughly studied (Yamada and Rogerson 2003). $I_h(d_{ij})$ is an indicator function defined by Equation 7.3:

$$I_h(d_{ij}) = 1 \quad \text{if} \quad d_{ij} \leq h, \quad 0 \quad \text{otherwise} \tag{7.3}$$

The K function increases as distance h becomes larger. To statistically test whether the observed point pattern follows a regular, clustered, or random pattern, the K function is evaluated for a large number (M) of Monte Carlo simulations. For each simulation, a number (n) of events are generated (e.g., randomly) within the study area. If the observed K value is larger than the upper simulation envelope, spatial clustering for that distance is statistically significant. Observed K values smaller than the lower simulation envelope show that point patterns exhibit regularity that is statistically significant for the corresponding distance. K(h) is then evaluated against distance (h) to identify the scales at which the point pattern follows a regular, clustered, or random pattern. For a given value of h, if the K function is above, between, or below the upper and lower envelopes, the point pattern is clustered, random, or regular, respectively.

The K function can be transformed to the L function using Equation 7.4 to obtain constant variance with respect to a benchmark of zero, which facilitates the comparison of L values across all h:

$$L(h) = (K(h)/\pi)^{1/2} - h \tag{7.4}$$

where $L(h) = 0$ if the pattern conforms to CSR, $L(h) > 0$ if clustered, and $L(h) < 0$ for regular patterns.

Recent methodological advancements of Ripley's K function improve the ability to find the appropriate scale of clustering by computing K for distance increments (Tao, Thill, and Yamada 2015). In addition, the K function was adapted for network-constrained data (Yamada and Thill 2007) which violate the planar space assumption that underlies many spatial point pattern analysis methods. Although the K function is popular because it

evaluates levels of clustering at different scales, its computation is time-consuming, especially for large datasets. Recent work has underscored the capability of high performance computing, for instance through the use of Graphics Processing Units (GPU) for its acceleration (Tang, Feng, and Jia 2015).

7.3 Ripley's K Function for Spatiotemporal Point Pattern Analysis

In this study, we investigate the use of Ripley's K function in the analysis of spatiotemporal point patterns. We focus on the combination of global and local forms of Ripley's K function. While the former evaluates the spatiotemporal characteristics of a point pattern at the aggregated level (i.e., the entire dataset), the local form of Ripley's K function quantifies the characteristics of the point pattern, as well as its deviation from what would be expected, locally. In this section, we discuss in detail global and local forms of Ripley's K function for spatiotemporal point pattern analysis.

7.3.1 Global Ripley's K Function for the Analysis of Spatiotemporal Point Pattern

Tests for spatial patterns fail at evaluating the dynamics of the point process. When point events have a temporal attribute, we can investigate whether two events are space and time dependent, suggesting the presence of a space-time link. There are several techniques to evaluate patterns among spatiotemporal point events. The Knox test for space-time interaction evaluates the presence of a space-time cluster at given spatial and temporal distances (Knox 1964). Knox's test method is limited due to its arbitrary definition of closeness and the critical distance does not account for population heterogeneity (Jacquez 1996). The Mantel test (Mantel 1967) incorporates the notion of distance decay in which nearby pairs of events are more important than distant pairs. Jacquez's k-Nearest Neighbor k-NN statistic (Jacquez 1996) addresses the weaknesses of the Knox and Mantel statistics by counting the number of pairs of events that are nearest neighbors in both space and time.

Equivalent to the purely spatial K function (discussed above), the space-time Ripley's K function is a global statistic computed on the entire set of space-time point events. The theoretical space-time K function, given a set of point events S, is calculated by dividing E, the number of events that are expected to fall within spatial distance d and temporal distance t, by the intensity λ of S (first-order property):

$$K(d,t) = E(d,t) / \lambda \tag{7.5}$$

Equation 7.5 characterizes the pattern of S within the space-time cube framework (Nakaya and Yano 2010), where a cylinder of base πd^2 and height t is centered on each sampling point to compute the number of events falling within. Again, the sampling points are equal to the event locations. Then the total number of events n is divided by the volume of the irregular prism formed by the study area/period, which results in intensity λ. Space-time Ripley's K function is the cumulative distribution of observed point events S with increasing space and time distance. It is expected that $K(d,t) = \pi d^2 t$ if the point distribution conforms to complete spatiotemporal randomness (CSTR), $K(d,t) > \pi d^2 t$ if the points cluster within spatial and temporal distance d and t, and $K(d,t) < \pi d^2 t$ for regular space-time patterns. Using Equation 7.6, the space-time K function is formulated as (Bailey and Gatrell 1995):

$$K(h,t) = ((L * R)/n^2) * \sum_{i}^{n} \sum_{j}^{n} (I_{h,t}(d_{ij}, t_{ij})/w_{ij}) \tag{7.6}$$

where t_{ij} is the time that separates two events i and j. d_{ij} is the distance between events i and j. L denotes the area of the study region and R is the duration of the study period. The product of L and R results in the volume of the irregular prism that is formed by the study area (base) and the study period (height). $I_{h,t}(d_{ij},t_{ij})$ is an indicator function defined in Equation 7.7:

$$I_{h,t}(d_{ij}, t_{ij}) = 1 \quad \text{if} \quad d_{ij} \leq h \quad \text{and} \quad t_{ij} \leq t, 0 \quad \text{otherwise} \tag{7.7}$$

Larger time t and distance h intervals will contribute to an increase in the space-time K function. For the case that no space-time interaction exists, Equation 7.6 becomes the product of the spatial and temporal K functions K(h)*K(t). Testing for space-time dependence is achieved by subtracting K(h)*K(t) from the combined space-time I function K(h,t). Methods for handling edge effects of the space-time K function have been studied by Gabriel (2014).

The space-time Ripley's K function is transformed to the space-time L function by Equation 7.8:

$$L(h,t) = (K(h,t)/\pi t)^{1/2} - h \tag{7.8}$$

where $L(h,t) = 0$ under CSTR, $L(h,t) > 0$ for clustered patterns, and $L(h,t) < 0$ for regular patterns.

7.3.2 Local Ripley's K Function for the Analysis of Spatiotemporal Point Patterns

While Ripley's K evaluates the spatial pattern at the global level (i.e., indicating whether clustering is present in the entire study area or not), the same

measure can be considered in its local form to pinpoint where the clustering actually occurs (Getis and Franklin 1987):

$$K_i(h) = (A/n) * \sum_{j}^{n} (I_h(d_{ij})/w_{ij}) \tag{7.9}$$

Here, the local K function is evaluated at each sampling point i, which either is part of (1) a regularly spaced grid drawn over the study area, (2) the events themselves, or (3) a number of random points. The indicator function I is equivalent to Equation 3. Several meaningful extensions to local Ripley's K function have been suggested, such as the local K function for network-constrained space to study transportation-related cases (Okabe and Yamada 2001; Yamada and Thill 2007), as well for characterizing patterns in flow data, thereby upgrading the classic hot spot detection paradigm to the stage of "hot flow" detection (Tao and Thill 2016).

Equivalent to the purely spatial case, the local K function for spatiotemporal point patterns identifies the location and time of clusters within the study area/period. The local space-time K function is evaluated at any sampling point i:

$$K_i(h,t) = ((L * R)/n^2) * \sum_{j}^{n} (I_{h,t}(d_{ij}, t_{ij})/w_{ij}) \tag{7.10}$$

Using Equation 7.10 for each sampling point i, we can estimate the local level of space-time clustering and its statistical significance using Monte Carlo simulations. Further, we can identify the scale at which space-time clustering is the greatest. This information can be very valuable when conducting spatial analysis over a non-homogeneous population of events. Despite these attractive outcomes, the local version of the space-time K function is computationally very demanding, and the execution time depends on (1) the number of data points, (2) the number of sampling points, and (3) the number of spatial and temporal bandwidths, for which K is computed.

7.4 Case Study

To gain insights into the mechanisms of spatiotemporal point pattern analysis, we now illustrate our implementation of the global and local Ripley's K function spatiotemporal explicit set of dengue fever cases in Colombia for the years 2010–2011.

7.4.1 Study Area and Data

The city of Cali is located in the southwest of Colombia. It is the third largest metropolitan area in the country with a total population of around 2.3 million and a population density of 4140/km^2 in 2013 (Cali 2014). Cali experiences two rainy seasons: the first from April to July and the second from September to December. Located at approximately 1000 m above sea level, it has an average temperature of 26°C and an average precipitation of 1000 mm over most of the metropolitan area (Cali 2014). The city is administratively divided into 22 communes covering 120.9 km^2, and composed of 340 neighborhoods (see Figure 7.1). A commune is a grouping of neighborhoods based on homogeneous demographic and socioeconomic characteristics. Neighborhoods are classified using a stratification system composed of six classes, one being the lowest and six the highest. The strata are developed by evaluating the type of housing, urban environment, and context. The city, as in most colonial cities in Latin America, grew from its central core, following the city spine, and toward the periphery. Peripheral neighborhoods are typically characterized

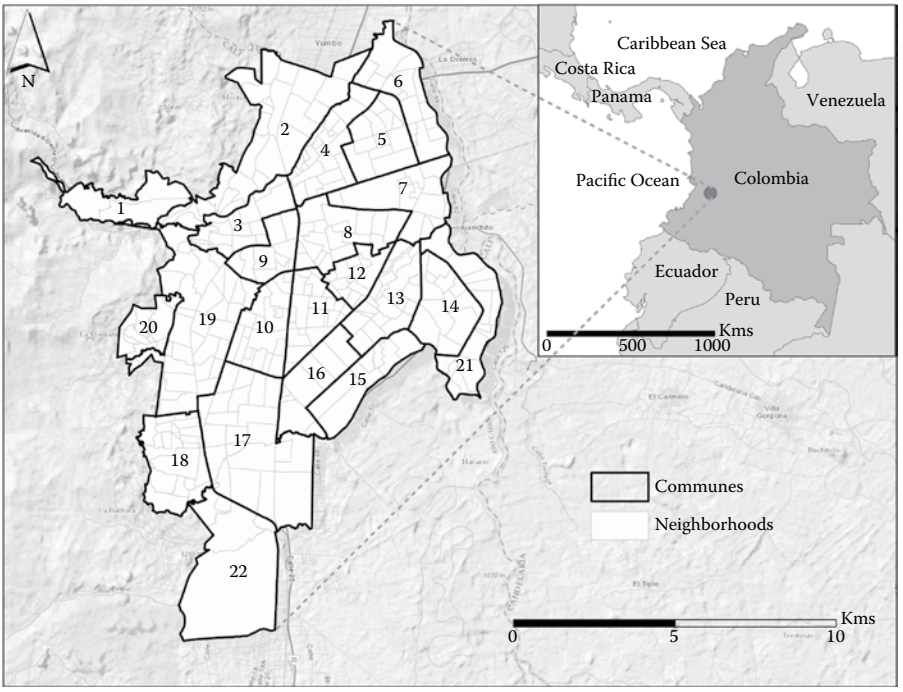

FIGURE 7.1
The city of Cali, Colombia. (Esri, HERE, DeLorme, Intermap, increment P Corp., GEBCO, USGS, FAO, NPS, NRCAN, GeoBase, IGN, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), swisstopo, MapmyIndia, © OpenStreetMap contributors, and the GIS User Community.)

by high density and low income since they have been the result of squatter settlements and poor urban planning (Restrepo 2011).

We use a dataset of dengue fever cases within the city of Cali in this study. The data are extracted from the "Sistema de Vigilancia en Salud Pública (SIVIGILA)" (English: Public Health Surveillance System) for the city of Cali for the years 2010 and 2011. The SIVIGILA system has as a main responsibility to observe and analyze health events with the objective of planning, follow-up, and evaluation of public health practices (Colombia 2017). Reported cases of dengue fever are entered into the system daily. Each case includes personal information about the patient such as their home address and when they were diagnosed. A total of 11,056 cases were geocoded to the closest intersection to guarantee a level of privacy for both years. There were 9606 cases in 2010 and 1562 in 2011. The difference in the number of cases is explained by the fact that 2010 was identified as an epidemic year (Varela et al., 2010).

7.4.2 Analysis

7.4.2.1 Global Space-Time K Function

Since the epidemiological interest is to find clusters of disease occurrence, we evaluated the magnitude of space-time clustering within the dengue fever dataset (n = 11,056) by computing the global space-time Ripley's K and corresponding L functions (see Section 3.1). We used spatial bandwidths from 50 m to 1000 m in 50 m increments and temporal bandwidths from 0 to 14 days in 1-day increments. Using Equation 7.11, we assessed statistical significance of the observed K function by comparison with 100 population-adjusted random simulations and finding the spatial and temporal scales at which the difference between the observed function and the upper simulation envelope (noted as $L_{diff\_upper}(h,t)$) was maximal (also see Hohl et al. 2016):

$$L_{diff\_upper}(h,t) = L_{obs}(h,t) - L_{upper\_envelope}(h,t) \tag{7.11}$$

where $L_{obs}(h,t)$ is the observed L value and $L_{upper\_envelope}(h,t)$ represents the L value of the upper simulation envelope at spatial bandwidth h and temporal bandwidth t.

7.4.2.2 Local Space-Time K Function

Once we determined the presence of clusters in the dengue fever dataset, we illustrated the locations and times at which the clusters may occur by computing the local space-time Ripley's K function. We imposed a regularly spaced grid of sampling points on our study area/period using a space-time resolution of 250 m and 7 days. This results in a total of 202,755 sampling points at which the local space-time K functions were evaluated (although more accurate results can be obtained at a finer scale, e.g., 100 m and 1 day,

estimating local space-time clusters every 250 m and 7 days is computationally more accessible). Equivalent to our estimation of the global space-time K function (see Section 4.2.2), we used spatial bandwidths of 50 m to 1000 m in increments of 50 m and temporal bandwidths of 0–14 days in 1-day increments at each sampling point. Again, using Equation 11, we assessed statistical significance of the observed local K function by comparison with the upper simulation envelope of 100 population-adjusted Monte Carlo simulations. To illustrate the effects of scale on space-time point pattern, we show significantly clustered sampling locations at two different scales by drawing a point cloud within the space-time cube (Delmelle et al. 2014): (1) 500 m and 7 days, (2) 750 m and 10 days.

7.4.2.3 Global Space-Time K Function of Local Settings

For illustration purposes, we assess the magnitude and statistical significance of space-time clustering at various scales by selecting three distinct locations from the space-time grid of sampling points (see section 4.2.2). Each of the three locations is representative of a particular space-time pattern. We chose Location 1 in the center of the dengue fever cluster in the south-western part of the city during the first half of 2010. Location 2 is the same as Location 1, but has a much later time stamp, during which the infectious outbreak is in its declining stage. It can be seen that Locations 1–2 are sites where the virus is present throughout the endemic period. It is a constant focal point of infection for more than 150 days. This information is valuable to health authorities in order to target the location to stop the spread of the disease. This area corresponds to a military base where the municipality spraying cycles are not as regular as in other areas in the city. Location 3 lies in the eastern part of the city, which never exhibits a clustered pattern during the entire study period. Their space-time coordinates (x, y, t), using the Bogota Transverse Mercator coordinate system and Julian date [0–730], are: Location 1: (1,058,498.1, 864,811.9, 35); Location 2: (1,058,498.1, 864,811.9, 210); Location 3: (1,064,998.1, 870,311.9, 35). For each of the three locations, we identified surrounding dengue fever cases within distance 1000 m and 14 days, and computed the global space-time K and L functions for this local setting (the same way as we compute the global K and L functions for the entire study area/period as in Section 4.2.1). To distinguish between clustered, random, and regular space-time patterns, we compared the observed L functions with an upper simulation envelope from the population adjusted simulations using Equation 11, as well as a lower simulation envelope using Equation 7.12:

$$L_{diff\_lower}(h,t) = L_{obs}(h,t) - L_{lower\_envelope}(h,t) \tag{7.12}$$

where $L_{obs}(h,t)$ is the observed L value and $L_{lower\_envelope}(h,t)$ represents the L value of the lower simulation envelope at spatial bandwidth h and temporal bandwidth t.

7.4.2.4 Implementation

All programs were written using Python and R (package: stpp) and we used a high performance computing cluster with 32 nodes connected through an infiniband network switch (Pfister, 2001) to accelerate the spatiotemporal point pattern analysis using Ripley's K function. Each computing node of the high performance computing cluster has 12 CPUs and 12 GBs of memory, in total 384 CPUs (Intel Xeon processor with a 2.67 GHz clock speed). Similar to Delmelle et al. (2014), we used Voxler, an interactive 3D modeling environment (Golden Software, Colorado), for the visualizations of the local space-time K functions.

7.5 Results

7.5.1 Global Space-Time K Function

Figure 7.2 shows the difference between L values of the observed data and the upper simulation envelope of 100 population-adjusted Monte Carlo

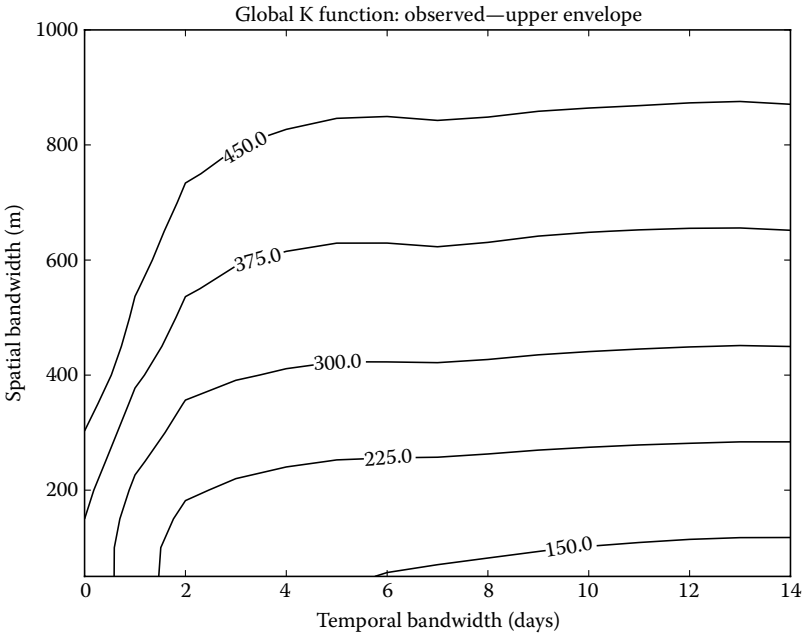

FIGURE 7.2
Global space-time K function. Absolute difference between observed L values and upper simulation envelope.

simulations. Note that we did not account for edge effects in our case study, therefore setting parameter w_{ij} to 1 in all calculations (same for the K function analysis of local settings). Observed L values greater than simulated ones (positive difference) indicate clustering at the corresponding scale: the greater the difference, the stronger the clustering is. Figure 7.2 shows positive values across all bandwidths, with the overall trend of stronger clustering for larger spatial bandwidths. However, the magnitude of the difference in clustering decreases with increasing temporal bandwidth, suggesting that dengue fever cases tend to occur shortly after one another, but do not exhibit strong temporal clustering beyond a week. Thus, the change in clustering intensity is mainly driven by the spatial scale, meaning that, as opposed to changing the temporal bandwidth, the difference becomes larger when increasing the spatial bandwidth.

7.5.2 Local Space-Time K Function

We estimate the local space-time K function at 202,764 regularly spaced grid points (250 m, 7 days intervals). For each grid point, we report the absolute difference in L values of the observed data and the upper simulation envelope of 100 population-adjusted Monte Carlo simulations at different spatial and temporal bandwidths.

Figure 7.3 illustrates a map that visualizes the difference in L-values for each grid point (or voxel, volumetric pixel), using a spatial and temporal bandwidth of 500 m and 7 days (from two different perspectives: southeast and northwest). Negative values (where observed counts are less than expected) are not shown on the map. Colored dots denote regions where the number of observed cases is greater than what is expected; the magnitude of this difference is illustrated with tones of red (darker red dots are on the end

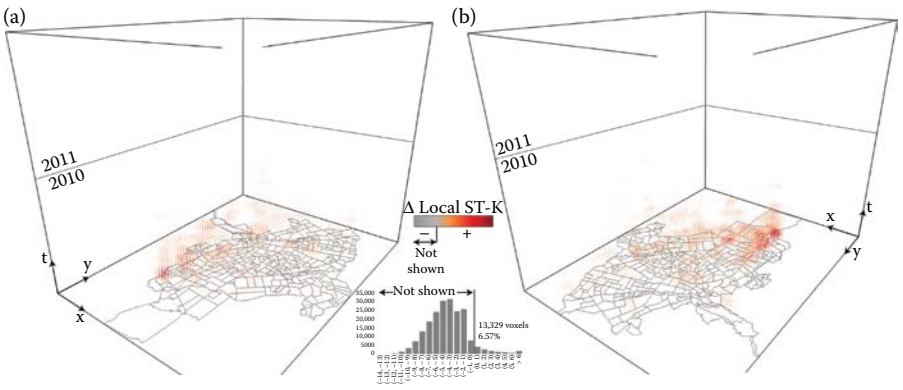

FIGURE 7.3
Local space-time K function using spatial and temporal bandwidths of 500 m and 7 days, respectively. Only positive differences between observed data and the upper simulation envelope are shown (a) from the Southeast, and (b) from the Northwest.

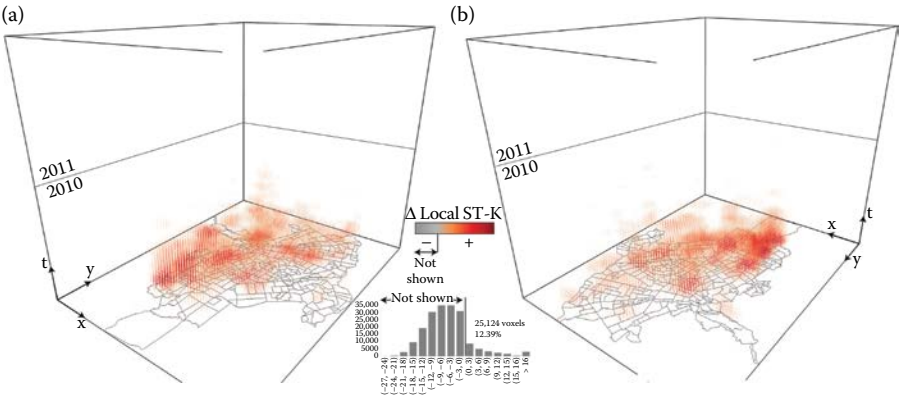

FIGURE 7.4
Local space-time K function using spatial and temporal bandwidths of 750 m and 10 days, respectively. Only positive differences between observed data and the upper simulation envelope are shown (a) from the Southeast, and (b) from the Northwest.

of that spectrum). We note the presence of strong clusters at the beginning of the year 2010, coinciding with an increase in cases during the first few months of the year (see Hohl et al. 2016). Figure 7.4 is similar to Figure 7.3, but uses larger spatial and temporal bandwidths (750 m and 10 days). Using the same legend as in Figure 7.3, we observe a much greater number of grid points where the difference between the observed and expected L-values is large (n = 25,124 voxels or 12.39%, compared to 13,329 or 6.57% in the former scenario with spatial and temporal bandwidths of 500 m and 7 days).

Figure 7.5 shows the interpolated variation in the space-time K function, with bandwidths of 500 m and 7 days. Essentially, this map uses data from

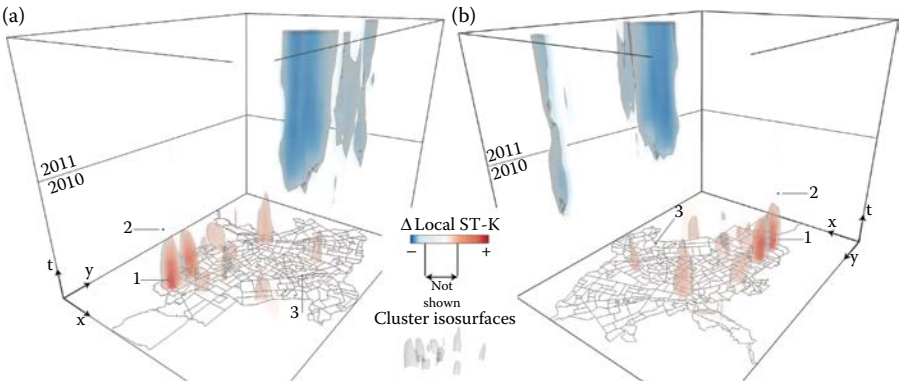

FIGURE 7.5
Local space-time K function using spatial and temporal bandwidths of 500 m and 7 days, respectively. Absolute difference between observed data and the upper simulation envelope. Locations 1, 2, and 3 for global space-time K functions of local settings are annotated in the map (a) from the Southeast, and (b) from the Northwest.

Figure 7.3 as an input, but shows a smooth continuous volume. We use a combination of visualization techniques (volume rendering, transparency; see Delmelle et al. (2014)) to render the strength of the clustering, while isosurfaces reinforce the extent of such clusters. For this map, of more interest are regions of strong positive clustering (represented in red). Although we observe strong, positive clusters at the beginning of 2010, we note that at the end of 2011, some regions are showing negative values, suggesting a tendency toward regularity.

7.5.3 Global Space-Time K Functions of Local Settings

Figure 7.6 depicts the absolute difference between observed Ripley's K and the upper simulation envelope for Location 1 within 1000 m and 14 days of the space-time bandwidth. Location 1 lies within a space-time cluster of dengue fever cases. Local Ripley's K values suggest clustered patterns of dengue fever cases within 1000 m and 14 days with respect to Location 1 since all the values of the absolute difference with respect to the upper simulation envelope are positive. As the spatial bandwidth increases, the clustering response

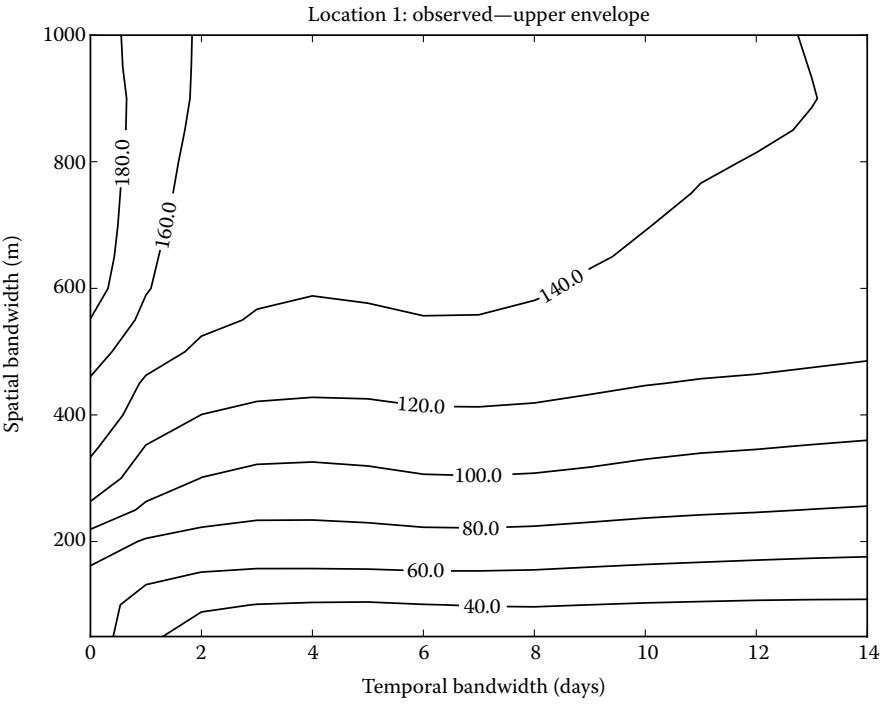

FIGURE 7.6
Global space-time K function of a local setting (Location 1 in Figure 7.5). Absolute difference between observed data and the upper simulation envelope.

becomes stronger, but weakens with decreasing temporal bandwidth values. Strongest clustering response concentrates within a space-time range of 500–1000 m and 1 day.

The difference between observed data and the upper simulation envelope for Location 2 is shown in Figure 7.7. For this location, there is no clustering pattern that we can observe. However, when the temporal bandwidth is from 0 to 1 day or 6 to 14 days within 150 m of the spatial bandwidth, the dengue fever data around that location exhibit a weaker clustering pattern—only a small-scale cluster is observed. To have a better understanding of the dengue fever pattern at Location 2, we compare the difference between observed data and the lower simulation envelope (see Figure 7.8). When the difference between the observed and simulated data is higher than the lower envelope, a completely spatiotemporally random (CSTR) pattern is suggested. Otherwise, space-time regularity is observed when the difference with respect to the lower envelope is negative. As we see in Figure 7.8, positive values of the difference with the lower envelope are observed across all space-time bandwidths. Therefore, we cannot reject the null hypothesis that the spatiotemporal pattern of dengue fever incidents is completely random.

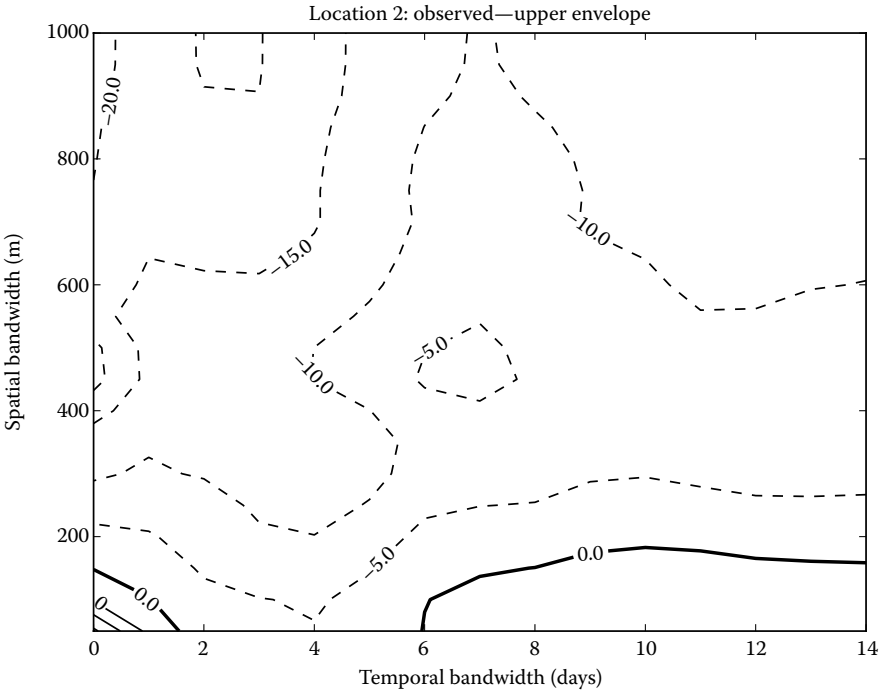

FIGURE 7.7
Global ST K function of a local setting (Location 2 in Figure 7.5). Absolute difference between observed data and the upper simulation envelope. Dashed lines denote negative differences.

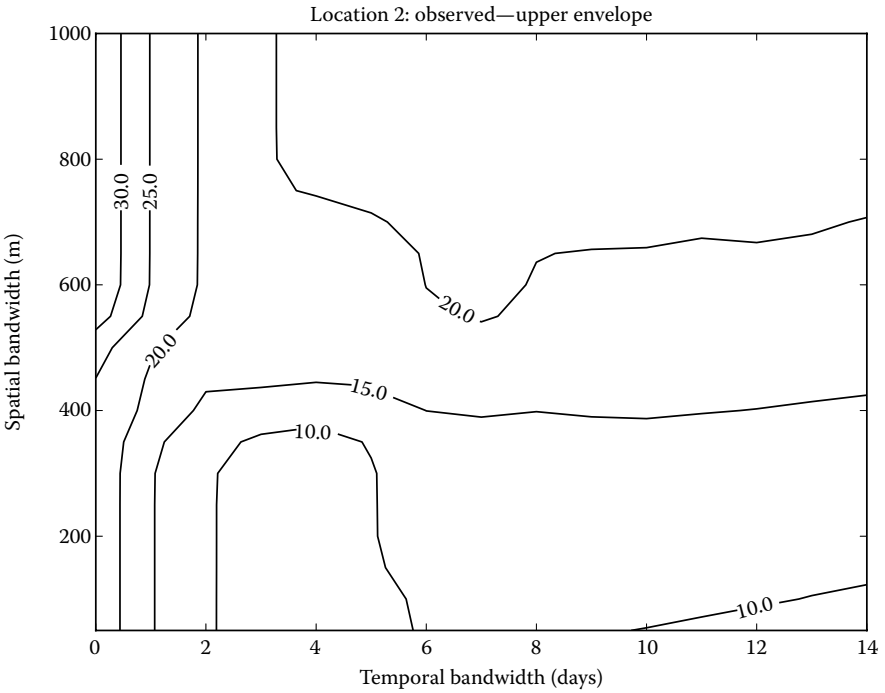

FIGURE 7.8
Global space-time K function of a local setting (Location 2 in Figure 7.5). Absolute difference between observed data and the lower simulation envelope.

For Location 3, all the differences in L-values between observed and upper simulation envelope are negative (see Figure 7.9). When we plot the difference between observed values and the lower simulation envelope (see Figure 7.10), we observe a random pattern throughout all scales, especially for spatial bandwidths from 150 to 750 m and temporal bandwidths between 0 and 1 day. For spatial bandwidths from 250 to 650 m and temporal bandwidths longer than 1 day, the dengue fever data exhibit spatiotemporal regularity with respect to Location 3.

7.6 Conclusions

In this study, we investigated the use of Ripley's K function for the analysis of spatiotemporal point patterns. Using a combination of global and local Ripley's K functions allowed us to discover the space-time characteristics of dengue fever in Cali, Colombia for the years 2010 and 2011.

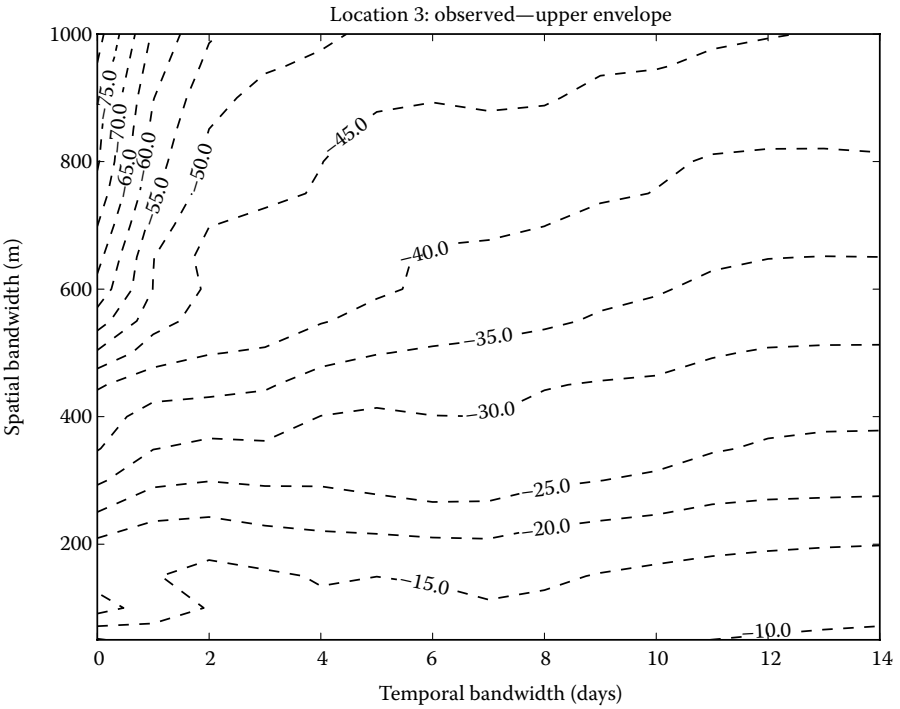

FIGURE 7.9
Global space-time K function of a local setting (Location 3 in Figure 7.5). Absolute difference between observed data and the upper simulation envelope. Dashed lines denote negative differences.

In the case of dengue fever and other vector-borne diseases, being able to identify the space-time location of potential clusters of infection can make a difference in controlling and stopping the spread of the virus. It will help health authorities to better design and plan control strategies in a timely way to stop an epidemic from happening. It will also provide insight into understanding the timeline of the infectious process.

The 3D visualization approach is able to map the shape of each cluster, while giving a clear understanding of the presence of clusters of dengue fever over space and time. Our future work will focus on a number of threads. First, we will perform edge correction to improve global and local forms of space-time Ripley's K function. Second, once a fine spatial-temporal resolution is used, the 3D visualization approach will map more accurate shapes and forms of each cluster. Third, more years of dengue fever data will be added in our study to better understand and explain the space-time complexity of the infectious process.

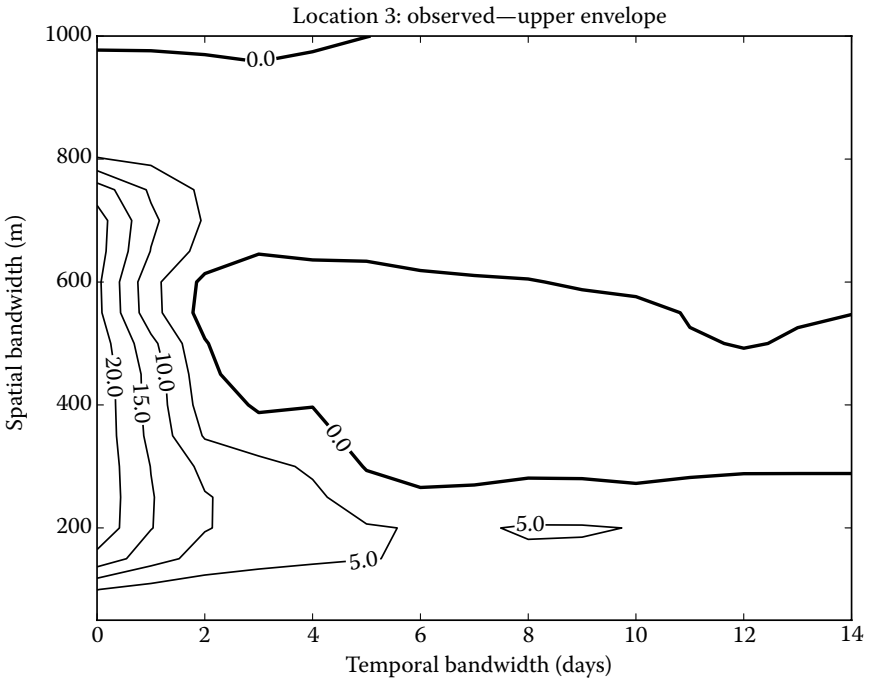

FIGURE 7.10
Global space-time K function of a local setting (Location 3 in Figure 7.5). Absolute difference between observed data and the lower simulation envelope.

Acknowledgment

The authors would like to thank the Public Health Secretariat of the city of Cali for the dengue fever surveillance data. Support of computing resources from University Research Computing (URC) at the University of North Carolina at Charlotte, U.S. NSF XSEDE Supercomputing Resource Allocation (SES170007) is acknowledged.

References

Aigner, W., S. Miksch, W. Müller, H. Schumann, and C. Tominski. 2007. Visualizing time-oriented data—A systematic view. *Computers & Graphics* 31 (3): 401–409.

An, L., M.-H. Tsou, S. E. Crook, Y. Chun, B. Spitzberg, J. M. Gawron, and D. K. Gupta. 2015. Space–time analysis: Concepts, quantitative methods, and future directions. *Annals of the Association of American Geographers* 105 (5): 891–914.

Anselin, L., J. Cohen, D. Cook, W. Gorr, and G. Tita. 2000. Spatial analyses of crime. *Criminal Justice* 4 (2): 213–262.

Bailey, T. and Q. Gatrell. 1995. *Interactive Spatial Data Analysis*. Edinburgh Gate, England: Pearson Education Limited.

Cali, A. d. S. d. ed. 2014. *Cali en cifras*. Cali, Colombia: Alcaldía de Santiago de Cali.

Chen, B. Y., H. Yuan, Q. Li, S.-L. Shaw, W. H. Lam, and X. Chen. 2016. Spatiotemporal data model for network time geographic analysis in the era of big data. *International Journal of Geographical Information Science* 30 (6): 1041–1071.

Colombia, M. d. S. 2017. Sistema de vigilancia en salud pública. Ministerio de Salud Colombia 2017 [cited February 2017]. Available from https://www.minsalud.gov.co/salud/Paginas/SIVIGILA.aspx.

Delmelle, E. 2009. Point pattern analysis. *International Encyclopedia of Human Geography* 8: 204–211.

Delmelle, E., C. Dony, I. Casas, M. Jia, and W. Tang. 2014. Visualizing the impact of space-time uncertainties on dengue fever patterns. *International Journal of Geographical Information Science* 28 (5): 1107–1127.

Demšar, U., K. Buchin, F. Cagnacci, K. Safi, B. Speckmann, N. Van de Weghe, D. Weiskopf, and R. Weibel. 2015. Analysis and visualisation of movement: An interdisciplinary review. *Movement Ecology* 3 (1): 5.

Diggle, P. J. 2013. *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns*. Boca Raton: CRC Press.

Diggle, P. J., J. Besag, and J. T. Gleaves. 1976. Statistical analysis of spatial point patterns by means of distance methods. *Biometrics* 32: 659–667.

Dixon, P. M. 2013. *Ripley's K function*. Encyclopedia of Environmetrics.

Gabriel, E. 2014. Estimating second-order characteristics of inhomogeneous spatio-temporal point processes. *Methodology and Computing in Applied Probability* 16 (2): 411–431.

Gatrell, A., T. Bailey, P. Diggle, and B. Rowlingson. 1996. Spatial point pattern analysis and its application in geographical epidemiology. *Transactions of the Institute of British Geographers* 21 (1): 256–274.

Getis, A. 1984. Interaction modeling using second-order analysis. *Environment and Planning A* 16 (2): 173–183.

Getis, A. and J. Franklin. 1987. Second-order neighborhood analysis of mapped point patterns. *Ecology* 68 (3): 473–477.

Goodchild, M. F. 2013. Prospects for a space–time GIS: Space–time integration in geography and GIScience. *Annals of the Association of American Geographers* 103 (5): 1072–1077.

Gould, P. R. 1969. *Spatial Diffusion. Resource Paper 4*. Washington, DC: Association of American Geographers.

Haase, P. 1995. Spatial pattern analysis in ecology based on Ripley's K-function: Introduction and methods of edge correction. *Journal of Vegetation Science* 6 (4): 575–582.

Hohl, A., E. Delmelle, W. Tang, and I. Casas. 2016. Accelerating the discovery of space-time patterns of infectious diseases using parallel computing. *Spatial and Spatio-Temporal Epidemiology* 19: 10–20.

Illian, J., A. Penttinen, H. Stoyan, and D. Stoyan. 2008. *Statistical Analysis and Modelling of Spatial Point Patterns*. New York: John Wiley & Sons.

Jacquez, G. M. 1996. A k nearest neighbour test for space–time interaction. *Statistics in Medicine* 15 (18): 1935–1949.

Kiskowski, M. A., J. F. Hancock, and A. K. Kenworthy. 2009. On the use of Ripley's K-function and its derivatives to analyze domain size. *Biophysical Journal* 97 (4): 1095–1103.

Knox, G. E. 1964. The detection of space-time iterations. *Journal of the Royal Statistical Society* 13: 25–29.

Kwan, M.-P. and T. Neutens. 2014. Space-time research in GIScience. *International Journal of Geographical Information Science* 28 (5): 851–854.

Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Research* 27 (2 Part 1): 209–220.

Mountrakis, G. and K. Gunson. 2009. Multi-scale spatiotemporal analyses of moose–vehicle collisions: A case study in northern Vermont. *International Journal of Geographical Information Science* 23 (11): 1389–1412.

Nakaya, T. and K. Yano. 2010. Visualising crime clusters in a space-time cube: An exploratory data-analysis approach using space-time kernel density estimation and scan statistics. *Transactions in GIS* 14 (3): 223–239.

Okabe, A., B. Boots, and T. Satoh. 2010. A class of local and global K functions and their exact statistical methods. In *Perspectives on Spatial Data Analysis*, edited by L. Anselin, and S. Rey, 101–112. Berlin: Springer.

Okabe, A., and I. Yamada. 2001. The K-Function Method on a Network and Its Computational Implementation. *Geographical Analysis* 33 (3): 271–290.

Perry, G. L., B. P. Miller, and N. J. Enright. 2006. A comparison of methods for the statistical analysis of spatial point patterns in plant ecology. *Plant Ecology* 187 (1): 59–82.

Pfister, G. F. 2001. An introduction to the infiniband architecture. *High Performance Mass Storage and Parallel I/O* 42: 617–632.

Restrepo, L. D. E. 2011. El plan piloto de cali 1950. *Bitácora Urbano Territorial* 1 (10): 222–233.

Ripley, B. D. 1976. The second-order analysis of stationary point processes. *Journal of Applied Probability* 13 (2): 255–266.

Tang, W., W. Feng, and M. Jia. 2015. Massively parallel spatial point pattern analysis: Ripley's K function accelerated using graphics processing units. *International Journal of Geographical Information Science* 29 (3): 412–439.

Tao, R. and J. C. Thill. 2016. Spatial Cluster Detection in Spatial Flow Data. *Geographical Analysis* 48 (4): 355–372.

Tao, R., J.-C. Thill, and I. Yamada. 2015. Detecting clustering scales with the incremental K-function: comparison tests on actual and simulated geospatial datasets. In *Information Fusion and Geographic Information Systems (IF&GIS'2015)*, edited by V. Popovich, C. Claramunt, M. Schrenk, K. Korolenko, and J. Gensel, 93–107. Heidelberg: Springer.

Varela, A., E. G. Aristizabal, and J. H. Rojas. 2010. *Analisis epidemiologico de dengue en Cali*. Cali: Secretaria de Salud Publica Municipal.

Wiegand, T. and K. A. Moloney. 2004. Rings, circles, and null-models for point pattern analysis in ecology. *Oikos* 104 (2): 209–229.

Yamada, I. and P. A. Rogerson. 2003. An Empirical Comparison of Edge Effect Correction Methods Applied to K-function Analysis. *Geographical Analysis* 35 (2): 97–109.

Yamada, I. and J. C. Thill. 2007. Local Indicators of Network-Constrained Clusters in Spatial Point Patterns. *Geographical Analysis* 39 (3): 268–292.

Yamada, I. and J.-C. Thill. 2010. Local indicators of network-constrained clusters in spatial patterns represented by a link attribute. *Annals of the Association of American Geographers* 100 (2): 269–285.