A WEB-BASED GEOGRAPHIC FRAMEWORK TO DETECT AND VISUALIZE SPACE-TIME CLUSTERS OF COVID-19

Yu Lan*and Eric Delmelle

Department of Geography & Earth Science, The University of North Carolina at Charlotte, Charlotte, US
Center for Applied GIScience, The University of North Carolina at Charlotte, Charlotte, US
* ylan1@uncc.edu

Keywords: COVID-19, space-time cluster, geovisualization, WebGIS, diseases surveillance

Introduction

Space-time clustering detection is essential in spatial epidemiology, especially for infectious diseases (Pfeiffer et al., 2008). Local indicators of spatial association (LISA) and space-time scan tests are two popular methods to detect local space and space-time clusters for infectious diseases.

The space-time geovisualization of these two methods is crucial to infectious diseases like COVID-19, but many current dashboards neglect this temporal component (Lan et al., 2021). Moreover, the space-time scan statistic requires an application (SaTScan) or libraries (through R, for instance) to estimate the presence and magnitude of clusters, while a commercial GIS is still needed to visualize the space-time extent of these clusters. A tight-coupling system with different modules connected into one system can address this issue. However, to the best of our best knowledge, this type of tight coupling system is not available.

This study proposes a web-based geographic framework for detecting and visualizing space-time clusters for infectious diseases. To illustrate the framework, a prototype based on this framework using COVID-19 data in the conterminous US at the county level is developed. Here we further elaborate on the objectives of this prototype, namely: (1) to implement automatic and customized space-time clustering detection for a selected state; (2) to generate (3D) interactive geovisualization of space-time clustering; (3) a tight coupling system that incorporated daily data updated and components for the objective one and two.

Method

In this study, the COVID-19 Data Repository by the Center for Systems Science and Engineering at Johns Hopkins University is used to extract daily US COVID-19 data (JHU CSSE COVID-19 Data) at both state and county levels. The time range of the data is from January 22nd to the latest date on which the data is updated. Figure 1 introduces the framework and workflow for detecting and visualizing space-time clusters for the prototype. The framework is based on a tight coupling system to detect space-time

clustering with customized spatial and temporal settings, and it incorporates data extraction capabilities, clustering detection, and geovisualization in a web-based GIS environment, with a server side and a client.
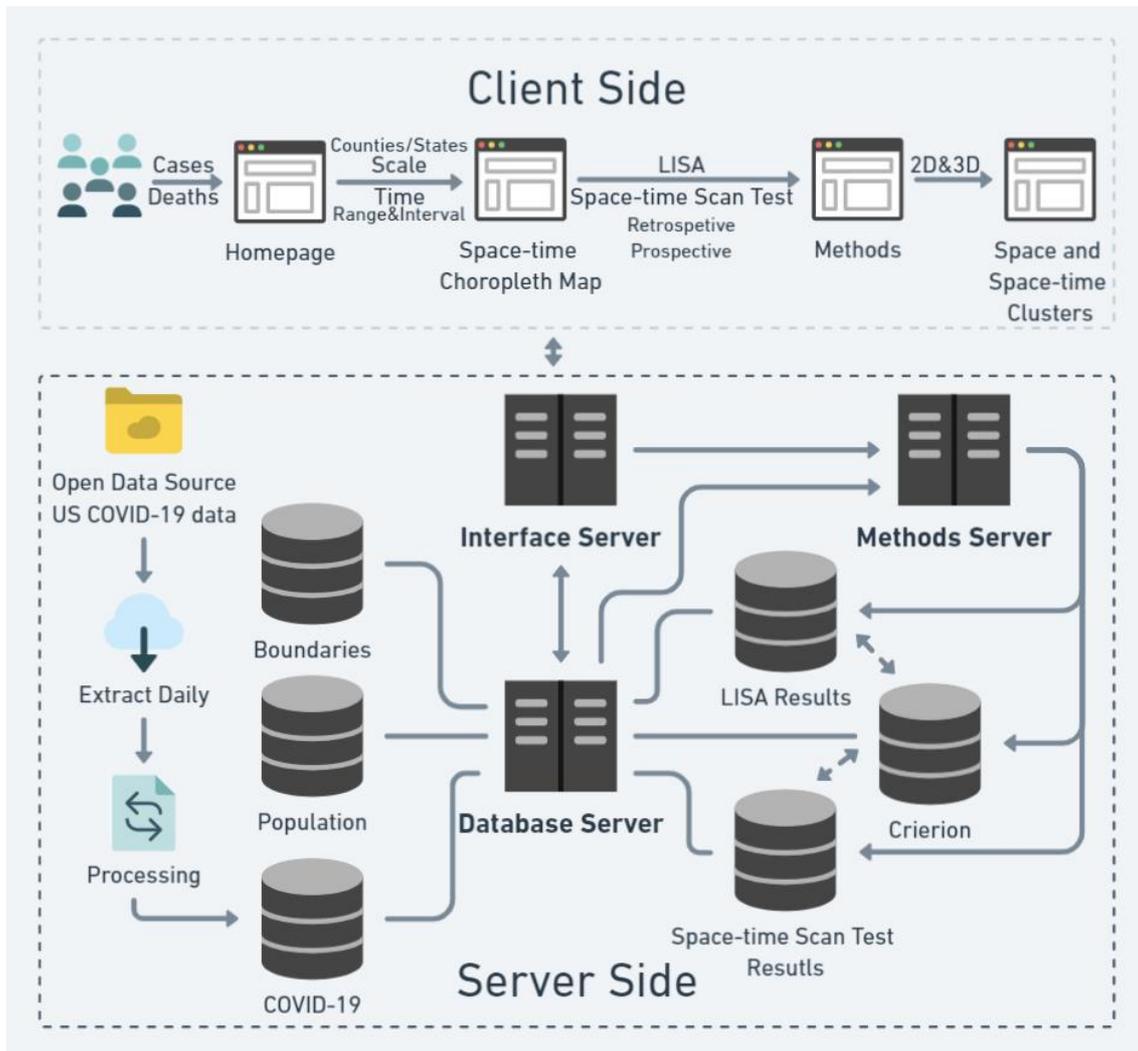
Figure 1: The framework of a tight coupling system to detect space-time clustering of COVID-19.

The **server side** contains three types of servers for different purposes: a database server, an interface server, and a method server. The database server stores all the relevant input data and output results, and each type of data is stored in an independent database. The disease detection algorithms are processed on the methods server, and the users' selection of methods is transferred from the interface server, which hosts the interface of the prototype.

The **client side** is the graphical user interface (GUI) of the prototype. It consists of basic online map functions including zoom, pan, search, etc. Users can customize their maps by select states and time ranges to generate space-time cluster maps. The workflow of communication from the client side to the server side is shown in Figure 2.
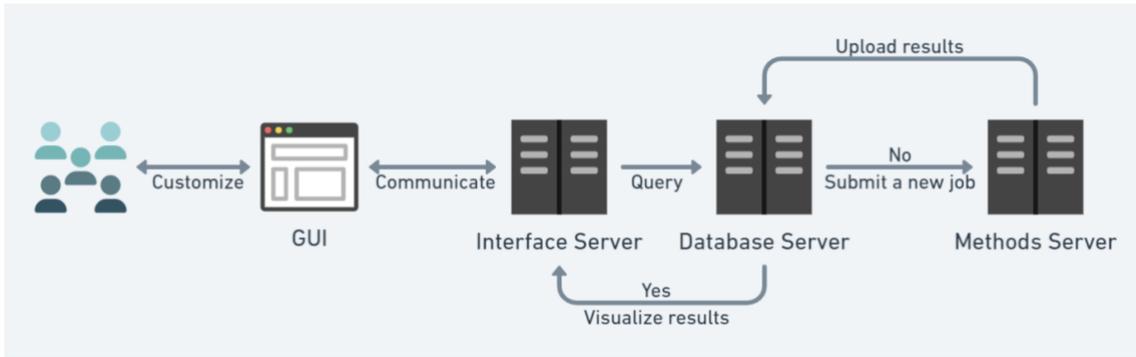
Figure 2: The workflow of communication from the client side to the server side.

Analysis module

Local indicators of spatial association-LISA. Anselin (1995) introduce LISA as decomposition of global indicators of each individual observation, which can detect significant local clustering around an individual location and recognize the spatial nonstationary with outliers. For region i, the local indicators of spatial association I_i defined as:

$$I_i = \frac{(n-1)(x_i - \bar{x})}{\sum_{j=1}^{n}(x_j - \bar{x})^2} \sum_{j=1}^{n} w_{ij}(x_j - \bar{x})$$

Eq. 1

where x_i is the attribute of the variable of interest (here, the disease rate) in region i, $\bar{x}$ is the mean of x_i (i=1…, n), w_{ij} is the spatial weight between regions i and j.

Space-time scan statistics. Kulldorff and Nagarwalla (1995) introduce the spatial scan statistic as a test for detecting clusters by assessing the likelihood ratios of events inside and outside of circular scanning windows, adjusted for spatial variations in the density of the population. The radii of circular scanning windows are varied continuously from zero to the maximum bandwidth for each point center. Kulldorff et al. (1998) further expand the spatial scan statistics into space-time scan statistics by expanding the circular window into the cylindrical window by adding the circle's height to represent the time. Figure 3 depicts the mechanism of space-time scan statistics.
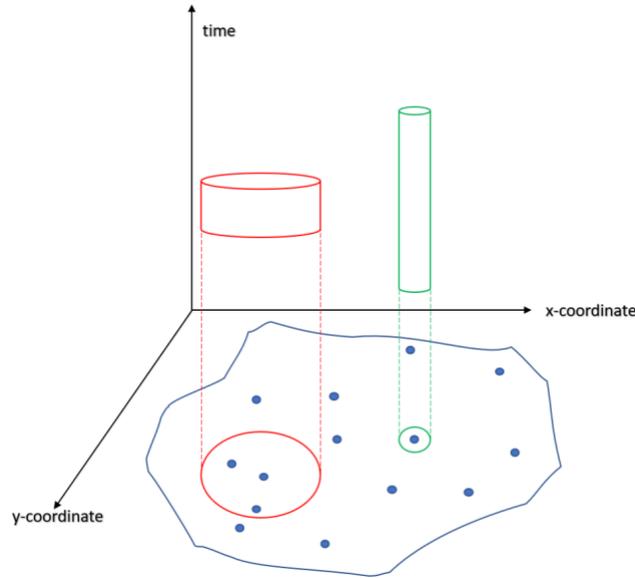
Figure 3: The illustration of space-time scan statistics.

As the space-time scan statistic is designed to detect disease clusters, the null hypothesis of the space-time scan statistic H_0 is that the probability of being a case within a cylinder Z (a) is same to the probability of being a case outside this cylinder (b); the alternative hypothesis H_a is that the probability of being a case within a cylinder Z is larger than the probability of being a case outside this cylinder (a > b). Accordingly, the expected number of cases (μ) based on the null hypothesis is

$$\mu = p \times \frac{N}{P}$$
Eq. 2

with p the population in the cylinder, N the total number of cases within the cylinder Z, and P the total population within the study area. Thus, the maximum likelihood ration to identify space-time clusters is defined as:

$$\frac{L(Z)}{L_0} = \frac{\left(\frac{n_Z}{\mu(Z)}\right)^{n_Z} \left(\frac{N - n_Z}{N - \mu(Z)}\right)^{N - n_Z}}{\left(\frac{N}{\mu(T)}\right)^N}$$
Eq. 3

where $L(Z)$ is the likelihood function for the cylinder Z, and L_0 is the likelihood for the null hypothesis H_0, n_Z is the number of cases in the cylinder Z, $\mu(Z)$ is the number of expected cases in cylinder Z, and $\mu(T)$ is the total number of expected cases within all time periods in the study area.

Analysis procedure. In this prototype, LISA is repeatedly conducted to detect space-time clusters for different time ranges, using the python library from spatial data analysis, pygeoda. As to space-time scan statistics, SaTScan is used to run analysis in a batch mode approach. Both LISA and space-time scan statistics are automated conducted every night when latest data are updated and retrieved.

Visualization module

The visualization module is consisted of basic cluster maps and advanced maps displaying other variables (e.g., relative risk) from cluster detection results. Basic cluster maps will combine results from LISA and space-time scan statistics

implemented with animation. Advanced maps will include multiple visualization, such as animated bivariate maps. In addition, 3D web-based geovisualization will also be developed based on WebGL, a JavaScript API for rendering interactive 2D and 3D graphics. All web-based visualization will use multiple JavaScript libraries, including Data-Driven Documents (D3) (Bostock et al., 2011), and React, as well as other web-based technologies including HTML (Hypertext Markup Language), CSS (Cascading Style Sheets), JS (JavaScript), and SVG (Scalable Vector Graphics).

Data processing and WebGIS

Data processing for this prototype contains daily data retrieving and processing, data analysis, and data storing, and these steps are connected in the WebGIS environment. All the data are stored databases created and managed using PostGIS. Daily COVID-19 cases and other related data are extracted daily, and retrieved data are processed and imported into related databases on the server.

Results

Figure 4 is the proposed geovisualization of the space-time cluster. This visualization consists of three visual components: Bivariate Map (top left), Time Charts (right), and Spiral Map (bottom left). Bivariate Map is an animated and interactive bivariate map of two visualization variables, the relative risk of the county and the relative risk of the cluster if the county belongs to a cluster on that day. Time Charts shows the relative risk of counties within the selected state during the study time, and it is aimed to visualize the trend of relative risk of each county. Unlike two other visual components showing space-time patterns at the county level, Spiral Map shows the dynamic of relative risk at the state level by calculating the average of this value daily. With the combination of these three components, users can explore the same space-time pattern with different focuses on spatial (Bivariate Map), temporal (Time Charts), and both county and state scales (Spiral Map). Figure 5 is the example of possible 3D visualization.
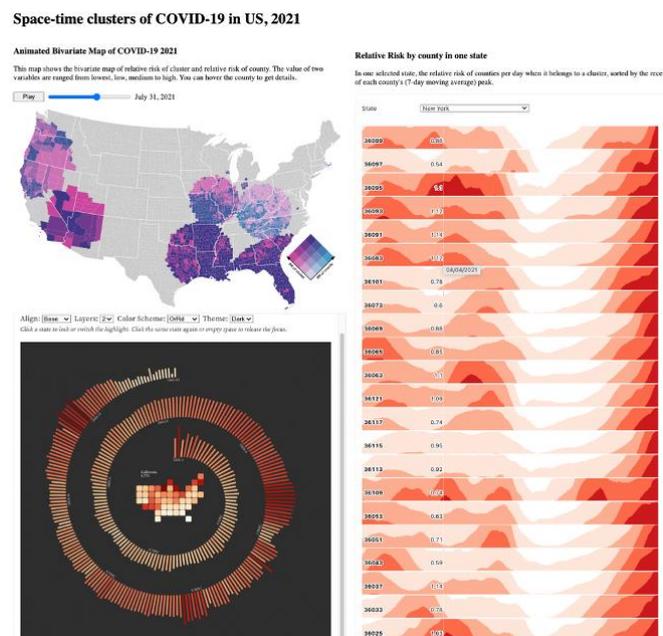
Figure 4: The proposed geovisualization of space-time clusters.
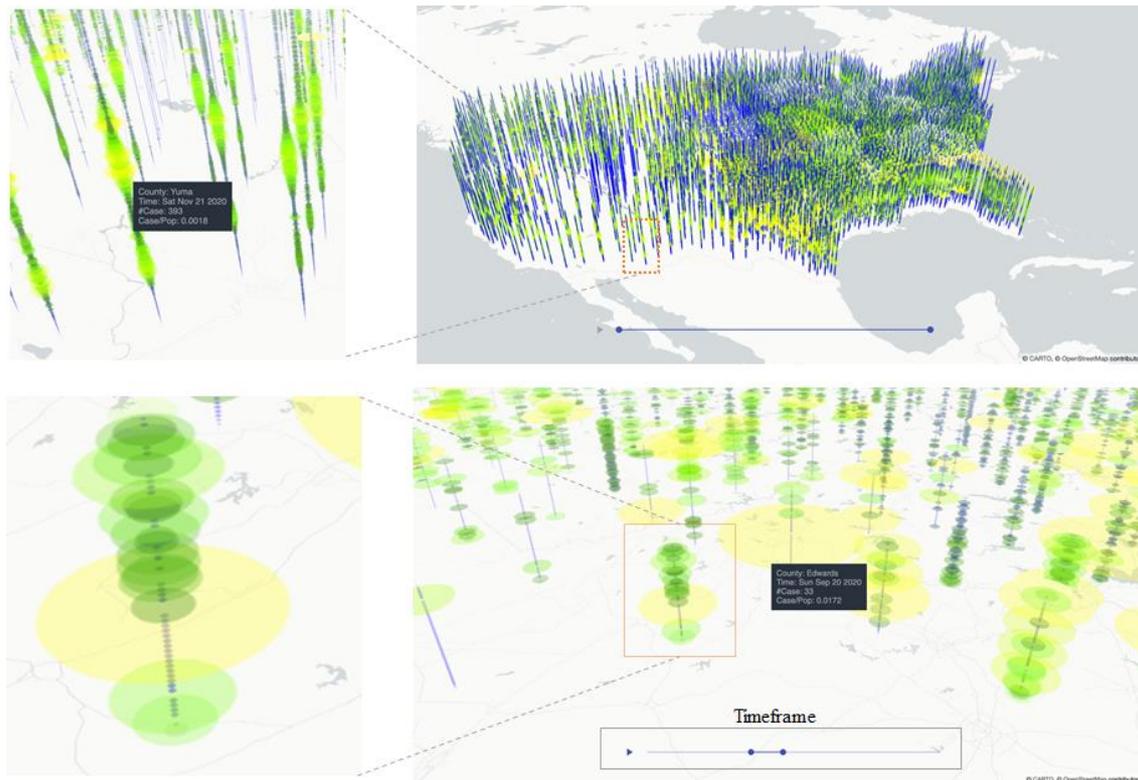
Figure 5: 3D geovisualization using the example of case incidence rates.

Discussion and Conclusion

In this study, we present a web-based geographic framework for automatic detecting and visualizing space-time clusters for infectious diseases in a daily basis. This framework integrates techniques, including 3D, WebGIS, and interactivity, to robust visualize space-time patterns, which has not been done before. It also integrates data retrieving and preparation, analysis, and visualization into a connected and automatic manner, facilitating timely surveillance.

Early cluster detection, such as daily spatiotemporal clusters detection, can facilitate "outbreak recognition and investigation, provider and community outreach, and timely intervention (Greene et al., 2016)." However, it could take a significantly longer time to obtain results when each step is loosely connected and not automated. Therefore, for daily surveillance, an automatic and routine process is needed, and this process should include 1) downloading and preparing data from the data source, 2) running the statistical analysis analyses on freshly daily updated data, and 3) generating the visualization results for that for that analysis results.

The major limitations of this study are as follows. First, the analysis module focuses on detecting both exact and inexact space-time clusters implemented with SaTScan and LISA statistics, but other methods could   also be implemented depending on the research questions. Second, the case dataset cannot be uploaded from the client-side to the server-side because data security is most important. Third, only cases are used in this prototype, but deaths or hospitalizations could also be used.

References

Anselin, L. (1995). Local indicators of spatial association—LISA. *Geographical analysis*, *27*(2), 93-115.

Bostock, M., Ogievetsky, V., & Heer, J. (2011). D³ data-driven documents. *IEEE transactions on visualization and computer graphics*, *17*(12), 2301-2309.

Greene, S. K., Peterson, E. R., Kapell, D., Fine, A. D., & Kulldorff, M. (2016). Daily reportable disease spatiotemporal cluster detection, New York City, New York, USA, 2014–2015. *Emerging infectious diseases*, *22*(10), 1808.

Kulldorff, M., Athas, W. F., Feurer, E. J., Miller, B. A., & Key, C. R. (1998). Evaluating cluster alarms: a space-time scan statistic and brain cancer in Los Alamos, New Mexico. *American journal of public health*, *88*(9), 1377-1380. https://doi.org/10.2105/ajph.88.9.1377

Kulldorff, M., & Nagarwalla, N. (1995). Spatial disease clusters: detection and inference. *Statistics in medicine*, *14*(8), 799-810.

Lan, Y., Desjardins, M. R., Hohl, A., & Delmelle, E. (2021). Geovisualization of COVID-19: State of the Art and Opportunities. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *56*(1), 2-13.

Pfeiffer, D., Robinson, T. P., Stevenson, M., Stevens, K. B., Rogers, D. J., & Clements, A. C. (2008). *Spatial analysis in epidemiology* (Vol. 142). Oxford University Press Oxford.