



# Detecting space–time patterns of disease risk under dynamic background population

Alexander Hohl<sup>1</sup> · Wenwu Tang<sup>2,3</sup> · Irene Casas<sup>4</sup> · Xun Shi<sup>5</sup> · Eric Delmelle<sup>2,3,6</sup>

Received: 30 April 2021 / Accepted: 15 February 2022 / Published online: 20 April 2022  
© The Author(s) 2022

## Abstract

We are able to collect vast quantities of spatiotemporal data due to recent technological advances. Exploratory space–time data analysis approaches can facilitate the detection of patterns and formation of hypotheses about their driving processes. However, geographic patterns of social phenomena like crime or disease are driven by the underlying population. This research aims for incorporating temporal population dynamics into spatial analysis, a key omission of previous methods. As population data are becoming available at finer spatial and temporal granularity, we are increasingly able to capture the dynamic patterns of human activity. In this paper, we modify the space–time kernel density estimation method by accounting for spatially and temporally dynamic background populations (ST-DB), assess the benefits of considering the temporal dimension and finally, compare ST-DB to its purely spatial counterpart. We delineate clusters and compare them, as well as their significance, across multiple parameter configurations. We apply ST-DB to an outbreak of dengue fever in Cali, Colombia during 2010–2011. Our results show that incorporating the temporal dimension improves our ability to delineate significant clusters. This study addresses an urgent need in the spatiotemporal analysis literature by using population data at high spatial and temporal resolutions.

**Keywords** Kernel density estimation · Dengue · Background · Adaptive · Population

---

✉ Alexander Hohl  
alexander.hohl@geog.utah.edu

<sup>1</sup> Department of Geography, University of Utah, Salt Lake City, UT 84112, USA

<sup>2</sup> Center for Applied Geographic Information Science, University of North Carolina at Charlotte, Charlotte, NC 28223, USA

<sup>3</sup> Department of Geography and Earth Sciences, University of North Carolina at Charlotte, Charlotte, NC 28223, USA

<sup>4</sup> School of History and Social Sciences, Louisiana Tech University, Ruston, LA 71272, USA

<sup>5</sup> Department of Geography, Dartmouth College, Hanover, NH 03755, USA

<sup>6</sup> Department of Geographical and Historical Studies, University of Eastern Finland, Joensuu Campus, P.O. Box 111, FI-80101, Finland

## 1 Introduction

Novel methodologies for spatial and temporal analysis of geographic phenomena have emerged due to an abundance of geospatial information at the individual level (Anselin 2011). Domain examples include crimes (Wang et al. 2017; Malleson and Andresen 2015a; Koo et al. 2020) and disease events (Hohl et al. 2016; Delmelle et al. 2014), which typically cluster near city centers or exhibit seasonal cyclic patterns. Knowledge about the intensity, scale, location, and time of such clusters is important. In the case of disease outbreaks, this information is critical to inform authorities on their decision to allocate resources, such as staff for disease prevention efforts (Casas et al. 2010). Spatial and spatiotemporal statistics are a set of popular analytical methods for identifying and quantifying inherent patterns in the data, as they capture geospatial phenomena and their variability in space and time (Bailey and Gatrell 1995; Cressie and Wikle 2015) and across multiple scales (Fotheringham et al. 2017). Among the palette of exploratory statistics used to characterize a given spatiotemporal point pattern, space–time kernel density estimation (STKDE; Nakaya and Yano 2010) stands out. It allows for visualizing the occurrence of events in space and time by computing the localized intensity of the point process at hand and hence, summarizing the distribution of a spatial variable through time. STKDE has been employed as a key analytical procedure for identifying clusters of crime (Nakaya and Yano 2010), exploring human mobility patterns (Gao 2015), as well as discovering outbreaks of dengue fever (Delmelle et al. 2014).

Disease risk can be understood as the ratio between the number of disease cases and the population-at-risk within a given area (Desjardins et al. 2020). Therefore, it is imperative to consider density estimates of disease cases in relation to the local population-at-risk (a.k.a. ‘the background’). Otherwise, the estimates are a proxy of the population distribution and we might observe a cluster of high density merely due to a large local population. While conventional STKDE does not consider the background at all, several approaches incorporate spatially varying backgrounds (Shi 2010; Davies and Hazelton 2010; Davies et al. 2016; Tiwari and Rushton 2005). Hence, adjusting for *spatial* variation of the background is a common practice to date. However, adjusting for *spatial and temporal* variation of the background is a novelty, to the best of our knowledge. By omitting time in their models, existing methods ignore temporal dynamics (temporal variation) of the background, and are therefore unable to capture rapid population change.

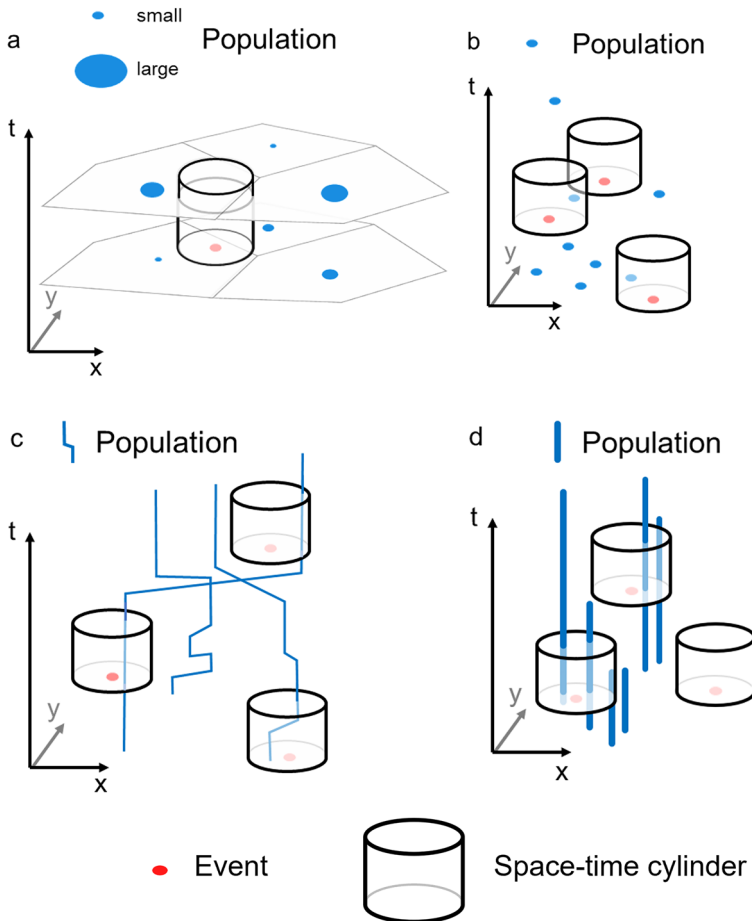
However, we currently find ourselves in the age of migration (Castles et al. 2013), where individuals move from their residential location for many reasons: forced migration due to climate change (Martin 2001), conflicts (Mitchell 2011), or to find labor (Münz 2007). For instance, cities experience waves of urbanization (Meentemeyer et al. 2013), suburbanization (Lang and Simmons 2003), reurbanization and counter-urbanization (Champion 2001). Hence, it is no longer acceptable to ignore temporal population dynamics, which is especially relevant for longitudinal studies. In addition, population data are becoming available at

finer spatial and temporal resolutions, and given current technological advances, it is foreseeable that this development will continue (Bhaduri et al. 2007; Huang et al. 2020; Kang et al. 2020). This calls for an extension of the current kernel methods for computing disease risk to address background populations that change over space *and* time.

Adaptive kernels allow for variation in bandwidth (a.k.a search radius) across the study area, while fixed-bandwidth kernels do not. Therefore, fixed-bandwidth kernels may not properly capture human dynamics at varying geographic scales (Yuan 2018). They tend to oversmooth, leading to a loss of spatial detail (Goodchild 2001) and conceal regions of interest (Shi 2010), especially when analyzing large-scale human geographic phenomena, where events like disease cases emerge out of the background. While fixed-bandwidth kernels establish constant areal support, a kernel that adapts its bandwidth to the background is useful to establish constant population support (Carlos et al. 2010). This allows for comparison of disease rates across regions, which is especially suited for analyzing chronic diseases, such as lung cancer (Shi 2010). On the other hand, a kernel that adapts to surrounding cases sacrifices the constant population support property and establishes constant case support. This is suited for analyzing communicable disease, as distance between cases is a reasonable approximation of interaction between infected individuals, a common cause of disease spread (Bhopal 2016). Hence, adaptive kernels allow for measuring the scale over which a point process operates (Fotheringham et al. 2017; Xu et al. 2017).

It is important to recognize that there are many ways to obtain a representation of the background (Fig. 1), i.e. from population data at high spatial and temporal resolution. Apart from census data (Fig. 1a), scientists have used social media posts such as tweets (Fig. 1b) as a proxy for population (Malleson and Andresen 2015b). Alternatively, trajectories of individuals (Fig. 1c) may be created through retrospective activity diaries (Chen et al. 2011; Kwan 2000, 2004), or migration history datasets (Shaw et al. 2008). Lastly, residential location may be used to establish the population at risk (Fig. 1d). In addition to the spatial coordinates  $x$  and  $y$ , the temporal coordinates  $t_1$  and  $t_2$  may be known, which represent start and end date of residence. In summary, the representations of population in Fig. 1 are profoundly different from each other. Besides availability, the following principle should guide the choice of population data: The scale of the population data should match the scale of the case data. For instance, if we use patient residential locations, we may want to use control data, as activity diaries population information would be too detailed. As the background can be represented in different ways, adaptations to the kernel density method for computing risk estimates are necessary to accommodate such diversity.

In this study, we address spatially and temporally varying backgrounds in kernel density estimation, a key omission of many existing applications. We introduce ST-DB, a space–time kernel density estimator that considers a spatially *and* temporally dynamic background. We compare ST-DB with its purely spatial counterpart and address the question of whether adding time to our analysis yields different estimates of disease risk. While our study focuses on introducing ST-DB and its application to spatiotemporal analysis of infectious disease, it can be applied to any geographic phenomenon that involves point events that emerge out of a background.



**Fig. 1** Different representations of the background: **a** centroids (e.g. census tracts), **b** social media posts (e.g. tweets), **c** trajectories of individuals, **d** residential locations

This article is organized as follows: We explain our methodology, as well as our dataset in Sects. 2–4, illustrate our results in Sect. 5, and finally present discussion and conclusions in Sects. 6 and 7.

## 2 Methods

### 2.1 Kernel density estimation for static backgrounds

Kernel density estimation (KDE) is an essential method for analyzing spatial point (a.k.a. ‘event’) patterns (Silverman 1986). It results in a smooth surface of density estimates by imposing a regular grid of points (‘pixels’) on the study

area. The density for each grid point is computed based on neighboring events, as follows: The kernel, a circular window with radius  $h_s$  (‘bandwidth’,  $s$  denotes ‘spatial’) is centered on a data point. Any grid point located within the kernel receives a contribution (a.k.a. ‘weight’) toward its density estimate. The contribution is determined by their distance to the data point in the center, which is plugged into the kernel function ( $k_s$ , closer proximity results in higher contribution). Lastly, the weights are summed for each grid point, as multiple data points can contribute to a given grid point. We repeat the procedure for each data point and hence, create a density surface based on the observed point data. For a given grid point  $(x, y)$ , kernel density  $\hat{f}(x, y)$  is calculated as follows (Eq. 1):

$$\hat{f}(x, y) = \frac{1}{nh_s^2} \sum_{i=1}^n k_s \left( \frac{d_{i(x,y)}}{h_s} \right) \tag{1}$$

where  $n$  is the number of data points within the study area,  $k_s$  is the kernel function, and  $d_{i(x,y)}$  is the distance between the grid point and data point  $i$ . Kernel functions like Epanechnikov, Gaussian, or Biweight are widely used and accepted (Bowman and Azzalini 1997). We use the Epanechnikov kernel function (Epanechnikov 1969) in all of our analyses due to its popularity for spatial- and spatiotemporal analysis.

### 2.2 Kernel density estimation for dynamic backgrounds

For many geographical research questions, density as the distance-weighted number of points per *unit area* may not provide a suitable answer (Bithell 1990). In the case of estimating disease risk, mapping the distance-weighted number of disease cases per *unit population-at-risk* rather than per *unit area* might be more realistic. The latter assumes that geographic distance is the sole determinant of the contribution of a case to the disease risk at a grid point (Shi 2010). As a result, an area of elevated risk identified by KDE might merely reflect a large local background population (Bithell 2000). Depending on the phenomenon under study, the population-at-risk can exhibit an uneven distribution in space and time, include all or only certain segments of the population (i.e. for COVID-19, elderly individuals and people with comorbidities are at increased mortality risk), and may be a sample of the full population-at-risk. This is referred to as the background population, or simply ‘the background’ (Carlos et al. 2010). A generic method to deal with a spatially varying background is to compute the risk ( $\hat{r}$ ) at location  $(x, y)$  by dividing the density of cases ( $c$ ) by the density of the background population ( $p$ ), shown in Eq. 2 (Davies and Hazelton 2010).

$$\hat{r}(x, y) = \frac{c}{p} \tag{2}$$

We use Eq. (2) to compute the risk at any location  $(x, y)$  by centering the kernel on each data point  $i$  (disease case). We then compute  $i$ ’s contribution to risk at surrounding grid points by factoring in the population within the kernel. Therefore, the

contribution of  $i$  to the risk at  $(x, y)$  is determined by the population near  $i$ , and not by the population near  $(x, y)$ . In other words, the contribution of a case to the disease risk at a particular grid point is controlled by the local population surrounding that case, rather than the population surrounding the grid point. This distinction is relevant as it can result in different risk estimates under spatially varying backgrounds (Shi 2010).

Fixed kernels have constant bandwidth (search radius), whereas adaptive kernels allow the bandwidth to adapt to local conditions (Sain 2002; Brunson 1995). In our case, the kernel can either adapt to the background (Fig. 2a) or neighboring cases (Fig. 2b). While a fixed kernel results in constant areal support for each case, an adaptive kernel establishes either a constant population or case support. Areas that exhibit a high density of disease events ('clusters'), for example, are often sought out for prevention efforts (Coleman et al. 2009). As infectious diseases spread between individuals in close proximity (Salathé et al. 2010), an area where cases cluster may be characterized as 'high risk,' whereas an area where cases are sparse may be 'low risk' (Bhopal 2016; Riley 2007). Therefore, we choose a kernel that adapts to neighboring disease cases and adjusts the estimate to the background within.

We achieve this by centering the kernel on a disease case and start increasing the bandwidth until it encircles a specified number of neighboring cases (the *support* threshold). Note that as the kernel expands, the case in its center will expand the spatial range of its contribution to disease risk. In other words, as the circle grows outward, seeking support, more grid points will receive contribution from the disease case in its center.

Shi (2010) proposes an adaptive bandwidth kernel density estimator (Eq. 3), which corresponds to Fig. 2a):

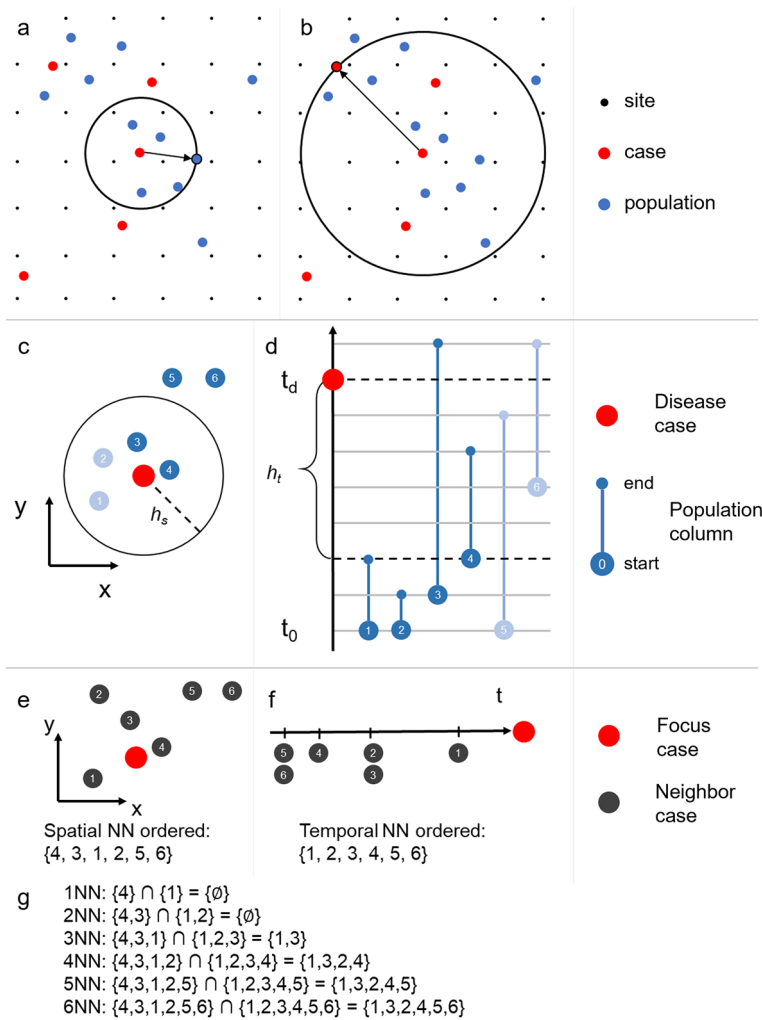
$$\hat{r}(x, y) = \sum_{i=1}^n k_s \left( \frac{d_{i,(x,y)}}{h_s [p(x_i, y_i)]} \right) \quad (3)$$

where the bandwidth  $h_s$  is a function of the local population density  $p$  at the location  $(x_i, y_i)$  of case  $i$ . The weight of  $i$  for  $(x, y)$  is divided by the population in the kernel. This method results in disease risk values that are defensible in health studies, while also being more statistically comparable (Carlos et al. 2010; Shi and Wang 2015; Shi 2010).

Equation (4) denotes S-DB, the purely spatial kernel density estimator for dynamic backgrounds that adapts to the neighboring cases, which corresponds to Fig. 2b):

$$\hat{r}_{\text{S-DB}}(x, y) = \sum_{i=1}^n k_s \left( \frac{d_{i,(x,y)}}{h_s [c(x_i, y_i)]} \right) \quad (4)$$

Here the bandwidth  $h_s$  is a function of the local case density  $c$  at the location  $(x_i, y_i)$  of case  $i$ . The density contribution of  $i$  to the locations within bandwidth is then divided by the population in the kernel. Note that both estimators (Eqs. 3, 4) require



**Fig. 2** Top: adaptive bandwidth kernel with spatially varying background, **a** kernel adapts to population, **b** kernel adapts to cases. Note that *support* in **a** is 5 people and in **b** is 3 cases (the case in the center of the circle is not counted toward the *support* threshold). Middle: Quantifying the population within the kernel, **c** spatial view, **d** temporal view. In this example, the population within the kernel is 8 people-days. Only population columns 3 and 4 are inside the kernel spatially and temporally and the sum of their lengths inside the kernel is  $5 + 3 = 8$ . Bottom: spatiotemporal nearest neighbors (NN), **e** spatial NN, **f** temporal NN, **g** intersection of spatial and temporal NN

choosing the *support* threshold value. We circumvent this choice by analyzing the sensitivity of the resulting risk estimates to the *support* threshold value.

### 2.3 Space–time kernel density estimation for static backgrounds

So far, we have ignored the temporal dimension in the discussion. Many geographic studies do not consider the temporal dimension or employ time-flattening: collapsing the temporal dimension into a single 2D map, which represents the entire study period (Bach et al. 2016). Another approach discretizes time into a number of time slices, which can be displayed as small multiples (Boyandin et al. 2012). However, both methods are limited in their ability to depict patterns of spatiotemporal point events: Time-flattening ignores any temporal variation in the data, and the small multiples approach is not scalable (Delmelle et al. 2014).

Space–time kernel density estimation (STKDE) extends traditional bivariate KDE with the temporal dimension and is suited for characterizing spatiotemporal patterns of spatial point events with a timestamp (Nakaya and Yano 2010). STKDE produces density estimates for a spatiotemporal grid of points (‘voxels’) based on the proximity and number of surrounding point data (Delmelle et al. 2014; Brunson et al. 2007). We can visualize the density estimates within a space–time cube (Hagerstrand 1970; Bach et al. 2016; Hohl et al. 2016; Desjardins et al. 2019; Gao 2015; Nakaya and Yano 2010; Demšar et al. 2015) that has two spatial ( $x, y$ ) and a temporal dimension ( $t$ ). STKDE is computed as follows: We center the bottom of the kernel, a cylindrical window defined by its circular base with radius  $h_s$  (spatial bandwidth) and height  $h_t$  (temporal bandwidth) on a data point. Any voxel located within the kernel receives a contribution (or weight) toward its density estimate, as a case imposes risk only to the time-period after the event. The contribution is determined by the distance between voxel and data point in the center, which is plugged into the spatial and temporal kernel functions ( $k_s, k_t$ ). Lastly, the weights are summed for each voxel, as multiple data points can contribute to a given voxel. We repeat the procedure for each data point and hence, create a density volume based on the observed point data. For a given voxel ( $x, y, t$ ), density is calculated as follows (Eq. 5):

$$\hat{f}(x, y, t) = \frac{1}{nh_s^2 h_t} \sum_i k_s \left( \frac{d_{i(x,y)}}{h_s} \right) k_t \left( \frac{d_{i(t)}}{h_t} \right) \quad (5)$$

Every voxel  $s$  with coordinates ( $x, y, t$ ) receives a density estimate  $\hat{f}(x, y, t)$ , which is determined by distance and number of neighboring data points  $i$ . Data points in the neighborhood of  $s$  are weighted by the spatial and temporal kernel functions,  $k_s$  and  $k_t$ , which are computed as separate components, and then multiplied to calculate the contribution of a given data point to the density estimates on surrounding grid points. Lastly,  $d_{i(x,y)}$  and  $d_{i(t)}$  are the spatial and temporal distances between voxel and data point, respectively.



### 2.4 Space–time kernel density estimation for dynamic backgrounds

Here, we discuss the extension of STKDE to account for spatially and temporally varying backgrounds (Eq. 6). It is the temporal extension of Shi’s case-side adaptive bandwidth kernel density estimator:

$$\hat{r}(x, y, t) = \sum_i k_s \left( \frac{d_{i,(x,y)}}{h_s [p(x_i, y_i)]} \right) k_t \left( \frac{d_{i,(t)}}{h_t [p(t_i)]} \right) \tag{6}$$

The spatial- and temporal bandwidths  $h_s$  and  $h_t$ , respectively, are a function of the local population density  $p(x_i, y_i)$ ,  $p(t_i)$  at space–time location  $(x_i, y_i, t_i)$  of data point  $i$ . The background is assessed within a half cylinder moving through 3D space, which means that we consider the population within the kernel until the disease case occurs, but not after. A kernel that adapts to the background population is useful to establish constant population support (constant  $p$  in Eq. 2), rather than constant areal support, which is the case with fixed bandwidth kernels.

As seen in Sect. 2.2, it may make sense to adapt the bandwidth to the surrounding cases  $c(x_i, y_i, t_i)$  instead of the local population  $p(x_i, y_i, t_i)$  (Eq. 7). Therefore, we define the kernel density estimator for spatially and temporally dynamic backgrounds (ST-DB) as follows:

$$\hat{r}_{\text{ST-DB}}(x, y, t) = \sum_i k_s \left( \frac{d_{i,(x,y)}}{h_s [c(x_i, y_i)]} \right) k_t \left( \frac{d_{i,(t)}}{h_t [c(t_i)]} \right) \tag{7}$$

Here, the spatial and temporal bandwidths  $h_s$  and  $h_t$  expand until a specified number of neighboring disease cases is found within the cylindrical kernel. The density contribution of the disease case  $i$  to the voxels within bandwidth is then divided by the population in the kernel. As population information might be available in different formats and conceptualizations (see Sect. 1), we pick the population columns model (Fig. 1d) to illustrate the utility of our approach. The within-kernel population is computed by summation of the segment length of all population columns within the cylinder (Fig. 2c, d). The sum represents the number of individuals and their length of exposure to the disease case. It is measured in people-days (an analogy to the term ‘man-hours’ used to quantify the amount of work that can be done by one person within this period).

The *support* can be achieved in multiple ways. In search for neighbors, we could either exclusively expand the spatial bandwidth, or exclusively the temporal bandwidth, or both in an alternating pattern. Therefore, ambiguity arises by the choice of search strategy. To solve this problem, we need to unify the spatial and temporal dimensions, allowing us to expand the bandwidths simultaneously. We employ the  $k$ -nearest neighbors (kNN) method (Jacquez 1996) for this task as follows:

1. Generate two ordered sets for each disease case: (1) the spatial  $k$ -nearest neighbors (Fig. 2e) and (2) the temporal  $k$ -nearest neighbors (Fig. 2f) of case  $i$ .
2. Compute the cardinality  $card()$  of the intersection between the two sets (Fig. 2g).

Starting with  $k=1$ , we increase  $k$  and apply the procedure until  $card()$  equals the *support* threshold. We then compute the spatial and temporal bandwidths  $h_s$ ,  $h_t$ , respectively, as the spatial and temporal distance of the farthest point in the intersection set to the case. Using this procedure, we unify the spatial and temporal dimensions, enabling search for the support in adaptive-bandwidth kernel density estimation for spatially and temporally dynamic backgrounds. Therefore, we solve the multiway problem for ST-DB.

### 3 Data

#### 3.1 Case data

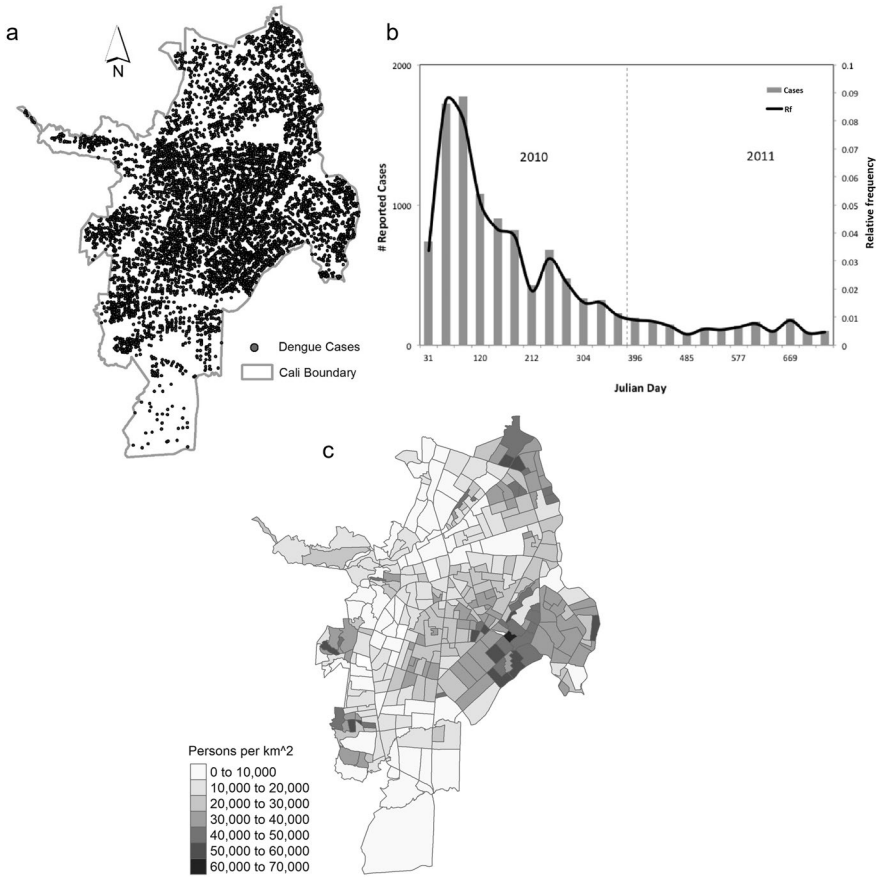
For the case study we use dengue virus data from the city of Santiago de Cali, Colombia ('Cali'). Cali is located in the southwest of Colombia, which exhibits very suitable conditions for the *Aedes Aegypti* mosquito (the principal mosquito vector for dengue virus). The city of Cali has a population of around 2.5 million and it is considered an endemic area for dengue fever (Cali 2010). The dengue fever dataset includes individual case records, for which  $x$ - and  $y$ -coordinates of the patient residential address, as well as the time of diagnosis are available (Delmelle et al. 2014). A total of 11,056 cases were observed in our 2010–2011 study period. Figure 3 presents the spatial distribution of cases within the city (a) and the temporal distribution through the two years of study (b).

#### 3.2 Population Data

Population data are obtained from the Administrative Planning Department from the city of Cali (Cali 2019), which include population projections by neighborhood from 2006 to 2036. We eliminated six neighborhoods that had zero population (parks, sports complexes, military facilities), which resulted in a total of 334 neighborhoods in our study area. We computed summary statistics for dengue fever case counts and population at the neighborhood level, as well as their change, for the years 2010 and 2011 (Table 1) and mapped the 2010 population density by neighborhood (Fig. 3c). Peripheral areas, especially to the east of the city, have a high concentration of population, while neighborhoods in the central part of the city which constitute an extension of the city core have lower population density levels.

We put forward a simple procedure to disaggregate population data from their spatial and temporal units (neighborhoods, years) to individual level (Fig. 4), similar to Jacquez and Jacquez (1999), Shi (2009) and Luo et al. (2010):

1. We distribute the population of the first year (2010, 'the initial population') within each of the 334 neighborhoods at random locations, which are noted in our population table as  $[x, y, t_1, t_2]$ -tuple ( $x$ -coordinate,  $y$ -coordinate, first day of residence,

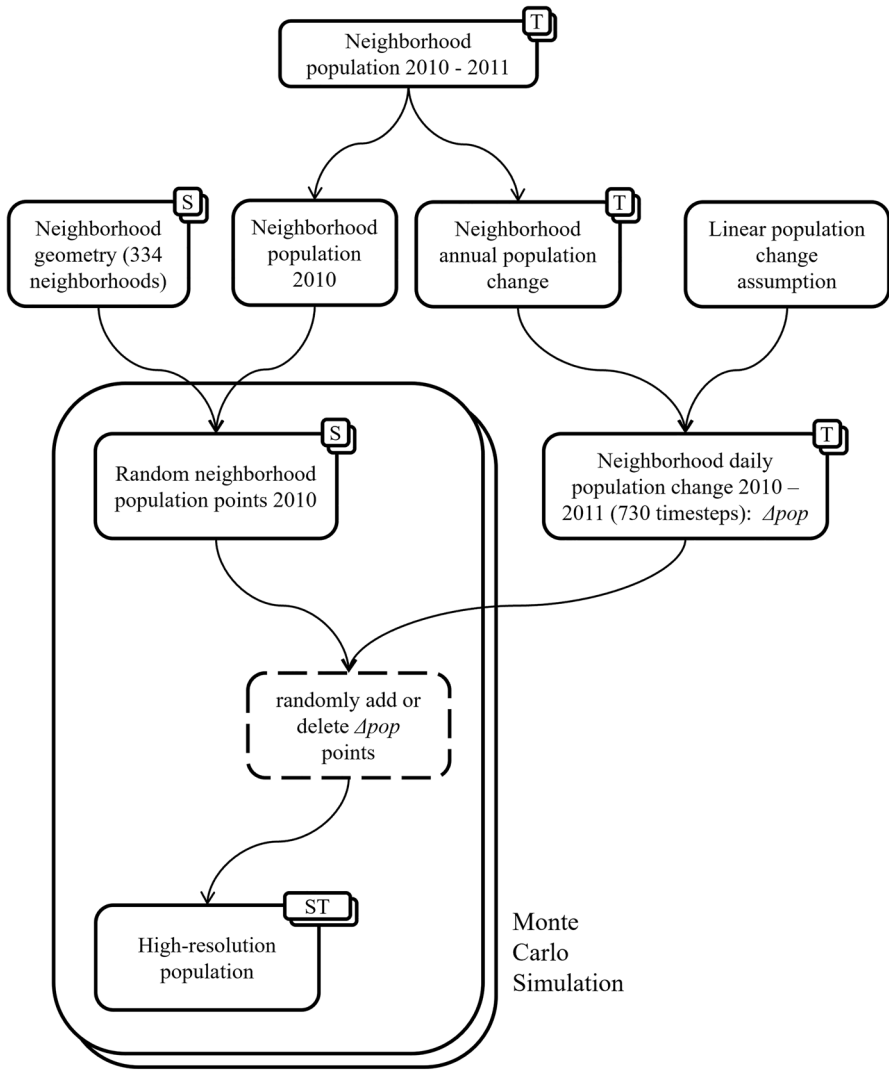


**Fig. 3** The dengue fever dataset: **a** spatial and **b** temporal distributions. **c** Population density of Cali, Colombia in 2010

**Table 1** Summary statistics of neighborhood-level dengue fever case counts and population 2010–2011

	Dengue fever cases		Population	
	2010	2011	2010	2011
Min	0	0	59	58
Max	143	34	49,978	50,581
Mean	28.4	4.6	6,556.9	6,629.3
Standard deviation	25.3	4.4	6,369.2	6,478.8

and last day of residence). For instance, a person who lived in Cali from 1/1/2010 to 31/12/2011 would have  $t_1 = 1$  and  $t_2 = 730$ . Every neighborhood receives a number of  $[x, y, t_1, t_2]$ -tuples commensurate with its total population of 2010.



Legend:

- ST Spatial & temporal units
- S Spatial units
- T Temporal units
- Datasets

Fig. 4 Flowchart for population disaggregation

2. Using the yearly neighborhood population counts 2010–2011, we compute annual population change (increase/decrease) for each neighborhood. We scale down the annual change to daily values, assuming linear change. Hence, we compute the daily change in population by dividing the annual change by 365.
3. For each day within 2010–2011 (which amounts to 730 timesteps), we add random points commensurate with the population increase and set their  $t_1$  to the current day. In case of population decline, we randomly pick existing points commensurate with the population decline and set their  $t_2$  to the current day.

Hence, we create a spatiotemporal, individual-level population dataset, equivalent to the population ‘columns’ in Fig. 1d.

## 4 Analysis

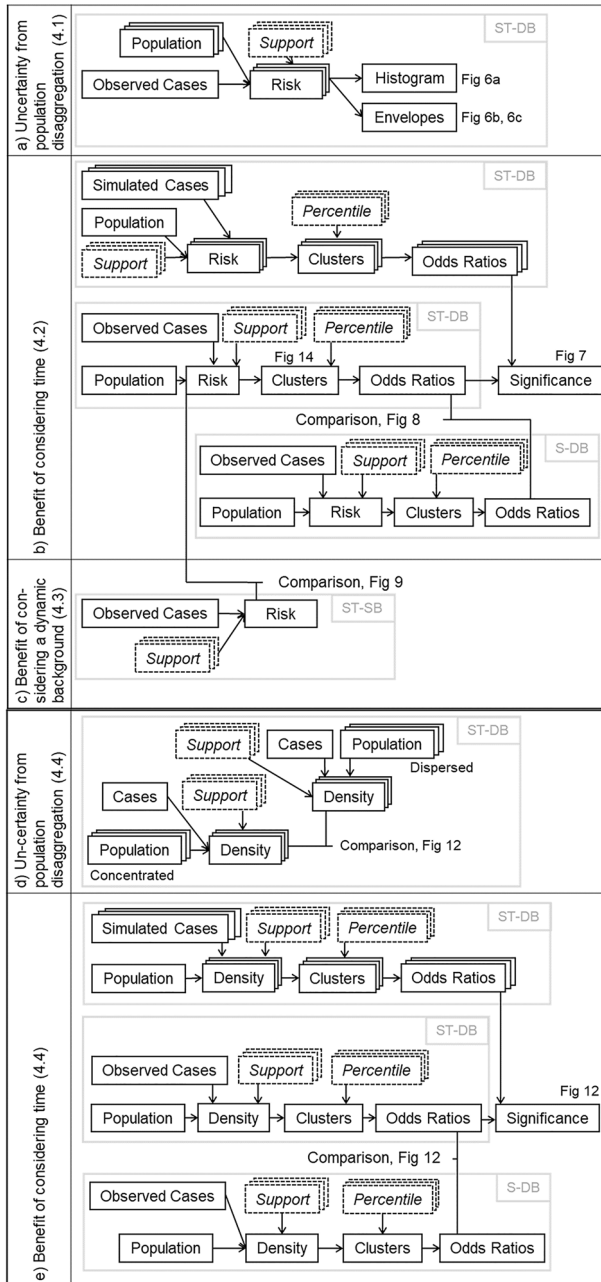
We conduct analyses within our methodological framework (Fig. 5) that summarizes our analytical steps. It also contains references to sections in the text where the corresponding steps are detailed, as well as to figures that show the results.

### 4.1 Uncertainty from population disaggregation

We quantify uncertainty from population data disaggregation (Sect. 3.2), as the process includes random elements. Therefore, we create 99 disaggregated population datasets, leaving the disease data unchanged, and compute 99 grids of risk estimates, which allows us to extract upper and lower envelopes as the maximum and minimum value for each grid point (Fig. 5a). This results in variance of risk, which is a measure of the uncertainty resulting from the random elements involved in disaggregating population data. To quantify the uncertainty, (1) we compute a histogram of the differences between the upper and lower envelope, and (2) visualize them within the space–time cube. If the histogram indicates that the difference is mostly small, we conclude that uncertainty from population disaggregation is small. In addition, the depiction within the space–time cube enables for detecting patterns of where and when the results may be subject to high uncertainty.

### 4.2 Benefit of considering time

In this study, we compare ST-DB (Eq. 7) with its purely spatial counterpart, S-DB (Eq. 4) and assess whether incorporating a temporally dynamic background improves the ability to detect areas/periods of high disease risk (a.k.a. ‘clusters’, Fig. 5b). We apply the following procedure using both approaches (ST-DB and S-DB):



**Fig. 5** Unified methodological framework. Case study (a–c), simulation study (d, e). Black boxes denote data items while arrows in between denote processes on the data items; multiple overlapping boxes denote multiple simulated datasets; dashed line boxes are parameters; gray boxes are methods

Step 1: We compute disease risk estimates.

Step 2: We denote the grid points with  $n$ -th percentile of risk or higher as clusters (*percentile* threshold).

Step 3: We compute the strength of the clustering using odds ratios (within-cluster risk vs. out-of-cluster risk).

Step 4: We compute cluster significance using Monte Carlo simulation.

We run both methods with the same data but ignore the temporal dimension for S-DB (Step 1). After computing disease risk estimates (ST-DB produces a 3D grid, S-DB a 2D grid), we delineate disease clusters as follows (Step 2): We apply the *percentile* threshold and pick the grid points with the  $n$ th-percentile of risk or higher and label them as disease cluster. We measure cluster strength by computing odds ratios (Step 3), which is the ratio between risk inside and outside of the clusters (Bland and Altman 2000; Kulldorff 1997). A high odds ratio means that high-risk disease areas/periods have been delineated well from low-risk ones, as the ratio between cases and controls inside the cluster is much higher than outside. When comparing any two methods  $A$  and  $B$ , we say that method  $A$  delineates clusters better than method  $B$  if it produces a higher odds ratio. Lastly, we use Monte Carlo simulation to measure the statistical significance of clusters (Step 4). Method  $A$  is only better than  $B$  if it produces a higher odds ratio that is statistically significant. For each Monte Carlo simulation run, we randomize the locations of the observed disease cases by sampling from an inhomogeneous Poisson distribution with intensity that follows the population distribution of Cali, while using one realization of the population disaggregation procedure (as it turns out, uncertainty from population disaggregation is small). We compute odds ratios by applying Steps 1–3 for each simulation run, as well as for the observed dataset. The rank of the observed odds ratio among the simulated ones constitutes its  $p$ -value for testing the null hypothesis of Poisson distributed cases. We chose 99 simulation runs, which strikes a balance between computational feasibility and level of statistical confidence.<sup>1</sup>

To address the sensitivity of the resulting odds ratios to various parameter configurations, we apply Steps 1–4 for all combinations of different *percentile* threshold values for cluster delineation, and *support* threshold values for bandwidth selection (see Sect. 2.2). The *support* values are {5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85}, whereas the *percentile* values are {90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.9, 99.99}. This allows us to draw odds ratio surfaces across different parameter configurations, and to compute the difference between ST-DB and S-DB. With the goal of illustrating the utility of our approach, we perform significance testing for all parameter configurations (ST-DB only).

### 4.3 Benefit of considering a dynamic background

We illustrate the utility of adjusting kernel density estimates to a dynamic background by comparing ST-DB with space–time kernel density estimation for static

<sup>1</sup> Note that we would need to increase the number of simulations substantially to achieve confidence levels acceptable in health studies (1,000 – 10,000 simulations).

background (ST-SB), where the background is ignored or assumed to be distributed homogeneously in space and time (Fig. 5c). We visualize density estimates from ST-DB and ST-SB within the space–time cube and expect high density areas/times to differ between the methods: while ST-SB allows for identifying clusters of high case density, ST-DB allows for identifying clusters of high disease risk.

#### 4.4 Simulation study

As Cali had little change in population during our study period, we conduct a simulation study in addition to our case study in Cali, Colombia, to compare ST-DB and S-DB (Fig. 5). In our simulation study, we implement a hypothetical 10% population growth within two scenarios. Each scenario consists of one realization of a random process, where we distribute the 2010 population (initial population) within the neighborhoods of Cali, but add the additional population (population increase) in different ways: (1) dispersed population increase following the existing population distribution, (2) concentrated population increase, where we distribute the population increase within a circle of radius of 3.5 km in the southern part of the city. The temporal signature of the population increase is a homogeneous Poisson process with  $\lambda=2$  for both scenarios. The results from the simulation study illustrate the benefits of ST-DB under a high population increase scenario. To reduce the computational cost, we use a random subset ( $N=550$ ) of the original case data in our simulation study. As with our case study, we assess the uncertainty from population disaggregation (Fig. 5d), as well as the benefit of considering time (Fig. 5e), while using the following parameter configurations: support = {9, 12, 15, 18, 21}, and percentile = {90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.9, 99.99}.

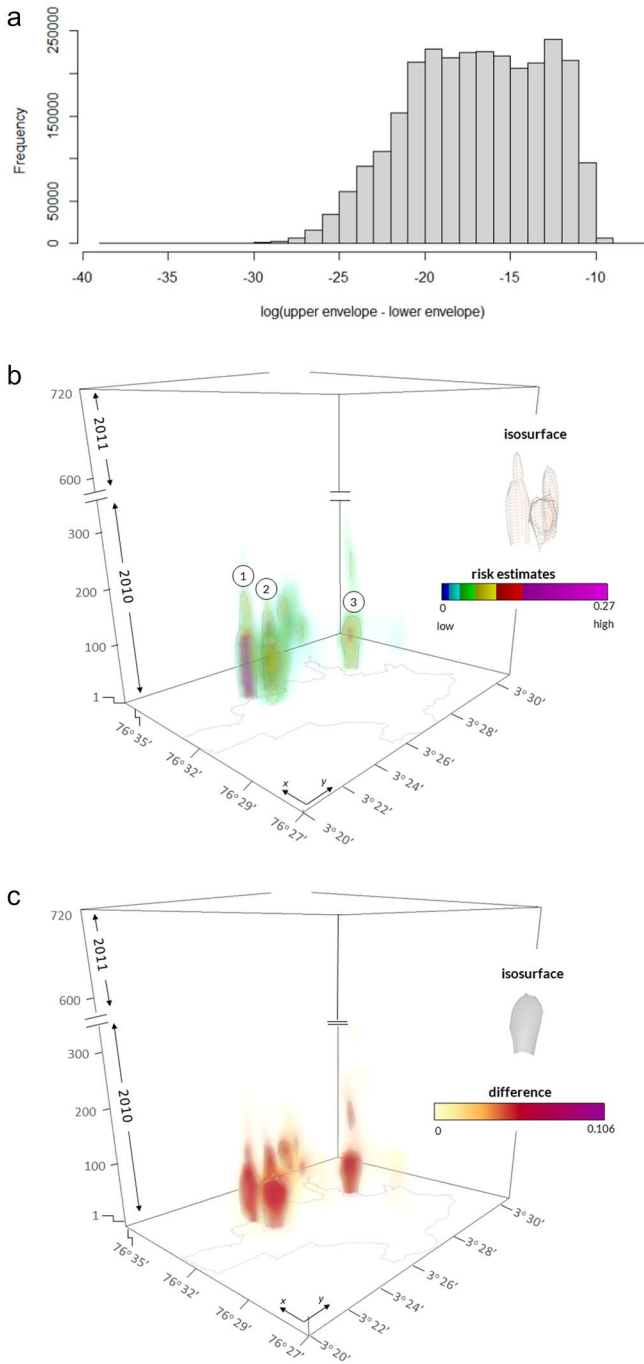
## 5 Results

### 5.1 Uncertainty from population disaggregation

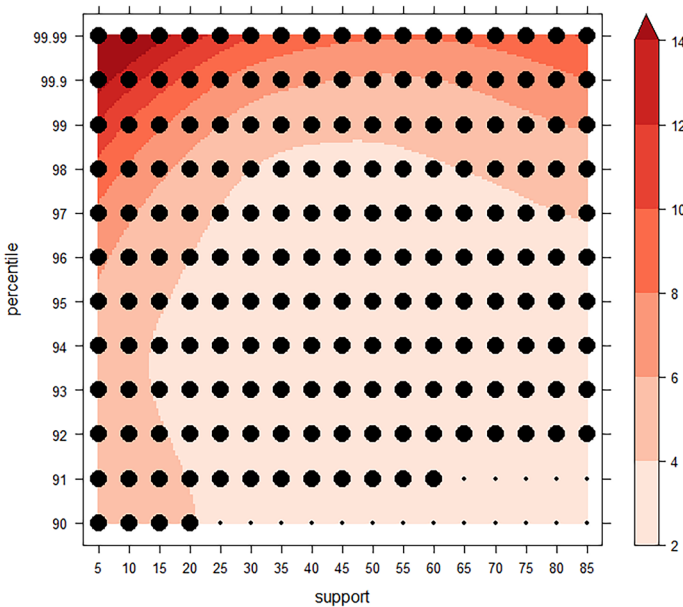
In our case study, using the observed population change and the entire dengue fever dataset, the 99 population simulations resulted in 99 risk estimates for each grid point. To illustrate the uncertainty from these simulations, we computed the difference between maximum and minimum risk value for each grid point (upper and lower envelope, support=5) and plotted their frequency within a histogram (Fig. 6a). The range of differences is 0.0–0.00037, which is a very small deviation, considering a range of risk values within 0.0–0.27. Therefore, the uncertainty from population simulation is rather small.

We plotted the upper envelope within the space–time cube (Fig. 6b) to provide a spatiotemporal depiction of the risk estimates. The lower envelope is not distinguishable from the upper envelope when viewing the scene at full extent, because the differences are very small. We can clearly see the two clusters of increased disease risk within the southwestern part of the city (Fig. 6b, points 1 & 2), commensurate with findings of Delmelle et al. (2014) and Hohl et al. (2016). These clusters are





**Fig. 6** **a** Histogram of differences between upper and lower envelope (support=5); **b** spatiotemporal distribution of the upper simulation envelope (uncertainty from population simulation); **c** difference between upper and lower simulation envelope



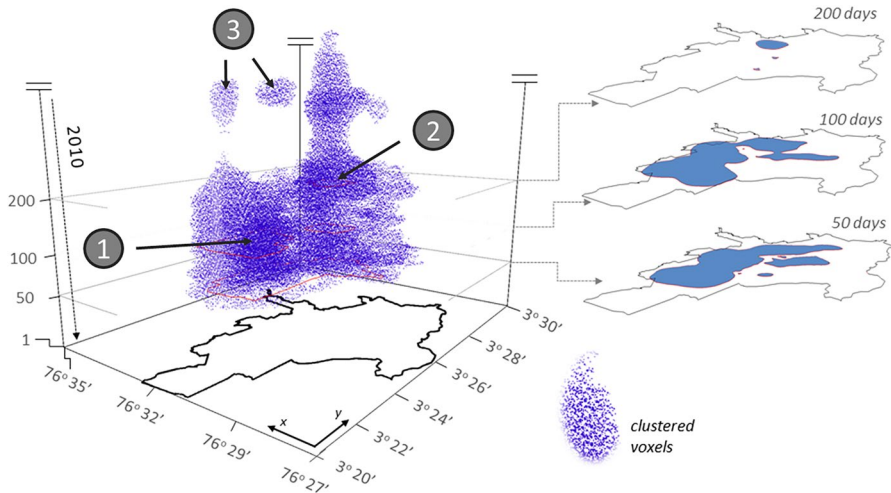
**Fig. 7** Difference between odds ratios S-DB—ST-DB. Y-axis: percentile threshold values. X-axis: support parameter values

active from the very beginning of the study period and remain so for the first quarter of the study period. We also see another risk zone within the more central part of the city (Fig. 6b, point 3) which exhibits elevated disease risk estimates for approximately the first half of the study period.

The spatiotemporal distribution of the difference between upper and lower envelopes shows higher values where risk estimates are high as well (Fig. 6c). Hence, the differences follow the distribution of risk estimates. This result is expected and confirms that the uncertainty from population simulation is relatively small while following the spatiotemporal distribution of risk.

## 5.2 Benefit of considering time

Using the actual population of Cali during our study period (2010–2011), we compare S-DB and ST-DB and find that ST-DB performs better for the parameter space we assessed. Figure 7 indicates the difference between odds ratios produced by S-DB and ST-DB. The entire parameter space in Fig. 7 shows positive values, i.e. ST-DB has a higher odds ratio than S-DB. The differences range from 2.4 to 26.9 and increase toward higher *percentile* and lower *support* threshold values. Significance testing of clusters generated using ST-DB shows that all parameter combinations are significant except toward lower *percentile* and higher *support* values, i.e.  $\text{percentile}=91$  &  $\text{support} \in \{65, 70, 75, 80, 85\}$ , as well as  $\text{percentile}=90$  &  $\text{support} \in \{25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85\}$ .

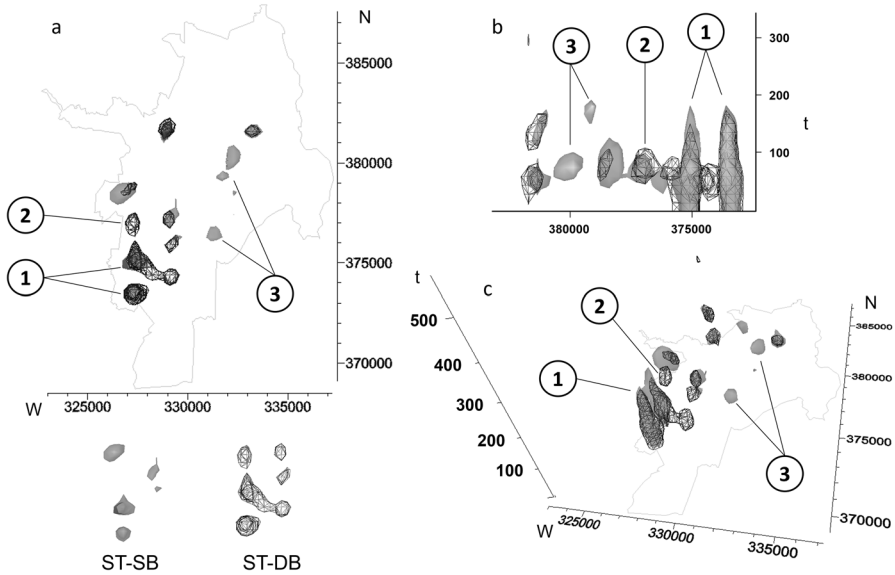


**Fig. 8** Voxels that form a significant cluster at the 0.01-level. Support=45, percentile=95

We found clusters of elevated dengue fever risk that are significant at the 0.01-level. Here, we illustrate an example using the parameter values of support=45 and percentile=95 (Fig. 8). The clustered voxels are distributed within the center of the city and within the first 314 days of the study period. The cluster has a large base at the beginning of the study period, which becomes thinner as time progresses. Therefore, distinct patterns of cluster shape are visible toward the upper end of the 314-day period. For instance, the cluster seems to consist of two parts: one in the South (Fig. 8, point 1) and in the North (Fig. 8, point 2). The Northern section of the cluster lasts substantially longer than its Southern counterpart. We are also able to make out detached ‘clouds’ of voxels that have been identified as clusters (Fig. 8, point 3). These ‘clouds’ indicate regions that experienced a resurgence of dengue risk after a period of little activity.

### 5.3 Benefit of considering a dynamic background

The density estimates produced by ST-SB exhibit an interesting pattern (Fig. 9). The isovalues are chosen to equalize the volume enclosed by the isosurfaces between the density grids of ST-DB and ST-SB. Figure 9 shows that there are three distinct areas: Point 1 has high values for both, ST-DB and ST-SB. This area corresponds to the well-known dengue clusters in the southwestern part of the city during the first 200 days of our study period (Delmelle et al. 2014). It means that these areas have high case density and a high population density, and therefore, conform to the expected spatiotemporal pattern of a disease epidemic. Point 2 is found just north of Point 1, where we observe high values for ST-DB, but not for ST-SB. It means we observe more cases than we would expect given the lower population density. Lastly, areas belonging to Point 3 are located in the central part of the city at around 50–200 days and exhibit high values for ST-SB, but not for ST-DB. They are on the



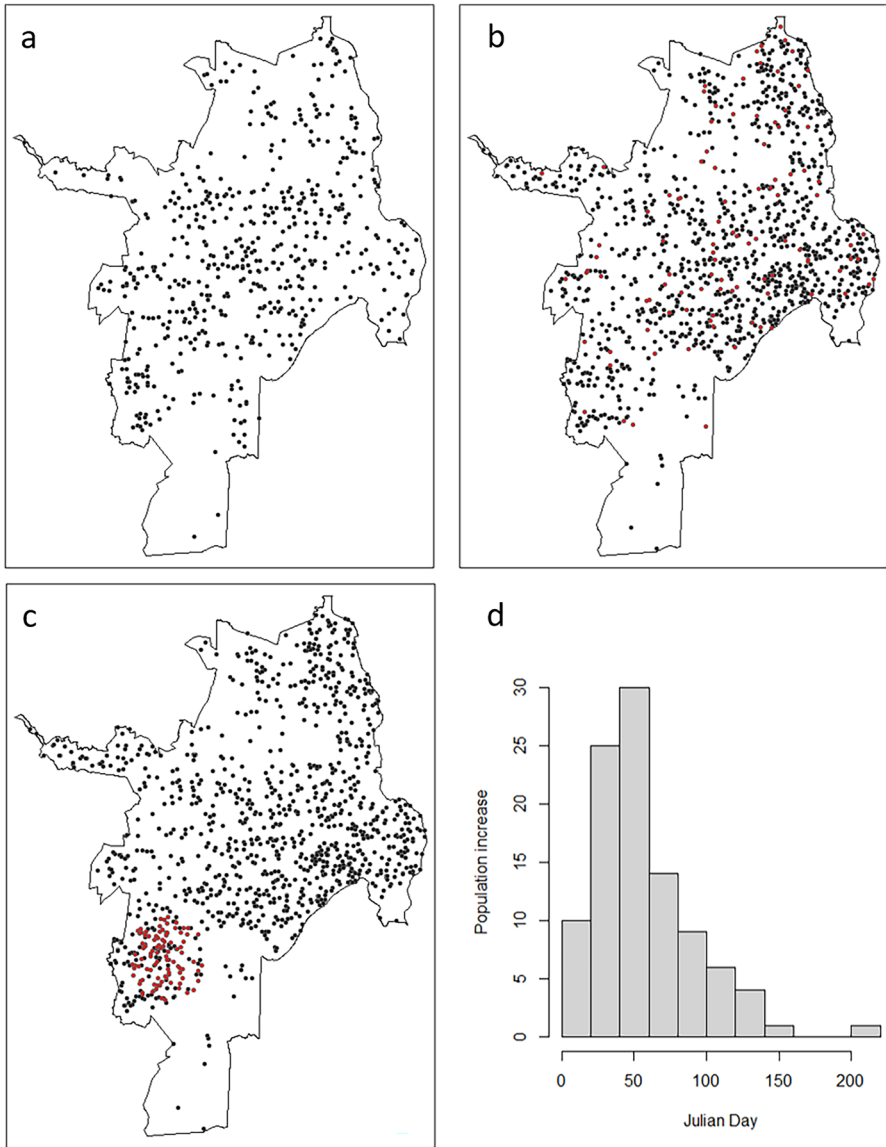
**Fig. 9** Difference between ST-DB and ST-SB

fringe of the high population density areas in the eastern part of the city. As they are not visible when adjusting for the underlying population (ST-DB), these areas are within or below the expected pattern, yet using ST-SB alone, one would identify them as high-density areas.

#### 5.4 Simulation study

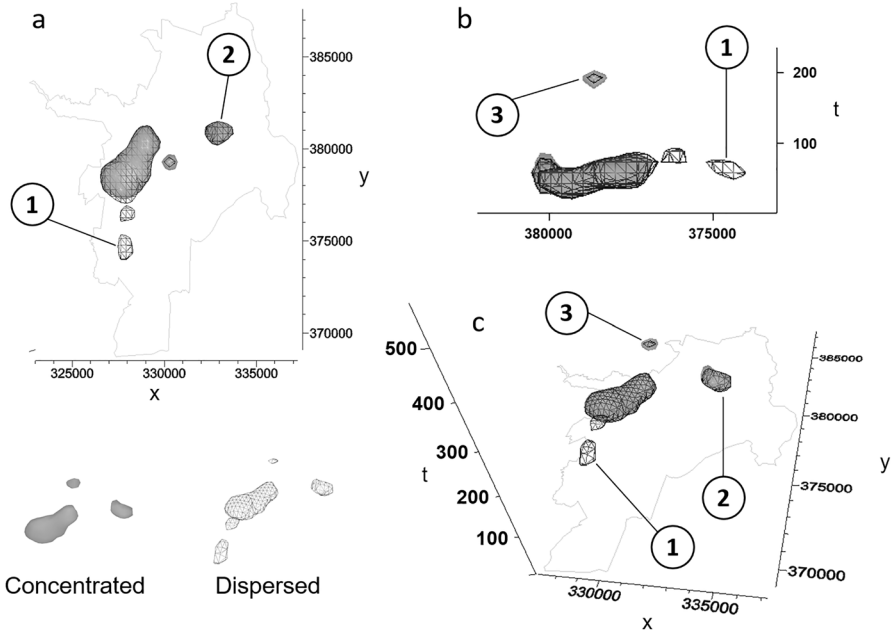
The simulation study uses a random subset of the original dengue fever dataset of 550 points (Fig. 10a). We employ two scenarios that differ in the spatial pattern of population increase: dispersed (Fig. 10b), and concentrated (Fig. 10c). The temporal signature of population change is the same for both scenarios (Fig. 10d), exhibiting the strongest growth in the first half of the first year of our study period (2010).

It is apparent that the risk estimates of dengue fever cases resulting from ST-DB differ between the two simulation scenarios (Fig. 11). Figure 11 includes an isosurface for each scenario, where we chose isovalues in such a way that both isosurfaces enclose the same volume. This is different from choosing the same isovalue because the ranges of risk values differ. The concentrated population increase leads to lower risk estimates in the southwestern part of the city (see Fig. 11, Point 1) as compared to the estimates of the dispersed population increase scenario. This is evident from the isosurface of the dispersed scenario in Point 1, while the isosurface of the concentrated scenario is absent in that location. For both scenarios, the highest densities are found during the first 100 days of the study period, with a spatial concentration on the western parts of the city. Notable exceptions are two detached clusters in the central area of Cali, one emerging around day 100 (Fig. 11, point 2), one around day 200 (Fig. 11, point 3).

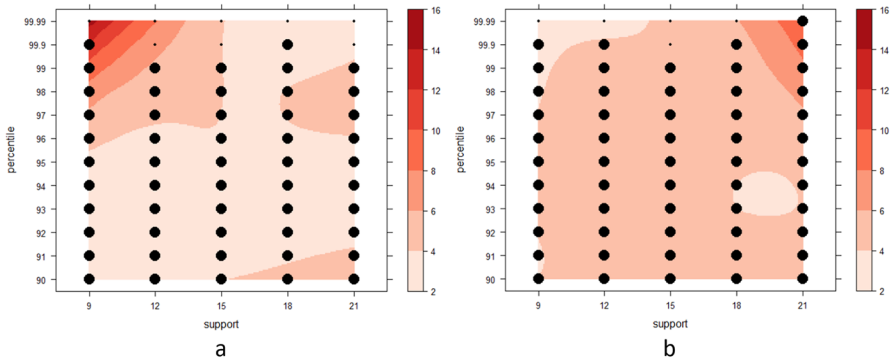


**Fig. 10** Simulation study—population simulation. **a** 550 cases (random sample from dengue fever dataset), **b** Simulated population dataset with initial population (black dots) and dispersed population increase (red dots), **c** Simulated population dataset with initial population (black dots) and concentrated population increase (red dots), **d** temporal signature of population increase, common to both population increase scenarios

Comparing odds ratios between ST-DB and S-DB, we find that ST-DB outperforms S-DB for the entire parameter space assessed (Fig. 12), as indicated by the positive difference all throughout. Significance testing shows that most



**Fig. 11** Risk estimates from ST-DB (support=9). Isosurfaces of concentrated (gray color) and distributed population increase (mesh). View from **a** top, **b** West, **c** Southeast (elevated viewpoint). One simulation run out of 99 (population simulation)



**Fig. 12** Simulation study—case simulation. Heatmap of difference between odds ratios of ST-DB and S-DB. Large dots indicate parameter configurations that resulted in significant clusters, small dots denote insignificant clusters. **a** Concentrated population increase, **b** dispersed population increase. X-axis: support parameter values. Y-axis: percentile threshold values

clusters are significant, with few exceptions i.e. for extreme values of the *percentile* threshold parameter (percentile = 99.99).

## 6 Discussion

Our experiments indicate that ST-DB yields higher odds ratios than S-DB for the parameter configurations assessed in both, our simulation and case studies (Figs. 7, 12). This is a positive result, suggesting a benefit of considering the temporal dimension in spatial analysis, as it may better allow us to capture the scale at which a point process operates. This implies that ST-DB, the key innovation of this study, improves our ability to delineate clusters of disease occurrence under spatially and temporally dynamic backgrounds. It is important to note that the ‘winner’ of the comparison between ST-DB and S-DB is not ‘better’ in a universal sense as it may merely result in different conclusions.

The choice of parameters is admittedly subjective, but we confirmed the validity of the resulting clusters by significance testing. Therefore, we created two measures of describing clusters: (1) We quantify the strength of a cluster by its odds ratio. The higher the ratio, the greater the difference in odds of contracting the disease inside vs. outside the cluster. (2) We quantify the significance of the cluster by its  $p$ -value. Therefore, the clustering of observed dengue cases by arbitrary parameter values generates higher odds ratios compared to all the Poisson distributed simulated datasets. The ability of choosing parameter values allows for ‘tuning’ the resulting significant clusters (Fig. 8). This way, authorities can choose the case support and risk threshold based on their needs and resources. If resources for disease prevention and mitigation are scarce, the risk threshold can be increased to produce smaller areas/ periods of significant clusters.

We clearly saw that uncertainty from population disaggregation was low in our case study (Fig. 6a, c). This result is due to the small rates of population change within the study period (Table 1). To clearly illustrate the utility of ST-DB we carried out a simulation study with two different scenarios of rapidly changing background population (concentrated vs dispersed population increase). The results of the simulation study demonstrate that the spatiotemporal distribution of the background does have a substantial effect on the resulting risk estimates. This effect is visually detectable and apparent from Fig. 11. In addition, the distribution of the background has an effect on the benefit of considering the temporal dimension in our analysis: In both simulation scenarios, ST-DB produces higher odds ratios than S-DB across all parameter configurations, where the difference between odds ratios increases toward extreme parameter values. However, while differences increase toward higher percentile and lower *support* threshold values in the concentrated scenario, they increase toward higher percentile and higher *support* in the dispersed scenario. This shows us which parameter configuration we may choose to maximize the benefit of considering the temporal dimension of our data for different distributions of the background.

The comparison between space–time kernel density estimation for dynamic (ST-DB) and static (ST-SB) backgrounds underlines the usefulness of ST-DB. Health officials may be interested in the ability to identify areas that conform to the expectation of *high population density equals high case density* versus areas that do not.

High risk areas/times identified through ST-DB could be targeted for interventions like mosquito control or awareness campaigns in the case of dengue fever.

The results obtained here point toward the following weaknesses and discussion points, some of which need to be addressed in the future. First, the observed population change in the city of Cali was moderate, therefore we simulated two high population growth scenarios to show how ST-DB performs under such conditions. However, more simulation studies are needed to demonstrate the benefit of considering the temporal dimension, e.g., under a population decrease scenario. Second, ST-DB assumes that cases and population are distributed on an infinitely continuous planar space, which justifies the use of Euclidean distance. However, as people and goods move along the road network, it is necessary to adapt ST-DB toward network distance, drawing from existing research about kernel density estimation for networks (Okabe et al. 2009), space–time hotspot detection for street-level incidents (Shiode and Shiode 2013), and local indicators of network-constrained clusters (Yamada and Thill 2007). Third, we use epidemiological data under the assumption that people contracted the disease at their residential location. However, this is not necessarily true, as people move around the city for daily commutes or leisure time activities. Therefore, uncertainty in the spatiotemporal location of disease cases could undermine our results. Lastly, our use of the ‘people-days’ metric to quantify the population within the kernel may face the following weakness: a given number of people-days may be achieved by multiple ways that represent different conditions for disease transmission. For instance, 12 people present in a kernel for one hour each has the same result like one person present for 12 h. Hence, domain knowledge is key to establish a measure that best suits the purpose at hand.

We think that ST-DB is most useful for scenarios where the phenomenon of interest, as well as the background varies considerably in space and time, which depends on the scale and spatiotemporal extent of the study area/period, as well as data availability. Analyzing infectious disease at the municipal or regional level is a prime example of such a situation. Other application examples include chronic disease in rapidly urbanizing regions, or traffic accidents under varying traffic volumes. On the other hand, we think that ST-DB is likely not useful if the background is relatively static. For instance, many western cities are at a stage in their development where population density has not changed substantially for years, which would diminish the benefits of ST-DB. Lastly, while ST-DB is suitable for municipal- or regional-level analyses over multiple years, it is less suitable for global- or continental-level analyses over decades, because spatiotemporal data tend to be coarse at such scales, thereby reducing variation in geographic phenomena across space and time.

## 7 Conclusions

In this study, we presented ST-DB, a kernel density estimator for spatially *and* temporally dynamic background populations. Our approach demonstrated the utility of accounting for a dynamic background for kernel density estimation by comparing it to an estimator that ignores the temporal dimension (S-DB). To the best of our knowledge, this approach is the first within the GIScience domain



that explicitly accounts for a changing background population through time. We showed that accounting for temporal variation in the background in addition to spatial variation (ST-DB) may produce stronger clusters (as expressed in odds ratios) than an approach that only accounts for spatial variation (S-DB) for some parameter configurations.

Our approach addresses the limitations of current methods by adjusting kernel density estimates for dynamic backgrounds. We need to revisit and update current methods for analyzing spatiotemporal phenomena to explicitly incorporate temporal variation in background population. In the face of current migration and urbanization trends, as well as the availability of population data at high spatiotemporal resolution, this step has been long overdue.

Future work includes various improvements and applications of kernel density estimation for spatially and temporally dynamic backgrounds: first and foremost, it should be a priority to address some of the limitations mentioned here, including additional simulation studies to confirm the general validity of our findings. Second, we plan to develop a parallel version of ST-DB, which allows harnessing the processing power of high-performance computing and increases applicability of the method. Third, new methods for kernel density estimation under dynamic backgrounds are needed, which are able to accommodate various formats and conceptualizations of population data (Fig. 1). This will require further thinking about issues of scale and granularity of spatiotemporal analyses, as data types range from individual trajectories to population counts that are aggregated to the spatiotemporal units of census tracts and decades.

**Authors' contributions** Alexander Hohl: Conception, literature review, code implementation, data analysis, results interpretation, manuscript writing, project oversight. Eric Delmelle: Artwork, consulting, manuscript editing. Irene Casas: Data provider, artwork, manuscript writing, manuscript editing. Wenwu Tang: Consulting. Xun Shi: Consulting.

**Funding** Open access funding provided by University of Eastern Finland (UEF) including Kuopio University Hospital. This research has not been funded.

**Availability of data and material** The dengue fever dataset cannot be made publicly available as we do not have an agreement with the provider to share it. A detailed description of its content and format, which should be sufficient for applying ST-IB to other spatiotemporal point datasets, as well as all other data and codes used in this study are available in [<https://figshare.com/s/7ff3db21fe2a10a39356>].

**Code availability** All codes for this research are available in [<https://figshare.com/s/7ff3db21fe2a10a39356>].

## Declarations

**Conflict of interest** No conflict of interest/competing interests are declared.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the

material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Anselin L (2011) From SpaceStat to CyberGIS: twenty years of spatial data analysis software. *Int Reg Sci Rev* 35:131–157
- Armitage Cadavid M (2017) Cali en cifras, at Cali, Colombia
- Bach B, Dragicevic P, Archambault D, Hurter C, Carpendale S (2016) A descriptive framework for temporal data visualizations based on generalized space–time cubes. *Comput Graph Forum* 36(6):36–61. <https://doi.org/10.1111/cgf.12804>
- Bailey T, Gatrell A (1995) *Interactive spatial data analysis*. Pearson Education Limited, Edinburgh Gate
- Balk D, Deichmann U, Yetman G, Pozzi F, Hay S, Nelson A (2006) Determining global population distribution: methods, applications and data. *Adv Parasitol* 62:119–156. [https://doi.org/10.1016/S0065-308X\(05\)62004-0](https://doi.org/10.1016/S0065-308X(05)62004-0)
- Bhaduri B, Bright E, Coleman P, Urban ML (2007) LandScan USA: a high-resolution geospatial and temporal modeling approach for population distribution and dynamics. *GeoJournal* 69(1–2):103–117. <https://doi.org/10.1007/s10708-007-9105-9>
- Bhopal RS (2016) *Concepts of epidemiology: integrating the ideas, theories, principles, and methods of epidemiology*. Oxford University Press, Oxford
- Bithell J (2000) A classification of disease mapping methods. *Stat Med* 19(17–18):2203–2215. [https://doi.org/10.1002/1097-0258\(20000915/30\)19:17/18%3C2203::AID-SIM564%3E3.0.CO;2-U](https://doi.org/10.1002/1097-0258(20000915/30)19:17/18%3C2203::AID-SIM564%3E3.0.CO;2-U)
- Bithell JF (1990) An application of density estimation to geographical epidemiology. *Stat Med* 9(6):691–701. <https://doi.org/10.1002/sim.4780090616>
- Bland JM, Altman DG (2000) The odds ratio. *BMJ* 320(7247):1468. <https://doi.org/10.1136/bmj.320.7247.1468>
- Bowman AW, Azzalini A (1997) *Applied smoothing techniques for data analysis: the kernel approach with S-Plus illustrations*. Oxford University Press, Oxford
- Boyandin I, Bertini E, Lalanne D (2012) A qualitative study on the exploration of temporal changes in flow maps with animation and small-multiples. *Comput Graph Forum* 31(3):1005–1014. <https://doi.org/10.1111/j.1467-8659.2012.03093.x>
- Brunsdon C (1995) Estimating probability surfaces for geographical point data: an adaptive kernel algorithm. *Comput Geosci* 21(7):877–894. [https://doi.org/10.1016/0098-3004\(95\)00020-9](https://doi.org/10.1016/0098-3004(95)00020-9)
- Brunsdon C, Corcoran J, Higgs G (2007) Visualizing space and time in crime patterns: a comparison of methods. *Comput Environ Urban Syst* 31:52–75. <https://doi.org/10.1016/j.compenvurbsys.2005.07.009>
- Cali DAdPdMdSd (2019) *Proyecciones de población Cali 2006–2036 escenario bajo 2017* [cited April 2019]. <https://planeacion.cali.gov.co/informacionestadisticacali/?dir=Demografia>
- Cali S (2010) *Historia del dengue en Cali. Endemia o una continua epidemia*. Cali: Secretaria de Salud Publica Municipal de Cali
- Carlos H, Shi X, Sargent J, Tanski S, Berke E (2010) Density estimation and adaptive bandwidths: a primer for public health practitioners. *Int J Health Geogr* 9(1):39. <https://doi.org/10.1186/1476-072X-9-39>
- Casas I, Delmelle E, Varela A (2010) A space–time approach to diffusion of health service provision information. *Int Reg Sci Rev* 33(2):134–156. <https://doi.org/10.1177/0160017609354760>
- Castles S, De Haas H, Miller MJ (2013) *The age of migration: International population movements in the modern world*. Palgrave Macmillan
- Champion T (2001) Urbanization, suburbanization, counterurbanization and reurbanization. *Handb Urban Stud* 160:1
- Chen J, Shaw S-L, Yu H, Lu F, Chai Y, Jia Q (2011) Exploratory data analysis of activity diary data: a space–time GIS approach. *J Transp Geogr* 19(3):394–404. <https://doi.org/10.1016/j.jtrangeo.2010.11.002>

- Coleman M, Coleman M, Mabuza AM, Kok G, Coetzee M, Durrheim DN (2009) Using the SaTScan method to detect local malaria clusters for guiding malaria control programmes. *Malar J* 8(1):68. <https://doi.org/10.1186/1475-2875-8-68>
- Cressie N, Wikle CK (2015) *Statistics for spatio-temporal data*. Wiley, New York
- Davies TM, Hazelton ML (2010) Adaptive kernel estimation of spatial relative risk. *Stat Med* 29(23):2423–2437. <https://doi.org/10.1002/sim.3995>
- Davies TM, Jones K, Hazelton ML (2016) Symmetric adaptive smoothing regimens for estimation of the spatial relative risk function. *Comput Stat Data Anal* 101:12–28. <https://doi.org/10.1016/j.csda.2016.02.008>
- Demšar U, Buchin K, Cagnacci F, Safi K, Speckmann B, Van de Weghe N, Weibel R (2015) Analysis and visualisation of movement: an interdisciplinary review. *Mov Ecol* 3(1):1–24
- Delmelle E, Casas I, Rojas JH, Varela A (2013) Spatio-temporal patterns of Dengue Fever in Cali, Colombia. *Int J Appl Geospat Res (IJAGR)* 4(4):58–75. <https://doi.org/10.4018/jagr.2013100104>
- Delmelle E, Dony C, Casas I, Jia M, Tang W (2014) Visualizing the impact of space–time uncertainties on dengue fever patterns. *Int J Geogr Inf Sci* 28(5):1107–1127. <https://doi.org/10.1080/13658816.2013.871285>
- Desjardins MR, Hohl A, Delmelle EM (2020) Rapid surveillance of COVID-19 in the United States using a prospective space–time scan statistic: Detecting and evaluating emerging clusters. *Appl Geogr* 118:102202. <https://doi.org/10.1016/j.apgeog.2020.102202>
- Desjardins MR, Hohl A, Griffith A, Delmelle E (2019) A space–time parallel framework for fine-scale visualization of pollen levels across the Eastern United States. *Cartogr Geogr Inf Sci* 46(5):428–440
- Dobson JE, Bright EA, Coleman PR, Durfee RC, Worley BA (2000) LandScan: a global population database for estimating populations at risk. *Photogramm Eng Remote Sens* 66(7):849–857
- Eicher CL, Brewer CA (2001) Dasymmetric mapping and areal interpolation: Implementation and evaluation. *Cartogr Geogr Inf Sci* 28(2):125–138. <https://doi.org/10.1559/152304001782173727>
- Epanechnikov VA (1969) Non-parametric estimation of a multivariate probability density. *Theory Prob Appl* 14(1):153–158. <https://doi.org/10.1137/1114019>
- Fotheringham AS, Yang W, Kang W (2017) Multiscale geographically weighted regression (MGWR). *Ann Am Assoc Geogr* 107(6):1247–1265. <https://doi.org/10.1080/24694452.2017.1352480>
- Gao S (2015) Spatio-temporal analytics for exploring human mobility patterns and urban dynamics in the mobile age. *Spat Cogn Comput* 15(2):86–114. <https://doi.org/10.1080/13875868.2014.984300>
- Goodchild MF (2001) Metrics of scale in remote sensing and GIS. *Int J Appl Earth Obs Geoinf* 3(2):114–120. [https://doi.org/10.1016/S0303-2434\(01\)85002-9](https://doi.org/10.1016/S0303-2434(01)85002-9)
- Hagerstrand T (1970) What about people in regional science? *Pap Reg Sci Assoc* 24:7–21
- Hohl A, Delmelle E, Tang W (2015) Spatiotemporal domain decomposition for massive parallel computation of space–time kernel density. *ISPRS Ann Photogramm Remote Sens Spatial Inf Sci*. DOI: 10.5194/isprsannals-II-4-W2-7-2015
- Hohl A, Delmelle E, Tang W, Casas I (2016) Accelerating the discovery of space–time patterns of infectious diseases using parallel computing. *Spatial Spatio Temporal Epidemiol* 19:10–20. <https://doi.org/10.1016/j.sste.2016.05.002>
- Huang X, Li Z, Jiang Y, Li X, Porter D (2020) Twitter reveals human mobility dynamics during the COVID-19 pandemic. *PLoS One* 15(11):e0241957
- Jacquez GM (1996) A k nearest neighbour test for space–time interaction. *Stat Med* 15(18):1935–1949. [https://doi.org/10.1002/\(SICI\)1097-0258\(19960930\)15:18%3C1935::AID-SIM406%3E3.0.CO;2-I](https://doi.org/10.1002/(SICI)1097-0258(19960930)15:18%3C1935::AID-SIM406%3E3.0.CO;2-I)
- Jacquez GM, Jacquez JA (1999) Disease clustering for uncertain locations. In: Lawson A, Bertollini R (eds) *Disease mapping and risk assessment for public health decision making*. Wiley, London
- Kang Y, Gao S, Liang Y, Li M, Rao J, Kruse J (2020) Multiscale dynamic human mobility flow dataset in the US during the COVID-19 epidemic. *Sci Data* 7(1):1–13
- Koo H, Lee M, Chun Y, Griffith DA (2020) Space–time cluster detection with cross-space–time relative risk functions. *Cartogr Geogr Inf Sci* 47(1):67–78. <https://doi.org/10.1080/15230406.2019.1641149>
- Kulldorff M (1997) A spatial scan statistic. *Commun Stat Theory Methods* 26(6):1481–1496. <https://doi.org/10.1080/03610929708831995>
- Kulldorff M (2010) SaTScan-software for the spatial, temporal, and space–time scan statistics. Harvard Medical School and Harvard Pilgrim Health Care, Boston
- Kulldorff M, Heffernan R, Hartman J, Assunção R, Mostashari F (2005) A space–time permutation scan statistic for disease outbreak detection. *PLoS Med* 2(3):e59. <https://doi.org/10.1371/journal.pmed.0020059>

- Kwan M-P (2000) Interactive geovisualization of activity-travel patterns using three-dimensional geographical information systems: a methodological exploration with a large data set. *Trans Res Part C Emerg Technol* 8(1–6):185–203. [https://doi.org/10.1016/S0968-090X\(00\)00017-6](https://doi.org/10.1016/S0968-090X(00)00017-6)
- Kwan M-P (2004) GIS methods in time-geographic research: Geocomputation and geovisualization of human activity patterns. *Geogr Ann B* 86:205–218. <https://doi.org/10.1111/j.0435-3684.2004.00167.x>
- Lang RE, Simmons PA (2003) Boomburbs: the emergence of large, fast-growing suburban cities. In: *Redefining urban and suburban America: evidence from census 2000*
- Li L, Losser T, Yorke C, Piltner R (2014) Fast inverse distance weighting-based spatiotemporal interpolation: a web-based application of interpolating daily fine particulate matter pm<sub>2.5</sub> in the contiguous us using parallel programming and kd tree. *Int. J. Environ. Res. Public Health* 11(9):9101–9141. <https://doi.org/10.3390/ijerph110909101>
- Luo L, McLafferty S, Wang F (2010) Analyzing spatial aggregation error in statistical models of late-stage cancer risk: a Monte Carlo simulation approach. *Int J Health Geogr* 9(1):51. <https://doi.org/10.1186/1476-072X-9-51>
- Malleson N, Andresen MA (2015a) Spatio-temporal crime hotspots and the ambient population. *Crime Sci* 4(1):1–8. <https://doi.org/10.1186/s40163-015-0023-8>
- Malleson N, Andresen M (2015b) The impact of using social media data in crime rate calculations: shifting hot spots and changing spatial patterns. *Cartogr Geogr Inf Sci* 42(2):112–121. <https://doi.org/10.1080/15230406.2014.905756>
- Martin SF (2000). Forced migration and the evolving humanitarian regime. The UNHCR EPUA Working Papers, 20. <https://www.unhcr.org/en-us/research/working/3ae6a0ce4/forced-migration-evolving-humanitarian-regime-susan-f-martin.html>
- Meentemeyer RK, Tang W, Dorning MA, Vogler JB, Cunniffe NJ, Shoemaker DA (2013) FUTURES: multilevel simulations of emerging urban–rural landscape structure using a stochastic patch-growing algorithm. *Ann Assoc Am Geogr* 103(4):785–807. <https://doi.org/10.1080/00045608.2012.707591>
- Mennis J (2003) Generating surface models of population using dasymetric mapping. *Prof Geogr* 55(1):31–42. <https://doi.org/10.1111/0033-0124.10042>
- Mitchell MI (2011) Insights from the cocoa regions in Côte d’Ivoire and Ghana: rethinking the migration–conflict nexus. *Afr Stud Rev* 54(2):123–144
- Münz R (2007) Migration, labor markets, and integration of migrants: an overview for Europe: HWWI policy paper. <http://hdl.handle.net/10419/47671>
- Nakaya T, Yano K (2010) Visualising crime clusters in a space–time cube: an exploratory data-analysis approach using space–time kernel density estimation and scan statistics. *Trans GIS* 14(3):223–239. <https://doi.org/10.1111/j.1467-9671.2010.01194.x>
- Okabe A, Satoh T, Sugihara K (2009) A kernel density estimation method for networks, its computational method and a GIS-based tool. *Int J Geogr Inf Sci* 23(1):7–32. <https://doi.org/10.1080/13658810802475491>
- Riley S (2007) Large-scale spatial-transmission models of infectious disease. *Science* 316(5829):1298–1301. <https://doi.org/10.1126/science.1134695>
- Sain SR (2002) Multivariate locally adaptive density estimation. *Comput Stat Data Anal* 39(2):165–186. [https://doi.org/10.1016/S0167-9473\(01\)00053-6](https://doi.org/10.1016/S0167-9473(01)00053-6)
- Salathé M, Kazandjieva M, Lee JW, Levis P, Feldman MW, Jones JH (2010) A high-resolution human contact network for infectious disease transmission. *Proc Natl Acad Sci* 107(51):22020–22025. <https://doi.org/10.1073/pnas.1009094108>
- Shaw SL, Yu H, Bombom LS (2008) A space–time GIS approach to exploring large individual-based spatiotemporal datasets. *Trans GIS* 12(4):425–441. <https://doi.org/10.1111/j.1467-9671.2008.01114.x>
- Shi X (2009) A geocomputational process for characterizing the spatial pattern of lung cancer incidence in New Hampshire. *Ann Assoc Am Geogr* 99(3):521–533. <https://doi.org/10.1080/00045600902931801>
- Shi X (2010) Selection of bandwidth type and adjustment side in kernel density estimation over inhomogeneous backgrounds. *Int J Geogr Inf Sci* 24(5):643–660. <https://doi.org/10.1080/13658810902950625>
- Shi X, Wang S (2015) Computational and data sciences for health-GIS. *Ann GIS* 21(2):111–118. <https://doi.org/10.1080/19475683.2015.1027735>
- Shiode S, Shiode N (2013) Network-based space–time search-window technique for hotspot detection of street-level crime incidents. *Int J Geogr Inf Sci* 27(5):866–882. <https://doi.org/10.1080/13658816.2012.724175>

- Silverman BW (1986) Density estimation for statistics and data analysis, vol 26. CRC Press, London
- Thakur GS, Kuruganti T, Bobrek M, Killough S, Nutaro J, Liu C, Lu W (2016) Real-time urban population monitoring using pervasive sensor network. In: Proceedings of the 24th ACM SIGSPATIAL international conference on advances in geographic information systems. <https://doi.org/10.1145/2996913.2996937>
- Tiwari C, Rushton G (2005) Using spatially adaptive filters to map late stage colorectal cancer incidence in Iowa. In: Developments in spatial data handling. Springer, pp 665–676. [https://doi.org/10.1007/3-540-26772-7\\_50](https://doi.org/10.1007/3-540-26772-7_50)
- Wang K, Zhou X, Li L (2017) Disentangle crime hot spots and displacements in space and time: an analysis for Chicago from 2001 to 2016. In: Proceedings of the 1st ACM SIGSPATIAL workshop on geospatial humanities. <https://doi.org/10.1145/3149858.3149860>
- Wright JK (1936) A method of mapping densities of population: with cape cod as an example. *Geogr Rev* 26(1):103–110
- Xu L, Kwan MP, McLafferty S, Wang S (2017) Predicting demand for 311 non-emergency municipal services: an adaptive space–time kernel approach. *Appl Geogr* 89:133–141. <https://doi.org/10.1016/j.apgeog.2017.10.012>
- Yamada I, Thill JC (2007) Local indicators of network-constrained clusters in spatial point patterns. *Geogr Anal* 39(3):268–292. <https://doi.org/10.1111/j.1538-4632.2007.00704.x>
- Yuan M (2018) Human dynamics in space and time: a brief history and a view forward. *Trans GIS* 22(4):900–912. <https://doi.org/10.1111/tgis.12473>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.