Taylor & Francis
Taylor & Francis Group

Journal of Location Based Services

ISSN: 1748-9725 (Print) 1748-9733 (Online) Journal homepage: https://www.tandfonline.com/loi/tlbs20

Understanding the relationship between complicated crossings and frequently visited locations – a case study with boro taxis in Manhattan

Andreas Keler & Jukka M. Krisp

To cite this article: Andreas Keler & Jukka M. Krisp (2019): Understanding the relationship between complicated crossings and frequently visited locations – a case study with boro taxis in Manhattan, Journal of Location Based Services

To link to this article: https://doi.org/10.1080/17489725.2019.1588406

Published online: 26 Mar 2019.

JOURNAL OF LOCATION BASED SERVICES
https://doi.org/10.1080/17489725.2019.1588406

Understanding the relationship between complicated crossings and frequently visited locations – a case study with boro taxis in Manhattan

Andreas Keler [a] and Jukka M. Krisp [b]

[a]Chair of Traffic Engineering and Control, Technical University of Munich, Munich, Bavaria, Germany; [b]Institute of Geography, Applied Geoinformatics, University of Augsburg, Augsburg, Bavaria, Germany

ABSTRACT

Urban mobility has complex patterns and principles. Data of moving entities on the underlying transportation infrastructure can help understanding those complex patterns and principles. Therefore, we need static infrastructural information and knowledge on spatio-temporal movement patterns of public transport services and of various vehicle fleets. We focus on inspecting data partitions of individual taxi movement acquisitions in New York City (NYC), together with OpenStreetMap (OSM) data extracts, for gaining more knowledge about the complex daily mobility patterns in NYC. We select trip information of tracked boro taxi drivers, who are restricted to pick up customers at the airports and the southern part of Manhattan. By computing with taxi customer drop-off positions, we define drop-off clusters as the customer destination hotspots of selected Saturdays in June 2015. These hotspots are then related to the OSM road network, in particular to its derivatives: complicated crossings. By comparing with a previous assumption of detecting 'fast leaving' behaviour within the restricted zone, we receive characteristic matching results: only few destination hotspots appear at complicated crossings. Nearly all the matching intersections have nearby situated pedestrian zones and many are associated with previous construction measures. Finally, we reason on the usefulness of the proposed method.

ARTICLE HISTORY
Received 12 May 2018
Accepted 22 February 2019

KEYWORDS
Urban transportation; spatial analysis; mobility patterns; boro taxis; vehicle fleets; transportation infrastructure

1. Introduction

Installed in-vehicle sensors can record the movement of individual vehicles of a taxi fleet. These are often components of an established taxi dispatcher service with the intention to monitor operating taxis in one selected area. In urban environments, the taxi service is an important mode of transportation (Tang et al. 2015). In general, taxi movement records are available as listed time-stamped vehicle locations, which are represented as a number of individual movement trajectories. These taxi

CONTACT Andreas Keler andreas.keler@tum.de

trajectories are useful for deriving urban dynamics. The types of urban dynamics are manifold and find applications in different domains: reaching from taxi trip distribution modelling (Tang et al. 2018b) over inferring driving trajectories (Tang et al. 2018a), we can predict taxi trip destinations based on historical data (Zong et al. 2019). Depending on the availability of recorded attributes, which consist, in case of the Floating Car Data (FCD) technique, of tracked geographical positions, and, of measurements from on-board electronics for every time stamp. The latter might have a consistent sampling interval or follow sampling based on selected distance measures (Jun, Guensler, and Ogle 2006). One special attribute is, in the case of tracked taxis, the operational status of the vehicle, if it is vacant or with a customer (Krisp et al. 2012).

Extracted positions of taxi trajectories, where customers enter or leave the vehicle are useful information for a big variety of applications (Castro et al. 2013). These points imply semantic information that is useful for different analytical tasks and services in different domains. Extracts of taxi trajectory positions are, for example, useful for assessing the taxi service quality (Zhang, Peng, and Sun 2014), or for analysing taxi driver behaviour (Li et al. 2011). Examples of taxi driver behaviour include hunting or waiting for customers, which are different operational strategies. By finding numerous similar positions of taxi customer pick-ups and drop-offs, it is possible to assign selected parts of the investigation area as service hotspots. In general, there are many possibilities for detecting taxi pick-up and drop-off hotspots (Krisp et al. 2012; Pan et al. 2013; Li et al. 2011). These possibilities include different types of point clustering techniques, as well as point density estimations. The detected hotspots often correlate with functional parts of the transportation infrastructure, as, for example, the taxi waiting zones near airports (Ding, Yang, and Meng 2015). The specific patterns of taxi pick-up and drop-off hotspots are classifiable by spatial and temporal components. In many cases, operational spatio-temporal taxi patterns correlate with typical or periodical mobility patterns of the whole city, or of connected urban environments. The spatio-temporal distribution of taxi pick-up and drop-off points also depends on the influences from seasonal social events at selected locations or on traffic events at periodical rush hours on the inspected road network.

1.1. Definition and representation of taxi trip destination hotspots

Specific positions of complete vehicle trip trajectories may indicate the change of operational mode. Keler (2018) shows that origins and destinations of different taxi fleets can appear at the same time in a close spatial range, as for the example of yellow taxis and boro taxis in NYC. Nevertheless, due to the fact of having anonymous taxi customers, it is impossible to prove any transit behaviour with these two data extracts. In general, origin and destination points are easy to extract from taxi trajectories, since the collected attribute values for operational statuses originate from on-board sensor acquisitions. These points are the changing points of the operational taxi status: occupied and non-occupied.

There are also data extracts from operating taxi service vehicles that do not include the operational status of the taxi. One option for an alternative taxi trip origin and destination extraction may base on deriving stopping and parking positions. In general, critical issues for all acquired vehicle trajectories from GNSS positioning are connected with data quality, since positioning accuracy may vary with different positioning devices and built infrastructure that influences the spatial accuracies. Nevertheless, GNSS positioning in urban environments enables detecting manifold causes of travel time variations (Tang, Yang, and Qi 2018).

When extracted taxi pick-up and drop-off points are massive, since coming from thousands of tracked taxi trajectories, it is possible to assign generalised pick-up and drop-off hotspots for specific times of the day. For the definition of a taxi pick-up and drop-off hotspot, there are many different possibilities. Krisp et al. (2012), for example, use k-means clustering on both types of points for defining hour-wise time windows of the day. Despite its high efficiency for clustering massive location data, k-means clustering needs a-priori knowledge about the number of clusters and the expectancy of having convex shaped clusters in Euclidean space (Yue et al. 2009). By the visual representation of the k-means clusters within a space-time cube, it is possible to inspect changes of spatial distribution of the hotspots over time.

Besides partitioning clustering of points, Pan et al. (2013) use the DBSCAN clustering algorithm, which was firstly defined by the Kriegel group, LMU Munich, in 1996 (Ester et al. 1996). DBSCAN and its related algorithms like OPTICS (Ankerst et al. 1999) are possibly the most frequently used techniques for defining taxi drop-off hotspots. The reason is their general ability of adapting clustering parameters based on various specific data sets. OPTICS can, for example, be applied for estimating optimal density-based clusters, given by search distances and number of points.

One other option for clustering taxi trajectory data might be the usage of trajectory clustering, favourably by distinguishing between occupied and non-occupied trajectory segments.

OPTICS, or Ordering Points To Identify the Clustering Structure, is useful, when nothing is known about the data distribution, as the missing knowledge on density-based connections between the points. OPTICS delivers useful insights in distinguishing between reasonable drop-off clusters and outliers, by the option of inspecting the search distance (Epsilon) histogram.

For Pan et al. (2013), the further interpretation of pick-up and drop-off hotspots is the conveying. Those cues have a relation to the social functions of regions (Pan et al. 2013).

One focus of this work is on extracting taxi drop-off positions and subsequent point generalization into hotspots. We believe that agglomeration of drop-off positions may help representing the interests of individual customers. In particular, we focus on data extracts from Saturdays, when we expect many taxi customers following social interaction in connection with recreational activities.

In general, we can say that in the 2010s the taxi customer pick-up densities in Manhattan are significantly higher than in the outer boroughs of NYC (Qian and Ukkusuri 2015). Sayarshad and Chow (2016) inspect NYC taxi data extracts at different temporal scales showing, especially in the central part of NYC, specific variations between the distributions of trip origin points and trip destination points for selected time windows. The visual inspection of taxi trip origins and destinations from temporal partitions is a first indication technique for finding spatial patterns for further inspections as in the approach of Ferreira et al. (2013).

Alfeo et al. (2018) propose a stigmergy-based process of NYC taxi hotspot discovery. The discovered hotspots are similar to those found by other methods as by Keler and Krisp (2016b). Definite operational taxi hotspots on weekdays, for both yellow and boro taxis, are situated in northern Manhattan. Besides operational hotspots, it is feasible to model demand and supply of taxi trips at it is proposed by Yang (2015) with a data-driven modelling approach. Specific taxi trips can reveal more detailed information than in their aggregated form. Due to the various information and the massive data collection, Douriez et al. (2016) point out that the identification of specific taxi drivers with their daily income is inferable. Besides this fact, it is even possible to associate the origins of customers with properties of the payment itself.

The difference to other studies using the NYC taxi data sets consists of relating dynamic taxi trip information with static infrastructural information from OSM. We believe that specific road network designs and every composition of design elements, together with buildings and various other infrastructural elements, influence daily mobility patterns, especially in complex urban environments. Furthermore, we believe that taxi travel behaviour has not only connections to dynamic aspects of traffic quality and varying traffic states. There are also connections between taxi travel behaviour and level of infrastructural complexity, which might occur in specific spatio-temporal patterns. One contribution of this work is the attempt of comparing the static complexity of road intersections with dynamic taxi service hotspots without implying routing applications.

1.2. The complexity of road intersections

Transportation infrastructure elements are important elements in urban environments and often local knowledge is decisive for the properties of selected road elements. Even when not perceived superficially, selected parts of the transport infrastructure might imply danger for vehicle drivers as locations of traffic bottlenecks.

One typical type of locations in many different transportation infrastructures is the road intersection. People perceive some of these elements as complicated, depending on the group of people and their selected mode of travel. Krisp and Keler (2015) propose a data-driven method for estimating the level of complexity of a perceived complicated crossing. The case study bases on questionnaires, where driving beginners in driving schools in Munich evaluated their perceived complicated elements within traffic participation. The road network of Munich was input information for testing the method. Resulting from the method application, numerous perceived complicated crossings are extractable and show reasonable results, when comparing to local knowledge.

Perceived complexity of transportation infrastructure is not only dependent on the spatial configuration of road network segments. There are manifold dynamic components that can influence the perception of complexity. Nevertheless, the road design can be seen as one non-dynamic component in estimating the complexity of road intersections. Therefore, we try to introduce regularities that express that intersection consisting of curvy densely ordered segments of different road types are more complex and weighted higher in perceived complexity. One advantage of this estimation is that no local knowledge of the investigation area is needed.

In contrast to specific points of interests with functional attributes as airports and stations, we want to inspect taxi drop-off hotspots at complicated crossings. The idea is to find out any connection to perceived complicated crossings. This is based on an extension of the data-driven method proposed by Krisp and Keler (2015), which was already tested for Munich's road network with reasonable results. The extension consists of excluding road segments of specific road types while adapting perceived complexity in the investigation area. Within this procedure, we want to test the exclusion of pedestrian influence, which results in the absence of interaction points between vehicles and pedestrians on infrastructural level.

Additionally, it is to say that there are numerous possibilities for defining the complexity of road intersections. Sladewski, Keler, and Divanis (2017) propose a technique that respects the number of connected roads and of turning options, resulting in very differing test results when comparing with the technique by Krisp and Keler (2015). The numerous turning options on the magic roundabout in Swindon (UK) have, for example, the degree of two to three in the method by Sladewski, Keler, and Divanis (2017). This does not show a high complexity at any part of the magic roundabout, whereas the high density of extracted nodes of this roundabout indicates a high complexity in the method by Krisp and Keler (2015). Sladewski, Keler, and Divanis (2017) use a vehicle network of Le Havre without information on bicycle and pedestrian road segments.

This work applies a technique variation of Krisp and Keler (2015) on a road network extract from OSM of NYC with all available road types. The parameters for the method are adjusted based on observations of the present transportation infrastructure of the investigation area. Compared to previously conducted studies on NYC taxi data sets, we include the static component of the road network design together with the linkage to transport infrastructure of other modes of transport into the analysis of dynamic information. Exclusion of specifically classified road segments might influence assigning the complexity of selected road intersections.

2. Description of the case study and the test data sets

This work focus on testing a data analysis technique for extracting local knowledge. The input data consists of dynamic and static information for one selected investigation area.

In this work, we test taxi data sets of a taxi fleet with operational restrictions: the boro taxi service. The underlying data allow extractions of every boro taxi trip destination in June 2015, which equal the points where passengers leave the taxi. The second type of data are extracts from the road network of the OSM project for the administrative area of New York City (NYC). It is the base for estimating the complexity of crossings or the local transport infrastructure in general.

2.1. Boro taxi data

Boro taxis operate in NYC since 2013 and were introduced for coping the taxi passenger demand in the boroughs outside of Manhattan, namely Brooklyn, Queens, the Bronx and Staten Island. The reason for introducing this service results from previous analyses of spatial distribution of yellow taxi GPS data.[1] In particular, passenger pick-up positions of yellow taxis are a matter of investigation. Around 95% of the pick-ups occurred within Manhattan and the rest in the outer boroughs.

Having a cheaper taxi license fee, 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 consist of 20 attributes, and every record represents one trip. We select boro taxi data from all trips of the four Saturdays in June 2015, coming from the NYC Taxi & Limousine Commission (TLC).[2] 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 borrows 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 partition, we extract start and destination points and focus on the latter for inferring customer drop-off points. Additionally, there is a possibility to associate 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).

After inspecting the selected four Saturdays in June 2015, we state that there are in general more drop-offs on weekends than on working days, especially in southern Manhattan (yellow zone). The drop-off hotspots on Saturdays have less periodical spatio-temporal distribution than those on working days.

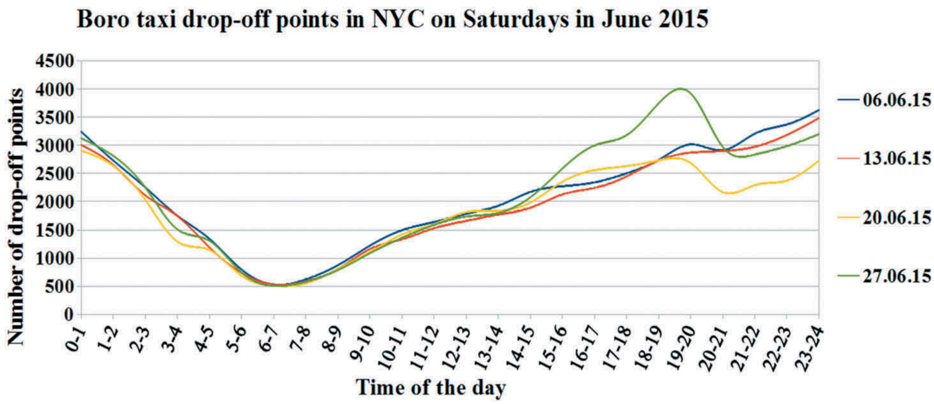

Figure 1. Number of boro taxi drop-off points for time of the day in NYC on the four Saturdays in June 2015.

Many drop-off hotspots can be associated with social events. Therefore, Figure 1 shows four curves for the selected Saturdays in June 2015. Whereas the first part of the Saturdays has a very similar distribution, evening hours have a high variation. Each variation can result from specific social events and have specific differences in spatial distributions. Weather events are besides social event information important components that can influence numerous taxi drivers in NYC (Camerer et al. 1997).

2.2. OSM road network of NYC

Road segments can have manifold representations. We can represent road networks as a number of non-connected line or polyline features or we can assure an accurate connectivity between the different road segments, which respect and describe the driving directions. Another option has far more applications: connected and directed network graphs. As in most routing applications, for pedestrians and vehicle drivers, the 1D network space is used for computing the path between two points. In its simplest case, the shortest path algorithm by Dijkstra (1959) is applied. In many cases of freely available road network data, connectivity of arcs and nodes, together with direction is not necessarily included. As in case of ATKIS (Harbeck 2001), and sometimes OSM, the data has more focus on higher visual representation quality of occupied space than the accurate connections (arcs) between the nodes, which can be road intersections of road networks.

Nevertheless, many routing applications make use of OSM road network information, which is a matter of testing for its practical use. One often needed step is conversion into 1-D network space, which sometimes comes along with problems in connectivity and has to be modified for practical routing applications. Within the project, there are differences in the mapping quality, especially concerning the road network information. In previous examples, where vehicle routing applications are designed (Karrais, Keler, and Timpf 2014; Keler and Mazimpaka 2016; Sladewski, Keler, and Divanis 2017) significant differences are detectable between OSM road network and the network from Google Maps. Additionally, there are possibilities of abstracting road networks, mainly by applying line generalization methods. Abstracted road networks may imply less spatial (position and size) and semantical (driving directions, restrictions, lane number and connectivity) information than usual OSM road networks, as for the case of the GraphStream network of Le Havre (Sladewski, Keler, and Divanis 2017). Compared to these examples, the OSM road network is relatively accurate and rich on additional information on the number of lanes, road type and speed limits (Keler and Krisp 2016a).

OSM road networks are for Stanica, Fiore, and Malandrino (2013) one of the most accurate, which are freely available. NYC has a very detailed and relatively consistent OSM road network, of which an extract is pictured in Figure 2.

3. Relating complicated crossing with frequently visited locations

Based on the previous findings, we want to introduce a simple and computationally efficient method for defining and extracting boro taxi drop off point hotspots and relate them to detected complicated crossings. The idea is to make use of the density connectivity of daily boro taxi drop off points by using the OPTICS algorithm (Ankerst et al. 1999). 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 3 shows this technique (in beige box), together with the OSM-based detection of complicated crossings (upper left blue box).

4. Results

By applying the method for detecting frequently visited locations as pictured in Figure 3, we receive a number of intersecting polygons for each of the four Saturdays in June 2015. The outcomes of detecting complicated crossings via OSM road network data and of defining and extracting boro taxi customer drop-off polygons are a matter of respective visual inspections.

4.1. Complicated crossings in NYC

The technique by Krisp and Keler (2015) consists of at least three different forms of data representation. Figure 4 shows the workflow of the technique for defining complicated crossings in NYC. Compared to Munich, NYC has longer

Figure 2. Inspected OSM road network data extract from NYC.

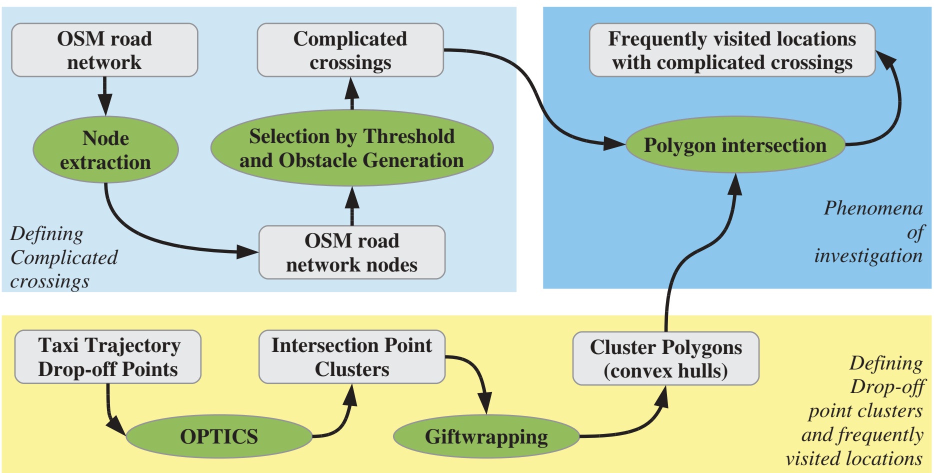

Figure 3. Workflow of the method for detecting frequently visited locations at complicated crossings.

and wider road segments with less curvature. Depending on the input road information from OSM (Figure 4(a)), different densities appear in its extracted nodes (Figure 4(b)). As in Krisp and Keler (2015) we can use an average diameter, as, for example, 60 m for the case of Munich, which can base on roundabouts within the respective investigation areas for selecting a threshold for creating obstacle polygons.

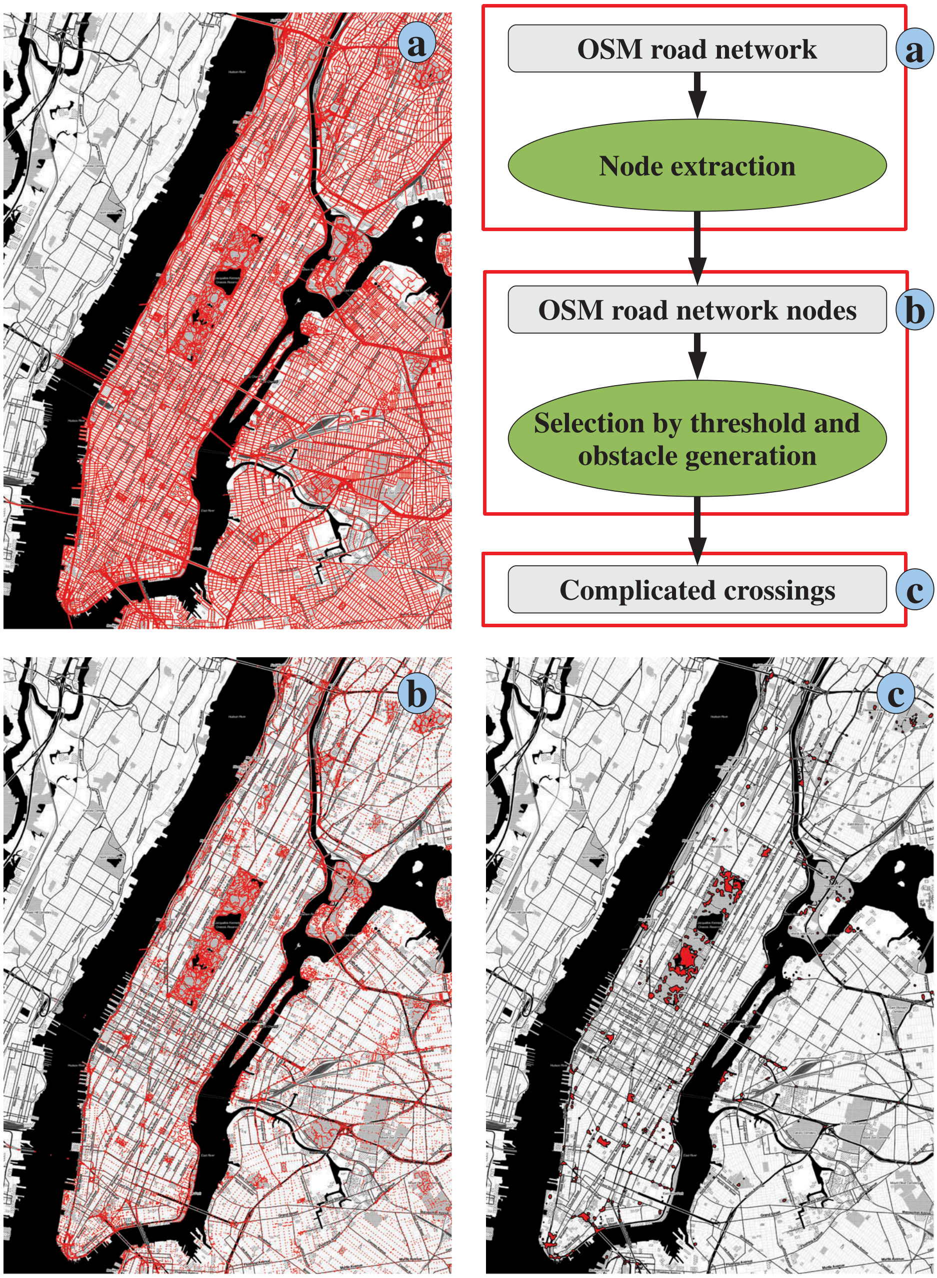

Figure 4. Technique for defining complicated crossings, with (a) node extraction, (b) selection by threshold and (c) extraction of polygons.

In general, NYC has lower node densities than Munich. The resulting complicated crossings in Figure 4(c) do not only depend on node densities but as well on intersection points between different road segments. In case the intersecting roads are from different types, the resulting intersection points gain a higher weight, which likely indicates a complicated crossing. Therefore, Figure 4(c) shows

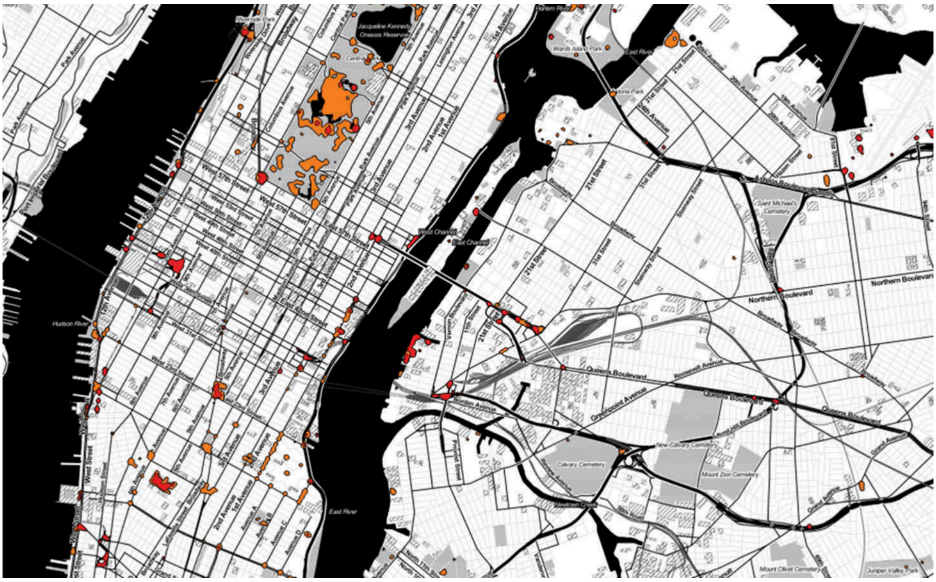

Figure 5. Complicated crossings in the central part of NYC with the inclusion of pedestrian pathways in the input information (orange polygons) and with its exclusion (red polygons).

unreasonable results of complicated crossings in the Central Park and in other parks in NYC, because road segments are intersecting bicycle lanes and pedestrian pathways. Other reasons for this misclassification at all parks in NYC are the highly curvy pathways with intersections to other road types. High curvature of pathways comes together with high node densities.

One additional step for immediately improving the results might be to exclude pedestrian pathways from the input information for detecting complicated crossings. After this improvement step, many complicated crossings disappear especially in the area of the Central Park. Figure 5 shows the difference between detected complicated crossings with pedestrian pathways as input information (orange polygons) and those without this road type (red polygons).

Due to the disappearance of many prominent crossings in NYC, especially those with pedestrian crossings, when excluding pedestrian pathways, we use the orange polygons in Figure 5 for the further matching with boro taxi drop-off hotspot polygons.

To our knowledge, the technique of Krisp and Keler (2015) implies all available road infrastructure information and introduces a static weighting scheme throughout the whole procedure. One novelty of this paper is the exclusion of the pedestrian influence on estimating the complexity of road intersections by subtracting one road type from the complete road segment data set. Whereas this is not reasonable when estimating the complexity or road intersections, since vulnerable road users (VRUs) are an important component, in reality, this shows that it has a direct influence on the complexity classification outcomes.

4.2. Boro taxi drop-off hotspot polygons

By applying the density-based technique OPTICS for boro taxi drop-off points of the four Saturdays in June 2015, we receive more than 4,000 clusters per day. After applying the gift wrapping algorithm on every cluster for a better visual representation, the respective convex hulls are the outcomes. The resulting drop-off hotspot polygons are pictured in Figure 6, with the selection of day wise polygons for every of the four Saturdays. These polygons result from specific adjusting of parameters values of OPTICS. After previous tests with subsequent visual inspection, we provide the drop-off hotspot detection for all four data partitions with the search distance (Epsilon) of 31.0896 meters, which equals the maximum Street width in Time Square (102 feet).

In parallel, we have to define taxi drop-off point clusters by using the OPTICS algorithm. In the case of this algorithm, we have to define a search distance (Epsilon) and a minimum number of points (MinPts) for estimating the density. Our minimum number of taxi drop-off points will be 2 (MinPts = 2). With those two parameters, we define our density connection between the drop-off points. The resulting clusters are then transformed by the gift wrapping algorithm (Jarvis 1973) into polygons.

4.3. Hotspots at complicated crossings

The final step of the method is to connect drop-off hotspots with complicated crossings. The idea is to provide a spatial overlay of each of the four hotspot polygon sets with detected complicated crossings in NYC. For the matching of two polygon layers, there are numerous possibilities, especially when following the

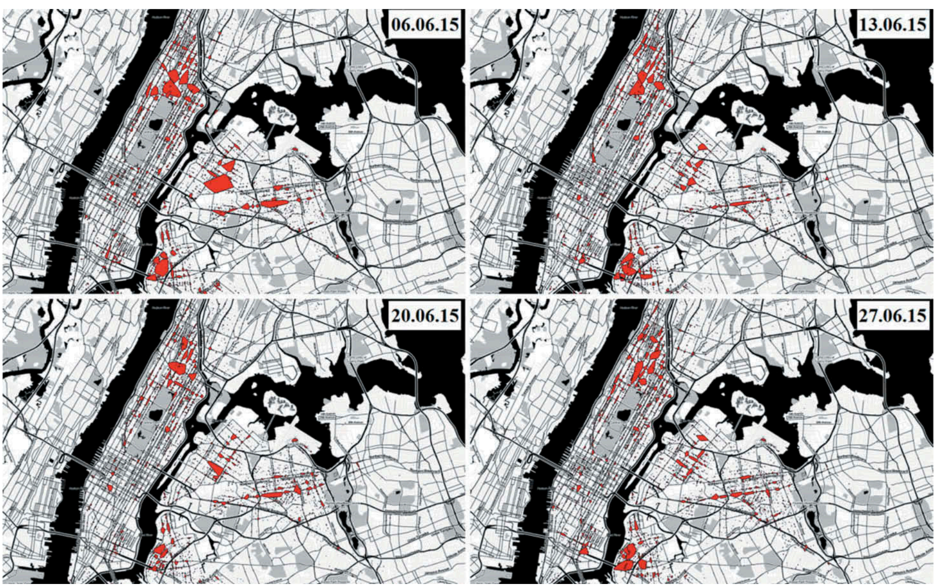

Figure 6. Boro taxi drop-off hotspot polygons for the four Saturdays in June 2015 in NYC.

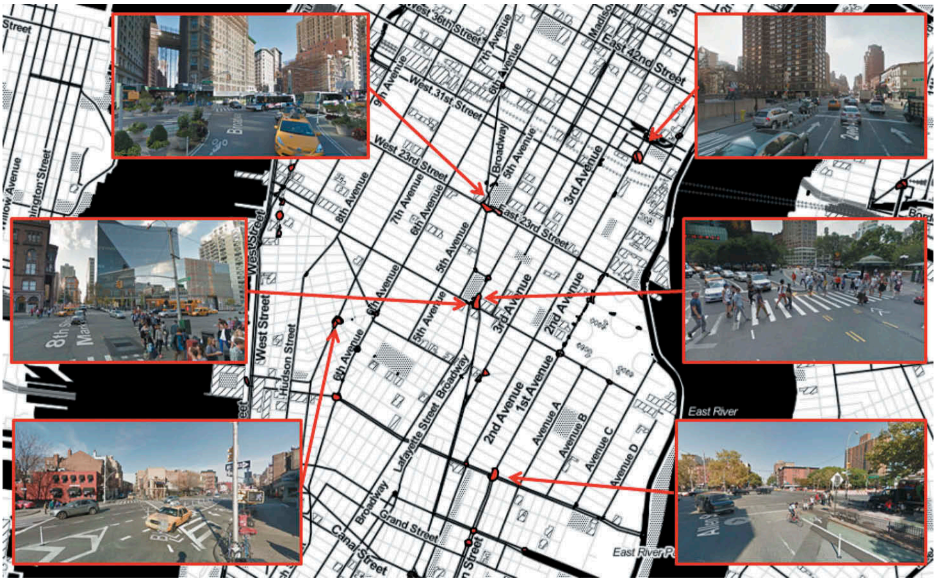

Figure 7. Six detected boro taxi drop-off hotspots at complicated crossings in southern Manhattan.

idea of representing the relations of the spatial matching, in terms of its spatial configuration. Numerous methods can handle a series of polygons. One option is to extract the intersection areas of matching polygons. Another option is to extract the information of how polygons are intersecting. One possibility for doing so is using the 9-intersection model as in Keler (2018). Each of the polygons matchings might be represented as an Egenhofer matrix with nine fields. We select the option of extracting the areas of spatial overlay, in case hotspot polygons and complicated crossings are matching. The idea behind this is to show only areas associated with both polygon types, without spatial tolerance areas.

In the last step of defining frequently visited locations, we match the complicated crossings with taxi drop-off point polygons in NYC. Figure 7 shows a cut-out of the defined frequently visited locations in NYC every Saturday in June 2015, which will be the base for our further inspections.

Figure 7 shows a selection of six detected boro taxi drop-off hotspots at complicated crossings in southern Manhattan. Additionally, every outcome of the polygon intersection step is pictured by Google StreetView images from the same location, showing the spatial configuration of roads at road intersections.

5. Conclusions

This work proposes a method for detecting taxi drop-off hotspots at complicated crossings. We are using drop-off points of a taxi fleet with specific mobility patterns: boro taxis. For the selected Saturdays, most of the boro taxi drop-off hotspots take place at the border to the restricted yellow zone. There are high

Table 1. Numbers of drop-off hotspots and those intersecting with complicated crossings.

Inspected Saturday in NYC	06.06.2015	13.06.2015	20.06.2015	27.06.2015
Total number of drop-off hotspots	5006	4919	4769	5233
Number of drop-off hotspots intersecting complicated crossings	168	169	137	148
Percentage of drop-off hotspots at complicated crossings from total number	3.356%	3.436%	2.872%	2.582%

numbers of hotspots in northern Manhattan and at Williamsburg Bridge. This shows the general trend of boro taxi drivers to avoid driving in the restricted yellow zone. The assumption that boro taxi drivers avoid dropping-off customers at complicated crossings has a decisive proof: a relatively small number of taxi drop-off hotspots appear at complicated crossings. In total, the boro taxi drop-off hotspots on the four Saturdays in June 2015 overlap to only 2% to 4% with complicated crossings. Table 1 shows the numbers of hotspots and those intersecting with complicated crossings.

Around 130 complicated crossings in NYC, mostly situated in southern Manhattan, always intersect with daily drop-off hotspots. Surprisingly, all modified complicated crossings in NYC fall into this category. Six of these are pictured in Figure 7. The modification of at least one road segment into a pedestrian zone makes it also reasonable to drop-off customers at these intersections. The different initiatives of the Department of Transportation in NYC, include, besides other types of street re-engineering, the approach type ‘choose quality over quantity’ for simplifying complex intersections (NYCDOT 2013). The complex intersections definition by NYCDOT (2013) includes the existence of odd angles between the streets, as intersections with five or more roads. The re-engineering strategy is to divert or to remove selected road segments from the intersection and to create new plaza space for pedestrians, which might serve as good locations for taxi customer drop-offs (NYCDOT 2013).

The technique for detecting complicated crossings by Krisp and Keler (2015) delivers reasonable results for the OSM road network of Munich. The results after applying the same technique for the OSM road network of NYC are less reasonable: many detected complicated crossings appear at park areas. The reason for this appearance results from the high number of intersections with different road types as pathways and minor roads. Additionally, NYC has a smaller density of complicated crossings, especially in southern Manhattan. This is connected with the generally wider and straighter street segments in NYC than in Munich. For delivering comparable results, input parameters and parts of the method itself need modifications. One key point of this assumption is that the inferred road network node density does not necessarily indicate the complexity. Therefore, exclusion of specific road types as bicycle lanes or pedestrian paths, as an extension of the technique, can show the direct influence on variations in number, shapes and sizes of classified road intersections.

The proposed method provides a connection between daily mobility and transportation infrastructure. This type of approach is extendable in many ways. The NYC boro taxi records are useful data for deriving information of appearing public events. The variations in drop-off numbers, travel times, and drop-off locations can be associated with events (changes in periodical patterns, outliers).

The contribution of the presented approach consists of providing the potential spatial analyst a number of indicators resulting from unusual relationships of static and dynamic urban data. Relating operational mobility hotspots from specific services with the used underlying transport infrastructure is not new. Nevertheless, qualitative data on various perceptions of urban space can provide even further revisions and approval to specific spatio-temporal patterns. Here, it is important not only to focus on the statically built infrastructure, since temporal variations are possible and in many urban areas of the world present, but imply differently varying temporal scales.

Additionally, we can say that manifold reasons can imply avoidance of customer pick-ups or drop-offs as, for example, traffic situations, events or the perceived complexity due to the number of signals of specific traffic control elements. Extracted polygons of the presented procedure might serve as indicators for present traffic bottlenecks, higher traffic volumes at specific time of the day or even specific traffic light signalling.

One application value of the presented technique for intelligent transportation systems might benefit concepts on autonomous vehicles or taxis. Due to the higher complexities of selected road intersections, autonomous vehicles might avoid the polygons that are extracted in this paper and route around them. Since those polygons are hotspot areas of other taxi fleets, have complex road infrastructure and imply numerous different traffic participants, interaction and communication of autonomous taxis with its environment might be expensive and time-consuming. The presented method by Krisp and Keler (2015) for estimating complexities of road networks might then be adapted to the requirements of autonomous vehicles. Estimating complexities for autonomous vehicles might imply sensor-specific properties that are difficult to model and possibly not visible for pedestrians. The simplest path for autonomous vehicles is possibly the one with the lowest interaction with other traffic participants, which indicates that just data on static infrastructure might not be enough. On the other hand, static information can serve for extracting conflict points and less intersecting road lanes may indicate less expected interactions. Taking this into account, we can say that our approach might be one option to guarantee traffic safety.

Another idea comes from achieving optimal operational quality by avoiding typical operational hotspots of other taxi fleets and mobility services. This idea comes with a paradox, since these mobility hotspots also indicate a higher activity of transit. On the other hand, autonomous vehicles can navigate on areas with a lower activity of change of transport mode, since less pedestrians are expected that switch vehicles. This strategy might contribute to higher traffic safety at specific urban locations in the future.

6. Outlook

Further steps motivated from our approach consist of reasoning about an automatic extraction of local knowledge. Therefore, the inclusion of social media data as geocoded Tweets might serve as a useful addition or a matter for result evaluations.

This might be useful for associating social events with varying drop-off numbers in selected parts of the city at selected times.

Another useful addition to the method is the inclusion of traffic flow information. The idea is to connect road segment complexity with road capacities. Understanding urban traffic with specific flow patterns on the transportation infrastructure is important for further analyses. Therefore, there is a need for differentiating between road types and road segments for analyzing mobility and road usage patterns.

One possible addition, when using OSM data, is enriching classified areas (as complicated crossings) with intersecting and nearby situated features (static and dynamic) that might consist of POIs or static technical devices as traffic lights.

One future direction of attempting comparing perceived static complexity with dynamic information might consist of testing the OSM-based complexity in a VR bicycle simulator application that evaluates only the used OSM information and not the present photorealistic depiction of a selected road intersection (Keler et al. 2018). This might reflect differences between the perceived complexities of the real world and the (geo-) data world.

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.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

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

References

Alfeo, A. L., M. G. C. A. Cimino, S. Egidi, B. Lepri, and G. Vaglini. 2018. "A Stigmergy-Based Analysis of City Hotspots to Discover Trends and Anomalies in Urban Transportation Usage." *IEEE Transactions on Intelligent Transportation Systems* 19 (7): 2258–2267. doi:10.1109/TITS.2018.2817558.

Ankerst, M., M. M. Breunig, H. P. Kriegel, and J. Sander. 1999. "OPTICS: Ordering Points to Identify the Clustering Structure." In *ACM SIGMOD international conference on Management of data*, 49–60. Philadelphia, PA: ACM Press.

Camerer, C., L. Babcock, G. Loewenstein, and R. Thaler. 1997. "Labor Supply of New York City Cabdrivers: One Day at a Time." *The Quarterly Journal of Economics* 112 (2): 407–441. doi:10.1162/003355397555244.

Castro, P. S., D. Zhang, C. Chen, S. Li, and G. Pan. 2013. "From Taxi GPS Traces to Social and Community Dynamics: A Survey." *ACM Computing Surveys (CSUR)* 46 (2). Article No. 17. doi:10.1145/2543581.2543584.

Dijkstra, E. W. 1959. "A Note on Two Problems in Connexion with Graphs." *Numerische Mathematik* 1: 269–271. doi:10.1007/BF01386390.

Ding, L., J. Yang, and L. Meng. 2015. "Visual Analytics for Understanding Traffic Flows of Transportation Hub from Movement Data." In *Proceedings of ICC 2015 - International Cartographic Conference*, August 23–28. Rio de Janeiro, Brazil.

Douriez, M., H. Doraiswamy, J. Freire, and C. T. Silva. 2016. "Anonymizing NYC Taxi Data: Does It Matter?" In *2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA*), 140–148. doi:10.1109/DSAA.2016.21.

Ester, M., H.-P. Kriegel, J. Sander, and X. Xu. 1996. "A Density-Based Algorithm for Sparse Representations." Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining, Portland, OR, 226–231.

Ferreira, N., J. Poco, H. T. Vo, J. Freire, and C. T. Silva. 2013. "Visual Exploration of Big Spatio-Temporal Urban Data: A Study of New York City Taxi Trips." *IEEE Transactions on Visualization and Computer Graphics* 19 (12): 2149–2158. doi:10.1109/TVCG.2013.226.

Harbeck, R. 2001. "15 Jahre ATKIS®, und die Entwicklung geht weiter." *Vermessung Brandenburg* 1/2001: 3–14.

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.

Jun, J., R. Guensler, and J. H. Ogle. 2006. "Smoothing Methods to Minimize Impact of Global Positioning System Random Error on Travel Distance, Speed, and Acceleration Profile Estimates." *Transportation Research Record: Journal of the Transportation Research Board* 1972: 141–150. doi:10.1177/0361198106197200117.

Karrais, N., A. Keler, and S. Timpf. 2014. "Routing through a Continuous Field Constrained by a Network." In *Extended Abstract Proceedings of the GIScience 2014*, edited by K. Stewart, E. Pebesma, G. Navratil, P. Fogliaroni, and M. Duckham, 57–61. Vienna, Austria.

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. Gülch, J. Schiewe, T. H. Kolbe, and U. Stilla, 397–407. Vol. 27. Munich: DGPF.

Keler, A., J. Kaths, F. Chucholowski, M. Chucholowski, G. Grigoropoulos, M. Spangler, H. Kaths, and F. Busch. 2018. "A Bicycle Simulator for Experiencing Microscopic Traffic Flow Simulation in Urban Environments." The 21st IEEE International Conference on Intelligent Transportation Systems (ITSC'18), Maui, Hawaii, US.

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 *Proceedings of the 13th International Symposium on Location-Based Services*, edited by G. Gartner and H. Huang, 186–190. Vienna, Austria.

Keler, A., and J. D. Mazimpaka. 2016. "Safety-Aware Routing for Motorised Tourists Based on Open Data and VGI." *Journal of Location Based Services* 10 (1): 64–77. doi:10.1080/17489725.2016.1170216.

Krisp, J. M., and A. Keler. 2015. "Car Navigation - Computing Routes That Avoid Complicated Crossings." *International Journal of Geographical Information Science* 29 (11): 1988–2000. doi:10.1080/13658816.2015.1058388.

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"*, 27–28 Sept. 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." In *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*, 63–68. doi:10.1016/j.tvjl.2010.05.029

NYCDOT. 2013 November. Making Safer Streets. Report by the New York City Department of Transportation. http://www.nyc.gov/html/dot/downloads/pdf/dot-making-safer-streets.pdf

Pan, G., G. Qi, Z. Wu, D. Zhang, and S. Li. 2013. "Land-Use Classification Using Taxi GPS Traces." *IEEE Transactions on Intelligent Transportation Systems* 14 (1): 113–123. doi:10.1109/TITS.2012.2209201.

Qian, X., and S. V. Ukkusuri. 2015. "Spatial Variation of the Urban Taxi Ridership Using GPS Data." *Applied Geography* 59: 31–42. doi:10.1016/j.apgeog.2015.02.011.

Sayarshad, H. R., and J. Y. J. Chow. 2016. "Survey and Empirical Evaluation of Nonhomogeneous Arrival Process Models with Taxi Data." *Journal of Advanced Transportation* 50 (7): 1275–1294. doi:10.1002/atr.1401.

Sladewski, R. S., A. Keler, and A. Divanis. 2017. "Computing the Least Complex Path for Vehicle Drivers Based on Classified Intersections." In *The 20th AGILE International Conference on Geographic Information Science (AGILE 2017)*. Wageningen, the Netherlands.

Stanica, R., M. Fiore, and F. Malandrino. 2013. "Offloading Floating Car Data." Proceedings of the 2013 IEEE 14th International Symposium and Workshops on a World of Wireless, Mobile and Multimedia Networks (WoWMoM), Madrid, Spain, 1–9.

Tang, J., J. Liang, S. Zhang, H. Huang, and F. Liu. 2018a. "Inferring Driving Trajectories Based on Probabilistic Model from Large Scale Taxi GPS Data." *Physica A* 506: 566–577. doi:10.1016/j.physa.2018.04.073.

Tang, J., F. Liu, Y. Wang, and H. Wang. 2015. "Uncovering Urban Human Mobility from Large Scale Taxi GPS Data." *Physica A* 438: 140–153. doi:10.1016/j.physa.2015.06.032.

Tang, J., Y. Yang, and Y. Qi. 2018. "A Hybrid Algorithm for Urban Transit Schedule Optimization." *Physica A* 512: 745–755. doi:10.1016/j.physa.2018.08.017.

Tang, J., S. Zhang, X. Chen, F. Liu, and Y. Zou. 2018b. "Taxi Trips Distribution Modeling Based on Entropy-Maximizing Theory: A Case Study in Harbin City - China." *Physica A* 493: 430–443. doi:10.1016/j.physa.2017.11.114.

Yang, C. 2015. Data-Driven Modeling of Taxi Trip Demand and Supply in New York City. doi:10.7282/T3NP26FZ.

Yue, Y., Y. Zhuang, Q. Li, and Q. Mao, 2009. "Mining Time-Dependent Attractive Areas and Movement Patterns from Taxi Trajectory Data." In *17th International Conference on Geoinformatics*, August 12–14, 1–6. Fairfax, Virginia, USA.

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.

Zong, F., Y. Tian, Y. He, J. Tang, and J. Lv. 2019. "Trip Destination Prediction Based on Multi-Day GPS Data." *Physica A* 515: 258–269. doi:10.1016/j.physa.2018.09.090.