

Exploring and visualizing spatial effects and patterns in ride-sourcing trip demand and characteristics

Bishoy Kelleny *, Sherif Ishak **

* Old Dominion University,
132B Kaufman Hall, Norfolk, VA 23529,
Civil and Environmental Engineering
bkell006@odu.edu

** Old Dominion University,
135 Kaufman Hall, Norfolk, VA 23529,
Civil and Environmental Engineering
sishak@odu.edu



Article history:

Received: October 26, 2021
1st Revision: October 30,
2021
Accepted: November 15,
2021

DOI:

[10.14254/jsdtl.2021.6-2.1](https://doi.org/10.14254/jsdtl.2021.6-2.1)

Abstract: The complex demand pattern of ride-sourcing remains to be a challenge to transportation modeling practitioners due to the infancy and the inherently dynamic nature of the ride-sourcing system. Spatial effects exploration and analysis protocols can provide informative insights on the underlying structure of demand and trip characteristics. Those protocols can be thought of as an opportunistic strategy to alleviate the complexity and help specifying the appropriate econometric models for the system. Spatial effects exploration is comparable to point pattern analysis, in which, signals from spatial entities, like census tracts, can be analyzed statistically to reveal whether a specific phenomenon respective signal distribution is a completely random process or if it follows some regular pattern. The results of such analysis help to explore the investigated phenomenon and conceptualize its causal forces. In this paper, we apply spatial pattern analysis edge methods integrated into a visual analytics framework to: (1) test the null hypothesis of system demand complete randomness; (2) further analyze and explain this demand in terms of the origin-destination (OD) flow and trips characteristics, i.e., length and duration; and (3) develop a pattern profile of the demand and trip characteristics to provide potential directions to modeling and predictive analytics approaches. This framework helps explain the ride-sourcing system demand and trip characteristics in space and time to fill the gap in integrating the system in multimodal transportation frameworks. We use the ride-sourcing trip dataset released from the City of Chicago, USA, for the year 2019 to showcase the proposed methods and their novelty in capturing such effects as well as explaining the

Corresponding author: *Bishoy Kelleny*
E-mail: bkell006@odu.edu

This open access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license.



underlying complexities in a streamlined workflow. The ride-sourcing demand hotspots were explored and identified in the city's central business district. A novel method to capture and analyze the origin-destination flowlines was developed and implemented. Finally, a complementary trip characteristics pattern analysis was conducted to fully comprehend the system and validate the findings from the system demand points and OD-flowlines.

Keywords: ride-sourcing, transportation network companies (TNCs), visual analytics, principal component analysis, Getis-Ord G_i^* and Local Anselin Moran's I.

1. Introduction

The transportation market is no exception to the evolving trend of sharing economy, as the relatively newly introduced family of transportation services known as shared mobility has witnessed a wide acceptance and adoption among users, especially in highly urbanized areas and metro cities (Roukouni & Homem de Almeida Correia, 2020). Amongst the controversial modes of shared mobility is ride-sourcing, which enables users to have access to the mobility-as-a-service model on an as-needed basis (Shaheen, Chan, Bansal, & Cohen, 2015). Ride-sourcing transportation services, provided by companies like Uber and Lyft, also known as Transportation Network Companies (TNCs), are considered a revolutionary trend in mobility and urban travel, involving a high level of real-time computations and matching algorithms, equipped with seamless payment and surge-pricing, solving a good number of mobility and travel problems (e.g., first and last trip mile). The service runs on top of a set of system enablers, including smartphone application that pairs riders and community drivers (in non-commercial vehicles) given their spatial and temporal constraints, online transactions (e-payment) to guarantee seamless payment method, and a cloud-based surge-pricing scheme matching the supply with the demand (Shaheen et al., 2015). The service is either offered in the form of sequential trips, i.e., serving individual riders, or concurrently to several riders in a ride- and fare-splitting fashion (e.g., Uber Pool and Lyft Line).

Since the inauguration of Uber early in the past decade, the implications of the newly introduced shared, on-demand, and potentially automated mobility services are significant, especially in the way people make their choices on: what activities to do and where and when to do such activities. The demand for TNCs ride-sourcing services is growing exponentially, with evidence from Uber rides that reached 1-billion in approximately six years in late 2015, and only six more months to reach the 2-billion rides (Uber, 2017). This positions ride-sourcing as a game-changer in the transportation market. However, there is a lack of information about the impact such smart mobility services will have on the planning, operation, and management of existing and future transportation networks. It is not yet made clear how people make their mode choice for ride-sourcing, nor whether this demand is induced (Rayle, Shaheen, & Chan, 2014) or shifted from other modes. In either case, research reveals that it has implications on congestion (Erhardt et al., 2019) and transit ridership (Erhardt et al., 2021). In their exploration of TNCs' ride-sourcing impact on congestion intensity and duration, a study (Diao, Kong, & Zhao, 2021) reported increases of 0.9% and 4.5%, respectively, in consistence with the findings of (Erhardt et al., 2019). A decline of 8.9% was reported as related to the ride-sourcing impact on transit ridership, again consistent with the finding in (Erhardt et al., 2021). Moreover, in (HENAO, 2017), the author drove for Uber and Lyft as part of the data collection course and concluded similar findings in terms of increased Vehicle Miles Traveled (VMT) and congestion.

Despite the relief TNCs offered to some accessibility and mobility chronic issues (e.g., impaired driving, first and last-mile connections to public transit, and late-night services), recently released reports and studies suggest that these newly introduced services accounted for approximately 50% of the increase in congestion in San Francisco between 2010 and 2016 (Castiglione et al., 2018). These figures on VMT and transit ridership can even get more significant when considering zero-occupied vehicles in an unmanned-autonomous-vehicles (UAV) ecosystem. Other research concluded that Uber boosts public transit ridership by five percent after two years of introduction in an average transit agency area (Hall, Palsson, & Price, 2018). Henao et al. (2019) (Henao, Marshall, & Janson, 2019) signal

other impacts on travel behavior and parking demand, and transportation equity. Confirming the ride-sourcing controversial nature as a disruptive transportation service with uncertain impacts, Jin et al., 2018 (Jin, Kong, Wu, & Sui, 2018) perceive ride-sourcing as positively impacting the economy and as a chance rather than a threat to transit. While they remain uncertain of its impact on car ownership, energy consumption and greenhouse gas emissions, they affirm the exacerbated digital division and exclusion from an equity standpoint.

Therefore, it was a non-trivial challenge when transportation scientists and practitioners were confronted with the newly introduced mode of TNCs ride-sourcing services with no sufficient data to integrate this mode in the current state of the practice and modeling framework. They relied on state-of-the-art conceptual and simulation-based frameworks to provide a preliminary understanding on the mode without resorting to real-world ride-sourcing trip data (F. Alemi & Rodier, 2016; Basu et al., 2018; Kelleny & Ishak, 2018; Winter, Cats, Martens, & van Arem, 2020). Other endeavors depend on data fusion approach between a relatively small ride-sourcing trip dataset or small to medium scale survey (Farzad Alemi, 2018; Hermawan, 2018) injected in the proposed modeling framework along with more generalized and more enormous datasets (Oviedo, Granada, & Perez-Jaramillo, 2020; Reck & Axhausen, 2020). Stated-preferences (Sperling, Young, Garikapati, Duvall, & Beck, 2019) and household travel revealed preferences surveys also contributed to form an initial understanding of ride-sourcing users' perceived utility (Azimi, Rahimi, Asgari, & Jin, 2020; Dias et al., 2017). The first real-world TNCs ride-sourcing trip data was released in 2014 on Uber along with other for-hire vehicle pick-ups from New York City. The data was limited to pick-up locations only and obfuscated for privacy, yet it was useful to conceptualize analytical spatial and temporal methods (Correa, Xie, & Ozbay, 2017; Faghih, Shah, Wang, Safikhani, & Kamga, 2020) on the first exposure to real-world ride-sourcing trip data.

At this stage of the system's maturation of understanding, a crucial dimension remains underrated and unexplored, which is the system's underlying spatial characteristics and effects. Given the inherent similarities between ride-sourcing and futuristic shared unmanned autonomous vehicles (SUAV), this spatial exploration and segmentation is so pertinent to SUAV-hubs planning and deployment and is very timely. Research should provide answers to questions on: (1) how ride-sourcing trip demand can be analytically and spatiotemporally analyzed; (2) do the demand patterns portrait any forms of clusters or urban pockets, or they are completely random processes? and (3) do the trip characteristics constituting this demand exhibit also a specific pattern of clusters? Capturing those patterns and their spatial effects significance is pivotal to integrating the system into regional multimodal transportation modeling frameworks, and to providing a coherent understanding of the way the system's trip demand and characteristics are spatially being shaped and growing.

Developing such spatial effects exploration protocols, perceived as a research gap in ride-sourcing, is essentially data-driven and would be meaningless if conducted without resorting to real-world data. Therefore, and to fill this gap on the ride-sourcing system research and analysis, we propose the spatial exploration and visualization framework developed in this paper, and we utilize the City of Chicago TNC trip dataset (City of Chicago, 2021) obtained at the census tracts level to showcase the workflow, results, and findings. The rest of this paper is organized in the following fashion to answer the afore-asked questions and attain those very objectives: (1) literature review on the spatial effects methods in different transportation modal domains; (2) case study data and the pipeline on preprocessing, handling and synthesizing their spatial structure and trend mining protocols; (3) methodological aspects of ride-sourcing spatial effects exploration proposed in this work; and finally (4) the implementation remarks and results. The conclusions section will set the future direction to advance this research.

2. Literature review

The term "spatial effects" refers to an observed pattern of dependence (or autocorrelation) in a spatial phenomenon (L. Anselin, 1988; Luc Anselin, 1999; Beron & Vijverberg, 2004). That heterogeneous pattern is typically governed by the underlying fabric, such as socioeconomic or demographic structure in urban phenomena. The integration of spatial effects in transportation planning and modeling remains to be an unbeaten track (Lopes, Brondino, & Silva, 2014), and a challenge to areas of advanced modeling, e.g., Spatial Agent-Based Models, in which contextualizing the spatial effect in the agents set of rules and relationships still remains a challenge (Heppenstall & Crooks,

2019), especially in shared mobility and ride-sourcing context (Kelleny & Ishak, 2018). Even for conventional transportation models, ecological fallacy and spatial dependence were amongst the key methodological issues discussed (Wang, Quddus, Ryley, Enoch, & Davison, 2012). Failing to detect, assess and account for those effects prior to modeling the phenomenon could lead to misspecified linear spatial regression models (Florax & Nijkamp, 2005). In a regression context, spatial heterogeneity is comparable to “non-constant error variances (heteroskedasticity)” (Luc Anselin, 1999). However, in an exploratory analysis context, as such in this research, it refers to agglomerations of spatial entities of phenomenal (high or low) values of the explored system, i.e., ride-sourcing pick-ups and drop-offs (Florax & Nijkamp, 2005).

Spatial effects exploration is comparable to point pattern analysis, in which, signals from spatial entities, like census tracts, can be analyzed statistically to reveal whether the signal distribution is a completely random process or if it follows some regular pattern (Bailey & Gatrell, 1995). The results of this analysis help to narrow down the investigated phenomenon “causal forces” (Boots & Getis, 1988). Correa et al. (2017) (Correa et al., 2017) is amongst the early endeavor to account for spatial effects in modeling ride-sourcing demand. Global Moran I (Moran, 1950) was used just to test for spatial dependence existence in the demand data of Uber and Taxi, which gave rise to adopting spatial lag and spatial error models along with linear models. However, no further exploration on such spatially dependent pattern was conducted in terms of highlighting the associated urban pockets or exploring the effects in trip characteristics. Yu and Peng (2019) (Yu & Peng, 2019) used the same Global Moran I metric on spatial autocorrelation to capture the heterogeneous pattern of ride-sourcing trip demand data prior to developing a Geographic Weighted Poisson Regression model. Soria and Stathopoulos (2021) (Soria & Stathopoulos, 2021), in their work on solo and pooled trips behavior modeling, implemented a spatial Durbin model to account for the so-called “spatial spillover”, or the impact of neighboring entities induced into their ride-splitting behavior.

In a transit ridership context and using GPS trajectories and smart card data, a study (Tu et al., 2018) highlighted the limitations of ordinary least square (OLS) globally fitted regression models in capturing spatially variant (heteroscedastic) demand, and therefore, proposed a Geographically Weighted Regression model (GWR). However, the GWR implementation was justified only by the clustered pattern of the dependent variable and regressors through visuals and quantified by the GWR outperformance at the expense of the OLS global model. No statistical protocols were utilized to assert such a pattern. Also (Liu, Sun, Sun, & Gao, 2020) adopted the same GWR modeling approach to control for the spatial effects in some of the regressors. In this work, the urban pockets breeding such heterogeneity, referred to as hotspots, were explored by means of Kernel density estimation (KDE), since the trip data was delivered in GPS trajectories format. However, no further elaboration on the statistical significance of those hotspots was provided in this work.

In a rail trip demand setting, a study (Cordera, Sañudo, dell’Olio, & Ibeas, 2018) introduced the Spatial Filtering approach to control for spatial effects in the demand driving variables to two models, namely, Poisson regression and gravity, to estimate the demand distribution between rail stations. The spatial effects accounted for by the gravity model exhibited significantly better performance, yet the models’ performance metrics were the only gauge on spatial dependence. In their “exploratory analysis on freight trip attraction” (Sánchez-Díaz, Holguín-Veras, & Wang, 2016), the authors utilized the traditional Global Moran’s I referred to earlier, but with a localized implementation using Local Indicators of Spatial Association, also known as Local Anselin Moran’s I (Luc Anselin, 1995) to reveal the pattern of association in the proposed freight model explanatory variables related clusters. The explorative framework for spatial effects then gave rise to adopting a spatial econometric model that outperformed the OLS fitted model.

The framework proposed in this paper makes two major contributions in the ride-sourcing research area: (1) conceptual contribution: on incorporating spatial effects and heterogeneity in trips data prior to model specification and development; (2) methodological contribution: on analyzing TNC trip data and providing guidance to researchers and practitioners on spatial effects exploration and identification in such type of data to integrate it into multimodal transportation frameworks. The implications of this research can be extended to informed policy-making, particularly congestion pricing. Identifying the demand hotspots and the heavily trafficked OD-pairs can help better implement measures like cordon and/or corridor-based congestion pricing mechanisms to regulate and mitigate the impacts of TNCs on congestion.

3. Research methods

The research framework presented in this paper is data-driven, i.e., we pair our proposed methodological aspects of ride-sourcing spatial effects exploration and identification with a case study to showcase the workflow, results, and findings. Therefore, it would be indispensable to discuss the background of the data we utilize in this section, along with the cleaning and rebuilding protocols. Thus, the first sub-section is organized to include: (1) trip ends, i.e., pick-up and drop-off, cleaning and preprocessing; (2) trip data spatial structure and pattern mining and exploration; and (3) origin-destination flowlines cleaning and preprocessing. In the second sub-section, we focus on the development of the methods.

3.1. Trip data description, cleaning, and preprocessing

Starting from November 2018, all TNCs operating in the City of Chicago were required by ordinance to report all their trips regularly, and the TNC trip data has been made publicly available through the Chicago Data Portal and updated on a quarterly basis. We chose to work with the year 2019 data to avoid any potential flaws in data reporting or packaging at the start of the initiative in 2018, and not to delve into odd travel behavior associated with the 2020 COVID-19 pandemic hit. The 2019 data has more than 97 million trip records. The variables relevant to this research are summarized in Table 1, along with their type, description, and cleaning and rebuilding protocols, if any. Trips start or end in either of the airports' census tracts, O'Hare International Airport and Chicago Midway International Airport, are eliminated from this analysis, too, since the airports' demand is not within the scope of this study due to their distinctive nature. Trips with a length of more than 50 miles are excluded, since they are not compatible with the extent of the study area.

Table 1: City of Chicago TNC Trip Dataset- Key Variables Glossary

Variable	Type and Description	Cleaning or Rebuilding Protocol
Trip ID	Text; a unique identifier to the trip	No missing data; no further cleaning was needed.
Trip Start Timestamp	Date & Time; when the trip started, rounded to the nearest 15 minutes	Eliminate entire trip record for missing record
Trip End Timestamp	Date & Time; when the trip ended, rounded to the nearest 15 minutes	Eliminate entire trip record for missing record
Trip Seconds	Number; time of the trip in seconds	Missing entries were estimated by querying similar trips of the same start timestamp of 1-Hour window and same pickup and dropoff census tracts and compute the average
Trip Miles	Number; distance of the trip in miles	Same estimation protocol used for trip seconds
Pickup Census Tract	Text; the Census Tract where the trip began.	Missing records were rebuilt by spatially join the pickup centroid location to the City of Chicago census tracts shapefile
Dropoff Census Tract	Text; the Census Tract where the trip ended.	Same rebuilding protocol used for pickup census tracts
Pickup Centroid Location	Point (Geometry); the location of the center of the pickup census tract (longitude, latitude)	If entry is missing along with the pickup census tract, the entire trip record is removed
Dropoff Centroid Location	Point (Geometry); the location of the center of the dropoff census tract (longitude, latitude)	If entry is missing along with the dropoff census tract, the entire trip record is removed

Source: (City of Chicago, 2021)

The trip characteristics distribution depicted in Figure 1-a and Figure 1-b shows more dense likeability of short trips. However, erroneous observations appear to exist in the data, as can be seen in the studied relation between trips' lengths and durations as shown in Figure 1-c. Treating these erroneous values as outliers would result in losing a considerable portion of the data on trip characteristics. Moreover, intuitive scrutinizing of those outliers leads to a strong belief that they are the outcome of errors in measurements or handling. For example, some observations have the length of several miles, yet the duration is a fraction of the second. This is a typical case of data contamination, in which data reduction can be a good remedy (Serneels & Verdonck, 2008), especially with the low-dimensionality space dealt with in this section. Since we are attempting to reduce the dimensions of the trip length-duration data

while refraining from distorting the variance in the pattern we are studying in the first place, the principal component analysis (PCA) can be a good candidate. The PCA has been a key player in this application area due to its protocol for minimizing the variance loss (Jolliffe & Cadima, 2016).

Figure 1-a: Trip Length Distribution

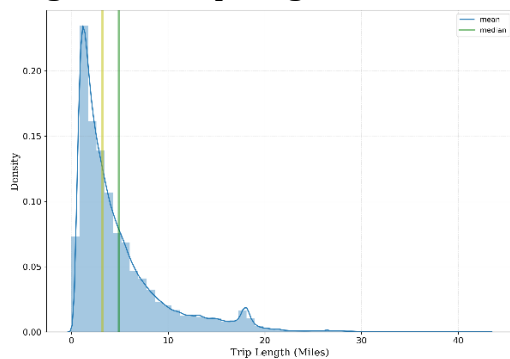


Figure 1-b: Trip Duration Distribution

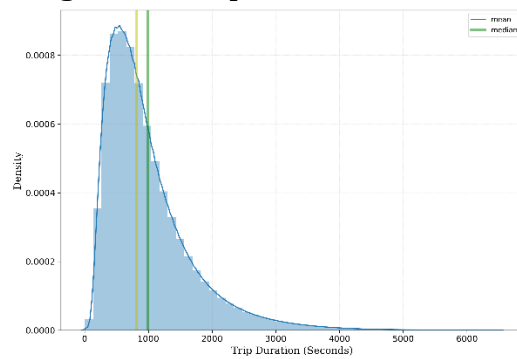


Figure 1-c: Trip Duration versus Length

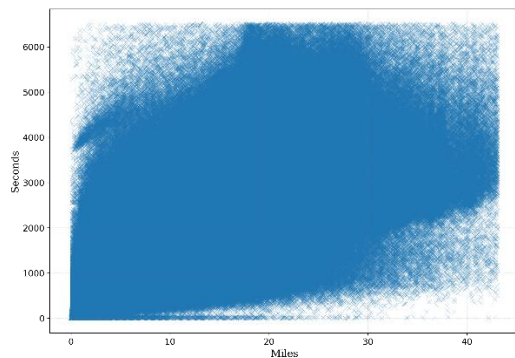
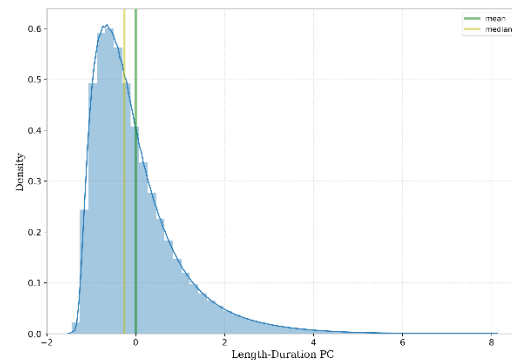
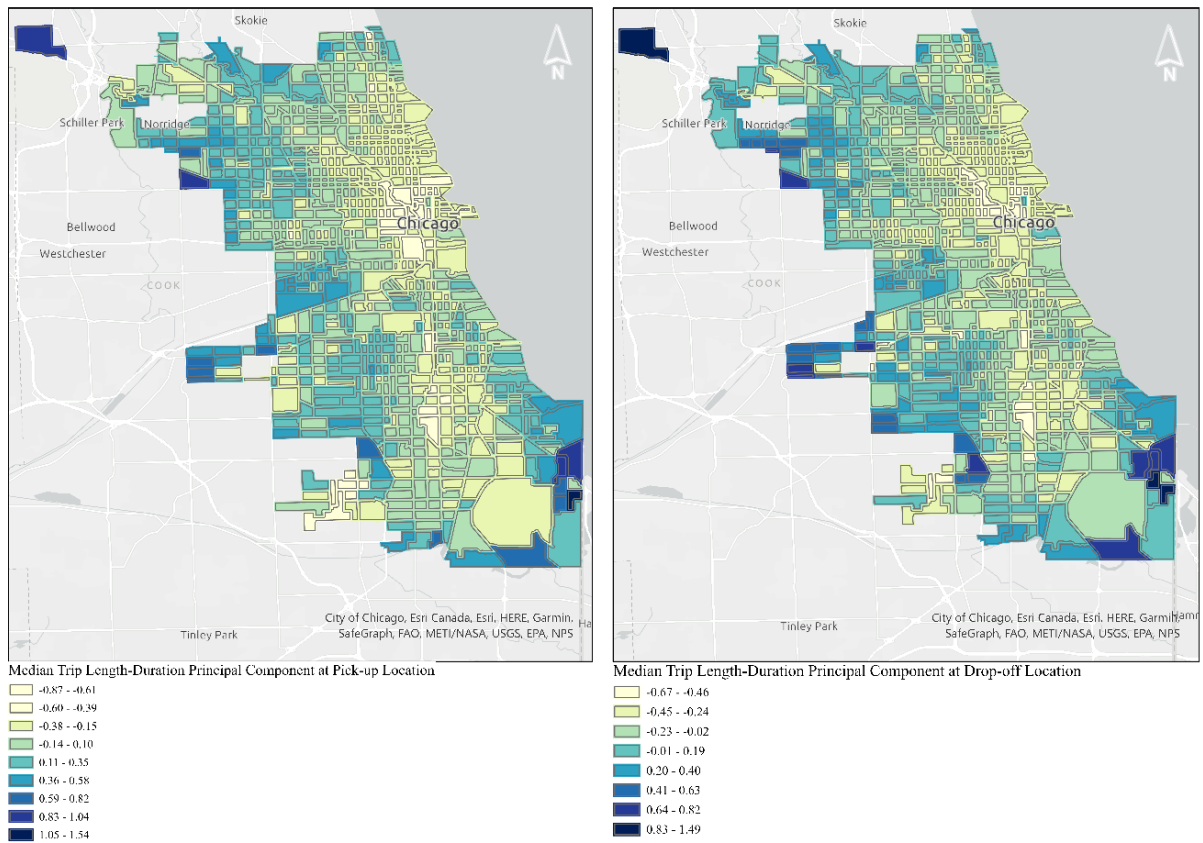


Figure 1-d: LDPC Distribution



A key concern for applying PCA to the trip characteristics is the impact of the outliers on the quality of the process as discussed in (Sapra, 2010), for which robust variants of the PCA were proposed and developed. This concern is magnified in high-dimensional applications, which is not typically the case tackled here. Moreover, the two features reduced here can be readily assumed to be correlated, and therefore, the impact of the outliers would be mitigated by the contribution of the common surface from correlation. Therefore, a classic PCA is experimented to reduce the trip lengths and durations into one component, while adopting the explained variance ratio as the determinant metric on the approach performance and validity. The trip characteristics data were not standardized prior to implementing the PCA since the PCA workflow in the exploited Python Machine Learning Library Scikit-learn (Pedregosa et al., 2011) already standardizes the data before decomposition. The resulting component is then scaled to the unit variance as this is found to be consistent with the downstream application of local Anselin's index. An explained variance ratio of 0.9999 is returned from the PCA, which indicates a minimal loss of variance in the decomposition, and the distribution of the trips' length-duration principal component (LDPC) is shown in Figure 1-d. The skewness in the distribution of the trips' LDPC suggests a higher frequency of shorter trips. The median, therefore, is chosen to represent the central location of the trips' LDPC of the analyzed census tracts. The geographical context of the trips' median LDPC is shown in Figure 2. The LDPC median range expands from negative fractional values representing very short trips in terms of length and duration to positive values greater than 1 for relatively longer trips.

Figure 2: Median TNC Trip LDPC at Pick-up and drop-off Census Tracts



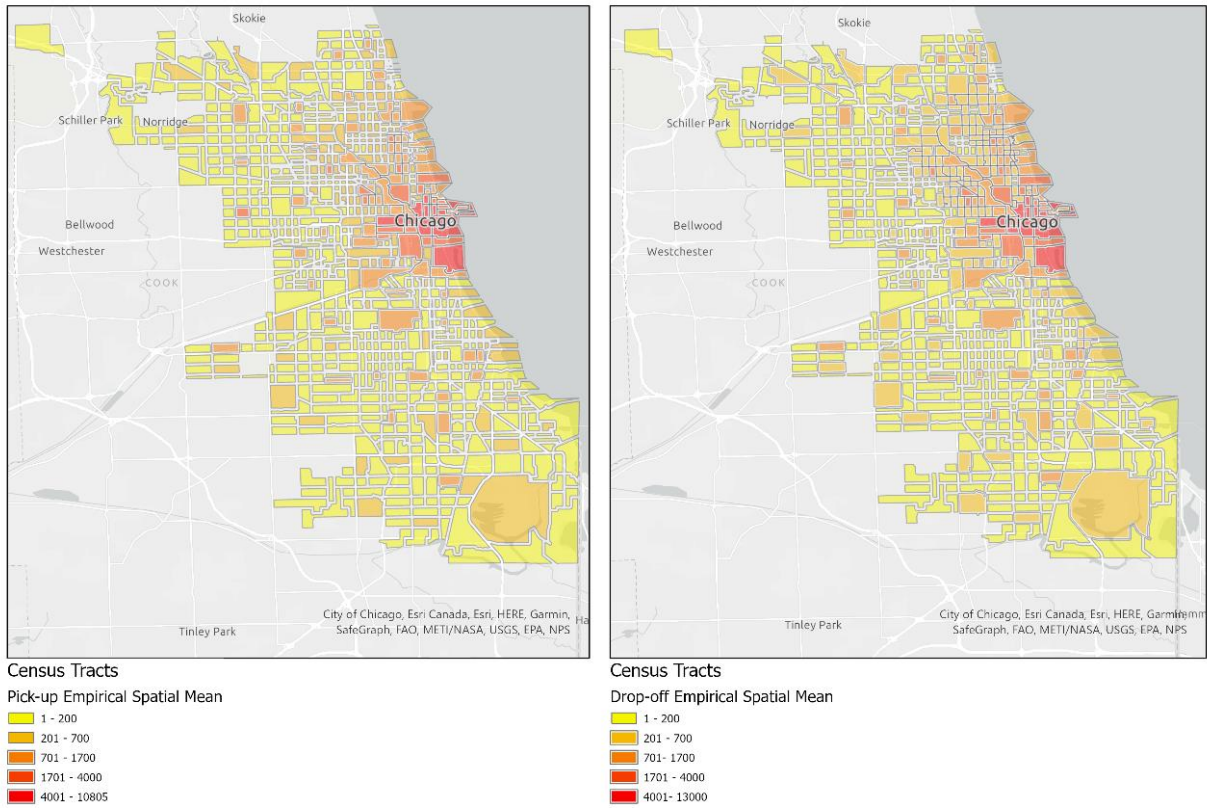
3.2. Trip Data Spatial Structure and Trends Mining

The empirical spatial means (ESM) for the daily pick-ups and drop-offs, $\hat{\mu}_{P,S}(S_i)$ and $\hat{\mu}_{D,S}(S_i)$, respectively as in (1) and (2), reveal the dominant magnitude of demand across the investigated geography during the analysis period. The denominator in the ESM equations, T , equals the number of days in which the corresponding census tract has pick-ups or drop-offs more than zero. As shown in Figure 3, daily pick-ups and drop-offs exhibit similar pattern of magnitude concentrated in and around the central business district (CBD) census tracts and vanishing in the direction approaching the peripheries. It should be denoted that ESM would not necessarily indicate a statistically significant hotspot, which is an exercise that should be worked out on its own, as will be shown in the next section.

$$\hat{\mu}_{P,S}(S_i) \equiv \frac{1}{T} \sum_{j=1}^T P(s_i; t_j) \quad \text{Pick-up Empirical Spatial Mean} \quad (1)$$

$$\hat{\mu}_{D,S}(S_i) \equiv \frac{1}{T} \sum_{j=1}^T D(s_i; t_j) \quad \text{Drop-off Empirical Spatial Mean} \quad (2)$$

Figure 3: Empirical Spatial Means for Daily Pick-ups and Drop-offs in the Year 2019



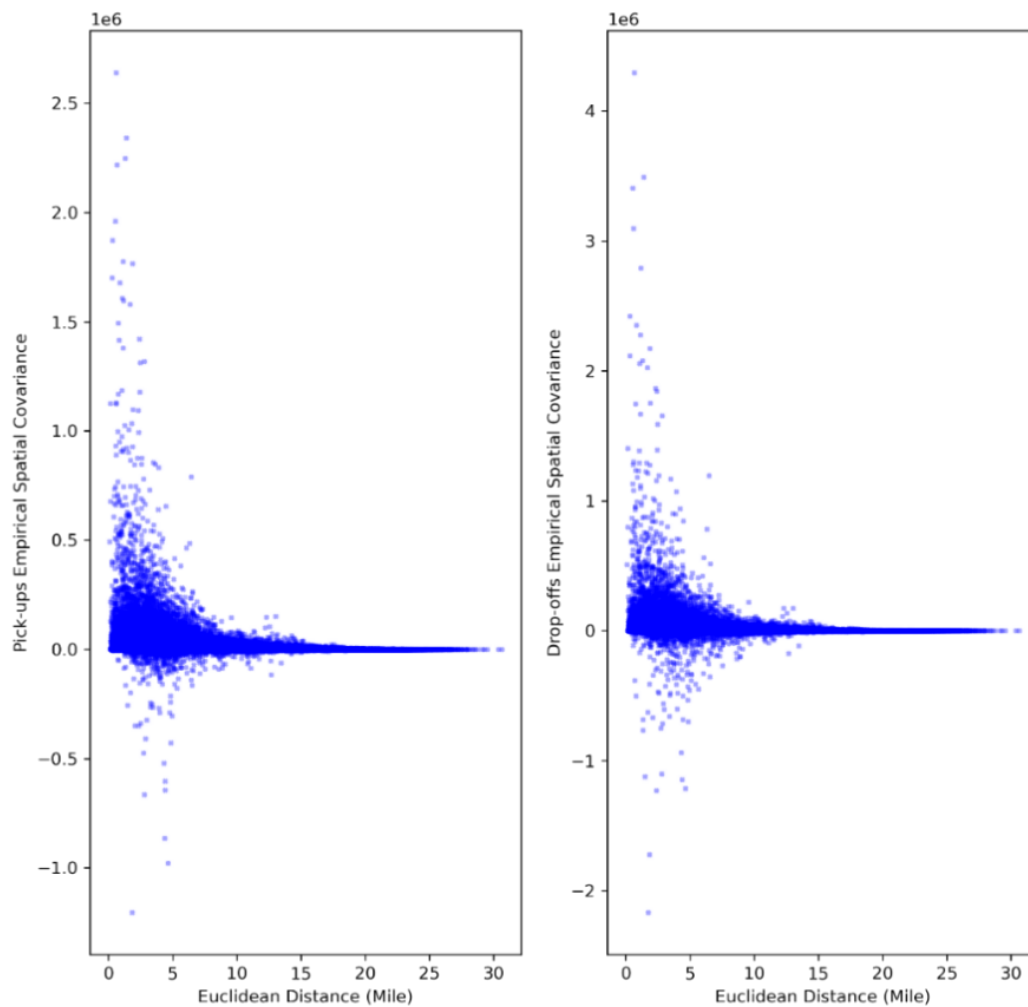
The demand patterns revealed by means of the ESM may suggest that ride-sourcing system demand is not a perfectly random process. Analyzing the empirical spatiotemporal covariance (ESTC) can provide a preliminary understanding in this regard. ESTC between two census tracts, S_i and S_j , is the covariance between the demand time series in both geographies. This ESTC matrix for pick-ups and drop-offs can be constructed as shown in (3 and (4, respectively). A pairwise census tracts' Euclidean Distance (ED) Matrix is constructed to study the interdependencies between the census tracts' demand co-variability, i.e., ESTC and the respective ED,. . One can notice an exponentially decaying trend between the ESTC and the ED shown in Figure 4, which may again suggest for spatial effects in the system.

$$\hat{C}_{P,S_k,S_l} \equiv \frac{1}{T-1} \sum_{j=1}^T (P(S_k; t_j) - \hat{\mu}_{P,S}(S_k))(P(S_l; t_j) - \hat{\mu}_{P,S}(S_l)) \tag{3}$$

$$\hat{C}_{D,S_k,S_l} \equiv \frac{1}{T-1} \sum_{j=1}^T (D(S_k; t_j) - \hat{\mu}_{D,S}(S_k))(D(S_l; t_j) - \hat{\mu}_{D,S}(S_l)) \tag{4}$$

Where: $l, k \in S_i$ (Census Tracts Set)

and \hat{C}_{P,S_k,S_l} : Pick-up Empirical Spatiotemporal Covariance & \hat{C}_{D,S_k,S_l} : Drop-off Empirical Spatiotemporal Covariance

Figure 4: TNC Demand Empirical Spatial Covariance and Euclidean Distance

3.3. OD Flow Data Cleaning and Preprocessing

The OD flow data had to undergo some cleaning and preprocessing protocols so that their magnitude could be analyzed and the underlying spatial effects could be explored and evaluated statistically. The study area comprised initially 638,401 OD-pairs, excluding the two airports census tracts. They are visualized by means of the total year trips encoded to the respective flowline in Figure 5 and weighted by 2×10^{-5} to reduce the visual clutter and emphasize the heavily trafficked OD flowlines. The trip ends of the census tracts' internal trips, i.e., trips that started and ended within the same census tracts and ended up having both trip ends assigned to the census tract centroid, were assigned to two random geolocations within the respective census tract polygon (see Handling Internal Trips in Figure 5(a)). Thus, heavily trafficked internal OD-flowlines would be captured in this analysis too. Figure 5(b) depicts high OD intensity near the CBD area.

Only 25% of the aggregated OD pairs accumulated more than 17 trips during the entire analysis year. Figure 6 offers a better depiction of the skewness and the kurtosis of the aggregated OD flow. This long-tailed thin distributional shape, along with the exponentially decaying trend of the relation between the flow and the Euclidean length of the OD-pairs depicted in Figure 7, both suggest a remarkable less likeability of flow between distant origin and destination census tracts. Therefore, to conduct a conclusive analysis on the magnitude of the OD flow between census tracts in the study area, an OD-flow sampling protocol is recommended in this study. We propose an arbitrary number of 2000 of aggregated trips as the threshold for OD-pairs to be further considered. After Filtering the data using that threshold, we remain only with 7142 OD-pairs.

Despite the significant portion of the pairs masked by that threshold, this approach should be acceptable in the context of capturing and analyzing the magnitude of the OD flow between census tracts. Following this step, a mesh of cell size of approximately 300 x 300 feet is constructed to summarize and compute the density of the OD flow. This visual analytic method is inspired by the work proposed in (Wood, Dykes, & Slingsby, 2010), and known as cell-based symbolization, in which the local number of flows is captured within each cell of the mesh, enabling further analytics to be performed on the flow. To elaborate, the scope of the analysis is to capture spatial heterogeneity and spatial dependence in the pattern of the OD flow; thus, any loss in the granularity of the pairs' flow magnitude would hinder such attainment. The OD map visual analytic method allows for reducing the visual clutter as well as capturing the analytical component on the flow. However, amongst the limitations of the method is the inability to discern the direction of the flow, yet that is not a barrier in this spatial analysis while focusing on the heterogeneity and dependence of the flow magnitude.

Figure 5: 2019 Ride-sourcing OD-flow

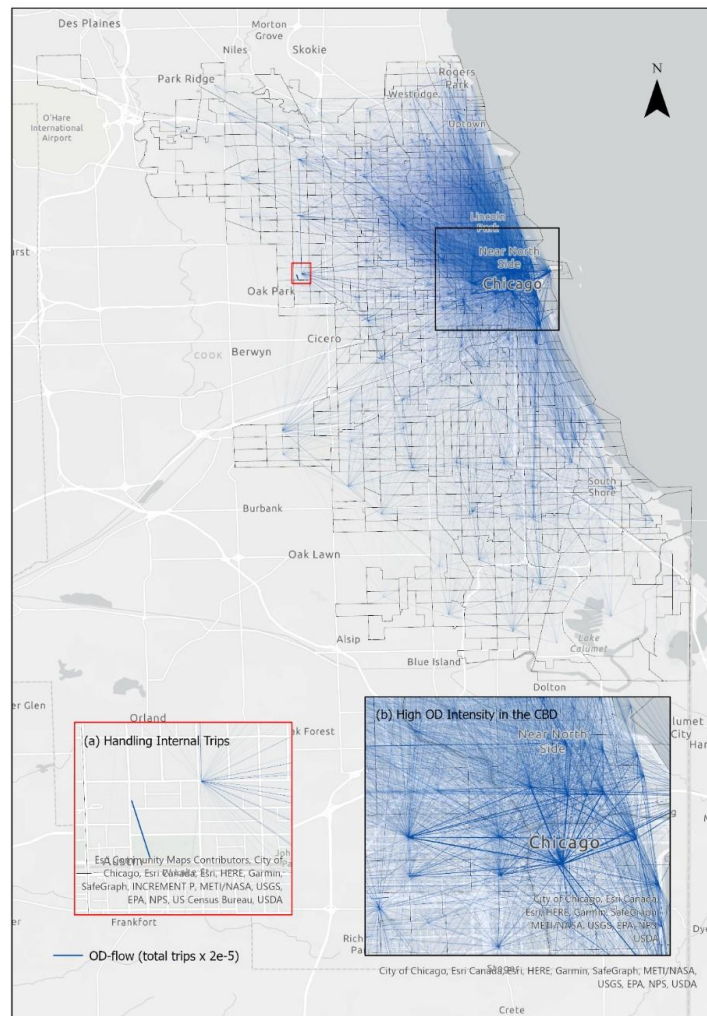


Figure 6: The Year 2019 Aggregated OD Flow Between Chicago Census Tracts

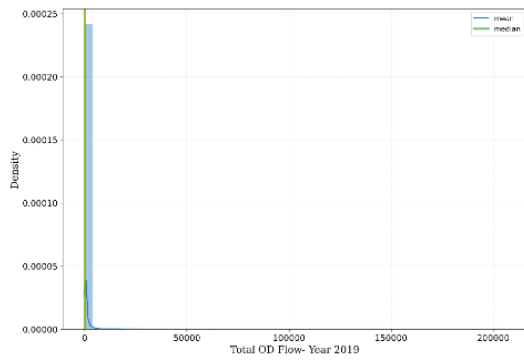
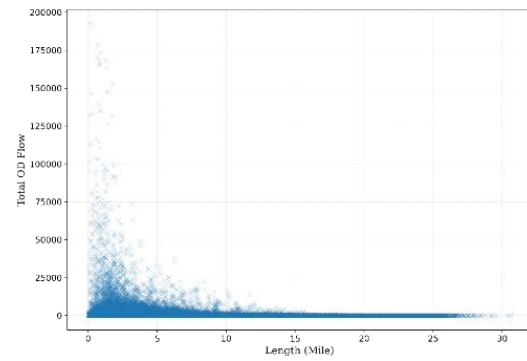


Figure 7: Relation between OD-pair Euclidean Distance and Total Year Flow



3.4. Methodological aspects of ride-sourcing spatial effects exploration

We focus on two aspects of the underlying pattern: (1) hotspot areas; and (2) clusters and outliers. Hotspot analysis can provide key insights on ride-sourcing trip demand and OD-flow spatial heterogeneity with rigorous statistical tests on significance as developed by Getis and Ord (Getis & Ord, 2010) and known as “Getis-Ord G_i^* ” (Esri Inc., 2021b). Getis and Ord developed the basic statistic to test the spatial association between significantly high and low weighted values with respect to their spatial relations. The definition of this spatial weight incorporated in the test statistic is pivotal. Based on the findings from the visual representation of the ESTC and ED in the previous section, an inverse Euclidean squared distance between the census tracts’ centroids is selected to tune the association tested. The Z-values, essentially the standard deviation of the Getis-Ord G_i^* (see (5) (Esri Inc., 2021b; Getis & Ord, 2010)), test the null hypothesis that the i^{th} tested spatial entity is not a significant hot or cold spot. High Z-values, i.e., above 1.65, 1.96, and 2.58, and their respective low p-values, i.e., 0.1, 0.05, and 0.01 will result in rejecting such hypothesis and identify the spatial entity as a hot or cold spot at confidence levels of 90%, 95%, and 99%, respectively.

$$Z_i = \frac{G_i(d) - E[G_i(d)]}{\sqrt{\text{Var } G_i(d)}} = \frac{\sum_{j=1}^n w_{i,j} x_j - \bar{X} \sum_{j=1}^n w_{i,j}}{S \sqrt{\frac{n \sum_{j=1}^n w_{i,j}^2 - (\sum_{j=1}^n w_{i,j})^2}{n-1}}} \tag{5}$$

Where: d is a defined distance; and $j \in \{1, \dots, n = 799\}$

$$\bar{X} = \frac{\sum_{j=1}^n x_j}{n}$$

$$S = \sqrt{\frac{\sum_{j=1}^n x_j^2}{n} - (\bar{X})^2}$$

The signal x_j is the pattern value in geography j ; $j \in \{1, \dots, n = 799\}$, namely the ESM $\hat{\mu}_{p,S}(S_i)$ for the daily pick-ups, and the ESM $\hat{\mu}_{D,S}(S_i)$ for drop-offs. As for the $w_{i,j}$, it is the spatial weight; the inverse Euclidean distance squared between the geographies’ centroids. It is noteworthy that the signal value of the target geography x_i is not included in the global format of Getis-Ord General G (Getis & Ord, 2010). This absence of the target geography from the scoring process allows for capturing the spatial effect, if any, on the adjacent geographies in a presumably exponentially decaying weight, i.e., to evaluate the signal from the hypothetically significant cluster against the global signal value. This should be emphasized here because the approach adopted to analyze the pick-ups and drop-offs pattern, although local, is adapted from a global one to reveal the existence of significant value clusters. Whilst for the analysis that will be conducted on the trips’ LDPC in a later section, an explicit local approach is adopted, in which the statistical test will be performed on a signal-to-signal basis.

The cluster and outlier analysis proposed for exploring the spatial effects in the trips’ LDPC is predicated on Anselin Local Moran’s I (Luc Anselin, 1995) and is proposed to reveal how the spatial pattern of trip characteristics complements the findings from the hotspot analysis and provides a full image on the system’s spatial pattern. This cluster and outlier analysis approach should be distinguished

from the hotspot analysis approach, though. Albeit both approaches evaluate local signals, the first one, i.e., Getis-Ord G_i^* does that against the average global signal, whilst the Anselin's Local index looks into the way the signals vary from one geography to another. This is found to be more harmonious with the tested signal, i.e., the trip LDPC, to see how hot or cold spots, if any, constituted by urban pockets of trips, are spatially clustered by their LDPC. The test statistic, therefore, is not calculated directly for the signal, but rather each geography is assigned an index, that is, the Anselin Local Moran's I_i , derived from the signal (see (6) (Luc Anselin, 1995; Esri Inc., 2021c), which is thereafter used for the Z test statistic (see (7). The null hypothesis in this analysis is that the system is an outcome of a complete spatial random process.

$$I_i = \frac{x_i - \bar{X}}{S_i^2} \sum_{j=1, j \neq i}^n w_{i,j} (x_j - \bar{X}) \quad (6)$$

Where: $i, j \in \{1, \dots, n = 799\}$;

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n}$$

$$S_i^2 = \frac{\sum_{j=1, j \neq i}^n (x_j - \bar{X})^2}{n - 1}$$

The z-score:

$$Z_{I_i} = \frac{I_i - E[I_i]}{\sqrt{V[I_i]}} \quad (7)$$

Where: $E[I_i]$ is the mean and $V[I_i]$ is the variance, computed as:

$$E[I_i] = \frac{\sum_{j=1, j \neq i}^n w_{i,j}}{n - 1}$$

$$V[I_i] = E[I_i^2] - E[I_i]^2$$

And $E[I_i^2] = A - B$; where:

$$A = \frac{(n - b_{2_i}) \sum_{j=1, j \neq i}^n w_{i,j}^2}{n - 1}$$

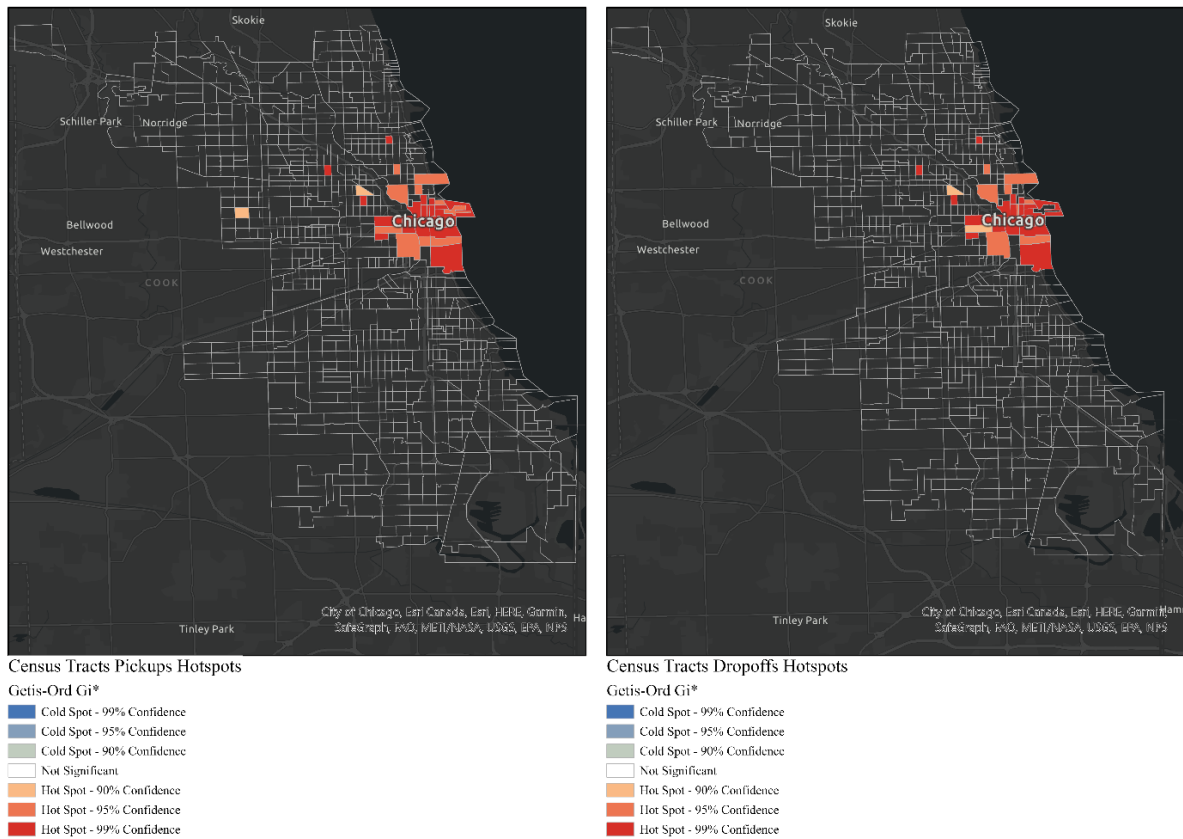
$$B = \frac{(2b_{2_i} - n) \sum_{k=1, k \neq i}^n \sum_{h=1, h \neq i}^n w_{i,k} w_{i,h}}{(n - 1)(n - 2)}$$

$$b_{2_i} = \frac{\sum_{i=1, i \neq j}^n (x_i - \bar{X})^4}{(\sum_{i=1, i \neq j}^n (x_i - \bar{X})^2)^2}$$

4. Implementation remarks, results and discussion

The proposed methods are implemented using the Hot Spot Analysis (Getis-Ord G_i^*) module in ArcGIS Pro spatial statistics toolbox (Esri Inc., 2021b). The results on TNC ESM pick-ups and drop-offs shown in Figure 8 support the existence of a hotspot of ride-sourcing demand in the CBD area, consistent with the early findings from the visuals in Figure 3. Nearly all census tracts outside the CBD cordon exhibit non-significant demand patterns comparable to the magnitude and concentration of the CBD one. The existence of such a hotspot flags heterogeneity in ride-sourcing demand.

Figure 8: Daily Pick-ups and Drop-offs Hotspot Analysis Using Getis-Ord Gi* Statistics

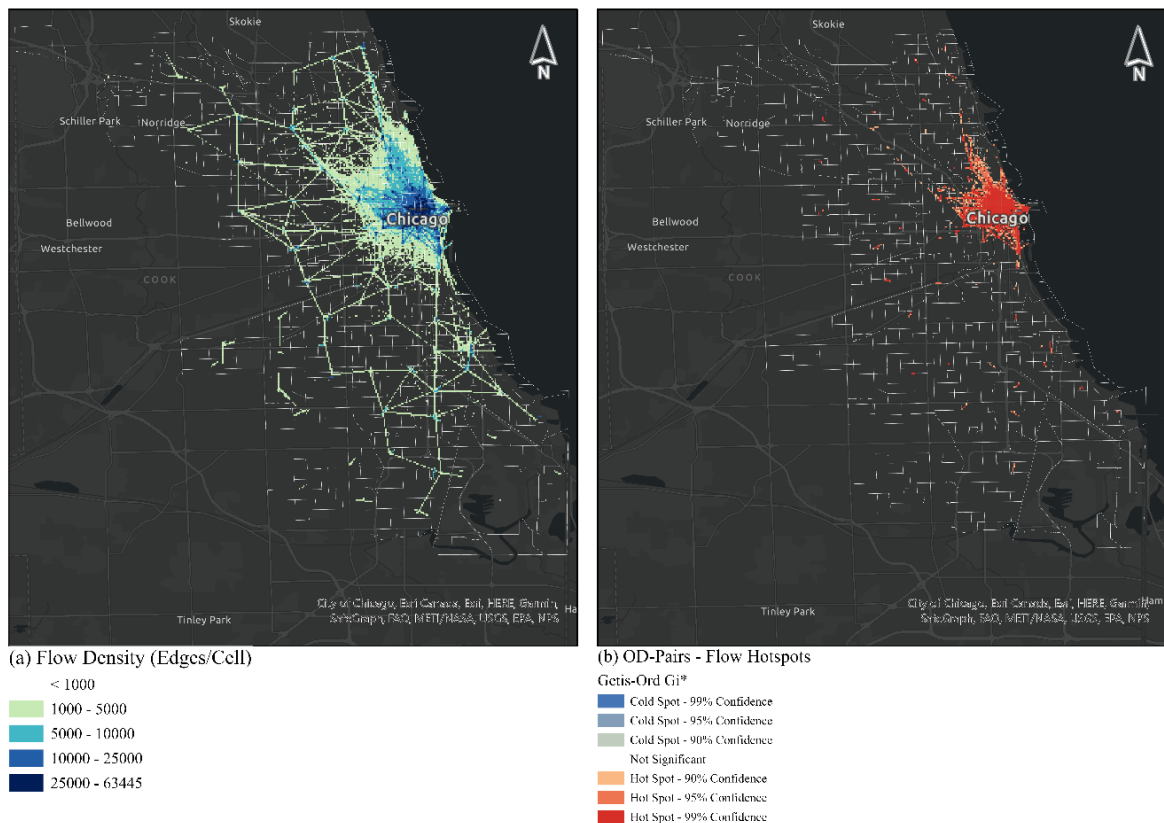


The hotspot analysis on the OD-flow is conducted on the OD-summarizing cells explained earlier by means of the Hot Spot Analysis (Getis-Ord Gi*) module in ArcGIS Pro spatial statistics toolbox (Esri Inc., 2021b). The spatial relationship between the OD-summarizing cells is conceptualized as an exponentially decaying relation, i.e., inverse Euclidean distance squared, as it is elicited from the trend observed in Figure 7. The cell-based OD mapping process shown in Figure 9 (a) shows prevailing OD flow activities in the CBD and extended into the northwest census tracts. The OD hotspot analysis results shown in Figure 9 (b) reveals similar findings with the hotspot cells concentrated around the CBD and dissipating into thin traces toward the northwest.

Examined visually, the trips' LDPC pattern shown in Figure 2 suggests the earlier elicited hypothesis on the hot demand spots in the CBD area that they are driven by longer trips from peripheral census tracts and relatively shorter trips from the CBD and the adjacent census tracts. Yet this remains a hypothesis until tested using the robust Anselin Local Moran's I process on cluster and outlier analysis. The conceptualization of the spatial relationship in this analysis is based on the inverse of the mere distance between the census tracts, unlike the inverse squared distance adopted in the hotspot analysis on pick-ups and drop-offs and the subsequent analysis on OD-flow, for which, there was a strong belief supported by evidence on an exponentially decaying trend with distance.

From the input signal on trips' LDPC, the Anselin Local Moran's I, and the respective Z-scores are calculated as explained earlier. The Z-score and p-value reveal the statistical significance in the context of the hypothesis tested. Higher Z-scores and p-values indicate the presence of spatial association of high or low signals and will lead to rejecting the null hypothesis, and the outcome on the cluster type will label the statistically significant geographies whether they are significant in high or low signal magnitude. This elaborated workflow highlights why this analysis should better answer the question on the existence of spatial effects in trip characteristics than other global or semi-global approaches, e.g., Getis-Ord Gi*, due to the locality of the test statistic of location-by-location basis. Thus, a propagation pattern from the CBD on trip characteristics, if existing, can be observed and highlighted.

Figure 9: Chicago TNC Trips OD Map (Pairs' Flow > 2000 trips): (a) Flow Density – Edges/Cell; (b) OD-Pairs- Flow Hot Spots



The cluster and outlier analysis process was conducted using the ArcGIS Pro module on cluster and outlier analysis (Esri Inc., 2021a). The null hypothesis presumes the complete spatial random property of the system’s examined signal, in this case, the trips’ LDPC. Thus, the pattern of the census tracts trips’ LDPC, and consequently the Anselin’s Local I, is rearranged permutationally around each signal location to test the null hypothesis on complete randomness against the hypothesis on pronounced cluster pattern. In this sequence, a pseudo p-value instead of the conventional p-value is the basis for statistical significance. A pseudo p-value is the proportion of the random permutations in which the respective Anselin’s Local Moran-I reflected a more pronounced clustering pattern than the real I. To elaborate, in a randomization process that comprises N-permutations, each census tract will develop a random distribution of Local Moran-I consisting of N-random variables computed during the permutations, and this will result in a “conditional permutation at each location” (Luc Anselin, 2016). The real I, obtained from the actual pattern, will then be compared against each conditional permutation I, using the Z_{I_i} , and in every time $Z_{I_{i_n}}$ from a permutation $n \in N$ exhibits a more significant clustering pattern than Z_{I_i} for the real I_i , this will count toward the pseudo p-value proportion. Lastly, at a confidence level of 95%, if pseudo p-value is less than 0.05, we reject the null hypothesis on complete randomness.

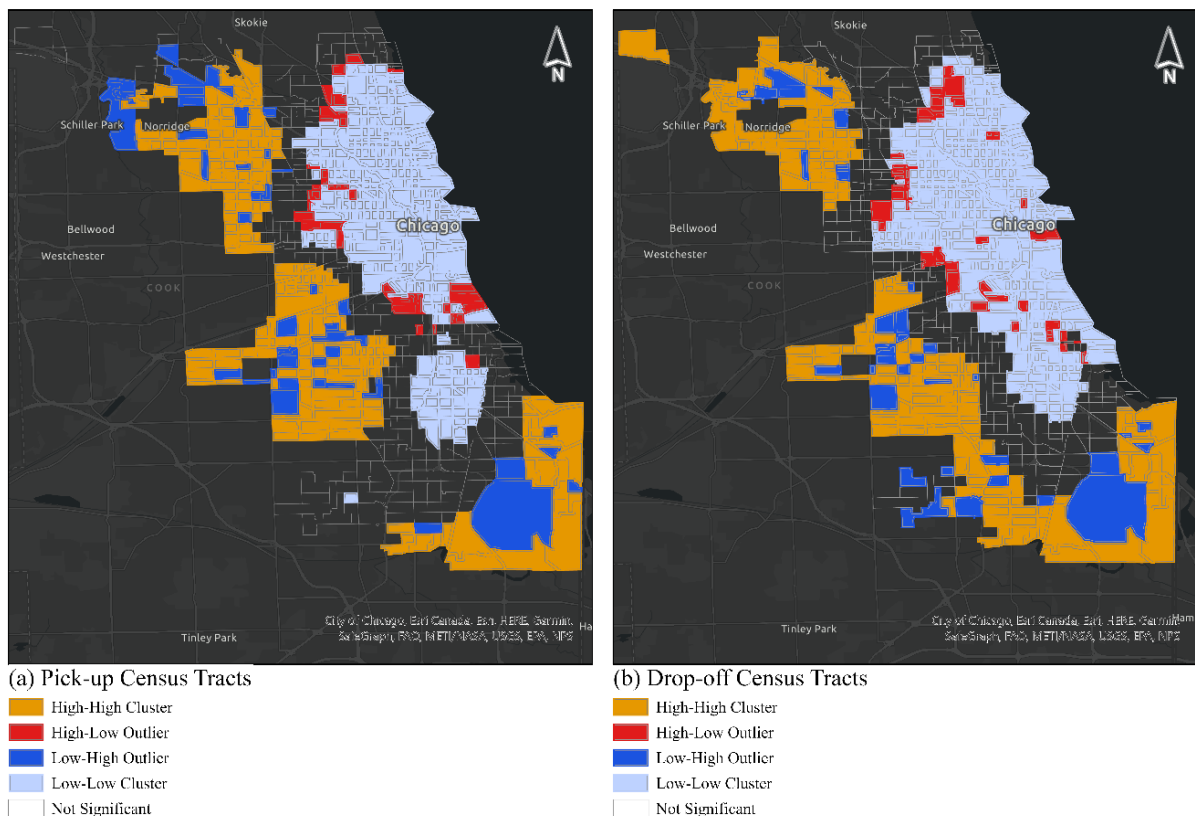
The precision of the experiment in accepting or rejecting the hypothesis is governed by the number of the permutations performed. Noting that the entire dependence upon the analytical approximation of the index (Local Moran-I) is expected to inferentially underperform (Luc Anselin, 2016), i.e., ill-conceived hypotheses. Thus, an arbitrary number of 799 permutations, equals to the number of census tracts in the study area, is chosen to conduct the analysis. However, testing multiple hypotheses raises concerns on alpha-error inflation, i.e., false-positive or type-I error. Therefore, a false discovery rate (FDR) correction is adopted, as developed in (Benjamini & Hochberg, 1995) to control for the false positives.

The cluster and outlier analysis results in five types of clusters and outliers, as shown in Figure 10, and can be interpreted as follows:

- 1) High-High Cluster: a statistically significant cluster of pronounced high signal values, i.e.,

- 2) High-Low Outlier: census tracts that have their trips' length-duration component comparatively low but fall within the significantly identified High-High cluster
- 3) Low-Low Cluster: statistically significant cluster of pronounced small signal values, i.e., census tracts with remarkably smaller trips' length-duration median component
- 4) Low-High Outlier: census tracts that have their trips' length-duration component comparatively high but fall within the significantly identified Low-Low cluster
- 5) Not Significant: census tracts with their trips' length-duration component median statistically neither significant high nor significant low and found to be unidentifiable in terms of signal value

Figure 10: Cluster and Outlier Analysis of TNC Trips' Length-Duration Principal Component



The results align exactly with the initial observation made on the short trips concentrated in and around the CBD, and longer trips in the peripheral census tracts. There is a clear buffer of not significant census tracts separating the two extremely significant clusters, with a trivial number of outliers as the exception that proves the rule.

Conclusions

The workflow of the analysis conducted in this paper adopted two spatial statistical methods to reveal the underlying effects in ride-sourcing trip demand, OD-flow, and trip characteristics; (1) Getis-Ord G_i^* ; and (2) Local Anselin Moran's I , in their respective domain of implementation. The key findings from the Getis-Ord G_i^* implementation on trip ends and OD-flow hotspot analysis can be summarized as follows: (1) TNC Trip Ends, i.e., pick-ups and drop-offs, both show spatial heterogeneity of statistically significant high rates concentrated within and around the CBD area; (2) TNC OD-flow analysis shows a strong exponentially decaying trend between the OD flow and the Euclidean distance (OD length); and (3) TNC OD-pairs show spatial heterogeneity, statistically significant too, with a larger footprint extending from the CBD toward the northwest census tracts. That agglomerations of demand in the

census tracts of Chicago's CBD and its surroundings, represented in the significantly identified hotspots of pick-ups and drop-offs, along with the hot OD-flowlines concentrated in the CBD and vanishing toward the peripheral areas, should be further explored in predictive analytics framework on ride-sourcing spatially contextualized pick-ups and drop-offs, as well as city-wide OD-flow modeling.

As for the Local Anselin Moran's I implementation on the clusters and outliers of trip length and duration characteristics, the results flag another key finding on the significantly shorter trips associated with that hot demand in the CBD, and longer trips, significant too, concentrated in the less demanding peripheral census tracts. Those findings on the system trip ends, and OD flow provide a clear image on the existence of spatial effects in the ride-sourcing system from a demand perspective as well as trip characteristics and would require further analysis to explore whether such pattern exists in willingness-to-share and mile-price. Moreover, the findings of this paper can typically be further utilized to fill two current practice and research gaps regarding: (1) standardizing ride-sourcing trip data handling protocols; and (2) incorporating spatial effects into agent-based modeling frameworks of ride-sourcing systems. The latter is considered in the critical component of social networking and interaction present in agent-based models, and their pertinence to spatial effects.

Funding

Not applicable

Conflicts of interest/Competing interests

Not applicable

Acknowledgements

The authors thank the anonymous reviewers for their suggestions.

Citation information

Kelleny, B., & Ishak, S. (2021). Exploring and visualizing spatial effects and patterns in ride-sourcing trip demand and characteristics. *Journal of Sustainable Development of Transport and Logistics*, 6(2), 6-24. doi:10.14254/jsdtl.2021.6-2.1

References

- Alemi, F. (2018). *What Makes Travelers Use Ridehailing? Exploring the Latent Constructs behind the Adoption and Frequency of Use of Ridehailing Services, and Their Impacts on the Use of Other Travel Modes*. (PhD Dissertation). University of California, Davis, Retrieved from <https://escholarship.org/uc/item/1cm0j0vm>
- Alemi, F., & Rodier, C. J. (2016). *Simulation of Ridesourcing Using Agent-Based Demand and Supply Regional Models: Potential Market Demand for First-Mile Transit Travel and Reduction in Vehicle Miles Traveled in the San Francisco Bay Area*.
- Anselin, L. (1988). *Spatial Econometrics: Methods and Models*.
- Anselin, L. (1995). Local Indicators of Spatial Association—LISA. *Geographical Analysis*, 27(2), 93-115. <https://doi.org/10.1111/j.1538-4632.1995.tb00338.x>
- Anselin, L. (1999). *Spatial econometrics*. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.439.3020&rep=rep1&type=pdf>
- Anselin, L. (2016). *Local Spatial Autocorrelation Clusters*. Retrieved from https://spatial.uchicago.edu/sites/spatial.uchicago.edu/files/6a_local_SA_r.pdf
- Azimi, G., Rahimi, A., Asgari, H., & Jin, X. (2020). Role of Attitudes in Transit and Auto Users' Mode Choice of Ridesourcing. *Transportation Research Record*, 2674(11), 1-16. <https://doi.org/10.1177/0361198120941841>
- Bailey, T., & Gatrell, A. (1995). *Interactive Spatial Data Analysis*.

- Basu, R., Araldo, A., Akkinapally, A. P., Nahmias Biran, B. H., Basak, K., Seshadri, R., . . . Ben-Akiva, M. (2018). Automated Mobility-on-Demand vs. Mass Transit: A Multi-Modal Activity-Driven Agent-Based Simulation Approach. *Transportation Research Record*, 2672(8), 608-618. <https://doi.org/10.1177/0361198118758630>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1), 289-300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Beron, K. J., & Vijverberg, W. P. M. (2004). *Probit in a Spatial Context: A Monte Carlo Analysis*. In L. Anselin, R. J. G. M. Florax, & S. J. Rey (Eds.), *Advances in Spatial Econometrics: Methodology, Tools and Applications* (pp. 169-195). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Boots, B. N., & Getis, A. (1988). *Point Pattern Analysis*. In G. I. Thrall (Ed.). Retrieved from <https://researchrepository.wvu.edu/cgi/viewcontent.cgi?article=1013&context=rri-web-book>
- Castiglione, J., Cooper, D., Sana, B., Tischler, D., Chang, T., Erhardt, G. D., . . . Mucci, A. (2018). *TNCs & Congestion* (10-2018). Retrieved from https://uknowledge.uky.edu/ce_reports/1
- City of Chicago. (2021). *Transportation Network Providers - Trips*. Retrieved from: <https://data.cityofchicago.org/Transportation/Transportation-Network-Providers-Trips/m6dm-c72p>
- Cordera, R., Sañudo, R., dell'Olio, L., & Ibeas, Á. (2018). Trip distribution model for regional railway services considering spatial effects between stations. *Transport Policy*, 67, 77-84. <https://doi.org/10.1016/j.tranpol.2018.01.016>
- Correa, D., Xie, K., & Ozbay, K. (2017). Exploring the Taxi and Uber Demands in New York City: An Empirical Analysis and Spatial Modeling. Paper presented at the *Transportation Research Board's 96th, Annual Meeting*, Washington, D.C.
- Diao, M., Kong, H., & Zhao, J. (2021). Impacts of transportation network companies on urban mobility. *Nature Sustainability*. <https://doi.org/10.1038/s41893-020-00678-z>
- Dias, F. F., Lavieri, P. S., Garikapati, V. M., Astroza, S., Pendyala, R. M., & Bhat, C. R. (2017). A behavioral choice model of the use of car-sharing and ride-sourcing services. *Transportation*, 44(6), 1307-1323. <https://doi.org/10.1007/s11116-017-9797-8>
- Erhardt, G. D., Mucci, R. A., Cooper, D., Sana, B., Chen, M., & Castiglione, J. (2021). Do transportation network companies increase or decrease transit ridership? Empirical evidence from San Francisco. *Transportation*. <https://doi.org/10.1007/s11116-021-10178-4>
- Erhardt, G. D., Roy, S., Cooper, D., Sana, B., Chen, M., & Castiglione, J. (2019). Do transportation network companies decrease or increase congestion?. *Science Advances*, 5(5), eaau2670. <https://doi.org/10.1126/sciadv.aau2670>
- Esri Inc. (2021a). *Cluster and Outlier Analysis (Anselin Local Moran's I) (Spatial Statistics)*. Retrieved from <https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/cluster-and-outlier-analysis-anselin-local-moran-s.htm>
- Esri Inc. (2021b). *Hot Spot Analysis (Getis-Ord Gi*) (Spatial Statistics)*. Retrieved from <https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/hot-spot-analysis.htm>
- Esri Inc. (2021c). *How Cluster and Outlier Analysis (Anselin Local Moran's I) works*. Retrieved from https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/h-how-cluster-and-outlier-analysis-anselin-local-m.htm#ESRI_SECTION1_C5005997CFA44DC7BCE05E19314EC613
- Faghih, S., Shah, A., Wang, Z., Safikhani, A., & Kamga, C. (2020). Taxi and Mobility: Modeling Taxi Demand Using ARMA and Linear Regression. *Procedia Computer Science*, 177, 186-195. <https://doi.org/10.1016/j.procs.2020.10.027>
- Florax, R. J. G. M., & Nijkamp, P. (2005). *Misspecification in Linear Spatial Regression Models*. In K. Kempf-Leonard (Ed.), *Encyclopedia of Social Measurement* (pp. 695-707). New York: Elsevier.
- Getis, A., & Ord, J. K. (2010). *The Analysis of Spatial Association by Use of Distance Statistics*. In L. Anselin & S. J. Rey (Eds.), *Perspectives on Spatial Data Analysis* (pp. 127-145). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Hall, J. D., Palsson, C., & Price, J. (2018). Is Uber a substitute or complement for public transit?. *Journal of Urban Economics*, 108, 36-50. <https://doi.org/10.1016/j.jue.2018.09.003>
- HENAO, A. (2017). *Impacts of ridesourcing –lyft and uber –on transportation including vmt, mode replacement, parking, and travel behavior*. (Doctor of Philosophy). University of Colorado, Colorado Retrieved from

- https://digital.auraria.edu/content/AA/00/00/60/55/00001/Henao_ucdenver_0765D_10823.pdf
- Henao, A., Marshall, W. E., & Janson, B. (2019). *Impacts of Ridesourcing on VMT, Parking Demand, Transportation Equity, and Travel Behavior* (MPC 19-379). Retrieved from <https://www.ugpti.org/resources/reports/downloads/mpc19-379.pdf>
- Heppenstall, A., & Crooks, A. (2019). Guest editorial for spatial agent-based models: current practices and future trends. *GeoInformatica*, 23(2), 163-167. <https://doi.org/10.1007/s10707-019-00349-y>
- Hermawan, K. (2018). *Transportation Network Companies' (TNC) Impacts and Potential on Airport Access*. (PhD dissertation). University of California, Irvine, Irvine. Retrieved from <https://escholarship.org/uc/item/01m726rr>
- Jin, S. T., Kong, H., Wu, R., & Sui, D. Z. (2018). Ridesourcing, the sharing economy, and the future of cities. *Cities*, 76, 96-104. <https://doi.org/10.1016/j.cities.2018.01.012>
- Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374.
- Kelleny, B., & Ishak, S. (2018, 25th-28th September 2018). *An Agent-Based Modeling Approach for Shared-Use Mobility -A Proof of Concept Study*. Paper presented at the MIPALCON, Stuttgart University.
- Liu, X., Sun, L., Sun, Q., & Gao, G. (2020). Spatial Variation of Taxi Demand Using GPS Trajectories and POI Data. *Journal of Advanced Transportation*, 2020, 7621576. <https://doi.org/10.1155/2020/7621576>
- Lopes, S. B., Brondino, N. C. M., & Silva, A. N. R. d. (2014). GIS-Based Analytical Tools for Transport Planning: Spatial Regression Models for Transportation Demand Forecast. *ISPRS Int. J. Geo Inf.*, 3, 565-583.
- Moran, P. A. P. (1950). Notes on Continuous Stochastic Phenomena. *Biometrika*, 37(1/2), 17-23. <https://doi.org/10.2307/2332142>
- Oviedo, D., Granada, I., & Perez-Jaramillo, D. (2020). Ridesourcing and Travel Demand: Potential Effects of Transportation Network Companies in Bogotá. *Sustainability*, 12(5). <https://doi.org/10.3390/su12051732>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . . Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.*, 12, 2825-2830.
- Rayle, L., Shaheen, S., & Chan, N. (2014). *App-Based, On-Demand Ride Services: Comparing Taxi and Ridesourcing Trips and User Characteristics in San Francisco*.
- Reck, D. J., & Axhausen, K. W. (2020). Subsidized ridesourcing for the first/last mile: how valuable for whom?. *European Journal of Transport and Infrastructure Research*, 20(4), 59-77. <https://doi.org/10.18757/ejtir.2020.20.4.5314>
- Roukouni, A., & Homem de Almeida Correia, G. (2020). Evaluation Methods for the Impacts of Shared Mobility: Classification and Critical Review. *Sustainability*, 12(24). <https://doi.org/10.3390/su122410504>
- Sánchez-Díaz, I., Holguín-Veras, J., & Wang, X. (2016). An exploratory analysis of spatial effects on freight trip attraction. *Transportation*, 43(1), 177-196. <https://doi.org/10.1007/s11116-014-9570-1>
- Sapra, S. K. (2010). Robust vs. classical principal component analysis in the presence of outliers. *Applied Economics Letters*, 17(6), 519-523. <https://doi.org/10.1080/13504850802046989>
- Serneels, S., & Verdonck, T. (2008). Principal component analysis for data containing outliers and missing elements. *Computational Statistics & Data Analysis*, 52(3), 1712-1727. <https://doi.org/10.1016/j.csda.2007.05.024>
- Shaheen, S., Chan, N., Bansal, A., & Cohen, A. (2015). *Shared Mobility: A Sustainability & Technologies Workshop: Definitions, Industry Developments, and Early Understanding*.
- Soria, J., & Stathopoulos, A. (2021). *Investigating Socio-spatial Differences between Solo Ridehailing and Pooled Rides in Diverse Communities*.
- Sperling, J., Young, S., Garikapati, V., Duvall, A., & Beck, J. M. (2019). *Mobility Data and Models Informing Smart Cities* (NREL/TP-5400-70734). Retrieved from Denver <https://www.nrel.gov/docs/fy20osti/70734.pdf>
- Tu, W., Cao, R., Yue, Y., Zhou, B., Li, Q., & Li, Q. (2018). Spatial variations in urban public ridership derived from GPS trajectories and smart card data. *Journal of Transport Geography*, 69, 45-57. <https://doi.org/10.1016/j.jtrangeo.2018.04.013>

- Uber. (2017). *Uber Hits 5 Billion Rides Milestone*. Retrieved from <https://www.uber.com/en-SG/blog/uber-hits-5-billion-rides-milestone/>
- Wang, C., Quddus, M. A., Ryley, T., Enoch, M., & Davison, L. (2012). *Spatial models in transport: a review and assessment of methodological issues*. Paper presented at the Transportation Research Board 91st Annual Meeting, Washington DC.
- Winter, K., Cats, O., Martens, K., & van Arem, B. (2020). *Relocating shared automated vehicles under parking constraints: assessing the impact of different strategies for on-street parking*. *Transportation*. <https://doi.org/10.1007/s11116-020-10116-w>
- Wood, J., Dykes, J., & Slingsby, A. (2010). Visualisation of Origins, Destinations and Flows with OD Maps. *The Cartographic Journal*, 47, 117-129.
- Yu, H., & Peng, Z.-R. (2019). Exploring the spatial variation of ridesourcing demand and its relationship to built environment and socioeconomic factors with the geographically weighted Poisson regression. *Journal of Transport Geography*, 75, 147-163. <https://doi.org/10.1016/j.jtrangeo.2019.01.004>

- Immediate, universal access to your article on publication
- High visibility and discoverability via the JSDTL website
- Rapid publication
- Guaranteed legacy preservation of your article
- Discounts and waivers for authors in developing regions

