# OD-Based Mobility Hotspot Pattern Exploration with Enhanced Spatial Clustering Methods

**Weiye Chen, Shaohua Wang, Shaowen Wang**

**Department of Geography and Geographical Information Science, CyberGIS Center for Advanced Spatial and Digital Studies, University of Illinois, Urbana-Champaign**

## Abstract

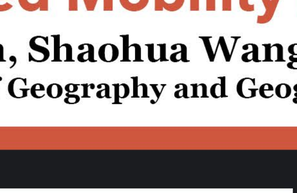
**Detections of spatial clusters of mobilities** are essential in helping people understand the patterns of movements within the cities. We proposed the Expansion-Based Spatial Clustering Method for Spatial Flows (ESCSF) based the ESCIP, to find *spatial clusters represented in OD-based mobility datasets*. Compared to the baseline multidimensional spatial scan statistics (MSSS) , the new ESCSF could effectively find the spatial clusters of more flexible shapes in a more computational sufficient way. We conducted case studies for the city of Haikou over two methods, where we looked into the travel pattern differences in the taxi trips for morning against evening rush hours within a large dataset. Interesting patterns have been discovered with both models, with the ESCSF more efficient and flexible.

### Keywords

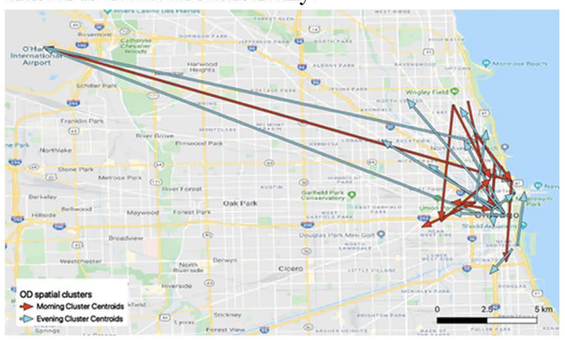
Spatial Data Mining, OD Flows, Human Mobility

## Concepts

Among all the forms of mobility data, origin-destination (OD) flows data remain the simplest, which are comprised of coordinates of origins and destinations of movement, and also non-spatial attributes describing the movement. A region in such hyperspace could be seen as a space for a set of OD flows, indicating a set of flows from one geographic space to the other.




**Spatial clusters of OD flows** would indicate clusters of movement from one place to the other, where a place could either be a location or an area. We often make a null hypothesis that the distribution of OD flows across space follows a random baseline process with a known density in the background, i.e., no local excess of OD occurrence at a particular origin-destination region should be spotted. If this null hypothesis is rejected in a local OD region, then a local OD cluster is detected statistically.

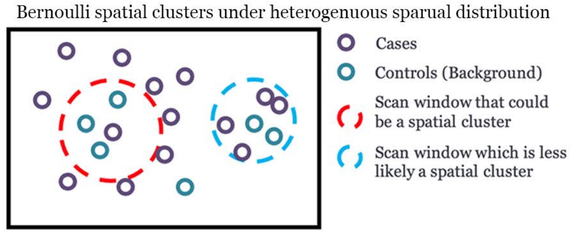



## Data and Modeling


We retrieved 12 million on-demand internet-based taxi services desensitized records from Didi (a Uber-like app in China), joined with the weather information from the open-source weather archive. This is a slice of dataset recording trips in Haikou City, Hainan Province, China, where the population is 2 million, from May 2017 to Oct 2017. We sampled a slice of dataset, and partitioned it into four groups of cases and controls with different attribution, as is shown in the following table. ODs in each group of partitions complys to a Bernoulli process, where each OD point is either a case or a control.

| Cases (Type 1) | | Controls (Type 2) | |
|---|---|---|---|
| **Description** | **Size** | **Description** | **Size** |
| Weekday Morning with Good Weather(Baseline) | 102175 | Weekday Evening with Good Weather (Baseline) | 137726 |
| Weekend Morning with Good Weather | 26505 | Weekend Morning with Good Weather | 36505 |
| Weekday Morning with Bad Weather | 33612 | Weekday Morning with Bad Weather | 43612 |
| Weekend Morning with Bad Weather | 7426 | Weekend Morning with Bad Weather | 12426 |

Bernoulli spatial clusters under heterogenuous sparual distribution




## References and Acknowledgement

Chen: I want to thank Prof. Shaowen Wang and Dr. Shaohua Wang at CyberGIS Center, University of Illinois, Urbana Champaign for supporting and guiding my academic life and research there. I further extends my gratitude to Prof. Zhenhong Du, Prof. Feng Zhang, Prof. Jie Lin and Prof. Lizhen Lu at Zhejiang University, China, my BSc. alma mater, for their encouragement and valuable suggestions in this research project.

Here's a list of the primary references and inspiration to this research.

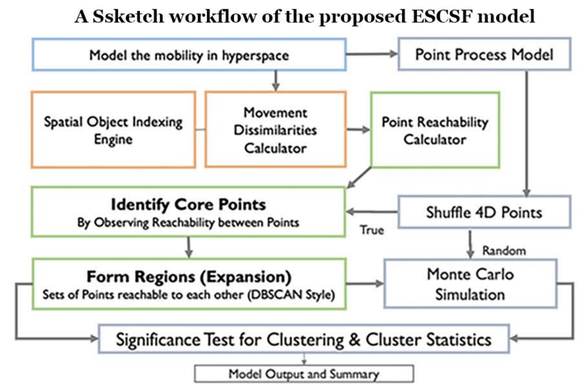
- Gao et al. 2018, Multidimensional Spatial Scan Statistics, International Journal of Geographic Information Science
- Li et al. 2019, ESCIP: An Expansion-Based Spatial Clustering Method for Inhomogeneous Point Processes, Annals of the American Association of Geographers
- Song et al. 2019, Detecting arbitrarily shaped clusters in origin-destination flows using ant colony optimization, International Journal of Geographic Information Science
- Yao et al. 2018, A Stepwise Spatiotemporal Flow Clustering Method for Discovering Mobility Trends, IEEE Access

## Methods and Benefits

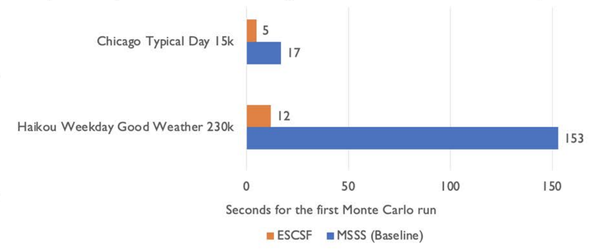
**Spatial Scan Statistics** (SaTScan, Kulldorff 1997) is a popular approach to detect spatial clusters. It generates a series of scan windows within which points fall form groups and use statistical analysis to determine if a group of points (a scan window) is a cluster. **Multidimensional spatial scan statistics (MSSS)** (Gao et al. 2018) extents the SaTScan to OD hyperspace. **ESCIP** (Li et al. 2020), or the Expansion-Based Spatial Clustering Methods for Inhomogeneous Processes, is an innovative enhancement over the SaTScan, which is capable of detecting flexible-shaped clusters while improving computational efficiency by incorporating a DBSCAN-like clustering expansion mechanism.

An extended ESCIP method for OD flows is proposed based on previous studies discussed above. Called Expansion-Based Spatial Clustering Method for Spatial Flows (**ESCSF**), it inherits the idea of abstracting OD flows as points from previous studies, and integrates the advantages of some models, including high-efficiency in ESCIP, geometry-based distance, and flexible shapes.

A Ssketch workflow of the proposed ESCSF model

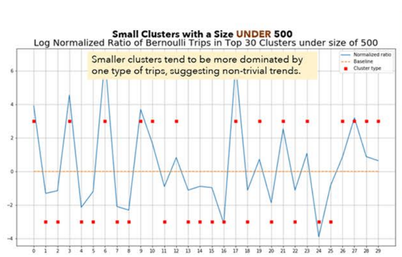


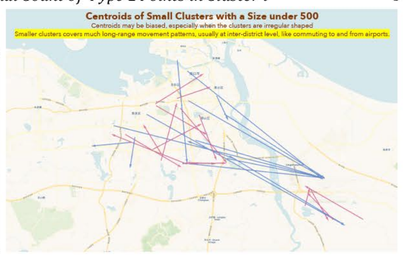

Computational performance of ESCSF compared to baseline MSSS over two differe§nt datasets

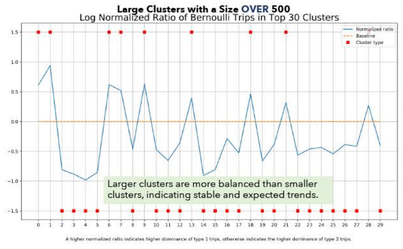


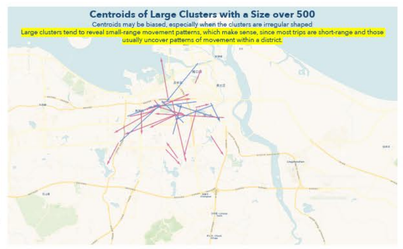

## Case Study - Haikou, China

$$\textit{Relative Dominance Ratio for } i^{th} \textit{ cluster} = \log\left[\frac{\textit{Actual Count of Type 1 Points in Cluster } i}{\textit{Actual Count of Type 2 Points in Cluster } i} \Big/ \textit{Expected Ratio}\right]$$

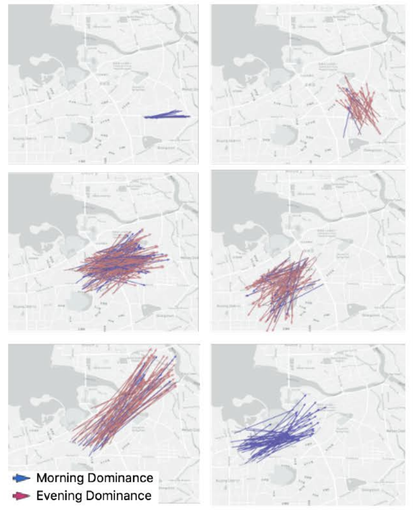


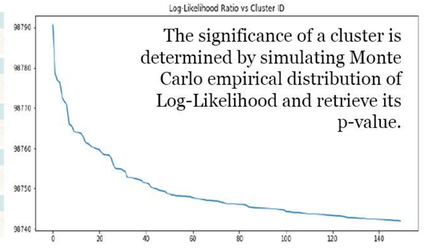




Top 6 Clusters on days of good weather




**Patterns are different between weekdays and weekends, Good and Bad Weather, and Morning and Afternoon.**




### Comments

- The model would provide trivial trends we may already know, but also helps us explore the unknown or unexpected trends.
- The greatest contribution that this research provides is that a quantitative method to measure and detect the pattern differences. The OD datasets are mapped in a novel 4D hyperspace where the regions are given special meaning.
- The spatial scan statistics adapted for multidimensional points are clever and innovative. The computation resources that this study requires are tremendous.
- The time complexity of the model is $O(N^2)$, which is sensitive to the size of the data. Handling large amount of data would be burdensome, regardless of the efficiency of the implementation.