Identifying Local Spatiotemporal Autocorrelation Patterns of Taxi Pick-ups and Drop-offs

Song Gao[1], Rui Zhu[1], Gengchen Mai[1]

[1]STKO Lab, Department of Geography, University of California, Santa Barbara, USA
Contact Email: sgao@geog.ucsb.edu

Abstract

Analyzing spatiotemporal autocorrelation would be helpful to understand the underlying dynamic patterns in space and time simultaneously. In this work, we aim to extend the conventional spatial autocorrelation statistics to a more general framework considering both spatial and temporal dimensions. Specifically, we focus on the spatiotemporal version of Getis-Ord's G^*. The proposed indicator STG^* can quantify the local association of adjacent features in space and time. As a proof of concept, the proposed method is then applied in a large-scale GPS-enabled taxi dataset to identify local spatiotemporal autocorrelation patterns of taxi pick-ups and drop-offs in New York City.

1. Introduction

Nowadays, large-scale spatiotemporal data (e.g., taxi trajectories, phone call records, social media posts) become available, which provide rich information to support research on human behaviors, transportation, urban landscape, and human-environment interactions. However, discovering patterns hidden in large-scale spatiotemporal datasets is challenging and thus attracts a lot of attention from GIScience community (Hardisty and Klippel 2010; Demšar and Virrantaus 2010; Scholz and Lu 2014; Claramunt and Stewart 2015).

Spatial autocorrelation statistics, like *Moran's I* and *Getis-Ord's G* are commonly designed for identifying spatial autocorrelation patterns (Fischer and Getis 2009). However, there is a gap in building corresponding measurements for spatiotemporal autocorrelations. For example, although human movements and activities may vary over time across different places, the observed activity hotspots and movement flow might exhibit a pattern of spatial dependence. Also, ignoring the temporal dimension would not be sufficient to discover underlying spatiotemporal dynamics. Therefore, our work aims to contribute to extend the conventional local spatial autocorrelation statistics to include both spatial and temporal dimensions. As a proof of concept, the proposed method is then applied in a large-scale GPS-enabled taxi dataset to identify local autocorrelation patterns of taxi pick-up points (PUPs) and drop-off points (DOPs) in New York City.

2. Methodology

Spatial autocorrelation measurements can be divided into two categories: *global* and *local* indices. Classic global indices of spatial autocorrelation include *Moran's I*, *Geary's C*, and *Getis-Ord's General G*, while local indices of spatial association (LISA) can be established by transforming the global indices into corresponding local measurements (Anselin 1995). Spatiotemporal autocorrelation concept refers to the relationship between some variable observed in each of space-time settings and the association with its neighbors. In a previous work, Gao (2015) proposed three global spatiotemporal autocorrelation indices but didn't describe how to decompose them into local versions. As an initial trial, this work focuses on extending *Getis-Ord's* G^* (Equation 1) (Getis and Ord 1992) by adding temporal

indexes into the formula, and we name this local spatiotemporal autocorrelation measure as *Spatiotemporal Getis-Ord's* G^* (STG^*) (Equation 2).

$$G_i^* = \frac{\sum_j w_{ij} x_j}{\sum_j x_j} \qquad (1)$$

where w_{ij} *indicates the spatial weight between location* i *and* j *and* x_j *is the attribute value at location* j.

$$STG_i^* = \frac{\sum_s \sum_t w_{ist} x_{st}}{\sum_s \sum_t x_{st}} \qquad (2)$$

where w_{ist} *is the extended weighting matrix regarding both spatial and temporal dimensions and* x_{st} *is the attribute value at space* s *and time* t.

The STG^*_i quantifies the spatiotemporal concentration of adjacent features associated with the target i, and works as an indicator for measuring local association in space and time simultaneously. To conceptualize the spatiotemporal neighbors, we implement a 3D-cube framework as shown in Figure 1, where each voxel consists of a geographic coordinate *S(x,y)* and a timestamp *T(t)*. The adjacency can be defined as the first-degree of "Queen" type, in which there are 26 spatiotemporal neighbors for a target voxel in the center cube. The weight w_{ist} for them in calculating STG^*_i is 1 otherwise is 0.

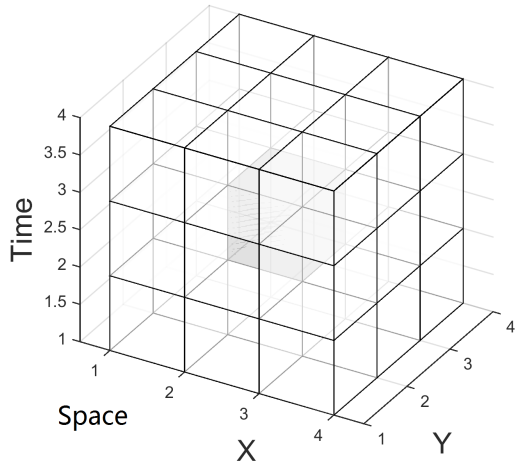

Figure 1. The 3D-cube visualization of spatiotemporal neighbors.

Furthermore, to statistically test the significance of the concentration of either high or low attribute values surrounding the target voxel, the Z-score of STG^*_i, as illustrated in Equation 3, is calculated. Thus, if a tested Z-score is significantly different from the corresponding expectation, the target feature would be a hot-spot (positive value) or a cold-spot (negative value).

$$Z_{G_i^*} = \frac{\sum_s \sum_t w_{ist} x_{st} - \bar{X} \sum_s \sum_t w_{ist}}{S \sqrt{\frac{\left[n \sum_s \sum_t w_{ist}^2 - \left(\sum_s \sum_t w_{ist} \right)^2 \right]}{n-1}}} \qquad (3)$$

where:

$$\bar{X} = \frac{\sum_s \sum_t x_{st}}{n} \text{ and } S = \sqrt{\frac{\sum_s \sum_t x_{st}^2}{n} - \bar{X}^2}$$

3. Case Study: Taxi Drop-offs and Pick-ups in New York City

3.1 Data and Processing

Taxi pick-up and drop-off locations in cities can reveal human movement patterns and thus playing an important role in urban informatics and transportation management. The data used in this study is downloaded from the NYC Taxi and Limousine Commission trip GPS records[1]. We extract one-week trips in five boroughs (*Manhattan, Brooklyn, Queens, The Bronx, and Staten Island*) of NYC from Jan. 3 to Jan. 9, 2015. As shown in Figure 2, by applying exploratory spatial autocorrelation analysis for the whole time period, we can find that three regions are significant hotspots (*Manhattan, JFK International Airport and LaGuardia Airport*) for both PUPs and DOPs. In order to further identify fine-scale local autocorrelation patterns, we spatially filter the original data to include only trips generated in *Manhattan* and there exist 2,548,952 PUPs and 2,462,199 DOPs in total. Figure 3 shows their temporal variations in different hours.

In order to further conduct local spatiotemporal autocorrelation analysis, we need to aggregate those points (PUPs or DOPs) into introduced space-time-cube structure One research question is how to find appropriate bin sizes in both spatial and temporal dimensions for defining neighbors. After calculating the nearest-neighbor distance for each point and the time difference for each pair, we found that spatial proximity is related to temporal closeness. Therefore, we suggest a strategy to find optimal bin sizes for defining spatiotemporal neighbors: Firstly, we spatially aggregate those points into regular grids or administrative polygons; the city block is taken in this study and the spatial bin can be set to one quarter of the average city-block size (about 520 meters) in *Manhattan*. Secondly, the average time difference of temporal adjacent points in each block is calculated and the median (or mean) of average time differences across all city blocks can be used as the temporal bin. Finally, the space-time cubes are constructed with a 130-meter spatial distance and about 20-minute temporal interval in this study. Figure 4 shows the visualization of aggregated PUPs in space-time cubes, in which the attribute value represents the count of PUPs in each voxel.

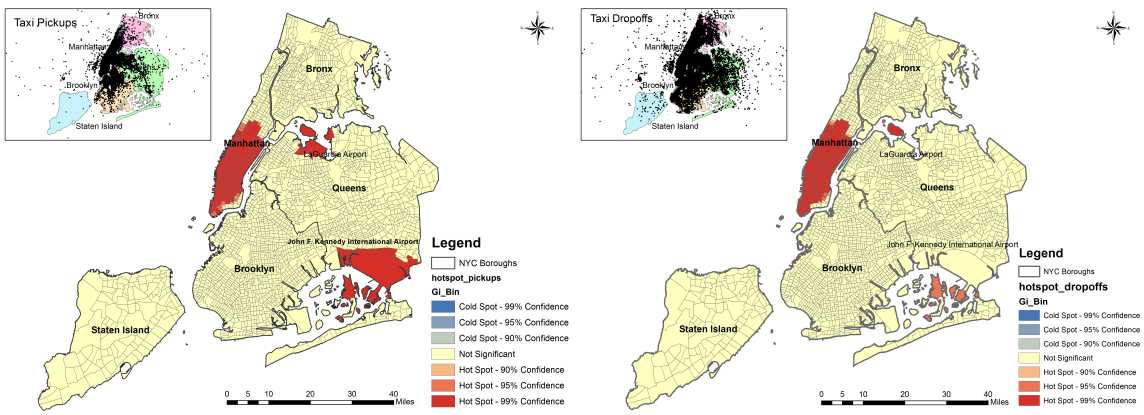

Figure 2. The spatial distributions of PUPs and DOPs and spatial autocorrelation results in NYC.

[1] http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml

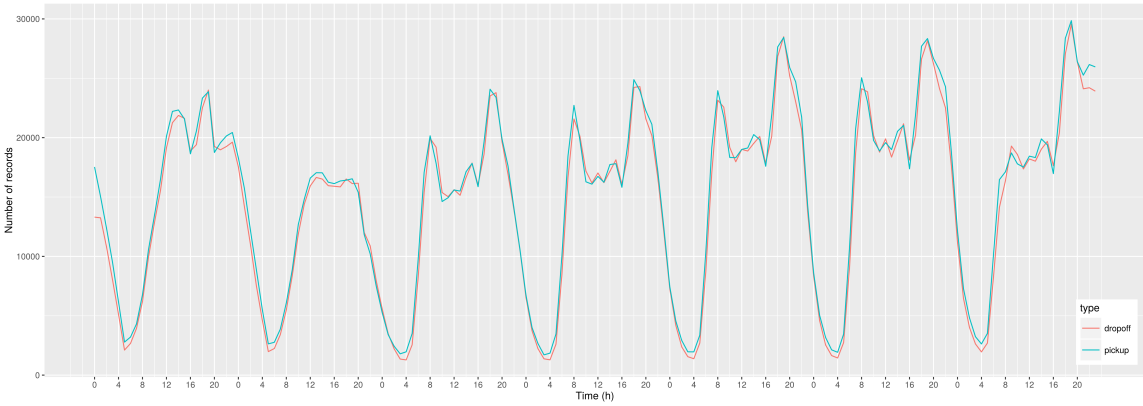

Figure 3. The temporal variations of PUPs and DOPs in different hours in Manhattan.

Figure 4. The spatiotemporal visualization of PUPs in Manhattan.

3.2 Results

By applying the proposed local spatiotemporal autocorrelation method, we calculate the STG^* statistic of PUPs (and DOPs) for each voxel and the corresponding z-score. Figure 5 shows different confidence levels (90%, 95%, and 99%) of spatiotemporal "hotspots" (red color: a large statistic value exists and its spatiotemporal neighbors also have large values) and "coldspots" (blue color: a small statistic value exists and small values for its spatiotemporal neighbors) for PUPs in *Manhattan*. We find statistically significant spatiotemporal hotspot clusters in the southern part and coldspot clusters in the northern part of *Manhattan*. Interestingly, those regions are spatially divided by the *Central Park*. Such spatiotemporal pattern of taxi trips usually links to the mixture land-use structure and human home-to-job activities, which has also been identified by other studies (Liu et al. 2012; Liu et al. 2015).

Table 1. The top ranked local hotspots for taxi pick-ups and drop-offs in Manhattan.

Rank	Pick-ups	Drop-offs
1	LocationID: 9536 Time: 1/8/2015 8:00-8:20 AM	LocationID: 9536 Time: 1/8/2015 8:00-8:20 PM
2	LocationID: 9536 Time: 1/4/2015 10:00-10:20 PM	LocationID: 7725 Time: 1/8/2015 8:20-8:40 AM
3	LocationID: 9535 Time: 1/4/2015 10:00-10:20 PM	LocationID: 7725 Time: 1/8/2015 8:40-9:00 AM
4	LocationID: 9535 Time: 1/4/2015 10:20-10:40 PM	LocationID: 7357 Time: 1/8/2015 9:20-9:40 AM
5	LocationID: 9536 Time: 1/4/2015 10:20-10:40 PM	LocationID: 7725 Time: 1/8/2015 8:00-8:20 AM

In addition, we can zoom to specific voxels in space-time cubes and compare their local STG^* values. Table 1 shows the top 5 hotspots of taxi PUPs and DOPs ranked by their z-scores of STG^*. It proves the existence of local spatiotemporal autocorrelation patterns.

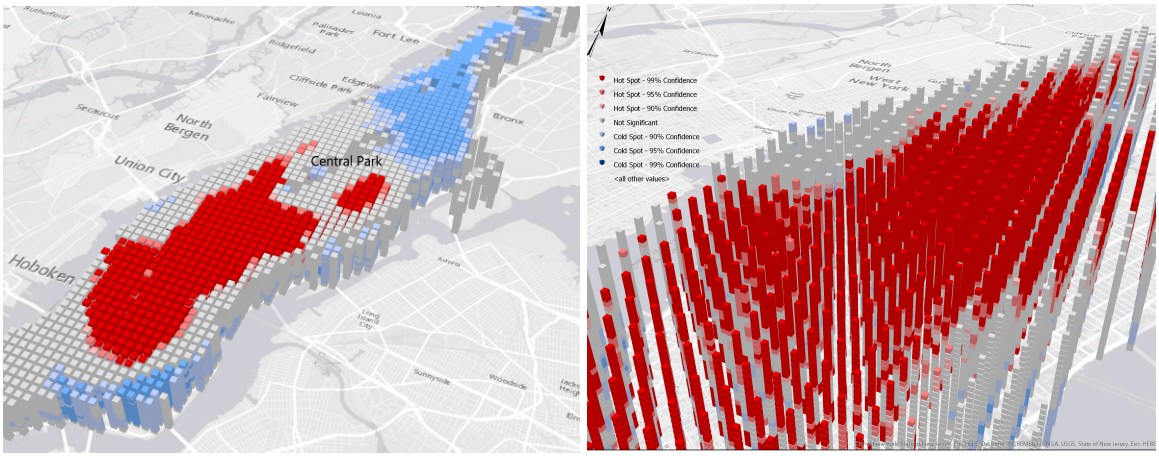

Figure 5. The visualization of spatiotemporal hotspots and coldspots in Manhattan.

4. Conclusions and Future Work

In this research, we extend the spatial association statistic *Getis-Ord's* G^* to the local spatiotemporal autocorrelation indicator STG^* which takes the adjacency in both space and time into consideration. The space-time-cube structure has been constructed to support spatiotemporal point pattern analysis and visualization. By performing the proposed method in a large-scale taxi trips, we find that STG^* can sufficiently identify local spatiotemporal autocorrelation patterns of taxi pick-ups and drop-offs in *Manhattan*. The proposed method can also be applied in other event data with spatiotemporal tags and thus has a broad impact.

In future work, more complex spatiotemporal weighting matrix rather than binary ones and its impact on autocorrelation structure will be studied. More empirical studies in other cities will also be conducted to find underlying association between space and time.

References

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

Claramunt C and Stewart K, 2015, Special issue on spatio-temporal theories and models for environmental, urban and social sciences: where do we stand?, *Spatial Cognition & Computation*, 15(2), 61-67.

Demšar U and Virrantaus K, 2010, Space–time density of trajectories: exploring spatio-temporal patterns in movement data. International Journal of Geographical Information Science, 24(10), 1527-1542.

Fischer MM and Getis A (eds.), 2009, *Handbook of applied spatial analysis: software tools, methods and applications.* Springer Science & Business Media.

Gao S, 2015, Spatio-temporal analytics for exploring human mobility patterns and urban dynamics in the mobile age. *Spatial Cognition & Computation*, 15(2), 86-114.

Getis A and Ord JK, 1992, The analysis of spatial association by use of distance statistics. *Geographical analysis*, 24(3), 189-206.

Hardisty F and Klippel A, 2010, Analysing spatio-temporal autocorrelation with LISTA-Viz. *International Journal of Geographical Information Science*, 24(10), 1515-1526.

Liu X, Gong L, Gong Y and Liu Y, 2015, Revealing travel patterns and city structure with taxi trip data. *Journal of Transport Geography*, 43, 78-90.

Liu Y, Wang F, Xiao Y and Gao S, 2012, Urban land uses and traffic 'source-sink areas': Evidence from GPS-enabled taxi data in Shanghai. *Landscape and Urban Planning*, 106(1), 73-87.

Scholz RW and Lu Y, 2014, Detection of dynamic activity patterns at a collective level from large-volume trajectory data. *International Journal of Geographical Information Science*, 28(5), 946-963.