

Detecting vehicle traffic patterns in urban environments using taxi trajectory intersection points

Andreas Keler, Jukka M. Krisp & Linfang Ding

To cite this article: Andreas Keler, Jukka M. Krisp & Linfang Ding (2017) Detecting vehicle traffic patterns in urban environments using taxi trajectory intersection points, Geo-spatial Information Science, 20:4, 333-344, DOI: [10.1080/10095020.2017.1399672](https://doi.org/10.1080/10095020.2017.1399672)

To link to this article: <https://doi.org/10.1080/10095020.2017.1399672>



© 2017 Wuhan University. Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 22 Nov 2017.



Submit your article to this journal 



Article views: 511



View related articles 



View Crossmark data 

Detecting vehicle traffic patterns in urban environments using taxi trajectory intersection points

Andreas Keler , Jukka M. Krisp and Linfang Ding

Institute of Geography, University of Augsburg, Augsburg, Germany

ABSTRACT

Detecting and describing movement of vehicles in established transportation infrastructures is an important task. It helps to predict periodical traffic patterns for optimizing traffic regulations and extending the functions of established transportation infrastructures. The detection of traffic patterns consists not only of analyses of arrangement patterns of multiple vehicle trajectories, but also of the inspection of the embedded geographical context. In this paper, we introduce a method for intersecting vehicle trajectories and extracting their intersection points for selected rush hours in urban environments. Those vehicle trajectory intersection points (TIP) are frequently visited locations within urban road networks and are subsequently formed into density-connected clusters, which are then represented as polygons. For representing temporal variations of the created polygons, we enrich these with vehicle trajectories of other times of the day and additional road network information. In a case study, we test our approach on massive taxi Floating Car Data (FCD) from Shanghai and road network data from the OpenStreetMap (OSM) project. The first test results show strong correlations with periodical traffic events in Shanghai. Based on these results, we reason out the usefulness of polygons representing frequently visited locations for analyses in urban planning and traffic engineering.

ARTICLE HISTORY

Received 30 October 2016
Accepted 9 September 2017

KEYWORDS

Floating Car Data (FCD); moving objects; transportation infrastructure; spatio-temporal patterns

1. Introduction

Tracked movement of objects is nowadays widely available and used for various applications in our society (Dodge et al. 2016). Detailed vehicle movement for example can benefit the prediction of short-term and long-term traffic situations.

One important domain of movement analysis is traffic and transportation. Research outcomes on the nature of traffic and transportation are important for most of the world's population. Understanding traffic and how traffic congestion or gridlocks appear and propagate in time and space is a task that has high importance to future generations. Besides tremendous emissions caused by traffic in air, land, and water, there is a gigantic loss of time and money every day in the world due to vehicle traffic congestion.

Moving object data-sets coming from mobile sensor devices such as Global Navigational Satellite Systems (GNSS), Global System for Mobile Communications (GSM), and Wireless Local Area Network (WLAN) are nowadays frequently discussed and analyzed. In case the moving objects are vehicles acquired with GNSS receivers, we talk about Floating Car Data (FCD). Due to its already available communication infrastructure for monitoring the vehicles, taxi fleets of urban environments deliver massive FCD collections.

In general, movement analysis includes detection of the first- and second-order effects (O'Sullivan and Unwin 2010). The first-order effects respect the context of movement, such as the visited environment, the mode of transportation, or any other domain specific information (Gschwend and Laube 2012). The second-order effects consist of detecting interactions of multiple entities, such as meeting points, convoys, and flocks (Gudmundsson, Laube, and Wolle 2012). These interactions are helpful for associating vehicle movement trajectories with traffic congestion events or stoplight dynamics, in case of analyzing FCD. The methods for detecting the second-order effects of movement include the usage of similarity measures, which can be temporal, spatial, or spatio-temporal (Ranacher and Tzavella 2014).

Our idea is to define specific taxi trajectory arrangement patterns by detecting areas with high vehicle flow rates. After defining these areas, we give an indication of an implied geographical context, as for example the function and grade of complexity of parts or the transportation infrastructure. The first step in this approach is to visually inspect taxi trajectories of selected time windows as intersected polylines. Figure 1 shows the idea of visual inspection with polyline layer ordering

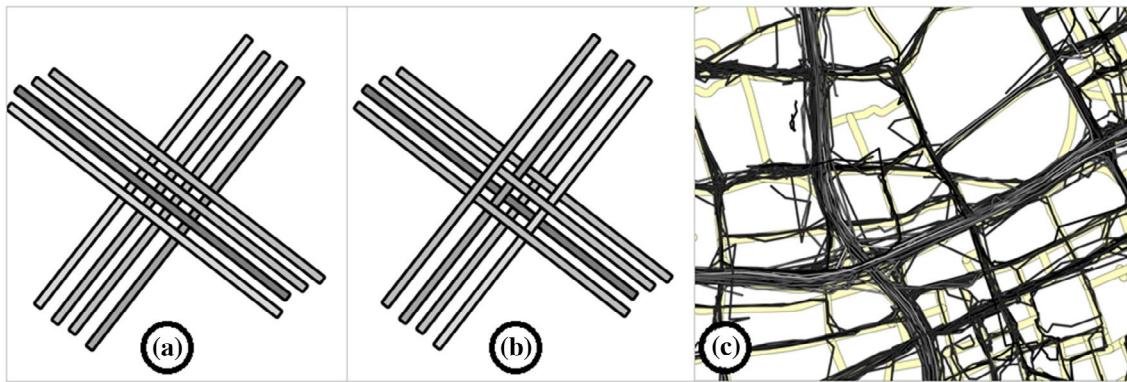


Figure 1. Imagined examples of trajectory polyline intersections with the example of (a) a detected road intersection with traffic lights, and of (b) elevated intersections, in comparison with (c) a trajectory polyline intersection patterns from real taxi trajectory data.

within a geographic information system (GIS). This approach bases on a previous study in Keler, Krisp, and Ding (2017). The ordering in Figure 1(a) and (b) might base on the starting time of each taxi trajectory or the total length of the polyline segments. Afterward, these patterns of specific polyline intersections can serve for extracting polyline intersection points. This has the idea to detect: (1) the varying point densities and (2) time and velocity differences in selected investigation areas.

Additionally, we detect self-intersections of the entities and provide statistics on what vehicles intersect each other. The last mentioned information may provide information on vehicle driving behavior and specific transportation infrastructures that are passed by the drivers. After defining locations that are frequently visited by taxis, we enrich those with information on the transportation infrastructure. The idea behind this is to inspect correlations between transportation infrastructure complexity and frequencies of passing taxis for the same locations. This reasoning is visually presented in Figure 1, where the trajectory polyline layers are ordered by starting times of the vehicle trips. This might result in patterns for road intersections with traffic lights (Figure 1a), and elevated intersections (Figure 1b), depending on the different intersection patterns. Figure 1(c) shows one example with a selected part of the transportation infrastructure in Shanghai.

High numbers of vehicle trajectories result in visual representations with high cluttering effects as in Figure 1(c). One idea to avoid these appearances is to extract only the intersection points of the polyline representations. Those points should imply additional information on the trajectory intersection, namely time and velocity differences.

2. Methods for analyzing traffic and mobility based on FCD

When analyzing traffic data, we have to differentiate between different data types and different research domains with different established methods. The

methods in the domain of analysing traffic and mobility will be our focus.

The analysis techniques are connected with vehicle trajectory representations, the representations of road networks, and how to match these two sources via algorithms. Another group of approaches include map inference techniques, which allow the generation of road segments based on FCD.

Analysing traffic data may consist of various approaches. Based on Chen, Guo, and Wang (2015), these approaches mainly include the preprocessing of data, derivation of patterns by various analysis methods and their representation. The way of traffic data analysis is always dependent on the underlying data-sets. Wang et al. (2014) named three different data input classes for traffic data analysis: static traffic sensors, mobile devices, and merged solutions (includes both classes). The most common form of traffic data is trajectory (Chen, Guo, and Wang 2015), which often represents movement of a concrete object. The great advantage of trajectories (derived from mobile devices) is the possibility of unbiased representations of the traffic density (Treiber and Kesting 2013).

One common challenge is the efficient handing FCD and other data from mobile devices due to its often massive size. Chen et al. (2016) solved this challenge using a compressed linear reference (CLR) technique for transforming network time geographic entities from spatiotemporal space (x, y, t) into 2D space. The outcome of the approach allows more feasible handling of the massive movement data in a classical spatial database. The subsequent step would be to define movement patterns that characterize urban traffic situations as urban traffic congestion. Bertini (2005) reviews the definition and measurement of urban traffic congestion using the results of a questionnaire for transportation experts. The answers are very differing, especially in the measurement of urban traffic congestion. In general, the groups of measuring delay (29%), level of service (20%), and volume over capacity (14%) are the biggest. Bertini (2005)

shows with this result that instead of highway traffic congestion measuring urban traffic congestion has many different possibilities. Besides describing movement in geographic space, which may be a two- or three-dimensional Euclidean movement space (Gudmundsson, Laube, and Wolle 2012; Wang et al. 2015a), we have the option to inspect vehicle movement in the network space (Ji, Luo, and Geroliminis 2014; Lan et al. 2014). Using average information of graphs with arcs and nodes, it is possible to detect congestion propagation and bottleneck identification in a computationally efficient way (Wang et al. 2015b). This allows modeling the street network for the car driver domain with weighting of road segments based on derived travel time or restrictions such as road closure, accidents or traffic congestion. Other approaches include the interactive selections based on road segments for deriving traffic congestion durations using traffic jam propagation graphs (Wang et al. 2013). We can enrich the network space, which consists of arcs and nodes, with real-time information or with averaged day-wise traffic information for the detection of traffic anomalies (Lan et al. 2014). The enriching process of road segments (arcs) with traffic data from mobile tracking devices is in most cases connected with an adapted Map Matching (MM) algorithm. MM is the task of connecting FCD with digital road segments and has become a frequently used group of methods with over 36 different algorithms in the year 2006 (Quddus, Ochieng, and Noland 2007; Zhao et al. 2012). Besides matching quality, the computational efficiency is important in designing MM algorithms. One difficult condition for MM in some urban road networks are elevated roads and on-ramps between multiple elevation levels (Li et al. 2007). This requires intelligent solutions for the case of no available height components within FCD records. Besides the mentioned MM approaches, there should be a decisive differentiation with those approaches handling low-frequency FCD, as in the case of the case study in this work. In general, the higher indistinctness occurs when sampling interval are higher within FCD set. Chen et al. (2014) focused on this challenge with the introduction of a new approach that can achieve accuracy and computational performance comparable to those MM algorithms developed for in-vehicle navigation.

Another group of methods includes movement aggregation by introducing irregular tessellation of space (Gudmundsson, Laube, and Wolle 2012), which may be based on FCD clustering. For the case of taxi FCD, often pick-up and drop-off points of individual taxis are clustered for detecting spatio-temporal hotspots. Besides partitioning clustering as k -means (Krisp et al. 2012), the most frequent clustering approaches include density-based methods (Rinzivillo et al. 2008; Yue et al. 2009; Pan et al. 2013) for taxi pick-up and drop-off points. When clustering all available FCD records or only those with certain extracted spatio-temporal pattern like low

velocities (Andrienko et al. 2011) different FCD constellations (Körner 2011) are possible such as points, lines, or polygons.

Instead of Andrienko et al. (2011), we inspect clusters of trajectory intersection points (TIPs) and not velocity-based clusters. Point clusters of low vehicle velocity are in general indicators for traffic congestion events in selected times of the day. Our idea is to extract trajectory intersections as an indicator of ongoing movement during locally known rush hours and thus represent the most efficient transportation infrastructure elements.

Based on FCD aggregation, we can create Thiessen polygons (Andrienko and Andrienko 2010) or introduce spatial grid cells (Andrienko and Andrienko 2007). For these features, which were also called summation places (Rinzivillo et al. 2008), we can introduce different average parameter classifications.

Another connected topic is road map inference from FCD. Several groups of methods are proposed for constructing road networks based on vehicle trajectories. For Duruisseau and Rouvoy (2014), there are three main groups of approaches for map inference, which are K -Means clustering, trace merging, and kernel density estimation. Within this research, one challenge is to detect road intersections, especially when they are missing in available road network data (Wang, Wei, and Forman 2013). One indicator for detecting road intersections can be direction changes in the vehicle trajectory, which is for Xie et al. (2015) not always helpful when working with real data. For inferring road maps from sparsely sampled trajectories Qiu, Wang, and Wang (2014) propose the use of density-based point clustering. Map inference often consists of numerous steps as an inference pipeline as in the case of Biagioni and Eriksson (2012) with five methods that are used subsequently.

3. Test taxi floating car data (FCD) and data preprocessing

In the following, we give an overview on our test data-sets and their respective preprocessing steps for the later analysis. Besides taxi Floating Car Data (FCD) sets, we include road network data from the OSM project in our analysis.

3.1. Taxi floating car data (FCD) from Shanghai

We inspect a taxi FCD set resulting from a survey (GPS) on a taxi fleet in Shanghai ("SUVnet-Trace Data").¹ There is a varying number of inspected taxis ranging time-dependent from around 7000 to around 10,000. The reason for this time-dependency might be caused by the taxi driver's behavior to switch off and turn on their respective tracking devices. Depending on the time of day some of the taxi drivers turn their tracking device off and some new appearing turn it on. The data structure of the inspected data-set is shown in Table 1.

Besides the selected attributes (car ID, longitude, latitude, time, and instantaneous velocity in Table 1), there are four further original attributes that are not used in our approach. In our testing, we use in our study a taxi FCD partition of one selected day (Thursday the 22 February 2007) from originally 15 days of acquisition between the 1 February and 1 March 2007. Each day of taxi FCD has over six million records. Throughout the data, there are differing sampling intervals in time, which in some cases results in time jumps between consecutive points. After detecting bigger time jumps to the mentioned operational status of the taxi tracking device by setting spatial buffers, we can inspect temporal jumps during continuous acquisition. The variation of these temporal jumps is between 1 and 30 s and has an average of around 14 s for all the data partition of the inspected day.

3.2. Derivation of taxi trajectory intersection points

Based on the explained taxi FCD, we first create individual trajectories for each entity as polyline representations and then extract the intersection points of all trajectory polylines. Figure 2 shows a workflow diagram of this process.

The calculation step of trajectory generation (see left part of Figure 2) results in taxi trajectory polylines for selected time windows as linear interpolation between movement positions of every taxi driver. Those trajectory polylines can be visualized as in Figure 3 by coloring every polyline based on the underlying total length. Brighter colors show longer polylines and indicate higher velocities. By inspecting the selected intersection

patterns in Figure 3, we can assign different types of transportation infrastructure elements with the ones in aerial imagery that are visually detectable.

We define every TIP by intersecting, respectively, two trajectories. Additionally, we calculate additional attributes for every taxi TIP based on the two respective trajectories. These attributes include interpolated time difference, average velocity difference as pictured in Figure 4. Besides this, it is possible to distinguish an intersection between two trajectories or self-intersection by comparing if the vehicle identifications are differing. The additional attribute travel time difference Δt results from created vectors between every consecutive point of individual trajectories. Subsequently, these vectors are enriched with averaged time stamps based on the time stamps of every two consecutive points. Our way of FCD inspection bases on Körner (2011) and his four constellation possibilities for FCD analysis: point, vectors, trajectories, and bundles of trajectories. Our method aims to extract point representations from the last mentioned and to compute differences in averaged attribute values.

The created TIPs are the base for our approach of defining locations with frequent taxi movement during rush hours. Every intersection point is created by intersecting every two trajectories as in Figure 4(a). The computed attributes are two different types of temporal information, namely $\Delta\bar{t}_v$, $\Delta\bar{t}$ and \bar{t} . Each of these attributes is assigned to line segments of the trajectory polylines. The attribute values that are assigned to TIP are the differences between intersecting line segments. Figure 4(a) shows the way of how average velocities differences $\Delta\bar{v}$ are computed based on averaged instantaneous measures \bar{v} in every movement position. Figure 4(b) consists of computing travel time based on \bar{v} multiplied by the distance between two movement positions. Figure 4(c) show how an average time stamp \bar{t} is estimated based on time stamps t .

3.3. OSM road network of Shanghai and extracted street nodes

According to Stanica, Fiore, and Malandrino (2013), the OSM project has one of the most accurate freely available digitized road networks. Additionally it has, depending

Fieldname	Details
Car ID	The unique ID of the car, in 5 digits
Longitude	In degree ($^{\circ}$); accurate to the 6th decimal place
Latitude	In degree ($^{\circ}$); accurate to the 6th decimal place
Instantaneous velocity	Accurate to 0.1 km/h
Record date	In form of YYYY-MM-DD
Record time	In form of hh:mm:ss

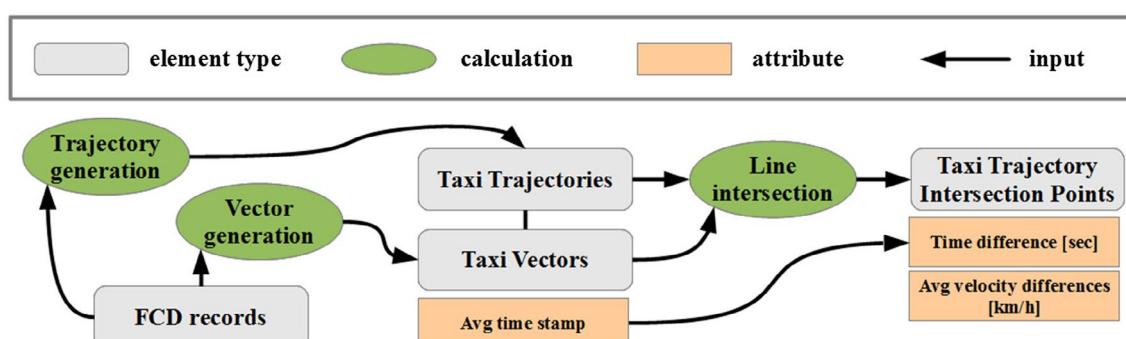


Figure 2. Workflow for creating taxi TIPs.

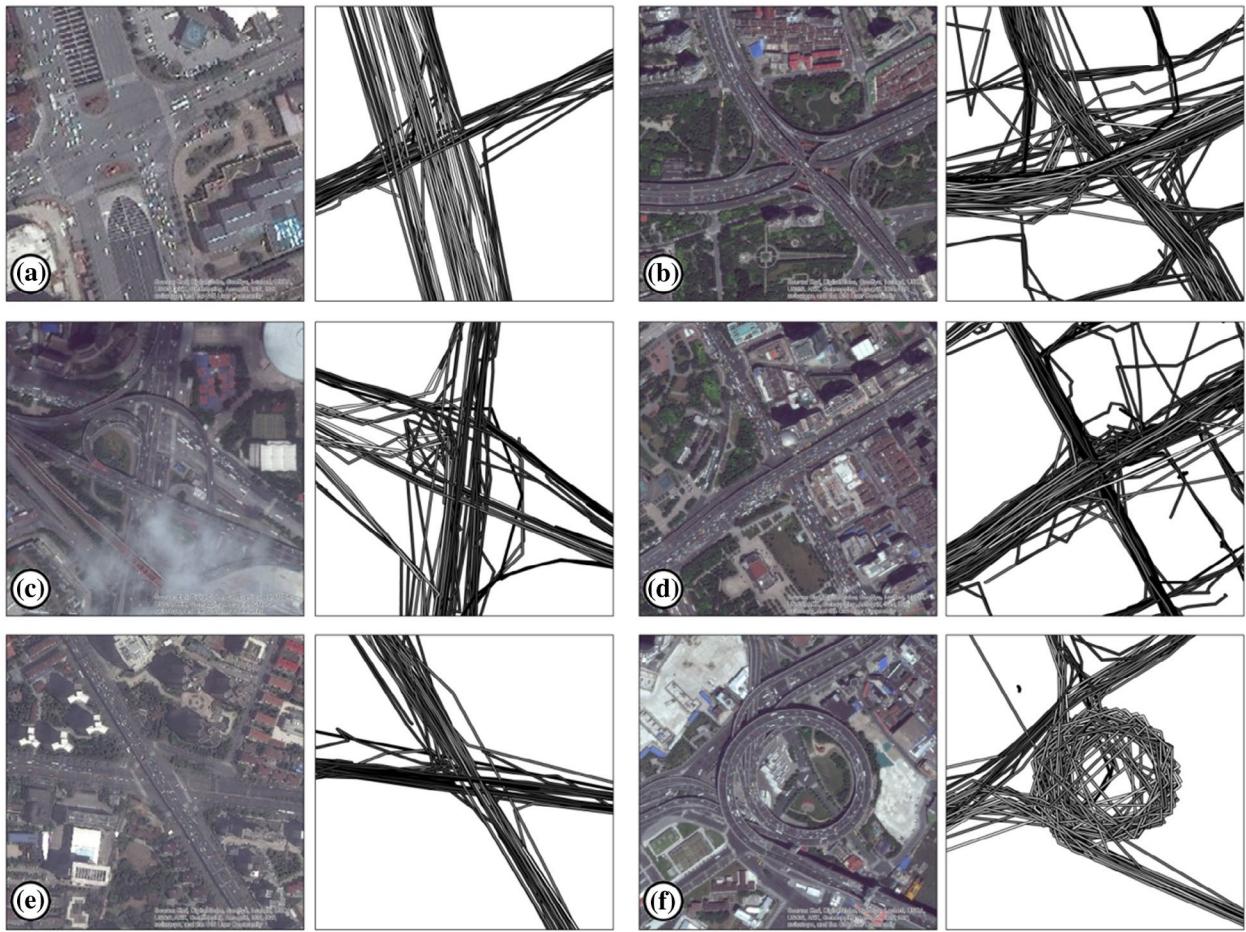


Figure 3. Different visual patterns of taxi trajectory polylines at selected infrastructure elements, such as (a) crossroads, (b) crossing with changes in elevation, (c) on-ramps, (d) elevated highway, (e) elevated highway, and, (f) Nanpu bridge on-ramp.

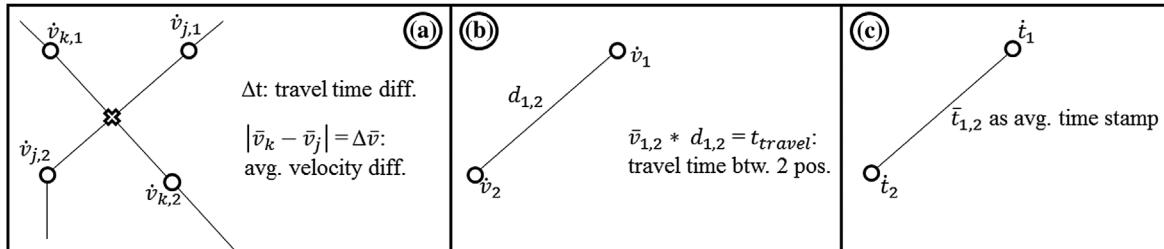


Figure 4. Attribute computation for trajectory polylines with (a) travel time based on average velocity differences, (b) travel time based on distances, and (c) travel time as averaged time stamps.

on the investigation area, a relatively high quality of road connectivity as it was evaluated by Graser, Straub, and Dragaschnig (2014) for vehicle routing quality in Vienna. In our approach, we used only a derivation of the original OSM road network, which is represented by polylines. Similar to Krisp and Keler (2015), we propose the abstraction of high node density for complicated crossings. The idea is to compare the frequency of taxi visits in crossings and in other road segments with its defined complexity. Therefore, we extracted the nodes of all available road segments (OSM class “highway”) within the administrative borders of Shanghai, except the pathways of pedestrians. Instead of the approach in

Krisp and Keler (2015), we would like not only to calculate the complexity of crossings, but as well respect other parts of the road network. We expect a correlation of high node densities in road crossings with frequent trajectory intersections.

4. Traffic analysis based on taxi TIP

We introduce the concept of analysing traffic with TIP. Individual taxi drivers are only a small portion of all traffic participants in urban environments. Nevertheless, we argue that the detected movement patterns are representative for the entire traffic situation within a city,

since there is a high coverage of the used taxi FCD on the urban road network of Shanghai, especially during rush hours. In the following, we present examples of the methodological steps of how TIP are processed and represented using test FCD sets.

4.1. Density-based clustering of taxi TIP

As we are interested in frequently visited areas during rush hours, we extract one selected rush hour from our test data: from 8:00 to 9:00 am. Within this time window, we have in total 283,141 records from which we calculate 4243 valid taxi trajectories (Figure 5a). After the taxi TIP extraction step (see Figure 3), we finally gained 15,741 TIP (Figure 5b). Additionally, we preselected self-intersecting trajectories (Figure 5c) and their extracted TIP (Figure 5d).

Places with high taxi intersection point densities indicate elevated transportation infrastructure elements (high taxi TIP density, high speed differences) or frequently visited crossings not influenced by traffic congestion (high TIP density, low speed differences). Following the last mentioned case, we used the typical rush hours of the selected investigation area in Shanghai for detecting taxi TIP densities. These densities were grouped by the density-based clustering method OPTICS (Ordering

Points To Identify Clustering Structure), which was introduced by Ankerst et al. (1999).

This algorithm connects points to clusters by calculated point densities resulting from the settings of search distance (Epsilon) and minimum point number (MinPts). After several tests, we introduced a search distance and minimum point number based on the aim of achieving more than 1000 TIP clusters. Therefore, we set the search distance to 150 m (Epsilon = 150 m) and the number of respected points to 4 (MinPts = 4). As a result, we define 1231 taxi TIP clusters for the time from 8:00 to 9:00 am on Thursday the 22 February 2007 in Shanghai.

4.2. Creation of polygons and subsequent enrichment with traffic and road information

From the 1231 defined taxi TIP clusters, we generate 1231 taxi TIP polygons (convex hulls) using the gift wrapping algorithm, which is also called Jarvis March (Jarvis 1973). This part of the workflow (see Figure 6) is a generalization step for facilitating further analyses. As stated in Section 3.2, we can detect correlations between complex road crossings and taxi TIP cluster polygons by visual inspection. Nevertheless, there are some exceptions due to the elevated roads in Shanghai.



Figure 5. Visualization of (a) used taxi trajectory polylines, (b) resulting TIP extracts, (c) trajectory polylines with self-intersections, and (d) resulting self-intersection point extracts.

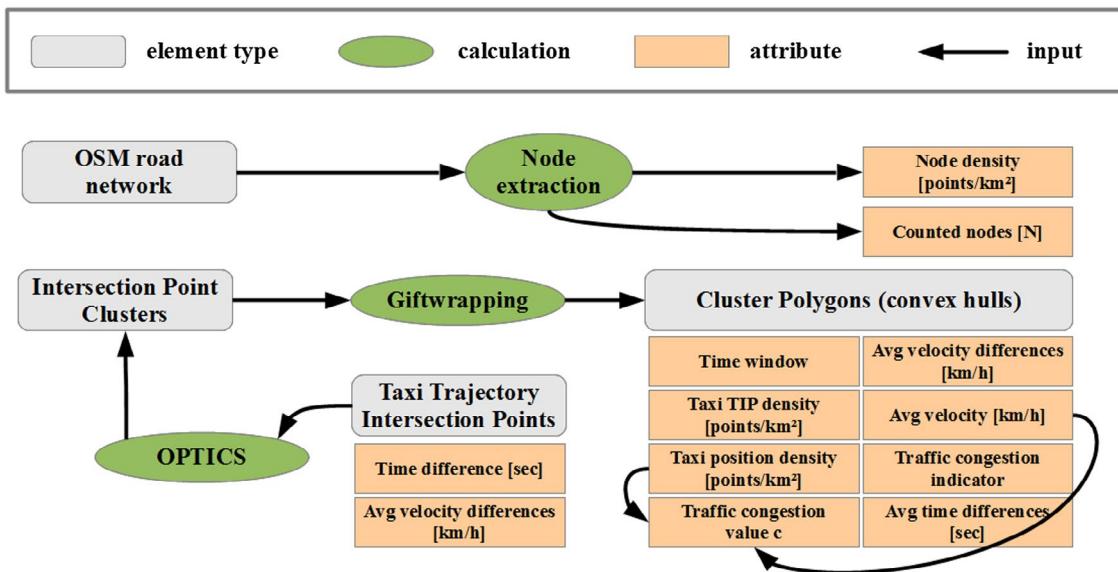


Figure 6. Workflow for creating and enriching taxi TIP polygons.

These taxi TIP cluster polygons are caused by frequent vehicle movement on both elevation levels.

After the definition of the shape and size of taxi TIP cluster polygons, we enrich them with average parameters calculated from taxi FCD partitions of one selected day (Thursday, 22 February 2007). Therefore, we selected different time windows and calculate average parameter information for each of the 1231 taxi TIP polygons. We enriched all polygons with average information of selected times of the day with the following average parameters (see Figures 4 and 6): taxi TIP density, taxi position density, average velocity, average velocity differences, average time differences, traffic congestion indicator, and traffic congestion value c .

For a visual representation of the taxi TIP polygons in Figure 7, we select the three parameters taxi density, average velocity, and congestion value c for certain hours of a working day. Those are hours in Shanghai that are typically influenced by high congestion events based on statistics from Sun et al. (2009): from 7:00 to 8:00 am and from 5:00 to 6:00 pm.

With the result in Figure 7, we have the historical average values for a typical working day. For reasons of comparison, we also use this procedure for the other hours of the working day.

Afterward, the 1231 created polygons are extended by information on the complexity of transportation infrastructure. As mentioned in Section 3.3 and pictured in Figure 6, we simply count selected OSM road network nodes within taxi TIP polygons. The additional attributes of taxi TIP polygons are the counted OSM road network nodes and the road network node density in each taxi TIP polygon. In our approach, the complexity of a transportation infrastructure element is dependent on the amount of its extracted nodes. This means that road partitions with multiple road lanes are more complex

than road partitions with a single lane. Crossings with on-ramps include more nodes than most of the other crossings with no on-ramps, and so on. The idea behind this definition is to inspect the proportion of the number of road network segment intersections. For the verification, this idea it is possible to inspect the density of taxi positions in selected areas.

An overview on the level of complexity is visualized in Figure 8 by varying coloration of the trajectory intersection polygons showing classes of counted road network nodes and the derived node density.

4.3. Analyzing time series of traffic information in defined cluster polygons

The created and enriched polygons have time-dependent average parameter values, which can be represented as time series. Time series are often visualized with interactive elements for visual analysis. One common solution for the visualization of temporal data is the use of time slider tools. Instead of these examples, we use slider tools for polygons and not for raster data or trajectory points. Following, our variation in space do not appear, but only the coloration of spatially fixed areas (polygons) varies in space. Furthermore, we can match the differing time-dependent attribute values for each polygon in the two-dimensional Euclidean space with road segments (arcs) and connection points (nodes) in the network space. The preferred way of inspection in this paper is the creation of attribute tables for selected taxi TIP cluster polygons without the connection to road segments.

5. Results and discussion

In the following, we apply the previous methodology for a data partition and inspect the attribute values for extracted taxi TIP cluster polygon for selected rush

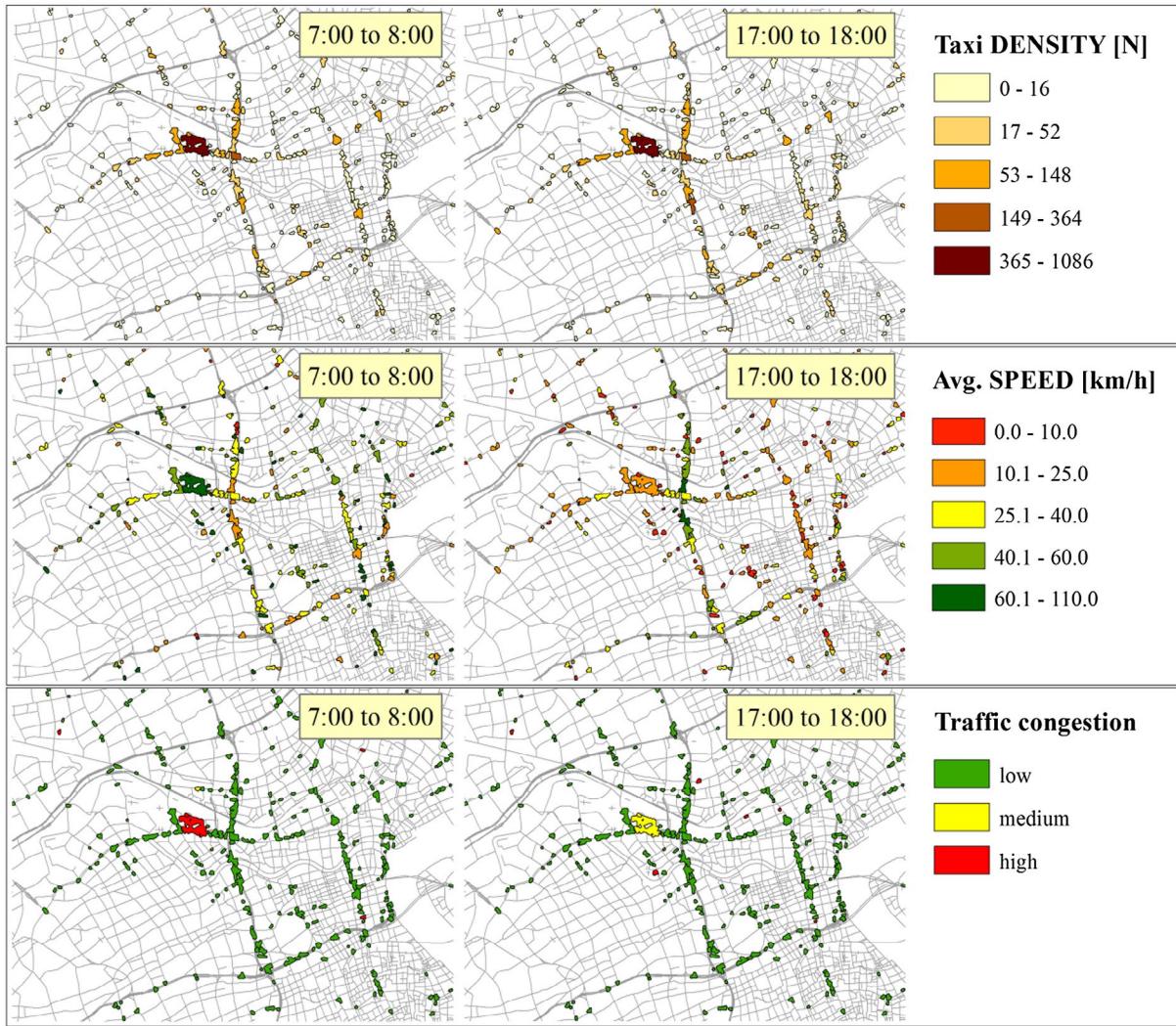


Figure 7. Comparison between the three quantities taxi density, avg. speed, and traffic congestion value c , respectively, for two selected time windows.

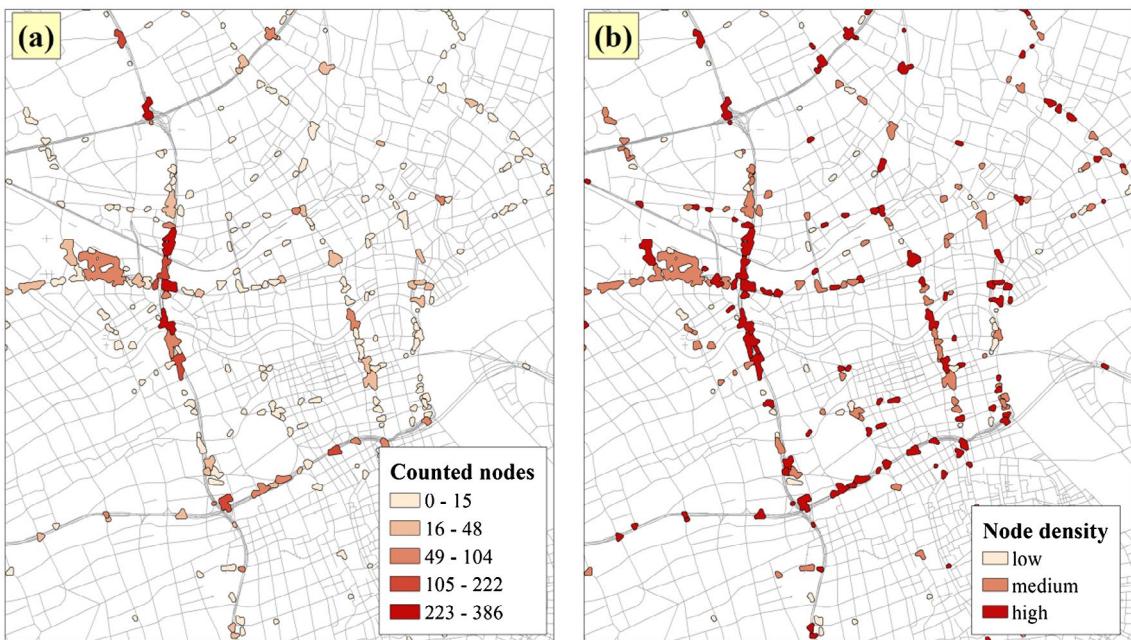


Figure 8. Level of complexity of transportation infrastructure elements for the center area of Shanghai, with (a) counted road network nodes and (b) calculated node density.

hours. Afterwards, we select one taxi TIP cluster polygon and enrich it with data from off-peak hours for 12 consecutive hours in Shanghai.

5.1. Inspection of enriched taxi TIP polygons (in space) for selected rush hour times

In Figure 7, we compare the enriched taxi TIP polygons with the quantities taxi density, average speed and traffic congestion value c for two selected times of a working day with typical rush hours in Shanghai (Sun et al. 2009). Figure 7 shows that the taxi density in both times of the day has low variations in the central part of Shanghai, especially in polygons with bigger surface areas. In contrast to this insight, there are great variations in average speed values and traffic congestion value c . This might indicate different traffic situations in the same location but on different elevation levels, as for example traffic congestion on the ground and free flowing traffic on an above elevated highway segment.

After comparing the detected density-based clusters with OSM street node densities, we may give answer on a more or less positive correlation. In Figure 8, it is detectable that the surface area of the trajectory intersection polygons is often not bindingly important for detecting complex infrastructure elements. It is often the case that simply the number of counted nodes (see Figure 8 a) shows more details than an overview about the node density distribution (see Figure 8 b).

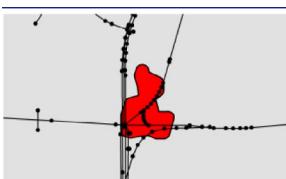
5.2. Inspection of enriched taxi TIP polygons for half a day

As an example, we select one TIP polygon with high taxi TIP densities during rush hours (Figure 9). Figure 9 shows an area in the northern part of Shanghai, where within the taxi TIP polygon at least one elevated highway segment is connected with multiple lower order road segments. Since this defined location has high taxi flow rates during rush hours, which is our indication resulting



Figure 9. Selection of TIP polygon in the north of Shanghai based on taxi FCD from 8:00 to 9:00 am on 22 February 2007.

Table 2. Distribution of the average traffic information in one selected area in Shanghai from 8:00 am to 8:00 pm (Thursday, 22 February 2007).

Selected location ^a	Time window	Counted taxi TIP	Counted taxi positions	$\Delta\bar{v}$ (km/h)	\bar{v} (km/h)	Traffic congestion value c
	8:00–10:00	1178	236	20.6	21.4	Heavy
	10:00–12:00	510	156	20.2	30.0	Mediocre
	12:00–14:00	483	144	18.1	25.3	Mediocre
	14:00–16:00	195	126	24.5	24.9	Mediocre
	16:00–18:00	52	68	29.9	28.6	Low
	18:00–20:00	55	86	30.1	35.3	Low
	(OSM nodes = 65; OSM node density = high)					

^a<http://sh.edushi.com/>.

from high taxi TIP densities; we inspect this polygon by its attribute value variations over time. Therefore, Table 2 shows the distribution of average traffic information for the time between 8:00 am and 8:00 pm on Thursday, 22 February 2007.

It is to mention that OSM road network node density of the selected taxi TIP polygon is high and includes in total 65 nodes. By inspecting Table 2, it is comprehensible that rush hours with heavy traffic congestion appear in the morning. From 8:00 to 10:00 am, there is the largest number of taxi positions and of taxi TIP. Whereas the average velocity \bar{v} in this time window is the lowest within 12 h of inspection with 21.4 km/h. Surprisingly, there is no traffic congestion in the evening, where also the lowest number of taxi TIP and taxi positions appear with higher average velocities. In general, the velocity differences $\Delta\bar{v}$ for all time windows show high values. This is indicating nearby road segments with different order and recommended speeds. Indeed this indication is compared using the 3D view of the digital mapping service Edushi, which is a component of the Baidu Maps² service. The service shows a complex road intersection with an intersecting elevated highway with an on-ramp. This can be the reason for the high average velocity differences $\Delta\bar{v}$.

6. Conclusions and outlook

The presented method in this work shows a simple way to detect specific traffic patterns based on taxi FCD. The extraction of polygons is based on the density of taxi TIP. This procedure may benefit spatial analysis and traffic visualization tasks of analysts in the fields of urban planning and traffic engineering. The first test results show that it is possible to gain new insights into the traffic situations of specific times of the day. Differences in computed average velocities $\Delta\bar{v}$ indicate densely ordered road segments of different types.

The approach might be tested for other taxi FCD sets, possibly with higher acquisition frequencies and a higher number of observed entities. Transportation infrastructures as on-ramps are associated with grade changes (as from highway to minor roads), can indicate static traffic bottlenecks (Wang et al. 2017). For a possible detection of these traffic bottlenecks, a transportation infrastructure classification scheme for different TIP densities could be tested. One challenge in these applications is to distinguish between different types of transportation infrastructure (as visualized in Figure 2) in an automated way by only using TIP extracts. Another direction for future work would be to test TIP extracts for improving established map matching techniques, especially for cases of low-frequency FCD. TIP extracts and polygons may facilitate matching FCD records on road intersections in providing knowledge on different road types together with assigned traffic states that are typical for the selected time of the day.

Notes

1. “SUVnet-Trace Data” was obtained from the Wireless and Sensor networks Lab (WnSN) at Shanghai Jiao Tong University, http://wirelesslab.sjtu.edu.cn/taxi_trace_data.html.
2. <http://map.baidu.com/>

Acknowledgments

The described taxi Floating Car Data set of Shanghai (SUVnet-Trace Data: http://wirelesslab.sjtu.edu.cn/taxi_trace_data.html) was obtained from the Wireless and Sensor networks Lab (WnSN) at Shanghai Jiao Tong University. We would like to thank the Laboratory for Wireless and Sensor Networks at Shanghai Jiao Tong University, especially Prof. Min-You Wu and Jia Peng, for providing access to this data.

Notes on contributors

Andreas Keler, PhD, is a postdoctoral researcher at the Technical University of Munich (TUM) and was a scientific staff member in Applied Geoinformatics (Institute of Geography) at the University of Augsburg. His research interests include the analysis and visualization of traffic congestion, air quality and massive movement data, routing applications for vehicle drivers and Location-Based Services (LBSs).

Jukka M. Krisp, PhD, is a 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 Chair 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

- Andrienko, N., and G. Andrienko. 2007. “Designing Visual Analytics Methods for Massive Collections of Movement Data.” *Cartographica: The International Journal for Geographic Information and Geovisualization* 42 (2): 117–138.
- Andrienko, N., and G. Andrienko. 2010. “Spatial Generalization and Aggregation of Massive Movement Data.” *IEEE Transactions on Visualization and Computer Graphics* 17 (2): 205–219.
- Andrienko, G., N. Andrienko, C. Hurter, S. Rinzivillo, and S. Wrobel. 2011. “From Movement Tracks through Events to Places: Extracting and Characterizing Significant Places from Mobility Data.” Paper presented at the IEEE Conference on Visual Analytics Science and Technology (VAST 2011), Providence, Rhode Island, USA, October 23–28, 161–170.
- Ankerst, M., M. M. Breunig, H.-P. Sander, and J. Kriegel. 1999. “OPTICS: Ordering Points to Identify the Clustering Structure.” Paper presented at ACM SIGMOD

- International Conference on Management of Data. Philadelphia, Pennsylvania, USA, May 31–June 3, 49–60.
- Bertini, R. L. 2005. "Congestion and Its Extent." Chap. 2 in *Access to Destinations*, edited by D. M. Levinson and K. J. Krizek, 11–38. Bingley: Emerald Group Publishing.
- Biagioli, J., and J. Eriksson. 2012. "Map Inference in the Face of Noise and Disparity." Paper presented at Proceedings of the 20th International Conference on Advances in Geographic Information Systems (SIGSPATIAL'12), Redondo Beach, California, November 6–9, 79–88.
- Chen, B. Y., H. Yuan, Q. Q. Li, W. H. K. Lam, S.-L. Shaw, and K. Yan. 2014. "Map Matching Algorithm for Large-scale Low-frequency Floating Car Data." *International Journal of Geographical Information Science* 28: 22–38.
- Chen, W., F. Guo, and F.-Y. Wang. 2015. "A Survey of Traffic Data Visualization." *IEEE Transactions on Intelligent Transportation Systems* 16 (6): 2970–2984.
- Chen, B. Y., H. Yuan, Q. Q. Li, S.-L. Shaw, W. H. K. Lam, and X. Chen. 2016. "Spatiotemporal Data Model for Network Time Geographic Analysis in the Era of Big Data." *International Journal of Geographical Information Science* 30: 1041–1071.
- Dodge, S., R. Weibel, S. C. Ahearn, M. Buchin, and J. A. Miller. 2016. "Analysis of Movement Data." *International Journal of Geographical Information Science* 30: 825–834. doi:10.1080/13658816.2015.1132424.
- Duruisseau, M., and R. Rouvoy. 2014. *Automatic Inference of Roadmaps from Raw Mobility Traces*. Research Report, RR-8585, INRIA. 19.
- Graser, A., M. Straub, and M. Dragaschnig. 2014. "Is OSM Good Enough for Vehicle Routing? A Study Comparing Street Networks in Vienna." In *Progress in Location-based Services. Lecture Notes in Geoinformation and Cartography*, edited by G. Gartner and H. Huang, 3–18. Berlin: Springer.
- Gschwend, C., and P. Laube. 2012. "Challenges of Context-aware Movement Analysis – Lessons Learned about Crucial Data Requirements and Pre-processing." Paper presented at the GIS Research UK 20th Annual Conference, Lancaster, April 11–13, 241–246.
- Gudmundsson, J., P. Laube, and T. Wolle. 2012. "Computational Movement Analysis." In *Springer Handbook of Geographic Information*, edited by W. Kresse and D. Danko, 423–438. Berlin: Springer.
- 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.
- Ji, Y., J. Luo, and N. Geroliminis. 2014. "Empirical Observations of Congestion Propagation and Dynamic Partitioning with Probe Data for Large-scale Systems." *Transportation Record* 2422 (2): 1–11.
- Keler, A.J. M. Krisp, and L. Ding. 2017. "Detecting Travel Time Variations in Urban Road Networks by Taxi Trajectory Intersections." Paper presented at GISRUK 2017 – 25th GIS Research UK Conference, Manchester, UK, April 18–21.
- Körner, M. 2011. "Nutzungsmöglichkeiten von Floating Car Data zur Verkehrsflussoptimierung." [Possible Uses of Floating Car Data for Traffic Flow Optimization.] In *Angewandte Geoinformatik 2011. Beiträge Zum 23. AGIT-Symposium Salzburg*, edited by J. Strobl, T. Blaschke and G. Griesebner, 381–386. Berlin/Offenbach: Wichmann.
- 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.
- 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." Paper presented at the 6th International Symposium "Networks for Mobility", Okinawa, Japan, September 27–28.
- Lan, J., C. Long, R. C.-W. Wong, Y. Chen, Y. Fu, D. Guo, S. Liu, Y. Ge, Y. Zhou, and J. Li. 2014. "A New Framework for Traffic Anomaly Detection." Paper presented at the 2014 SIAM International Conference on DATA MINING, Pennsylvania, USA, April 24–26, 875–883.
- Li, X., M. Li, W. Shu, and M. Wu. 2007. "A Practical Map-matching Algorithm for GPS-based Vehicular Networks in Shanghai Urban Area." Paper presented at the IET Conference on Wireless, Mobile and Sensor Networks (CCWMSN07), Shanghai, December 12–14, 454–457.
- O'Sullivan, D., and D. J. Unwin. 2010. "Geographic Information Analysis and Spatial Data." In *Geographic Information Analysis*, 1–32. Hoboken: John Wiley & Sons.
- 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.
- Qiu, J., R. Wang, and X. Wang. 2014. "Inferring Road Maps from Sparsely-sampled GPS Traces." In *Advances in Artificial Intelligence—AI 2014*. Vol. 8436 *Of Lecture Notes in Computer Science*, edited by M. Sokolova and P. van Beek, 339–344. Berlin: Springer.
- Quddus, M. A., W. Y. Ochieng, and R. B. Noland. 2007. "Current Map-matching Algorithms for Transport Applications: State-of-the Art and Future Research Directions." *Transportation Research Part C: Emerging Technologies* 15 (5): 312–328.
- Ranacher, P., and K. Tzavella. 2014. "How to Compare Movement? A Review of Physical Movement Similarity Measures in Geographic Information Science and beyond." *Cartography and Geographic Information Science* 41 (3): 286–307.
- Rinzivillo, S., D. Pedreschi, M. Nanni, F. Giannotti, N. Andrienko, and G. Andrienko. 2008. "Visually Driven Analysis of Movement Data by Progressive Clustering." *Information Visualization* 7 (3–4): 225–239.
- Stanica, R., M. Fiore, and F. Malandrino. 2013. "Offloading Floating Car Data." Paper presented at the 2013 IEEE 14th International Symposium and Workshops on a World of Wireless, Mobile and Multimedia Networks (WoWMoM), Madrid, Spain, June 4–7.
- Sun, J., H. Wen, Y. Gao, and Z. Hu. 2009. "Metropolitan Congestion Performance Measures Based on Mass Floating Car Data." Paper presented at the Second International Joint Conference on Computational Sciences and Optimization (CSO 2009), Computational Transportation Science, Sanya, China, April 24–26, 109–113.
- Treiber, M., and A. Kesting. 2013. "Trajectory and Floating-car Data." In *Traffic Flow Dynamics. Data, Models and Simulation*, 7–12. Berlin: Springer.
- Wang, Z., M. Lu, X. Yuan, J. Zhang, and H. Van De Wetering. 2013. "Visual Traffic Jam Analysis Based on Trajectory Data." *IEEE Transactions on Visualization and Computer Graphics* 19 (12): 2159–2168.
- Wang, Y., H. Wei, and G. Forman. 2013. "Mining Large-scale GPS Streams for Connectivity Refinement of Road Maps." Paper presented at the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (SIGSPATIAL'13), Orlando, FL, USA, November 5–8, 448–451.
- Wang, Z., T. Ye, M. Lu, X. Yuan, H. Qu, J. Yuan, and Q. Wu. 2014. "Visual Exploration of Sparse Traffic Trajectory Data." *IEEE Transactions on Visualization and Computer Graphics* 20 (12): 1813–1822.

- Wang, J., M. Duckham, and M. Worboys. 2015a. "A Framework for Models of Movement in Geographic Space." *International Journal of Geographical Information Science* 30: 970–992. doi:[10.1080/13658816.2015.1078466](https://doi.org/10.1080/13658816.2015.1078466).
- Wang, J., Y. Mao, J. Li, Z. Xiong, and W.-X. Wang. 2015b. "Predictability of Road Traffic and Congestion in Urban Areas." *Plos One* 10 (4): e0121825.
- Wang, Y., Y. Zou, K. Henrickson, Y. Wang, J. Tang, and B.-J. Park. 2017. "Google Earth Elevation Data Extraction and Accuracy Assessment for Transportation Applications." *Plos One* 12 (4): e0175756.
- Xie, X., K. B.-Y. Wong, H. Aghajan, P. Veelaert, and W. Philips. 2015. "Inferring Directed Road Networks from GPS Traces by Track Alignment." *ISPRS International Journal of Geo-Information* 4: 2446–2471.
- Yue, Y., Y. Zhuang, Q. Li, and Q. Mao. 2009. "Mining Time-dependent Attractive Areas and Movement Patterns from Taxi Trajectory Data." Paper presented at the 17th International Conference on Geoinformatics, Fairfax, Virginia, USA, August 12–14, 1–6.
- 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 IGARSS 2012 – IEEE International Geoscience and Remote Sensing Symposium, Munich, July 22–27, 5982–5985.