A Web-based geospatial toolkit for the monitoring of Dengue Fever

[Working Paper, March 2014]

Delmelle, E. M., Zhu, H., Tang, W., & Casas, I. (2014). A web-based geospatial toolkit for the monitoring of dengue fever. Applied Geography, 52, 144-152

Abstract

The rapid propagation of vector-borne diseases, such as dengue fever, poses a threat to vulnerable populations, especially those in tropical regions. Prompt space-time analyses are critical elements for accurate outbreak detection and mitigation purposes. Open access web-based geospatial tools are particularly critical in developing countries lacking GIS software and expertise. Currently, online geospatial tools for the monitoring of surveillance data are confined to the mapping of aggregated data. In this paper, we present a web-based geospatial toolkit with a user-friendly interactive interface for the monitoring of dengue fever outbreaks, in space and time. Our geospatial toolkit is designed around the integration of (1) a spatial data management module in which epidemiologists upload spatio-temporal explicit data, (2) an analytical module running an accelerated Kernel Density Estimation (KDE) to map the outbreaks of dengue fever, (3) a spatial database module to extract pairs of disease events close in space and time and (4) a GIS mapping module to visualize space-time linkages of pairs of disease events. We illustrate our approach on a set of dengue fever cases which occurred in Cali (659 geocoded cases), an urban environment in Colombia. Results indicate that dengue fever cases are significantly clustered, but the degree of intensity varies across the city. The design and implementation of the on-line toolkit underscores the benefits of the approach to monitor vector-borne disease outbreaks in a timely manner and at different scales, facilitating the appropriate allocation of resources. The toolkit is designed collaboratively with health epidemiologists and is portable for other surveillance data at the individual level such as crime or traffic accidents.

Keywords

Dengue Fever, Accelerated Kernel Density Estimation, Space-time Query, Spatial Point Patterns, Web GIS

1. Introduction

The rapid propagation of vector-borne diseases, such as dengue fever, poses a threat to vulnerable populations, especially in tropical regions (Gubler and Trent 1993; Gubler and Clark 1995; Bhatt et al. 2013; San Martín et al. 2010). Urban and suburban environments are particularly vulnerable due to rapid population movement and the abundance of potential breeding sites. In Colombia, South America, dengue fever reemerged in the 1970s after being eradicated in the 1950s and 1960s (Romero-Vivas, Leake, and Falconar 1998; Ocampo and Wesson 2004). Ever since, the disease has become endemic, presenting periodic outbreaks in 1991, 1994, 1998, 2001, and 2006. In 2010 alone, the city of Cali suffered one of its most significant outbreaks (11,760 cases), resulting in 16 reported deaths (population of Cali for the 2006 Census was close to 2.5 million, (Cali 2008)). To facilitate the monitoring of vector-borne disease outbreaks in space and time, we develop an interactive on-line GIS toolkit which was collaboratively designed and enhanced through consultation with spatial epidemiologists in the city of Cali, Colombia.

The contributions of exploratory spatial data analysis, including point pattern and kernel density estimation (KDE), to the monitoring of vector-borne diseases are well documented in the literature (Kulldorff 1997; Eisen and Eisen 2011; Delmelle et al. 2011; Cromley and McLafferty 2011). Prompt space-time analyses are critical for accurate outbreak detection and mitigation of vector-borne diseases (Vazquez-Prokopec et al. 2009; Kitron 1998; Kitron 2000; Eisen and Eisen 2011). Spatial analytical methods can generate disease distribution maps revealing significant information in terms of direction, intensity of a disease, as well as its likelihood to spread to new regions (Duncombe et al. 2012; Yoon et al. 2012). However, recent efforts to estimate the space-time signature of vector-borne diseases have primarily been focused at the aggregate level, mainly due to the scale at which data are generally reported (Young and Jensen 2012; Hsueh, Lee, and Beltz 2012).

As underscored by Boulos and Wheeler (2007), there is an increasing interest among health communities to disseminate analytical functionality over the internet, partly due to the availability of massive epidemiological datasets (e.g. social network such as twitter, Chunara et al. 2012). Second, the participation of volunteers in mapping health information has the inherent potential to promote community involvement, ultimately improving public health (Cromley and McLafferty 2011; Skinner and Power 2011; Eisen and Eisen 2011; Eisen and Lozano-Fuentes 2009; Delmelle et al. 2013; Dickin, Schuster-Wallace, and Elliott 2014). This latter is critical in developing countries with constrained financial capabilities and where GIS expertise is limited (Fisher and Myers 2011; Duncombe et al. 2012; Kienberger et al. 2013).

The importance of collaboration between health epidemiologists and research institutions has recently been underscored by Robinson, MacEachren, and Roth (2011) and Granell, Fernández, and Díaz (2013). Several agencies such as World Health Organization (WHO); the Center for Disease Control (CDC) and the European Center for Disease Prevention and Control (ECDC) have taken significant steps towards the development of infectious disease surveillance/tracking systems on the web. An example is CDC WONDER which allows individuals to query information of a disease while results are presented through Web browsers in multiple forms (EPI 2012). DengueNet was developed by the World Health Organization to compare disease burden between countries, but the quality of the underlying data remains a challenge (WHO 2009; Duncombe et al. 2012). Huang et al. (2012) propose an application that combines datasets on modeled diseases, vector distribution and air network traffic. This application is particularly useful in an educational setting when identifying the risk posed by transportation networks to the spread of an infectious disease.

Web-based GIS applications for the storing, analysis, and visualization of epidemiological data can potentially disseminate spatial analytical concepts (and their results) to virtually anyone (Boulos and Wheeler 2007; Boulos et al. 2011; Zook et al. 2010, Chapman et al. 2013). In spatial epidemiology, Gao et al. (2008) designed an interoperable service-oriented architecture framework based on Open Geospatial Consortium (OGC) standards to share spatio-temporal disease information. Newton, Deonarine, and Wernisch (2011) developed a web application interacting with an R web-user interface to map disease locations. Highfield et al. (2011) designed Community Health Information System (CHIS), an online mapping system using a Google mapping interface to facilitate the dissemination of health-related geospatial data. Foley et al. (2010) introduced MosquitoMap, a web-based spatial database of mosquito collection records and distribution models, which can integrate geographical data from different sources at various scales. Moncrieff et al. (2013) design and implement an open-source server-side web mapping framework for the analysis of health data, relying on Open Geospatial Consortium (OGC) web map service standard. Their framework, which can handle data query, was applied to the mapping of aggregated population distribution and disease rate.

A common characteristic of these applications is that their spatial analytical capabilities are restricted to the mapping of aggregated data. A notable exception is the work by Dominkovics et al. (2011) who used a commercial geoprocessing service to generate spatial density maps of turbeculosis based on individal observations. Despite these recent technological advances, there remain critical hurdles to the effective development of Web-based GIS applications in the field of spatial epidemiology. First, the functionality that is generally available over the web is restricted to aggregate data since individual-based analysis poses computational challenges. Second, there is a lack of spatial and temporal query capabilities (Boulos 2003, Thompson et al. 2009). Last, interaction between users and the system is generally passive in that individuals cannot analyze their own data.

In this article, we present an interactive web-based GIS toolkit (OnTAPP: an **O**n-line **T**oolkit for the **A**nalysis of **P**oint **P**atterns), collaboratively designed with epidemiologists from Colombia (South America) with the objective to monitor dengue fever outbreaks over the Internet and conduct spatial analysis in a limited timeframe. Our toolkit allows to 1) analyze epidemiological information at the individual level, 2) conduct temporal and spatial query, 3) generate spatial density distribution maps across a region to better determine the occurrence of hot spots, and 4) visualize space-time connections at a local scale. Our article is structured as follows. The implementation framework is described in section 2 (Methodology), and modules to conduct space-time query and visualize events close in space and time and density surface are presented. The effectiveness of the toolkit to identify clusters of dengue fever during an outbreak is illustrated in section 3 (Monitoring Dengue Fever Outbreaks). Section 4 is devoted to conclusions and future developments.

2. Methodology

OnTAPP (**O**nline **T**oolkit for the **A**nalysis of **P**oint **P**atterns) is a web-based geospatial toolkit, which is designed in a collaborative manner with spatial epidemiologists to assist them in identifying putative sources of the diseases, and facilitate the optimal allocation of resources. The functional features of the toolkit are designed to address the needs of epidemiologists: (1) functionality to upload surveillance information and download results from the analysis, (2) ability to map patterns of raised incidence of disease at large scale (metropolitan level) and at small scale (neighborhood level), (3) identification of space-time disease clusters, and (4) functionality to identify hotspots of disease outbreaks. These functions are organized into four functional modules: geospatial analysis, mapping and visualization, spatial data management, and time and space query.

2.1 Modular Design

Figure 1 summarizes the general framework which is adopted for the design of the OnTAPP toolkit and complies with generic Web GIS architecture (Peng and Tsou 2003). The four modules mentioned above comprise the primary functionality of the server side. On the client side, users which are intended to be health professionals in charge of decision making (public health agency or epidemiologists for instance) send requests to the server side, via a Web-based graphic user interface. The server carries out the corresponding spatial and temporal functionality and conveys the results to the client side for visualization. The framework is scalable and extensible so that additional functions or map layers from external data sources can easily be added to the application, similar to Beyer, Tiwari, and Rushton (2012).

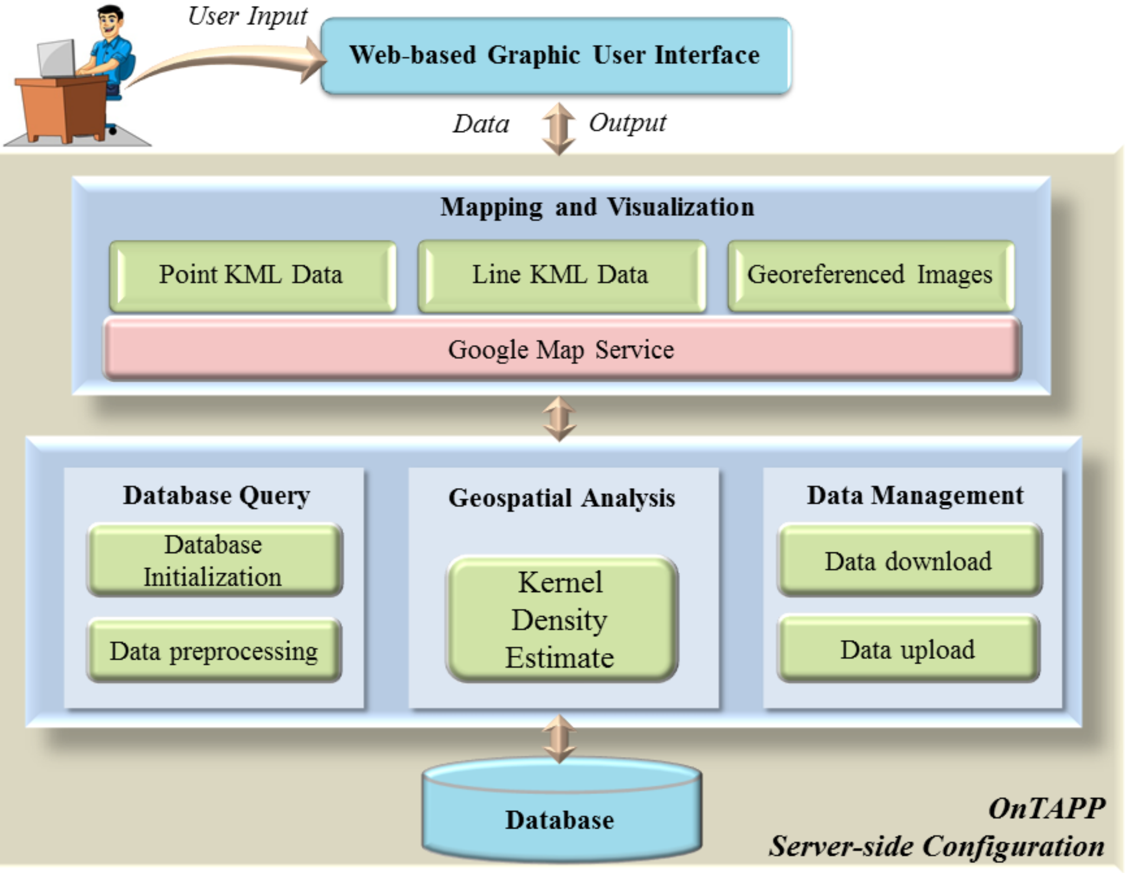


Figure 1: Web GIS Framework of OnTAPP for the analysis of spatial point patterns.

Data Management Module

Unlike other online GIS which are generally restricted to aggregated mapping and which follows a passive communication, an important feature of OnTAPP is its ability to allow users to upload their own data via a Web-based interface. Given privacy issues with health data, the users are expected to follow their organization's rules and geomask the geospatial data accordingly (see Kwan et al. 2004, Exeter et al. 2014). The data are generally composed of latitude, longitude, and time stamps. Typical format are text files, or CSV format. The data management module transforms longitude and latitude to projected coordinate system and converts the data to a KML format, which can then be downloaded by the users.

Geospatial Analysis Module

The objective of this module is to enable functionality to estimate the intensity of disease outbreaks and derive hot spots. Kernel Density Estimation (KDE) technology (Silverman 1986) is used to summarize the intensity of the geocoded disease events. The KDE algorithm generates a so-called heat map (a continuous surface image), in which the value of each grid cell reflects the intensity of the disease at that location (Bailey and Gatrell 1995; Kitron 1998; Kitron 2000; Morrison et al. 1998; Tran et al. 2004; Delmelle 2009; Peterson et al. 2009; Delmelle et al. 2011). For vector-borne diseases, KDE maps are used in conjunction with prevention and control programs to guide vector control or surveillance activities (Mammen et al. 2008; Eisen and Lozano-Fuentes 2009; Eisen and Eisen 2011).

Generic KDE. KDE is computed at each grid cell of the surface, which receives a higher weight if it has a larger number of observations in its surrounding neighborhood. Let ***s*** (located at (x, y)) be a grid location where the kernel density estimation needs to be estimated, and $\boldsymbol{s}_1 \ldots \boldsymbol{s}_n$ (located at x_i and y_i, $i = 1 \ldots n$), the locations of *n* observed events. Following Bailey and Gatrell (1995), the density $\hat{\mathrm{f}}(\mathrm{x}, \mathrm{y})$ at ***s*** is estimated by

$$\hat{\mathrm{f}}(\mathrm{x}, \mathrm{y}) = \frac{1}{h_{\mathbf{s}}^2} \sum_{\mathrm{i}} I(\mathrm{d}_{\mathbf{i}} < \mathrm{h}_{\mathbf{s}}) \mathrm{k}_{\mathbf{s}} \left(\frac{\mathrm{x} - \mathrm{x}_{\mathrm{i}}}{\mathrm{h}_{\mathbf{s}}}, \frac{\mathrm{y} - \mathrm{y}_{\mathrm{i}}}{\mathrm{h}_{\mathbf{s}}} \right) \tag{1}$$

where $I(\mathrm{d}_{\mathbf{i}} < \mathrm{h}_{\mathbf{s}})$ is an indicator function taking value 1 if $d_i < h_s$ and 0 otherwise. h_s is the search radius (or bandwidth), governing the strength of smoothing; and d_i is the distance between location ***s*** and event ***i***. The bandwidth can either be calibrated with a K-function or cross-validation (Bailey and Gatrell 1995; Delmelle 2009). The term k_s is a standardized kernel weighting function that determines the shape of the weighting function. Constraint $\mathrm{d}_{\mathbf{i}} < \mathrm{h}_{\mathbf{s}}$, indicates that only points falling within the chosen bandwidth contribute to the estimation of the kernel density at ***s***. The choice of a kernel density function may affect the computational time (Wand and Jones 1995). The KDE procedure is however more sensitive to the choice of the bandwidth and the granularity of the grid at which KDE is estimated. Larger bandwidths and smaller cell sizes (finer grid) will generate smoother surfaces, at the cost of a longer computational effort. A couple of concerns must be discussed. First, the computation time for the KDE is impacted by a larger bandwidth and larger datasets; as such; to accelerate this method we propose an accelerated KDE. Second, it is important to keep in mind that epidemiologists are generally interested in conducting the analysis with smaller bandwidths to extract locally varying patterns.

Accelerated KDE. Computational effort is a critical component to be considered when deploying a geospatial analytical tool over the web[1]. However, spatial analysis (such as KDE) can become computationally challenging in an Internet environment. Given a level of granularity (cell size), the numbers of rows and columns of a KDE image is determined when the study region is discretized along in a grid fashion. The estimated running time of the KDE algorithm is: ***row*column*n***, with ***n*** representing the total number of observed point events across the study region. Hence, a direct implementation from Equation (1) may result in an unacceptable performance within a Web environment when using large data sets since the computation requires a full search in order to compare the distance between location ***s*** and event ***i.*** To address this computational challenge, we developed an accelerated algorithm (hereafter "accelerated KDE") with the introduction of an additional constraint which reduces the computational burden. In the accelerated KDE, a virtual square window W is constructed around the center of location **s**. This window W has a bandwidth twice as large as the radius size h_s , and is tangent to the search circle. We determine whether an observation falls within W; if it does, this point is added to the estimation of the KDE function; otherwise it moves to the next point. Specifically, given the location ***s*** at $(x_s,\ y_s)$, we define the coordinates of W as:

$$left = x_s - bandwidth,\ bottom = y_s - bandwidth \tag{2}$$
$$right = x_s + bandwidth,\ top = y_s + bandwidth \tag{3}$$

For an arbitrary event (x_i, y_i), a simple comparison operation using the following logic criterion helps to determine whether the point falls within the square window

$$\mathrm{W}: \begin{cases} \boldsymbol{i} \in W & if\ (\ left < \ x_i < right)\ and\ \ (\ bottom\ < y_i < \ top\ ) \\ \boldsymbol{i} \notin \mathrm{W} & otherwise \end{cases} \tag{4}$$

To ensure that comparison is conducted in an efficient manner, data pre-processing is done in advance in which all points are sorted according to their x and y coordinates. This enables the comparison to be conducted within the window. This practice reduces the amount of computation, especially in the situation when bandwidth is small.

Mapping and Visualization Module

The objective of this module is to provide cartographic mapping support for the mapping of disease patterns. Different visuals are generated from the OnTAPP toolkit: point layers (original geocoded data and selected points from the spatial and temporal query), a line layer connecting events close with each other in space and time (shorter lines denote stronger connection among events), and an image layer reflecting the kernel density estimation. Similarly to Beyer, Tiwari, and Rushton (2012), the outputs are overlaid on geographical layers, which are Google map layers in the application (Figure 2). OnTAPP guarantees that each layer aligns with each other. All of these layers compose the output map, which allows epidemiologists to better understand the association between outbreaks of diseases and the environment.

Database Query Module

The objective of this module is to provide spatiotemporal querying functionality, following a SQL approach (Structured Query Language). The temporal query is defined by a start date and

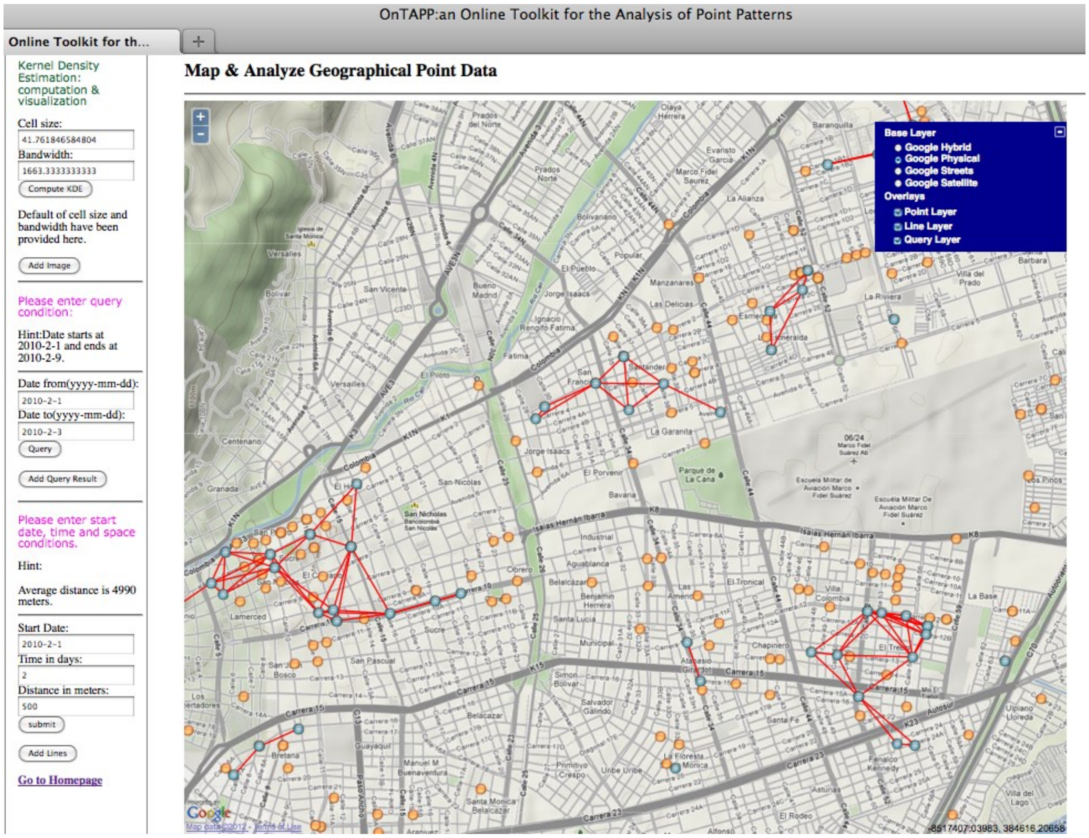
---

[1] Computation must be reduced, for instance when using Monte-Carlo simulations in a confirmatory setting.

an end date, while the spatial query identifies pairs of events separated by a certain distance. Point events that meet either condition are selected from the database and highlighted on the map. A combination of both queries (space-time) is also possible and is represented by straight line segments. This allows for the visualization of connections in space and time emphasizing locations where there is a high concentration of cases in a given space-time threshold.

2.2 Implementation

Motivated by Fisher and Myers (2011) and recommendations from epidemiologists, we use open-source technologies to implement OnTAPP. MySQL[2], an open-source relational database management system (RDBMS), manages spatial data. Most functions on the server side are implemented with PHP, a open source language for server scripting. These functions support different tasks, such as interaction with database and back-end computation. The accelerated KDE is implemented within the the Python environment on the server side. We use OpenLayers, which is a JavaScript library that provides functionality to display and render maps in web pages Figure 2 illustrates the primary user interface of the toolkit. The parameters available to conduct the spatial analysis of point patterns on the left of Figure 2 are related to KDE, temporal and spatial query.



Kernel Density Estimation: computation & visualization

Cell size: 41.761846584804

Bandwidth: 1663.3333333333

Default of cell size and bandwidth have been provided here.

Please enter query condition:

Hint:Date starts at 2010-2-1 and ends at 2010-2-9.

Date from(yyyy-mm-dd): 2010-2-1

Date to(yyyy-mm-dd): 2010-2-3

Please enter start date, time and space conditions.

Hint:

Average distance is 4990 meters.

Start Date: 2010-2-1

Time in days: 2

Distance in meters: 500

Figure 2: Client-side interface of the OnTAPP system (parameters on the left pane, and mapping environment on the right)

---

[2] MySQL was preferred to PostgreSQL and PostGIS since it comes as a default configuration on many host servers.

The performance of the accelerated KDE algorithm is tested against the generic KDE algorithm using datasets of different sizes (n=150, 250, 350, 650, 1250, 2500, 6000 and 12500 points), with 80% of the samples randomly generated in a 100*100 square, and the remaining 20% randomly generated in the lower left 0*50 square, simulating an artificial cluster. Both algorithms were tested in a Python environment on an Intel Duo Core 2.1 GHZ, and 1 GB of RAM. Running times are reported in seconds.

Figure 3 reveals a negative exponential relationship between KDE running time and grid cell size (in b). The accelerated KDE algorithm requires significantly less computational effort than the generic algorithm, especially so for smaller cell sizes. Table 1 summarizes numerical instances of the two types of KDE methods, containing the results from instances with fixed cell of size 1 but using different bandwidths and dataset size. The second section lists numerical instances when the cell size and dataset size vary but the bandwidth remains at a size of 10. The computation time of both KDE and accelerated KDE algorithms increases linearly with larger datasets, from 3.4 seconds (150 points) to 6 minutes (12,500 points). Time gain represents how much time is gained by using the accelerated KDE. The benefits of the accelerated algorithm (TimeGain) are mostly noticeable when using smaller bandwidths (Figure 3a). The column percentage improvement indicates that in some circumstances the running time can be reduced by nearly 70% (note the maximum reduction is 100% when the running time is equal to 0). When the bandwidth is larger than half the size of the study region[3], computational advantages of the accelerated algorithm tend to vanish, which is attributable to the cost of Equation (4) which compares the coordinates of a point to the grid cell where kernel density is estimated. Given these encouraging computational results, we implemented the accelerated KDE algorithm.

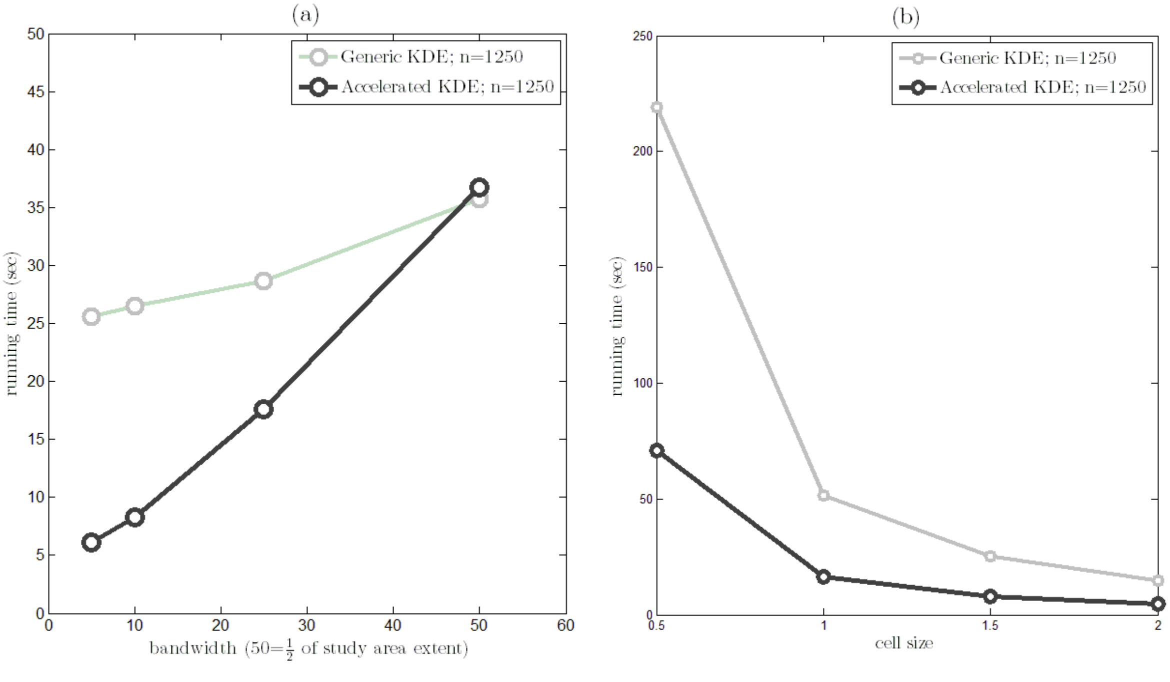


Figure 3: Running time (seconds) of the generic kernel density estimate and accelerated version, as a function of the bandwidth and the cell size for a set of *n*=1,250 points.

---

[3] Optimal bandwidths are generally determined from a spatial K-function. Large bandwidths may not be recommended as they blur the underlying point process due to an "over-smoothing" effect.

n	hs	cell size	Generic KDE (s)	Accelerated KDE (s)	TimeGain (s)	% improvement
150	5	1	3.4	1.1	2.3	67.6
150	10	1	3.7	1.5	2.2	59.5
150	25	1	3.8	2.5	1.3	34.2
150	50	1	4.6	4.7	-0.1	-2.2
1250	5	1	25.6	6.1	19.5	76.2
1250	10	1	26.5	8.3	18.2	68.7
1250	25	1	28.7	17.6	11.1	38.7
1250	50	1	35.7	36.7	-1.0	-2.8
2500	5	1	50.4	12.2	38.2	75.8
2500	10	1	51.3	16.5	34.8	67.8
2500	25	1	56.4	34.8	21.6	38.3
2500	50	1	70.1	74.7	-4.6	-6.6
12500	5	1	261.1	63.3	197.8	75.8
12500	10	1	265.3	87	178.3	67.2
12500	25	1	294.1	184	110.1	37.4
12500	50	1	362	383.5	-21.5	-5.9
150	10	0.5	14.5	6	8.5	58.6
150	10	1	3.7	1.5	2.2	59.5
150	10	1.5	1.7	0.7	1.0	58.8
150	10	2	1	0.4	0.6	60.0
1250	10	0.5	107.9	36.2	71.7	66.5
1250	10	1	26.5	8.3	18.2	68.7
1250	10	1.5	12.9	4.1	8.8	68.2
1250	10	2	7.2	2.3	4.9	68.1
2500	10	0.5	219.2	71	148.2	67.6
2500	10	1	51.3	16.5	34.8	67.8
2500	10	1.5	25.4	8	17.4	68.5
2500	10	2	14.6	4.5	10.1	69.2
12500	10	0.5	1146.3	358.4	787.9	68.7
12500	10	1	265.3	87	178.3	67.2
12500	10	1.5	130	40.8	89.2	68.6
12500	10	2	73.2	22.9	50.3	68.7

Table 1: Computational performance of the generic Kernel Density Estimation and its accelerated version, for different bandwidths (hs), cell sizes and sample sizes (n).

3. ***Monitoring Dengue Fever Outbreaks***

We illustrate OnTAPP's functionality for the visualization and exploration of vector-borne surveillance data in an urban environment. We use a geocoded dataset of dengue fever events in the city of Cali, Colombia, during an outbreak in 2010. During that time period, a total of *n*=11,760 cases were extracted from the Public Health Surveillance System (SIVIGILA). Currently, the city of Cali, Colombia produces a map on aggregated data annually at a very coarse geographical level (commune level), and a finer level (neighborhood) when deemed necessary.

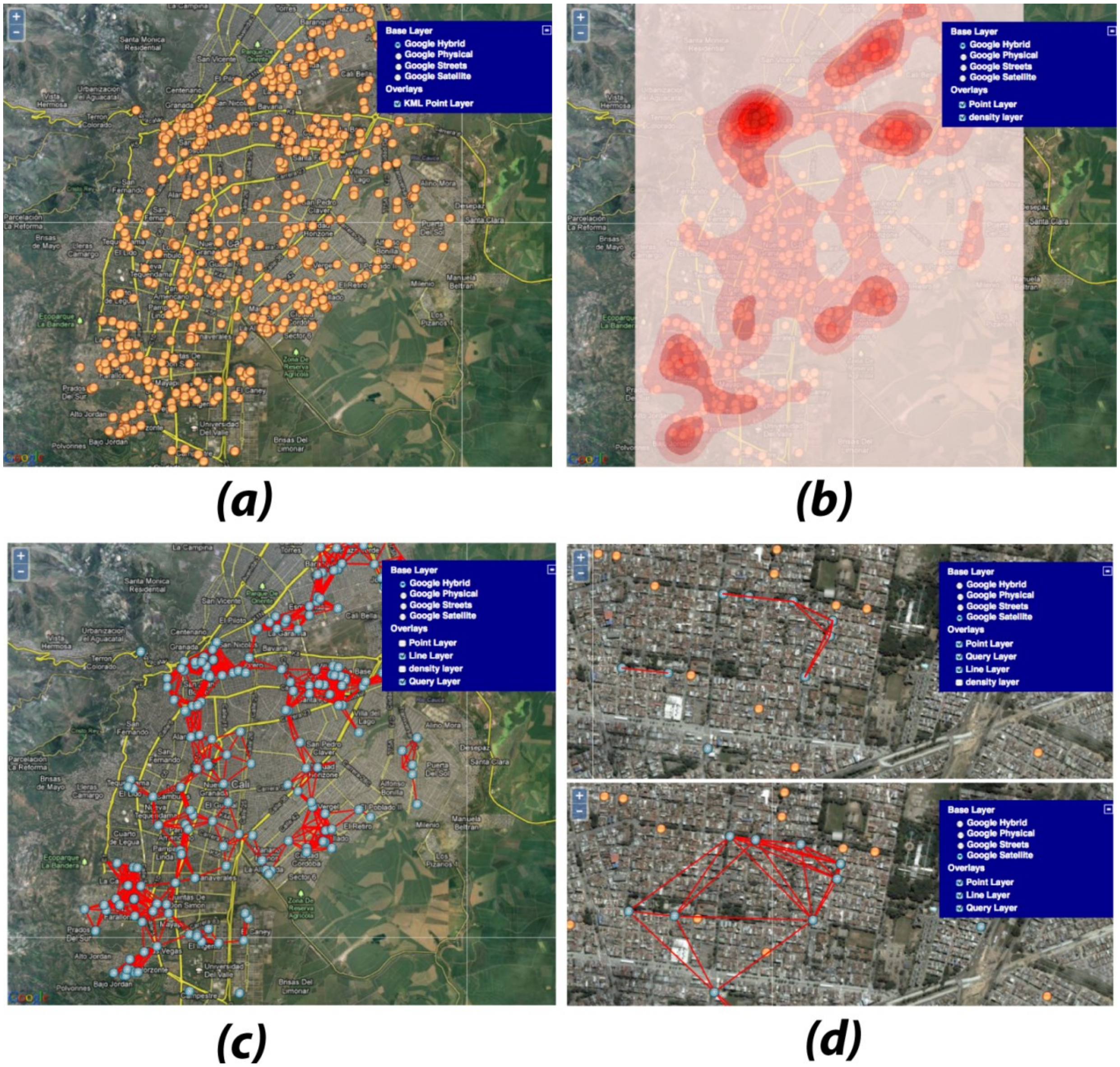


Figure 4: Illustration of spatial analysis results of OnTAPP. a): dengue fever cases for the city of Cali, Colombia for the first two weeks of February 2010 (n=659). b): kernel density estimation with a radius of 1,000 meters. c): space-time connections of 5 days and 1,000 meters. d): patients February 1-Februrary 2, 2010 are selected, with separation of 250 meters and 500 meters).

Individuals reported dengue fever symptoms at local hospitals on a daily basis (unit=Julian date). We use patients for the first two weeks of February 2010 (see Figure 4a), which corresponds to an outbreak with n=659 cases successfully geocoded and geomasked at the street intersection level for confidentiality purposes (Delmelle et al. 2013).

The Accelerated KDE algorithm is applied with spatial bandwidth h_s=1,000 meters (based on a K-function algorithm) and a cell size of 25 m. Different clusters become noticeable, in the northeast of the city along main arterial streets and populated areas (Figure 4b). Compactness and geographic extent of clusters varies significantly: a detailed description of these clusters is also given by Delmelle et al. (2013). Pairs of points separated by 5 days or less and within 1,000 meters of one another are connected together and visualized by straight line segments (Figure 4c). Epidemiologists are also interested to identify the space-time signature of a disease, for instance by using space-time linkages, at a local scale. This in turn allows identifying focal points from where the disease is spreading. The number of connections varies when the space-time constraint becomes stricter (Figure 4d). In our web application, the user can easily zoom in and out to identify local variability in the kernel density estimation and distribution of events.

4. Discussion and Conclusions

Our web-based GIS toolkit[4] can conduct exploratory spatial data analysis at individual-level point data and facilitates the discovery of underlying spatial patterns. The toolkit was specifically designed for health authorities with limited access to GIS software, and improved sequentially with feedback from epidemiologists. The results were used by local epidemiologists in the city of Cali in order to help them better understand the variation of dengue fever outbreaks at a neighborhood level. From a public health perspective, OnTAPP has a strong potential for health professionals with limited GIS access or capabilities to generate hypotheses based upon these patterns, and then conduct in-depth investigation to estimate the relationship between the spatial distribution of an infectious disease and socio-economic status after more spatially explicit information such as demographic information, income, age data etc. are added.

During the design phase of the application with epidemiologists, it became obvious that (1) there was a critical need to support cartographic solutions at a local scale level (not aggregated), which could potentially increase individual awareness of the spread of the disease. We therefore added the functionality of space-time linkages, and (2) the computational performance of the Kernel Density Estimation algorithm was accelerated by restraining the search locally. Our proposed framework complies with the generic Web GIS architecture where the client side sends requests to the server side, while the server carries out corresponding spatial analytical functions and returns the results to the client side for visualization. The benefits of our web-based toolkit were illustrated using an example of the outbreak of dengue fever, an infectious disease.
From a public health perspective, OnTAPP has a strong potential for health professionals with limited GIS access or capabilities to generate hypotheses based upon these patterns, and then conduct in-depth investigation to estimate the relationship between the spatial distribution of an infectious disease and environmental factors. In the case of the city of Cali, the Health Municipality has two objectives while monitoring and controlling dengue fever: environmental health and public health epidemiological surveillance. Currently, the environmental health dependency has an external contractor managing the system, while the public health

---

[4] A short movie highlighting the application is available at https://docs.google.com/open?id=0ByYHaCP6iioTdE1WUzFLVkl2aXM

epidemiological surveillance group consists of two to three individuals including the head epidemiologist. The staff of the epidemiological unit relies on Epi info™, an open source software distributed and developed by the Center of Disease Control (CDC) for disease surveillance. However, Epi info™ lacks strong spatial analytical methodologies, therefore shifting to the web-based toolkit presented here should be simple with basic training. The web-based toolkit will allow the health municipality staff conduct analysis that currently is not available, in real time, allowing them to make quick decisions that can aid in the control and spread of the dengue virus.

Our web-based toolkit addresses some of the critical limitations of current web-based GIS, specifically 1) improved (faster) spatial analytical capabilities at the individual level, 2) interaction between the user and the system is two-ways (users can contribute data to the server and access to the spatial analysis results) and 3) support for spatial and temporal querying capabilities. Our proposed framework is scalable and extensible; suggesting that integration for additional analytical functions or map layers from external data sources is feasible. More data layers can easily be integrated in our framework, from different sorts of data sources as Web map service. In addition, new analytical functions can be plugged in the application without modification of existing functions, such as confirmatory approaches. As an extension to our toolkit, several Monte Carlo simulations can easily be conducted to extract associated statistical significance, or p-values (Sabel et al. 2006; Delmelle et al. 2011) and these simulations can be exported to commercial GIS software. Although not reported here, we tested the significance of our results running multiple Monte-Carlo simulations of dengue fever cases. The set of Monte Carlo simulations represented dengue fever cases at the individual level, where neighborhoods of higher population density received a greater amount of cases. In all instances (except outside the city boundary), the observed KDE values were higher than the ones obtained for Monte-Carlo simulations, using similar bandwidth and cell size.

The set of online geospatial analytical tools presented in this study can be beneficial for public health professionals to conduct timely monitoring and mapping over the Web, which is particularly effective in developing countries when GIS resources are scarce. This information can be used in concert with predicted risk maps (Hongoh et al. 2012), for instance for cities that maintain a database with absence/presence of vectors. Since the toolkit only requires spatiotemporally explicit coordinates at the individual point level, its use can be applied to other disciplines such as criminology (Chainey and Ratcliffe 2005) and ecology, for instance, to determine animal home range.

We see several areas for future research. First, the classification procedure that is used to generate the image produced by the KDE algorithm assumes that the kernel density map is normally distributed. Second, other cartographic classifications algorithms (Jenks 1967) could easily be embedded. Along those lines, user interactivity can be enhanced by manually selecting the number of classes for KDE mapping. Third, introducing a bandwidth slide bar can give the user higher flexibility to discover and explore unknown patterns in the data. The fourth issue that merits further investigation is the computational performance of the accelerated KDE, for instance, through parallel computing or batch processing, which can be particularly useful when using adaptive bandwidth for non-homogeneous populations (Carlos et al. 2010). On-line geocoding tools could potentially be linked to the OnTAPP interface (Roongpiboonsopit and Karimi 2010). Additional avenues of future research include the implementation of the Accelerated Kernel Density Approach onto a network (Xie and Yan 2008). Finally, the Kernel Density Estimation does not include a temporally explicit component (temporal segments of the data can be queried, however). We are currently working on an extension of the KDE in time (Delmelle et al. 2014).

Acknowledgements

We thank both epidemiologists Dr. Alejandro Varela and Dr. Diego Calero, former and current Minister of Health for the municipality of Cali, respectively. Epidemiologist Dr. Jorge Rojas, in the Public Health Municipality City of Cali discussed the needs of the city in terms of spatial analytical functionality and provided valuable feedback throughout the design stage.

References

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

Beyer, K. M., C. Tiwari, and G. Rushton. 2012. Five Essential Properties of Disease Maps. *Annals of the Association of American Geographers* 102 (5):1067-1075.

Bhatt, S., P. W. Gething, O. J. Brady, J. P. Messina, A. W. Farlow, C. L. Moyes, J. M. Drake, J. S. Brownstein, A. G. Hoen, and O. Sankoh. 2013. The global distribution and burden of dengue. *Nature*.

Boulos, K. M. 2003. Geographic information systems and the spiritual dimension of health: a short position paper. *Int J Health Geogr* 2 (6).

Boulos, K. M., B. Resch, D. Crowley, J. Breslin, G. Sohn, R. Burtner, W. Pike, E. Jezierski, and K.-Y. Chuang. 2011. Crowdsourcing, citizen sensing and sensor web technologies for public and environmental health surveillance and crisis management: trends, OGC standards and application examples. *Int J Health Geogr* 10 (1):67.

Boulos, K. M., and S. Wheeler. 2007. The emerging Web 2.0 social software: an enabling suite of sociable technologies in health and health care education1. *Health Information & Libraries Journal* 24 (1):2-23.

Cali, A. d. S. d. 2008. Cali en cifras. Cali, Alcaldia de Cali.

Carlos, H., X. Shi, J. Sargent, S. Tanski, and E. Berke. 2010. Density estimation and adaptive bandwidths: a primer for public health practitioners. *Int J Health Geogr* 9 (1):39.

Chainey, S., and J. Ratcliffe. 2005. *GIS and crime mapping*.

Chapman, Ann LN, Thomas C. Darton, and Rachel A. Foster (2013). Managing and monitoring tuberculosis using web-based tools in combination with traditional approaches. *Clinical epidemiology* 5: 465.

Chunara, Rumi, Jason R. Andrews, and John S. Brownstein (2012). Social and news media enable estimation of epidemiological patterns early in the 2010 Haitian cholera outbreak. *The American Journal of Tropical Medicine and Hygiene* vol. 86 (1): 39-45.

Cromley, E., and S. McLafferty. 2011. *GIS and Public Health*. New York: Guilford Press.

Delmelle, E.M. 2009. Point Pattern Analysis. In *International Encyclopedia of Human Geography*, eds. R. Kitchin and N. Thrift, 204-211. Oxford: Elsevier.

Delmelle, E.M., I. Casas, J. Rojas, and A. Varela. 2013. Modeling Spatio-Temporal Patterns of Dengue Fever in Cali, Colombia. *International Journal of Applied Geospatial Research* 4 (4):58-75.

Delmelle, E.M., E. Delmelle, I. Casas, and T. Barto. 2011. H.E.L.P: a GIS-based health exploratory analysis tool for practitioners. *Applied Spatial Analysis and Policy* 4 (2):113-137.

Delmelle E.M., Jia M., Tang W., Dony C. and I. Casas (2014). Space-time visualization of dengue fever outbreaks. In: Kanaroglou P., Delmelle E. and A. Paez (Eds). *Spatial Analysis in Health Geography*, Ashgate.

Dickin, S. K., C. J. Schuster-Wallace, and S. J. Elliott. 2014. Mosquitoes & vulnerable spaces: Mapping local knowledge of sites for dengue control in Seremban and Putrajaya Malaysia. *Applied Geography* 46:71-79.

Dominkovics, P., C. Granell, A. Perez-Navarro, M. Casals, A. Orcau, and J. Cayla. 2011. Development of spatial density maps based on geoprocessing web services: application to tuberculosis incidence in Barcelona, Spain. *Int J Health Geogr* 10 (1):62.

Duncombe, J., A. Clements, W. Hu, P. Weinstein, S. Ritchie, and F. Esperanza Espino. 2012. Review: geographical information systems for dengue surveillance. *American Journal of Tropical Medicine and Hygiene* 86 (5):753.

Eisen, L., and R. Eisen. 2011. Using geographic information systems and decision support systems for the prediction, prevention, and control of vector-borne diseases. *Annual Review of Entomology* 56 (1):41-61.

Eisen, L., and S. Lozano-Fuentes. 2009. Use of mapping and spatial and space-time modeling approaches in operational control of aedes aegypti and dengue. *PLoS Negl Trop Dis* 3 (4):e411.

EPI, C.-. 2012. http://wwwn.cdc.gov/epiinfo/.

Exeter, D.J., Rodgers, S., and C.E. Sabel. 2014. "Whose data is it anyway?" The implications of putting smal area-level health and social data online. *Health Policy* 114 :88.

Fisher, R. P., and B. A. Myers. 2011. Free and simple GIS as appropriate for health mapping in a low resource setting: a case study in eastern Indonesia. *Int J Health Geogr* 10:15.

Foley, D., R. Wilkerson, I. Birney, S. Harrison, J. Christensen, and L. Rueda. 2010. MosquitoMap and the Mal-area calculator: new web tools to relate mosquito species distribution with vector borne disease. *Int J Health Geogr* 9 (1):11.

Gao, S., D. Mioc, F. Anton, X. Yi, and D. Coleman. 2008. Online GIS services for mapping and sharing disease information. *Int J Health Geogr* 7 (1):8.

Granell, C., Ó. B. Fernández, and L. Díaz. 2014. Geospatial information infrastructures to address spatial needs in health: Collaboration, challenges and opportunities. *Future Generation Computer Systems*.

Gubler, D., and D. Trent. 1993. Emergence of epidemic dengue/dengue hemorrhagic fever as a public health problem in the Americas. *Infectious Agents and Disease* 2 (6):383-393.

Gubler, D. J., and G. G. Clark. 1995. Dengue/dengue hemorrhagic fever: the emergence of a global health problem. *Emerging infectious diseases* 1 (2):55.

Highfield, L., J. Arthasarnprasit, C. Ottenweller, and A. Dasprez. 2011. Interactive web-based mapping: bridging technology and data for health. *Int J Health Geogr* 10 (1):69.

Hongoh, V., L. Berrang-Ford, M. Scott, and L. Lindsay. 2012. Expanding geographical distribution of the mosquito, *Culex pipiens*, in Canada under climate change. *Applied Geography* 33:53-62.

Hsueh, Y.-H., J. Lee, and L. Beltz. 2012. Spatio-temporal patterns of dengue fever cases in Kaoshiung City, Taiwan, 2003–2008. *Applied Geography* 34:587-594.

Huang, Zhuojie, et al (2012). Web-based GIS: the vector-borne disease airline importation risk (VBD-AIR) tool. *International journal of health geographics* 11.1: 33.

Jenks, G. 1967. The Data Model Concept in Statistical Mapping. *International Yearbook of Cartography* 7:186–190.

Kienberger, S., M. Hagenlocher, E. Delmelle, and I. Casas. 2013. A WebGIS tool for visualizing and exploring socioeconomic vulnerability to dengue fever in Cali, Colombia. *Geospatial health* 8 (1):313-316.

Kitron, U. 1998. Landscape ecology and epidemiology of vector-borne diseases: tools for spatial analysis. *Journal of Medical Entomology* 35 (4):435-445.

Kitron, U. 2000. Risk maps: transmission and burden of vector-borne diseases. *Parasitology Today* 16:324-325.

Kulldorff, M. 1997. A spatial scan statistic. *Communications in Statistics-Theory and methods* 26 (6):1481-1496.

Kwan, Mei-Po, Irene Casas, and Ben C. Schmitz (2004). Protection of geoprivacy and accuracy of spatial information: How effective are geographical masks?. *Cartographica: The International Journal for Geographic Information and Geovisualization* 39.2: 15-28.

Mammen, M., C. Pimgate, C. Koenraadt, A. Rothman, J. Aldstadt, A. Nisalak, R. Jarman, J. Jones, A. Srikiatkhachorn, C. Ypil-Butac, A. Getis, S. Thammapalo, A. Morrison, D. Libraty, S. Green, and T. Scott. 2008. Spatial and temporal clustering of dengue virus transmission in Thai villages. *PLoS Med* 5 (11):e205.

Moncrieff, S., West, G., Cosford, J., Mullan, N., & Jardine, A. (2013). An open source, server-side framework for analytical web mapping and its application to health. *International Journal of Digital Earth*: 1-22.

Morrison, A., A. Getis, M. Santiago, J. Rigau-Perez, and P. Reiter. 1998. Exploratory space-time analysis of reported dengue cases during an outbreak in Florida, Puerto Rico, 1991–1992. *American Journal of Tropical Medecine and Hygene* 58:287–298.

Newton, R., A. Deonarine, and L. Wernisch. 2011. Creating web applications for spatial epidemiological analysis and mapping in R using Rwui. *Source Code for Biology and Medicine* 6 (1):6.

Ocampo, C. B., and D. M. Wesson. 2004. Population dynamics of Aedes aegypti from a dengue hyperendemic urban setting in Colombia. *American Journal of Tropical Medicine and Hygiene* 71 (4):506-513.

Peng, Z., and M. Tsou. 2003. *Internet GIS: Distributed Geographic Information Services for the Internet and Wireless Networks*. Hoboken, NJ: John Wiley & Sons.

Peterson, I., L. Borrell, W. El-Sadr, and A. Teklehaimanot. 2009. A temporal-spatial analysis of malaria transmission in Adama, Ethiopia. *The American Journal of Tropical Medecine and Hygene* 81 (6):944-949.

Robinson, A. C., A. M. MacEachren, and R. E. Roth. 2011. Designing a web-based learning portal for geographic visualization and analysis in public health. *Health informatics journal* 17 (3):191-208.

Romero-Vivas, C., C. Leake, and A. Falconar. 1998. Determination of dengue virus serotypes in individual Aedes aegypti mosquitoes in Colombia. *Medical and veterinary entomology* 12 (3):284.

Roongpiboonsopit, D., and H. A. Karimi. 2010. Comparative evaluation and analysis of online geocoding services. *International Journal of Geographical Information Science* 24 (7):1081-1100.

Sabel, C. E., P. Bartie, S. Kingham, and A. Nicholson. 2006. Kernel Density Estimation as a spatial-temporal data mining tool: exploring road traffic accident trends. Paper read at GISRUK 2006, at University of Nottingham.

San Martín, J. L., O. Brathwaite, B. Zambrano, J. O. Solórzano, A. Bouckenooghe, G. H. Dayan, and M. G. Guzmán. 2010. The epidemiology of dengue in the Americas over the last three decades: a worrisome reality. *The American journal of tropical medicine and hygiene* 82 (1):128.

Silverman, B. W. 1986. *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.

Skinner, M. W., and A. Power. 2011. Voluntarism, health and place: Bringing an emerging field into focus. *Health & Place* 17 (1):1-6.

Thompson, J., Eagleson, S., Ghadirian, P. and A. Rajabifard (2009). SDI for collaborative health services planning, *Global Spatial Data Infrastructures World Conference*, Rotterdam, The Netherlands.

Tran, A., X. Deparis, P. Dussart, J. Morvan, P. Rabarison, F. Remy, L. Polidori, and J. Gardon. 2004. Dengue spatial and temporal patterns, French Guiana, 2001. *Emerging Infectious Diseases* 10 (4):615-621.

Vazquez-Prokopec, G. M., S. T. Stoddard, V. Paz-Soldan, A. C. Morrison, J. P. Elder, T. J. Kochel, T. W. Scott, and U. Kitron. 2009. Usefulness of commercially available GPS data-loggers for tracking human movement and exposure to dengue virus. *Int J Health Geogr* 8 (1):68.

Wand, M. J., and M. C. Jones. 1995. *Kernel smoothing*.

WHO. 2009. Dengue net. World Health Organization, Geneva, Switzerland.

Xie, Z., & Yan, J. (2008). Kernel density estimation of traffic accidents in a network space. *Computers, Environment and Urban Systems*, 32(5), 396-406.

Yoon, I.-K., A. Getis, J. Aldstadt, A. L. Rothman, D. Tannitisupawong, C. J. Koenraadt, T. Fansiri, J. W. Jones, A. C. Morrison, and R. G. Jarman. 2012. Fine scale spatiotemporal clustering of dengue virus transmission in children and Aedes aegypti in rural Thai villages. PLoS neglected tropical *diseases* 6 (7):e1730.

Young, S. G., and R. R. Jensen. 2012. Statistical and visual analysis of human West Nile virus infection in the United States, 1999–2008. *Applied Geography* 34:425-431.

Zook, M., M. Graham, T. Shelton, and S. Gorman. 2010. Volunteered Geographic Information and Crowdsourcing Disaster Relief: A Case Study of the Haitian Earthquake. *World Medical & Health Policy* 2 (2):7-33.