

Extracting commuter-specific destination hotspots from trip destination data – comparing the boro taxi service with Citi Bike in NYC

Andreas Keler, Jukka M. Krisp & Linfang Ding

To cite this article: Andreas Keler, Jukka M. Krisp & Linfang Ding (2019): Extracting commuter-specific destination hotspots from trip destination data – comparing the boro taxi service with Citi Bike in NYC, *Geo-spatial Information Science*, DOI: [10.1080/10095020.2019.1621008](https://doi.org/10.1080/10095020.2019.1621008)

To link to this article: <https://doi.org/10.1080/10095020.2019.1621008>



© 2019 Wuhan University. Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 10 Jun 2019.



Submit your article to this journal [↗](#)



Article views: 713



View related articles [↗](#)



View Crossmark data [↗](#)

Extracting commuter-specific destination hotspots from trip destination data – comparing the boro taxi service with Citi Bike in NYC

Andreas Keler^a, Jukka M. Krisp^b and Linfang Ding^{c,d}

^aDepartment of Civil, Geo and Environmental Engineering, Technical University of Munich, Munich, Germany; ^bApplied Geoinformatics, Institute of Geography, University of Augsburg (UniA), Augsburg, Germany; ^cThe KRDB Research Centre for Knowledge and Data, Free University of Bozen-Bolzano, Bozen-Bolzano, Italy; ^dDepartment of Cartography, Technical University of Munich, Munich, Germany

ABSTRACT

Taxi trajectories from urban environments allow inferring various information about the transport service qualities and commuter dynamics. It is possible to associate starting and end points of taxi trips with requirements of individual groups of people and even social inequalities. Previous research shows that due to service restrictions, boro taxis have typical customer destination locations on selected Saturdays: many drop-off clusters appear near the restricted zone, where it is not allowed to pick up customers and only few drop-off clusters appear at complicated crossing. Detected crossings imply recent infrastructural modifications. We want to follow up on these results and add one additional group of commuters: Citi Bike users. For selected Saturdays in June 2015, we want to compare the destinations of boro taxi and Citi Bike users. This is challenging due to manifold differences between active mobility and motorized road users, and, due to the fact that station-based bike sharing services are restricted to stations. Start and end points of trips, as well as the volumes in between rely on specific numbers of bike sharing stations. Therefore, we introduce a novel spatiotemporal assigning procedure for areas of influence around static bike sharing stations for extending available computational methods.

ARTICLE HISTORY

Received 15 November 2018
Accepted 28 April 2019

KEYWORDS

Urban transportation; spatial analysis; mobility patterns; boro taxis; vehicle fleets; bicycle-sharing service

1. Introduction

Human mobility in urban environments is complex and dynamically changing. One possibility for gaining more insights on urban human mobility is analyzing data from tracked entities, namely daily urban traffic participants. Vehicle movement trajectories of urban vehicle fleets can help predicting periodical travel time variations (Keler and Krisp 2016a) or classifying traffic congestion events by intensity (Keler, Ding, and Krisp 2016). Due to the massive size of data generated by vehicle or bicycle fleets, often only extracts are used in many analyses. In case of tracked taxis, these extracts might consist of the spatiotemporal positions, where events occur: a customer leaves or enters the taxi. These positions can reveal numerous useful information about operational effectiveness of the fleet (Zhang and He 2012; Zhang, Peng, and Sun 2014), driving behavior (Li et al. 2011), or the location-dependent service demand.

Our idea is to define hotspots of trip destination points of tracked taxis of a taxi fleet and tracked bicycles of a bicycle-sharing service. We apply our techniques for trips on selected Saturdays in June 2015 in NYC. By inspecting the spatiotemporal distribution of generated destination hotspots, we can gain various insights on the two different travel modes and their specific groups of users (or even

commuters). One important part of the latter is the connection of every generated destination hotspot to the origins or starting points of individual trips. The contribution of this paper consists of introducing a novel spatiotemporal assigning procedure for areas of influence around static bike sharing stations. Depending on the number of users, we aim to represent the space influenced by this specific mobility service by a new form of bike sharing station activity representation and visualization. For reasons of evaluation, we intersect those varying areas with boro taxi hotspots based on Keler (2018) and use these intersection areas for reasoning on possible relationships of the two modes. This enables associating commuter destinations of different mobility services.

2. Mobility analyses in urban environments

Mobility in urban environments has specific properties that are different from minor cities and rural areas (Miller, Wu, and Hung 1999). There are different modes of transport, private and public, that cause various different movement patterns of many individual traffic participants. Especially the public transport in urban environments shows a variety of possibilities including taxi services, trains, tramways, subways, buses, and bike sharing services. Besides

predefined lines or routes of train, subway and bus services, there are the services as taxi and bike sharing that may represent a more precise (spatially) destination of a certain passenger. Extracted trajectories of taxis can benefit mobility analyses by including information on the precise start and destination points of a taxi trip and additionally the taxi driver strategies for collecting customers. Analyzing taxi trajectory data has a relatively long tradition (since the early 2000s) and finds usage in many different research domains. Far less tradition have trajectory and OD pair analyses of active mobility modes (Grigoropoulos et al. 2019). The latter not only imply bicycling, but also using scooters, segways, e-bikes, pedelecs, and, of course walking. To a different extent, we can deduct the general flows on a higher level by using OD pair data from specific, often city-restricted, bike sharing services. Previous work on analyzing station-based bike sharing data, especially the CitiBike BSS in NYC, focusses on proving the importance of spatial and temporal effects on the service itself (Faghih-Imani and Eluru 2016) or differentiate between CitiBike stations and hubs (Gordon-Koven and Levenson 2014). The basics of vehicle trajectory and OD pair analyses and the differences to the ones focussing on bicycle-generated data are outlined in the following subsections.

2.1 Computing with taxi trajectories

Data coming from taxi fleets in urban environments finds usage in numerous industrial and university projects as Dmotion, CrowdAtlas or T-Drive. The purpose of these analyses is often finding geographical contexts for the patterns of different dynamics (Castro et al. 2013).

Besides this, computer science has numerous examples, where taxi trajectories serve as test data for evaluating the computational efficiency of different machine learning techniques, as extended techniques of clustering, regression, and outlier detection. Apart from knowing spatiotemporal clusters of taxi trip destinations, it is possible to infer traffic situations based on the travel patterns of selected vehicles. This is possible by segmenting the data into episodes in the way of a time series or by applying extended clustering techniques for spatiotemporal data as the spatiotemporal DBSCAN (Birant and Kut 2007).

Due to the often-immense size of daily data extracts resulting from thousands of tracked taxi drivers, often only abstracted movement information is available for analysis. This abstracted movement information is in many cases an extraction of movement points with specific attributes. It is possible to extract only those movement positions, which have an instantaneous velocity value of zero (Liu and Ban 2013). These points are useful for distinguishing

between free-flowing traffic and traffic congestion, parking vehicles or vehicles influenced by travel delays.

Keler, Krisp, and Ding (2017a) present another possibility of point extraction, where taxi trajectory intersection points are extracted. The resulting points are mainly situated at road intersections and reveal travel time variations and can even indicated the type of transportation infrastructure, mainly in the way how many elevation level appear at selected road intersections.

The possibly most frequently used type of taxi movement data extracts are the points where customer pick-ups and drop-offs occur. When tracking the whole taxi fleet, it is possible to infer functional zones of customer popularities or areas of interest. Taxi customer pick-up and drop-off points that form reasonably shaped clusters for selected time windows. There are different pick-up and drop-off hotspot clusters for the same time windows that are spatially distinguishable. Additionally, Yue et al. (2009) inspect the number of extracted points per hour in a temporal variation diagram for seven days. When comparing the number of pick-up and drop-off points for hour of the day, Weng et al. (2009) propose a loaded time rate diagram, which shows the hourly proportion of vacant taxis to the ones loaded with a customer. This allows estimating the peaks of the taxi service, where most taxi customers are available in the investigation area. Apart from density-based points clustering, Krisp et al. (2012) use k-means for hourly time windows. The number of clusters k results from visual inspection of the point distributions for all time windows. Visual inspection of the results appears in a space-time cube with hourly units on the time axis z .

Overall, we can say that origins and destinations of taxi trips deliver useful information on the taxi service and might indicate the rush hours of an investigation area. There are multiple possibilities to detect and represent the origin and destination hotspots. Especially, representing origins and destinations of vehicle trips in an understandable way is challenging in many ways. This might be, especially in comparison to station-based bicycles, a dynamic component that may changeover time, as general popularity of trip destination locations can change. Bike sharing stations are implemented, visible and non-displaceable, which means that the service is based on the knowledge of spatial locations of these stations.

2.2 Analyzing origins and destinations of vehicle and bicycle trips

Numerous approaches focus on the analysis of taxi trip origins and destinations. These time-stamped

positions might represent the interest of a large group of taxi users, together with daily commuters.

For example, Jahnke et al. (2017) proposed a geo-visual analytics application, where taxi traveler hotspots are detectable. These hotspots base on the temporal variations of taxi trip origins and destinations, together with a spatiotemporal association with points of interest (POIs). The aim of such an application is studying the taxi traveler's activity. In a similar way, Ding et al. (2016a, 2016b) focused on transportation hubs as airports for detecting further details of the taxi service via interactive pie charts, which represent taxi origin and destination points for each weekday with different colors together with size variations based on the varying numbers. The most reasonable form of this representation is an origin-destination matrix for selected days of the week, which is a form of an adjacency matrix with nodes as origins and destinations and weight values assigned to the connecting arcs.

On a different level of public transport, bike-sharing services gain increasing popularity since the early 2000s. Bicycles are frequently used in urban environments, especially in eastern and southeastern Asian cities. Besides private bicycle owners, there are bicycle fleets coming from different companies introduced as bike-sharing services. Similar to car sharing services, there is the option of having free-floaters or station-based services. In case of station-based services as in most of the cities in USA, origins and destinations are already known. Bike sharing services mostly rely on a reasonable distribution of bike sharing stations and trips occur between specific stations. The distribution of incoming and outgoing bicycle users can vary with time of the day and weekday and show typical commuter movements as well as detectable movement of tourists. The available data of tracked cyclists imply besides user anonymization via user identification, very detailed information of each trip. There are accurate origin and destination coordinates, together with trip lengths, type of user (subscriber or guest), or even age of the user.

2.3 Properties of urban transportation infrastructures and its influence on mobility patterns

There are numerous approaches of distinguishing between modes of urban traffic participants based on the change of the underlying movement parameters (Shafique and Hato 2016). The data input consists of tracks generated by smartphone users. One important topic towards the establishment of smart cities and intelligent transportation systems (ITS) are multimodal routing applications. The conception of such applications is motivated by assigning the spatial positions of transport mode changes. These positions are switching points and often the

important stations of subways, buses, and car-sharing services.

Spatiotemporal pattern consists in many cases of movement descriptions of individual moving objects. Additionally, there is the possibility of describing movement of temporally changing surfaces, such as surfaces of air quality. The latter might be associated with road networks and introduced as weights allowing to route on a network based on the best air quality (Karraï, Keler, and Timpf 2014).

Additionally, there were attempts to connect static geographic data with movement data of moving entities in the way of creating semantic trajectories (Yan 2009), intersecting complicated crossings with traffic congestion for inferring traffic bottlenecks (Keler, Krisp, and Ding 2017b). Other approaches, which combined static and dynamic geodata, consist of map matching movement positions onto road segments (Zhao et al. 2012) or inferring the road network based on vehicle trajectories (Ahmed et al. 2015). Besides polyline representations of road segments, it is possible to extract their nodes or street intersection points. Local knowledge of specific mobility services can enrich specific road representations in the way of defining typical operational patterns on parts of the road. The usage of boro taxi destination points, for example, for relating with complicated crossings of a road network can benefit the understanding of specific functions of the urban road network (Keler and Krisp 2016b). In particular, boro taxi operating areas with a restriction in customer pick-ups, as in the southern part of Manhattan, have more densities at the border to the restriction zone. Patterns of active mobility users are far more complex, since infrastructures of all available modes are possible. Amini, Twaddle, and Leonhardt (2016) showed this with an example of left turning of bicyclists at a signalized intersection in Munich: there are three main trajectory clusters that are characterized as (1) the expected bicycle left turn using bicycle infrastructure, (2) the pedestrian left-turn using (often illegally) pathways of pedestrians, and, (3) the vehicle turn as a partition of bicyclist participate within car traffic on the road segments assigned for vehicles. Resulting from this, we can say that exact routes between bike sharing stations have much more variations than between vehicle ODs, and, that the specific investigation area contributes to the number of accessible variations of routes of active mobility users. Despite this, the possibility of mode changes within transits in urban environments are much more definite for the case of station-based bike sharing services, since hotspots of trip origins and destinations have fixed locations of physically built stations. Keler (2018) shows that this is more complex for

the transit between and from taxi services and the subsequent association with other mobility services or modes of transport.

3. Data sets of the case study - borotaxi and Citi Bike in NYC

Our case study implies the abstraction of spatiotemporal destination points of two different urban public transport services into daily destination hotspot polygons: boro taxi and CitiBike. These abstractions are the destination hotspots of both services.

For the boro taxi service, we cluster the destination points of all tracked boro taxis for the whole day with specific parameter values resulting from previous spatial and visual data analysis. The CitiBike hotspots emerge from all available bike-sharing stations of the inspected time period. The spatial extents of Citi Bike destination hotspots result from a formula of calculating the radii of radial polygons around the stations based on the absolute number of incoming cyclist for whole days.

3.1 Boro taxi trajectories

Boro taxis operate since 2013, after inspecting the results of yellow taxi GPS data¹ analyses: 95% of all taxi customer pick-ups occurred within Manhattan and the remaining 5% in the outer boroughs. Having a cheaper license, boro taxi drivers are not allowed to pick up any customers at the two airports or in Manhattan (below East 96th and West 110th Streets).

The boro taxi data sets come from the NYC Taxi & Limousine Commission (TLC)². Every record consists of 20 attributes and represents one trip. It includes taxi trip records from all boro taxi trips and has very detailed information on travel times and routes. We select boro taxi data from the four Saturdays in June 2015. Keler and Krisp (2016b) show that it is possible to represent typical destinations of boro taxi users by density-based clusters. These users come in large part from the outer boroughs of NYC, outside of Manhattan. Most of the trips appear in the evening, one indicator that social events attract people to come to Manhattan. From the data partitions, we extract start and destination points and focus on the latter for inferring customer drop-off hotspots. Additionally, there is a possibility of associating boro taxi drop-off hotspots with road segments (as for example from the OpenStreetMap project) or, based on the previous, complicated crossings (Keler and Krisp 2016b).

3.2 Citi Bike trip data

The second data set of moving entities comes from a bicycle-sharing service in NYC: Citi Bike, introduced in 2013 and sponsored by Citigroup³. The service has fixed docking stations with recently

(2017) 603 location in the whole city. Therefore, the location of cluster centroids is already given. The important information here are spatiotemporal densities and their variations.

We extract only the Saturdays from a total number of 941,219 bike trips for June 2015. In total, there are around 320 visited stations in this period. The data sets have 11 attributes and each record represents one Citi Bike trip. Besides trip duration, there is no information on the trip length. Additional information comes from attributes including gender, year of birth and user type. The last mentioned allows distinguishing between subscribers, who are annual members, and customers, who have only a 24-h pass.

4. Methods for destination hotspot generation

We use for our approach the method by Keler and Krisp (2016b) for extracting boro taxi drop-off hotspots. The two main components of this technique are applying OPTICS (Ankerst et al. 1999) for the density cluster generation, and subsequently using the gift wrapping algorithm (Jarvis 1973) for convex hull generations. The selection of useful input parameters bases on previous inspection of drop-off point reachability and on appearances of the local transportation infrastructure. One example for the latter is the selection of search distance Epsilon based on the maximum street width in Time Square of 102 feet. A diagram in Figure 1 shows this technique (apple green box), together with the hotspot generation for Citi Bike destinations (blue box). The latter is considered the main contribution of this work, since the difficulty of defining hotspots of active mobility users relies on associating urban space that implies not only vehicle roads, but as well, pathways, bike lanes, green areas, and buildings.

Besides the two mentioned approaches for destination hotspot polygon generation, there is a third approach, which is pictured by the white box in Figure 1. It is the base for further reasoning, mainly by intersecting the hotspot polygon products. Afterwards, it is possible to intersect the resulting dual hotspots with polygons representing complicated crossing. These polygons are coming from OSM road network extracts by applying the technique of Krisp and Keler (2015).

4.1 Boro taxi drop-off hotspot generation

The apple green box in Figure 1 shows the definition of boro taxi drop-off hotspots with two calculation steps: OPTICS (Ankerst et al. 1999) and gift wrapping algorithm (Jarvis 1973). Ordering Points To Identify the Clustering Structure (OPTICS) is a density-based clustering algorithm, which has the two parameter MinPts and Epsilon for computing an unknown

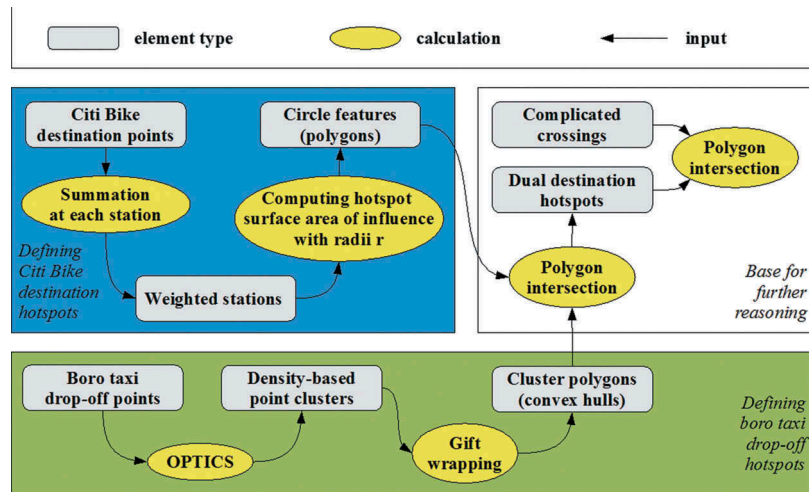


Figure 1. Workflow of the two methods for generating trip destination hotspots of the services boro taxi and city bike.

number of density-based clusters for daily data partitions. The minimum number of points $MinPts$ is selected as 2, and the search distance $Epsilon$ bases on the maximum street width at Time Square with the value of 31.0896 m or 102 feet. We convert every computed OPTICS cluster, except the noise cluster, of every Saturday into convex hulls by using the gift wrapping algorithm by Jarvis (1973). After applying both algorithms with the mentioned parameter values on the four boro taxi data partitions of Saturdays in June 2015, we receive varying numbers of drop-off hotspots as listed in Table 1.

The number of records in Table 1 is equal to the number of boro taxi customer drop-off points. By visual inspection of the spatial distribution of the resulting boro taxi drop-off hotspots, certain areas of NYC are showing hotspots with spatially large extents. These large hotspots are located near the border of the yellow taxi-operating zone, where it is restricted for boro taxi drivers to pick-up customers, as in Figure 2 for the 27th of June 2015.

Table 1. Total number of boro taxi data records for Saturdays in June 2015 together with the resulting total numbers of drop-off hotspots.

Inspected Saturday in NYC	06 June	13 June	20 June	27 June
Number of records	72,824	69,539	67,045	74,861
Number of drop-off hotspots	5006	4919	4769	5233

Figure 2 pictures these patterns in northern Manhattan and near the Williamsburg Bridge for the 27th of June. These spatially extended hotspots are detectable for all inspected data partitions of selected Saturdays. Findings in Keler and Krisp (2016b) show that besides this typical boro taxi behavior of leaving customers at the restricted zone, there is a fast leaving the zone behavior in the southern part of Manhattan. These patterns are more difficult to obtain via visual inspection, since there are small hotspots in southern Manhattan that imply higher drop-off point densities.

In general, the difficulty of defining taxi destination hotspots, especially in an urban environment,



Figure 2. Map view of boro taxi destination hotspots for the 27th of June 2015.

originates from defining suitable parameters for spatiotemporal clustering.

4.2 Citi Bike destination hotspot generation

Citi Bike data has origin and destination points of 320 specific stations in NYC. This allows adapting a suitable scale for analysis by respecting the spatial extent and distribution of the Citi Bike stations. The taxi drop-off hotspot generation and Citi Bike destination hotspot generation are generally different. This difference results from (a) the different traffic interaction patterns of active mobility users in comparison to vehicle drivers, (b) the more often used sidewalks by bicyclists, and, (c) the higher proportion of restricted areas accessible by vehicles than by bicycles.

Our idea is to introduce circular polygons with size variations dependent on the number of incoming bicyclists per day. Following the idea of having whether (1) different infrastructural elements that are being used (which is not always the case in NYC), (2) same infrastructural elements are being used differently (as riding along lane markings or usage of pedestrian paths), and, (3) bicycle parking possibilities are less strict observed (and allow the bicyclist to park in various spaces). Therefore, we do not rely on GNSS observations (as the case of boro taxi data) that show locations on highway segments, but, especially in connection with the 3rd point, we imply that any given transport infrastructure of all travel modes might serve as hotspot areas for bicycle usage. The technique for Citi Bike destination hotspots implies the definition of a surface area of influence for each hotspot. After several visual inspections of the incoming numbers of cyclists, one principle is designed of how to compute these surface areas of influence for each bicycle station. The number of Citi Bike trip destinations at a station is then eightfold the number of m^2 for each hotspot. Citi Bike hotspot radii r (in meters) are calculated as the following:

$$r = 8 \times \sqrt{\frac{n_{end} \times 1(m^2)}{\pi}}$$

After applying the formula on all 320 stations in NYC for the 6th of June, great variations are detectable between the areas of influence or destination hotspots of the individual stations, which is pictured in Figure 3. Every Citi Bike destination hotspot mostly influences at least one road intersection and sometimes intersects or nearly intersects neighboring destination hotspots. It is to mention that the larger hotspots also influence more than one road intersection and parts of different road segments.

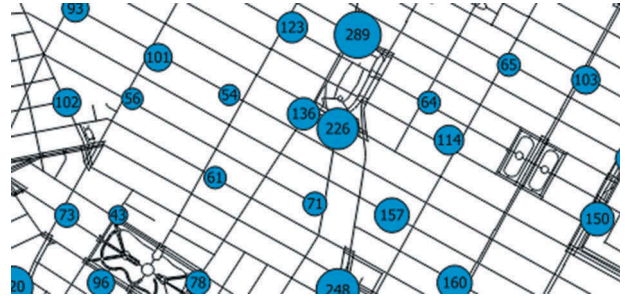


Figure 3. Distribution of Citi Bike destination hotspots for the 6th of June 2015 in central Manhattan, NYC.

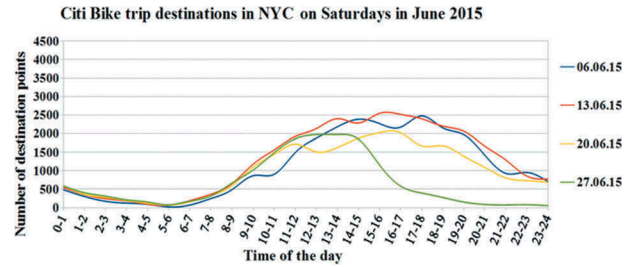


Figure 4. Hourly numbers of Citi Bike destination points in NYC for all Saturdays in June 2015.

5. Results and discussion

After applying the two techniques for hotspot inference on selected Saturdays in NYC, we are able to detect various spatiotemporal patterns. The first data inspection focuses on the hourly distribution of destination point numbers in both data sets. Figure 4 shows the hourly distribution of Citi Bike destination points. There are many similarities in the numbers of all four data partitions until around 10 AM. After 10 AM more variations appear between the hourly partitions. Especially the curve of the 27th appears different from the others. Comparing to the Citi Bike trip distributions on working days there are slightly less records on weekends. The highest numbers on Saturdays are usually in the late afternoon.

Figure 5 shows the four curves for the hourly distribution of boro taxi drop-off points. There are many similarities of the hourly numbers in the first half of the day. In general, we can say that there are more boro taxi drop-offs on weekends as on working days, especially in southern Manhattan (yellow zone). Additionally, Figure 5 shows more variations in the number of drop-offs in the late afternoon and evening. This appearance might correlate with the number of possible social events worthwhile to attract people from outer boroughs.

Both curves of data partitions from the 27th of June 2015 are outliers, since they differ in shape from the curves of the three previous Saturdays. One idea to gain more information from the two outliers is to relate the two curves in one diagram,

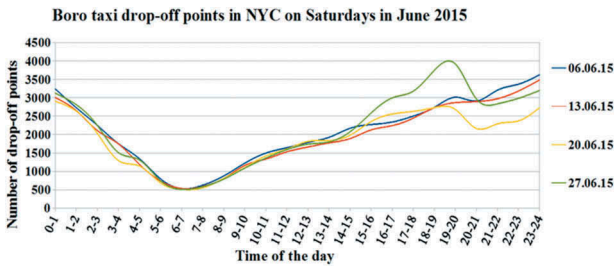


Figure 5. Hourly numbers of boro taxi drop-off points in NYC for all Saturdays in June 2015.

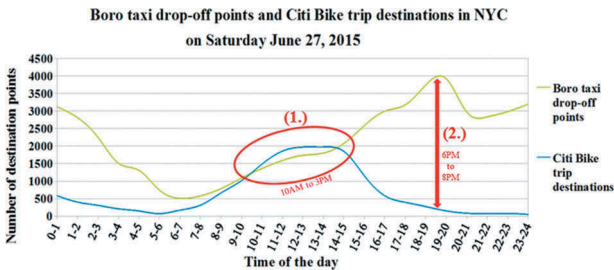


Figure 6. Hourly numbers of boro taxi drop-off points (apple green) and Citi Bike destinations (blue) in NYC for the 27th of June 2015.

which is pictured in Figure 6. The intention of relating selected distribution curves is finding eventually specific mobility patterns.

The two curves in Figure 6 show only similarities between 10 AM and 3 PM, highlighted via the red circle. More detailed in appearance (1), the blue Citi Bike destination points curve exceeds slightly the apple green curve of boro taxi destinations. Appearance (1) is a similarity, which might indicate the typical NYC rush hours on Saturdays.

After 3 PM, Citi Bike destinations are decreasing and boro taxi destinations are growing until reaching a peak at around 7 PM. This results in a second appearance (2), where a great difference in destination point numbers appears from 6 PM to 8 PM as pictured via the red arrow in Figure 6. One possible explanation of this appearance (2) might be a connection to weather events, which make the usage of taxis more attractive for usual customers.

By including the general weather information for the whole city on the 27th of June 2015, raining events appeared from around 3 PM until the early evening. This might show the dependency of bicycle usage attractiveness on weather changes.

In the next step, we inspect the spatial distribution of both types of hotspots for June 27. Figure 7 shows, for example, the border of the yellow zone, which is restricted for operating boro taxi drivers, situated along East 96th and West 110th Streets. Additionally, there is the boundary of Citi Bike availability in central Manhattan, starting from the southern part

of the Central Park. There are only few intersecting hotspots of both services.

In the southern part of Manhattan, the matching rate is even less. One reason for this appearance is the fact that it is economically unfavorable for boro taxi drivers to drop off a customer in southern Manhattan, since it is restricted for pick-ups. Another reason is the high density of Citi Bike stations in this restriction zone. In the following steps, we focus partially on the size variations of destination hotspots and their relation to each other. We intersect the polygon outcomes of both methods and spatially intersect Citi Bike and boro taxi hotspots.

From 12,752 Citi Bike trips, 4167 (32.6% in ratio) end at boro taxi hotspots, which include 109 out of 320 Citi Bike stations, as pictured in Figure 8(a).

Figure 8(a) shows the 4167 Citi Bike trips that end in boro taxi hotspots. Green dots show the destinations of Citi Bike trips and yellow their origin. By visual inspection, it is possible to see selected destination hotspots, especially in the eastern part of Manhattan and in Brooklyn. Most of the Citi Bike trips that end at boro taxi hotspots start in southern Manhattan. The red connection lines in Figure 8 show that many bike trips occur between Manhattan and Brooklyn, since the water areas are not visible. We provide this effect via setting 50% opacity to the red colored connection lines.

Figure 8(b) shows the boro taxi trips that end within Citi Bike hotspots. From 51,694 boro taxi trips, only 343 (0.7% in ratio) end at City Bike hotspots. Since it is not allowed to pick-up customers in the restricted zone, there are no origins in southern Manhattan. Instead, most of the 343 trips end in southern Manhattan. The low matching rate is visualized in a more detailed view in Figure 9(a), where the relative small matched Citi Bike hotspots are pictured via green circles and the matching boro taxi trip destinations in yellow.

From 5233 boro taxi polygons, only 103 (2% in ratio) intersect with Citi Bike destination hotspots, which are pictured in Figure 9(b). By visual inspection of Figure 9(b), the spatially larger boro taxi hotspots with Citi Bike destinations are in Brooklyn. These 2% have higher drop-off (boro taxi) and destination (Citi Bike) point densities.

These observations might indicate patterns of possible transit behavior between boro taxis and Citi Bike bicycles. This spatiotemporal assigning procedure for areas of influence around static bike sharing stations is one attempt for comparing two different mobility services (since different travel modes) operating at the same area within the same time windows. The outcomes of applying different types of polygon intersecting and assigning can deliver further insights on local knowledge, such as complex transit behavior at specific locations.



Figure 7. Distribution of destination hotspots for the 27th of June 2015 in central Manhattan with apple-green boro taxi drop-off hotspots and blue Citi bike destination hotspots (labeled with number of absolute trip destinations).

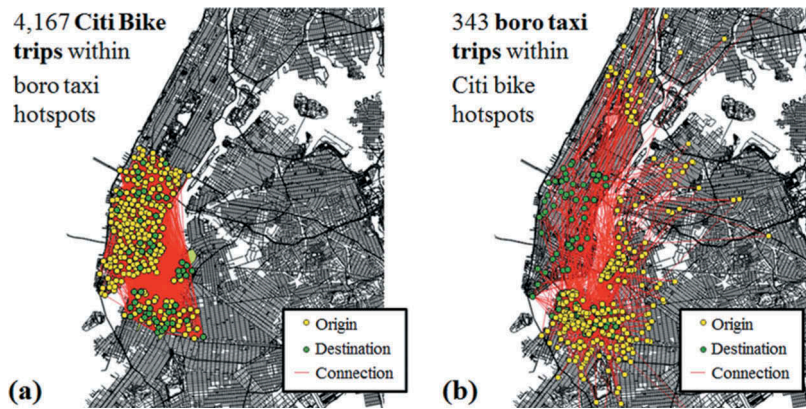


Figure 8. Connections of origins and destinations of (a) Citi bike trips ending in boro taxi hotspots and (b) boro taxi trips ending in Citi bike hotspots, for the 27th of June 2015 in NYC.

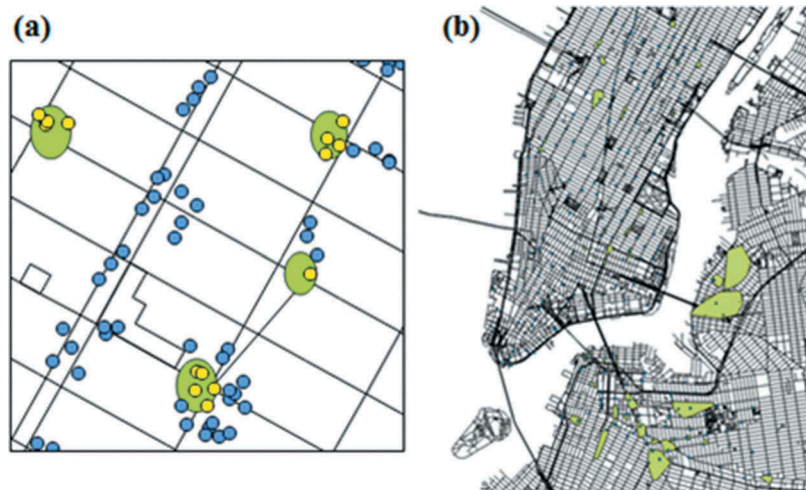


Figure 9. Descriptive map views on (a) boro taxi trips that end at Citi Bike hotspots, and (b) Citi bike trips that end at boro taxi hotspots.

6. Conclusions

The presented technique allows extracting frequently visited locations of two different mobility services, which are then connectable with events.

Furthermore, relating tracked movement of vehicles and bicycles from two different services allows formulating new insights based on intersecting areas of influence. Focusing on data extracts of Saturdays

might benefit, for selected investigation areas, the understanding of how social events influence movement patterns and how to relate them with interests of defined or known customers or customer groups. Especially those attributes of Citi Bike records that reveal further information on user groups, as their commuting behavior, and, as their average age, can support further insights on conducted analyses.

Nevertheless, it is challenging to define travel-mode- and commuter-specific destination hotspots. One problem here is finding a proper representation of the results, because information generalization might hide important details such as specific trip outliers. Commuters are detectable in the Citi Bike data, when comparing origins and destinations of trips on working days between certain stations via many available approaches. In general, some of these periodical movements are also detectable on Saturdays, which may result as the continuation of the working days. Other types of Citi Bike users are clearly tourists, which are indicated in the data as not registered users (user attribute value “customer” instead of “subscriber”), who usually have a very long time span of service usage. This results in not accessible information of the trip durations and spatial configurations of the trajectories.

The further findings of this work include also defining visualization techniques for representing trip destination hotspots, besides the computational procedure for station-based bike sharing destination hotspots.

When working with the intersection product of the two data sets, the technique allows detecting outliers or very specific details of trips, especially by inspecting the time components. This is a critical issue, since only few boro taxi and Citi Bike trips relate to each other in our case study. Consequently, it is important to find a suitable scale for performing the spatial analysis. One attempt might be inspecting suitable parameter values for the density-based clustering of boro taxi destinations together with appraising the radii for the Citi Bike hotspots. The method of the Citi Bike destination hotspot generation delivers questionable results, since the number of incoming cyclists at a station is generally not dependent on the surface area of influence at each station.

Figure 3 shows intersecting Citi Bike destination hotspots, which appear too large for analyses, whereas the matching results pictured in Figure 9 show that they possibly appear as too small. This concludes that the hotspot radii need a redefinition.

7. Future work

The presented approach is extendable in many ways. One direction of future work that would greatly benefit the quality value of the outcomes is an evaluation with local knowledge. Local

knowledge is extractable via social media posts on the service itself, which requires reasoning of possible indications of local knowledge. This is difficult to automatize and would require manual decisions. Another, possibly more reliable, method to extract local knowledge is via filling in questionnaires of Citi Bike users or boro taxi drivers. The knowledge of conditions typical for traveling in NYC would benefit the understanding of popular destinations for commuters, tourists, and locals on weekends.

Another direction of future work would be applying the same approach on another investigation area with comparable public transport services. It would be challenging to find similar or specific patterns that are investigation-area- and transport-service-dependent. Further steps that arise from proposed technique include the inclusion of points of interest (POIs), and, defining POIs-popularity measures. Additionally, routing algorithms may help evaluating travel delays of selected hours between specified locations. Based on the travel times and the lengths of the routes, we can estimate travel delays and eventually associate them with rush hours and different states of traffic. When thinking about the purpose of the resulting destination hotspots, it would be interesting to reason about the benefit of predicting destination hotspots of taxis and cyclists (Guo and Karimi 2017).

Another idea is connecting the influence of a destination hotspot with the underlying transportation infrastructure for including road network complexities. By including OpenStreetMap (OSM) road segments of NYC into our analyses it is possible to apply the method by Krisp and Keler (2015) for detecting complicated crossings. Additional information that may benefit further findings would be the locations and spatial extents of pedestrian zones. Based on previous research (Krisp and Keler 2015; Keler and Krisp 2016b), we want to estimate the complexity of road intersections and include the results into destination hotspot analyses as pictured in Figure 10. The two output polygons of this work are inspected together with the inferred complicated crossings by the approach of Krisp and Keler (2015) pictured in violet. The remaining problem in Figure 10 is finding a suitable scale for spatial analysis. This might facilitate the connection between destination hotspots and transportation infrastructure.

In case the destination hotspots are not intersecting inferred complicated crossings, we can include enriched information on road intersection complexity by using the approach of Sladewski, Keler, and Divanis (2017). All these approaches might be evaluated in a VR environment using a bicycle simulator (Keler et al. 2018). This means evaluation occurs on

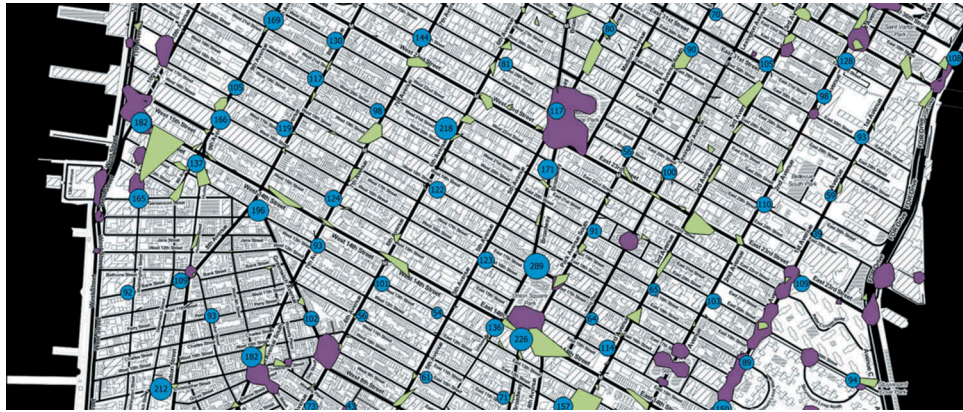


Figure 10. Distribution of destination hotspots for the 6th of June 2015 in central Manhattan with apple-green boro taxi drop-off hotspots and blue Citi bike destination hotspots (labeled with number of absolute trip destinations), together with violet polygons representing complicated crossings.

a microscopic level and might include simulated traffic of three groups of traffic participants: pedestrians, bicyclists and vehicle drivers.

After modeling transport infrastructure, we can include estimated or measured flows, and, record for every bike simulator test subject the perception and accessibility of modeled bike sharing stations. Trajectories and filled questionnaires of bicycle simulator test subjects traversing NYC transport infrastructure might benefit evaluating the outcomes of the previously presented approach.

Notes

1. Background on the Boro Taxi program. NYC Taxi & Limousine Commission. URL: http://www.nyc.gov/html/tlc/html/passenger/shl_passenger_background.shtml; Retrieved 18 December 2013.
2. http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml.
3. <https://www.citibikenyc.com/system-data>.

Notes on contributors

Andreas Keler, PhD, is a postdoctoral researcher at the Department of Civil, Geo and Environmental Engineering, Technical University of Munich. His research focus is currently on bicycle traffic and active mobility, including the analyses on efficient and safe urban infrastructure designs, interactions and communications with autonomous vehicles, and, bicycle simulator studies.

Jukka M. Krisp, PhD, is Professor of Applied Geoinformatics at the University of Augsburg. His research interests include Geovisualization and Location-Based Services (LBSs).

Linfang Ding, PhD, is a postdoctoral researcher in the KRDB Research Centre for Knowledge and Data at the Free University of Bozen-Bolzano and at the Department of Cartography at the Technical University of Munich (TUM). Her research interests include Geovisualization and Spatial Data Mining.

ORCID

Andreas Keler  <http://orcid.org/0000-0002-2326-1612>

References

- Ahmed, M., S. Karagiorgou, D. Pfoser, and C. Wenk. 2015. "A Comparison and Evaluation of Map Construction Algorithms." *GeoInformatica* 19 (3): 601–632. doi:10.1007/s10707-014-0222-6.
- Amini, S., H. Twaddle, and A. Leonhardt. 2016. "Modelling of the tactical path selection of bicyclists at signalized intersections." Paper presented at *Transportation Research Board 95th Annual Meeting*, Washington, DC, January 10–14.
- Ankerst, M., M. M. Breunig, H.-P. Kriegel, and J. Sander. 1999. "OPTICS: Ordering Points to Identify the Clustering Structure." *ACM SIGMOD International Conference on Management of Data*. ACM Press, Philadelphia, May 31–June 03. pp 49–60.
- Birant, D., and A. Kut. 2007. "ST-DBSCAN: An Algorithm for Clustering Spatial–Temporal Data." *Data & Knowledge Engineering* 60 (1): 208–221. doi:10.1016/j.datak.2006.01.013.
- Castro, P. S., D. Zhang, C. Chen, S. Li, and G. Pan. 2013. "From Taxi GPS Traces to Social and Commu-Nity Dynamics: A Survey." *ACM Computing Surveys (CSUR)* 46 (2): 17. doi:10.1145/2543581.2543584.
- Ding, L., M. Jahnke, S. Wang, and K. Karja. 2016b. "Understanding Spatiotemporal Mobility Pat-Terns Related to Transport Hubs from Floating Car Data." In: G. Gartner and H. Huang (eds.) Paper presented at *The 13th International Symposium on Location-based Services*, Vienna, November 14–16. pp 186–190.
- Ding, L., J. M. Krisp, L. Meng, G. Xiao, and A. Keler. 2016a. "Visual Exploration of Multivariate Movement Events in Space-Time Cube." Paper presented at *The 19th AGILE International Conference on Geographic Information Science*. Helsinki, Finland.
- Faghih-Imani, A., and N. Eluru. 2016. "Incorporating the Impact of Spatio-Temporal Interactions on Bicycle Sharing System Demand: A Case Study of New York CitiBike System." *Journal of Transport Geography* 54: 218–227. doi:10.1016/j.jtrangeo.2016.06.008.

- Gordon-Koven, L., and N. Levenson. 2014. *Citi Bike Takes New York*. New York: Rudin Center for Transportation Management and Policy.
- Grigoropoulos, G., A. Keler, H. Kath, M. Spangler, and F. Busch. 2019. "Using Geodata for Simulating Urban Traffic – Current Research in the Field of Traffic Engineering and Control." In *Geoinformationssysteme 2019 – Beiträge Zur 6. Münchner GI-Runde*, edited by T. H. Kolbe, R. Bill, and A. Donaubaauer, 44–49. Berlin/Offenbach: Wichmann.
- Guo, Q. L., and H. A. Karimi. 2017. "A Novel Methodology for Prediction of Spatial-Temporal Activities Using Latent Features." *Computers, Environment and Urban Systems* 62 (2017): 74–85. doi:10.1016/j.compenvurbsys.2016.10.006.
- Jahnke, M., L. Ding, K. Karja, and S. Wang. 2017. "Identifying Origin/Destination Hotspots in Floating Car Data for Visual Analysis of Travelling Behavior." In *Progress in Location-Based Services 2016*, edited by G. Gartner, M. Jobst, and H. Huang, 151–172. Cham: Springer International Publishing.
- Jarvis, R. A. 1973. "On the Identification of the Convex Hull of a Finite Set of Points in the Plane." *Information Processing Letters* 2: 18–21. doi:10.1016/0020-0190(73)90020-3.
- Karrais, N., A. Keler, and S. Timpf. 2014. "Routing through a Continuous Field Constrained by a Network." In: K. Stewart, E. Pebesma, G. Navratil, P. Fogliaroni, and M. Duckham (eds.) *Extended Abstract Proceedings of the GIScience 2014*, pp. 57–61. Vienna, Austria.
- Keler, A., and J. M. Krisp. 2016a. "Visual Analysis of Floating Taxi Data Based on Interconnected and Timestamped Area Selections." In *Progress in Cartography, EuroCarto 2015*, edited by G. Gartner, M. Jobst, and H. Huang, 115–131. Cham: Springer International Publishing.
- Keler, A., and J. M. Krisp. 2016b. "Is There A Relationship between Complicated Crossings and Frequently Visited Locations? – A Case Study with Boro Taxis and OSM in NYC." In: Gartner, G., And H. Huang (Eds.) *Proceedings of the 13th International Symposium on Location-based Services*, Vienna, November 14–16. pp. 186–190.
- Keler, A., J. M. Krisp, and L. Ding. 2017b. "Visualization of Traffic Bottlenecks: Combining Traffic Congestion with Complicated Crossings." In *Advances in Cartography and GIScience, Selections from the International Cartographic Conference 2017*, edited by M. P. Peterson, 493–505. Cham: Springer International Publishing.
- Keler, A. 2018. "Modeling and Visualizing the Spatial Uncertainty of Moving Transport Hubs in Urban Spaces - a Case Study in NYC with Taxi and Boro Taxi Trip Data." In 38. *Wissenschaftlich-Technische Jahrestagung Der DGPF Und PFGK18 Tagung in München – Publikationen Der DGPF*, edited by T. P. Kersten, E. Gülich, J. Schiewe, T. H. Kolbe, and U. Stilla, Band 27. 397–407. Munich: DGPF.
- Keler, A., J. Kath, F. Chucholowski, M. Chucholowski, G. Grigoropoulos, M. Spangler, H. Kath, and F. Busch. 2018. "A Bicycle Simulator for Experiencing Microscopic Traffic Flow Simulation in Urban Environments." Paper presented at *The 2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, IEEE, Maui, Hawaii, November 4–7. pp 3020–3023.
- Keler, A., J. M. Krisp, and L. Ding. 2017a. "Detecting Travel Time Variations in Urban Road Net-Works by Taxi Trajectory Intersections." Paper presented at *The GISRUK 2017-25th GIS Research UK Conference*. Manchester, United Kingdom.
- Keler, A., L. Ding, and J. M. Krisp. 2016. "Visualization of Traffic Congestion Based on Floating Taxi Data." *Kartographische Nachrichten* 66 (1): 7–13.
- Krisp, J. M., and A. Keler. 2015. "Car Navigation - Computing Routes That Avoid Complicated Crossings." *International Journal of Geographical Information Science (IJGIS)* 29 (11): 1988–2000. doi:10.1080/13658816.2015.1053485.
- Krisp, J. M., K. Polous, S. Peters, H. Fan, and L. Meng. 2012. "Getting in and Out of a Taxi: Spatio-Temporal Hotspot Analysis for Floating Taxi Data in Shanghai." Conceptual paper - 6th International Symposium "Networks for Mobility", Stuttgart, September 27–28. doi:10.1094/PDIS-11-11-0999-PDN.
- Li, B., D. Zhang, L. Sun, C. Chen, S. Li, G. Qi, and Q. Yang. 2011. "Hunting or Waiting? Discovering Passenger-Finding Strategies from a Large-Scale Real-World Taxi Dataset." Paper presented at *The 2011 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops), MUCS 2011: 8th IEEE International Workshop on Managing Ubiquitous Communications and Services, Managing Contexts*, Seattle, March 21–25. pp. 63–68.
- Liu, X., and Y. Ban. 2013. "Uncovering Spatio-Temporal Cluster Patterns Using Massive Floating Car Data." *ISPRS International Journal of Geo-Information* 2: 371–384. doi:10.3390/ijgi2020371.
- Miller, H. J., Y.-H. Wu, and M.-C. Hung. 1999. "Gis-Based Dynamic Traffic Congestion Modeling to Support Time-Critical Logistics." *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences 1999 (HICSS-32)*, Maui, January 5–8. doi: 10.1109/HICSS.1999.772611.
- Shafique, M. A., and E. Hato. 2016. "Travel Mode Detection with Varying Smartphone Data Collection Frequencies." *Sensors* 16: 716. doi:10.3390/s16050716.
- Sladewski, R. S., A. Keler, and A. Divanis. 2017. "Computing the Least Complex Path for Vehicle Drivers Based on Classified Intersections." Paper presented at *The 20th AGILE International Conference on Geographic Information Science (AGILE 2017)*. Wageningen, the Netherlands.
- Weng, J.-C., Y.-Q. Zhai, X.-J. Zhao, and J. Rong. 2009. "Floating Car Data Based Taxi Operation Characteristics Analysis in Beijing." In: M. Burgin, M. H. Chowdhury, C. H. Ham, S. Ludwig, W. Su, and S. Yenduri (eds.). *Proceedings of the 2009 WRI World Congress on Computer Science and Information Engineering (CSIE 2009)*, Los Angeles, March 31–April 2. Volume 5, Intelligent Transportation Systems. pp 508–512.
- Yan, A. 2009. "Towards Semantic Trajectory Data Analysis: A Conceptual and Computational Approach." Paper presented at *The 34th International Conference on Very Large Data Bases (VLDB 2009)*. Lyon, France, October 31– November 2.
- Yue, Y., Y. Zhuang, Q. Li, and Q. Mao. 2009. "Mining Time-Dependent Attractive Areas and Move-Ment Patterns from Taxi Trajectory Data." Paper presented at *The 17th International Conference on Geoinformatics*, Fairfax, August 12–14. pp 1–6.
- Zhang, D., and T. He. 2012. "pCruise: Reducing Cruising Miles for Taxicab Networks." Paper presented at *The Proceedings of the 2012 IEEE 33rd Real-time Systems Symposium (RTSS 2012)*, San Juan, PR, December 4–7. pp. 85–94. doi:10.1094/PDIS-11-11-0999-PDN.

Zhang, D.-Z., Z.-R. Peng, and D. J. Sun. 2014. "Comprehensive Taxi Assessment Index Using Float-Ing Car Data." *Journal of Harbin Institute of Technology* 21 (1): 7–16.

Zhao, Y., Q. Qin, J. Li, C. Xie, and R. Chen. 2012. "Highway Map Matching Algorithm Based on

Float-Ing Car Data." Paper presented at *The IGARSS 2012 - IEEE International Geoscience and Remote Sensing Symposium, THP-P: Image Information Extraction: Detection of Man-made Features*, Munich, July 22–27. pp 5982–5985. doi:[10.1094/PDIS-11-11-0999-PDN](https://doi.org/10.1094/PDIS-11-11-0999-PDN).