Universität Augsburg

Fakultät für Angewandte Informatik

Institut für Geographie

Traffic Pattern Analysis Framework with Emphasis on Floating Car Data (FCD)

Inauguraldissertation zur Erlangung des Doktorgrades

vorgelegt von

Andreas Keler

Gutachter:	Prof. Dr. Jukka M. Krisp
	Prof. Dr. Sabine Timpf
Tag der mündlichen Prüfung:	14.07.2017

Acknowledgements

During the time of writing the present work, I received a lot of support and inspiration from different people. First, I would like to thank Prof. Dr. Jukka M. Krisp for his advices and for giving me the chance of writing a dissertation. His encouragement helped me in developing my research. As a result, our collaboration resulted in numerous joined publications.

Additionally, I would like to thank Prof. Dr. Sabine Timpf for exciting discussions and for supporting me at my very first presentation at an international conference (GIScience'14), where I have talked about our joined research.

Furthermore, I thank the two other examiners, namely Prof. Dr. Peter Fiener and Prof. Dr. Jörg Hähner, for their challenging questions during my examination and for respectively representing the two disciplines geography and computer science.

A special thank goes to Dr. Linfang Ding, who supported me since the supervision of my masters' thesis, for being a nice colleague and friend. I want also to thank the second supervisor of my masters' thesis at Technical University of Munich Dr. Stefan Peters, who recently works in Australia.

Besides Dr. Linfang Ding, I am grateful to my other former colleagues at the geoinformatics group in Augsburg. I thank Nicole Bedacht (former Karrais) for joined research resulting in my first paper acceptance and for the nice time in our vislab. The other colleague during my time in the vislab is Dr. Jean Damascene Mazimpaka, who helped me with various insights on data processing issues and technical problems. Our joined research resulted in a journal paper, for which I am very grateful. For their mental support and for numerous humorous conversations, I thank Dr. Carolin von Groote-Bidlingmaier and Dr. David Jonietz. Additionally, I thank our visiting professor Prof. Dr. Xiang Li and his wife Lina Wang from Zhengzhou Institute of Surveying and Mapping for exciting conversations and showing me a little bit more of Augsburg by joined activities.

Finally yet importantly, I thank the Keler family including my parents Johann and Elena, my brother Eugen, my sister-in-law Katrin Keler and of course my sweet little niece Nina Lucia Keler. All of you supported me during the time of finishing my dissertation.

Abstract

Vehicle traffic in urban environments consists of a variation of traffic phenomena. Defining and measuring these traffic phenomena is challenging, since traffic sensors can still not observe the traffic situation of one city entirely over a period of time. One possibility to get general overviews is analyzing data coming from tracked vehicle movements. In the best cases, tracked vehicles are numerous and part of vehicle fleets that represent a big proportion of traffic participants in the investigation area. Traffic data in the form of movement trajectories is producible via the Floating Car Data (FCD) technology, which uses mobile devices that allow positioning and recording on-board information in every tracked vehicle. In case of operating taxis, these devices are part of already installed dispatcher systems and are able to produce Floating Taxi Data (FTD). One type of applications with FCD and FTD consists of inferring traffic situations with numerous different computational techniques. This thesis introduces a traffic pattern analysis framework for FCD with the emphasis on detecting specific vehicle traffic patterns. The extracted patterns should define urban traffic congestion as the detectable traffic phenomenon, which is the focus of this work. In general, tracking numerous moving entities participating in traffic is part of a large body of ongoing research. By reviewing traditional traffic data acquisition techniques from different domains, this work aims to provide a connection to various research disciplines connected with research on moving objects. Those fields are coming from physics, computer science, GIScience and geography to mention a few. In contrast to traffic phenomena on highways, which are well studied, this work focus on urban traffic in highly populated cities with dense transportation infrastructure. By selecting, modifying, and applying various methodological aspects, this work shows the establishment of a traffic pattern analysis framework that allows extracting typical periodical and unusual traffic patterns for each day of the week. Traffic congestion can be seen as a daily event, since it has starting and end points, that occurs on specific rush hours of the day, but as well as traffic anomalies that are caused by different events in the urban environment. The distinguishing between different types of traffic congestion events is challenging, especially when relying on classified movement patterns from FCD, which is only a fraction of all traffic participants. The first step is to clarify the various terminologies and to associate them with respective formalizations of each appearance, as the terms road capacity and traffic bottlenecks. Additionally, there are different aspects of traffic congestion detection, which includes reasoning on FCD representations, preprocessing and analytical possibilities. The last mentioned include map matching on road segments and density-based clustering of vehicle movement. Preceding steps of the framework consist of adjusted preprocessing of the data. The following six framework techniques aim to reveal specific traffic patterns from the preprocessed FCD by different forms of representing urban traffic congestion events. The underlying computational methods of the framework enable the possibility to apply various computations as a sequence that reveal an increasing number of details on urban traffic congestion events. The results of the framework computations include mainly three different products that are subsequently inferable: congestion polygons, congestion propagation polylines (CPP) and bundles of associated road segments. The affected road segments result from previous matching between road segments and congestion polygons, or congestion propagation polylines. The evaluation of the framework outcomes consists of visual analysis methods. A test FTD set from taxis in Shanghai from 2007 serves for the framework evaluation. The results show selected parts of the urban investigation area influenced by recurrent and non-recurrent traffic congestion, which conclude to expected travel time variations during rush hours. Afterwards, the test results serve for extensive discussions on the usefulness and reasonability of the framework methods. A concluding outlook outlines ideas on future work, which mainly consists of proposed methodical extensions and finding suitable applications for the traffic pattern analysis framework.

Zusammenfassung

Der Fahrzeugverkehr in städtischen Umgebungen besteht aus einer Variation von Verkehrsphänomenen. Die Definition und Messung dieser Verkehrsphänomene ist eine Herausforderung, denn Verkehrssensoren können die Verkehrssituation einer Stadt immer noch nicht gänzlich über einen längeren Zeitraum beobachten. Eine Möglichkeit, allgemeine Übersichten zu erhalten, besteht aus der Analyse von beobachteten Fahrzeugbewegungen. Im besten Falle sind die beobachteten Fahrzeuge zahlreich und Teil von Fahrzeugflotten, die einen großen Anteil der Verkehrsteilnehmer im Untersuchungsgebiet ausmachen. Verkehrsdaten in Form von Bewegungstrajektorien sind produzierbar über die Technologie Floating Car Data (FCD), die über mobile Geräte Positionierung und Aufzeichnung von fahrzeugeigenen Informationen in jedem beobachteten Fahrzeug ermöglichen. Im Falle von operativen Taxis sind diese Geräte Teil von bereits installierten Dispatchersystemen und können Floating Taxi Data (FTD) produzieren. Eine Art von Anwendung mit FCD und FTD besteht darin, Verkehrssituationen mit zahlreichen verschiedenen rechnerischen Methoden abzuleiten. Diese Arbeit stellt ein Verkehrsmusteranalyse-Framework für FCD vor mit dem Schwerpunkt auf der Erkennung spezifischer Fahrzeugverkehrsmuster. Die extrahierten Muster sollten städtische Verkehrsstaus als das nachweisbare Verkehrsphänomen definieren, das im Mittelpunkt dieser Arbeit steht. Im Allgemeinen ist die Beobachtung zahlreicher bewegter Objekte, die am Verkehr teilnehmen, Teil einer großen laufenden Forschungsaktivität. Durch die Bewertung traditioneller Verkehrsdatenerfassungstechniken aus verschiedenen Forschungsgebieten soll mit dieser Arbeit eine Verbindung zu verschiedenen Forschungsdisziplinen hergestellt werden, die sich mit der Forschung von sich bewegten Objekten beschäftigen. Diese Disziplinen kommen aus Physik, Informatik, GIScience und Geographie, um nur einige zu nennen. Im Gegensatz zu Verkehrsphänomenen auf Autobahnen, die schon gut erforscht sind, konzentriert sich diese Arbeit auf den städtischen Verkehr in hoch bevölkerten Städten mit dichter Verkehrsinfrastruktur. Durch die Auswahl, Adaption und Anwendung verschiedener methodischer Aspekte zeigt diese Arbeit die Etablierung eines Verkehrsmusteranalyse-Frameworks, das es ermöglicht, typische periodische und ungewöhnliche Verkehrsmuster für jeden Tag der Woche zu extrahieren. Verkehrsstau kann als tägliches Ereignis gesehen werden, da es Anfangs- und Endpunkte hat und zu bestimmten Stoßzeiten des Tages auftaucht. Verkehrsstaus können auch Verkehrsanomalien sein, die durch verschiedene Ereignisse in der städtischen Umgebung verursacht werden. Die Unterscheidung zwischen verschiedenen Arten von Verkehrsstauereignissen ist besonders dann schwierig, wenn man sich auf klassifizierte Bewegungsmuster von FCD stützt, die nur einen Bruchteil aller Verkehrsteilnehmer ausmachen. Der erste Schritt besteht darin, die verschiedenen Terminologien zu klären und sie mit den jeweiligen Formalisierungen jedes Erscheinungsbildes zu verknüpfen, wie beispielsweise die Begriffe Straßenkapazität und Verkehrsengpässe. Darüber hinaus gibt es verschiedene Aspekte der Verkehrsstauerkennung, die Darstellung, Vorverarbeitung und analytische Möglichkeiten von FCD beinhalten. Letztgenannte beziehen sich auch auf Map-Matching mit Straßensegmenten und das dichtebasierte Clustering von Fahrzeugbewegungen. Die vorangehenden Schritte des Frameworks bestehen aus einer angepassten Vorverarbeitung der Daten. Die folgenden sechs Methoden des Framework zielen darauf ab, spezifische Verkehrsmuster aus den vorverarbeiteten FCD aufzudecken durch unterschiedliche Darstellungsformen von städtischen Verkehrsstauereignissen. Die zugrunde liegenden Berechnungsmethoden des Frameworks ermöglichen es verschiedene Berechnungen als Sequenz anzuwenden, die eine ansteigende Anzahl von Details über städtische Verkehrsstauereignisse aufdecken kann. Die Ergebnisse der Framework-Berechnungen umfassen vor allem drei verschiedene Produkte, die nacheinander ableitbar sind: Staupolygone, Stauausbreitungspolylinien (CPP) und Bündel von zugehörigen Straßensegmenten. Die

betroffenen Straßensegmente resultieren aus einem vorherigen Matching zwischen Straßensegmenten und Staupolygonen oder Stauausbreitungspolylinien. Die Evaluierung der Ergebnisse der Framework-Anwendung besteht aus visuellen Analysemethoden. Ein Test-FTD-Satz von Taxis in Shanghai aus dem Jahre 2007 dient für die Evaluierung des Frameworks. Die Ergebnisse zeigen ausgewählte Teile des städtischen Untersuchungsgebietes, die durch wiederkehrende und nicht wiederkehrende Verkehrsstaus beeinflusst werden, die auf die zu den erwarteten Fahrzeitschwankungen während der Stoßzeiten rückschließen. Danach dienen die Testergebnisse für umfangreiche Diskussionen über den Nutzen und die Bedeutung der Framework-Methoden. Ein abschließender Ausblick skizziert Ideen für künftige Arbeiten, die vor allem aus methodischen Erweiterungen und geeignete Anwendungen für das Verkehrsmusteranalyse-Framework bestehen.

Contents

List of figures

List of tables

List of abbreviations

CMA	Computational Movement Analysis
CPP	Congestion Propagation Polyline
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
FCD	Floating Car Data
FTD	Floating Taxi Data
GDOP	Geometric Dilution Of Precision
GIS	Geographic Information System
GIScience	Geographic Information Science
GNSS	Global Navigation Satellite Systems
GSM	Global System for Mobile communications
GSM-GPS	Global System for Mobile communications via Global Positioning System
IoV	Internet of Vehicles
ITS	Intelligent Transportation Systems
KDE	Kernel Density Estimation
MM	Map Matching
MRR	Middle Ring Road (in Shanghai, China)
NRC	Non-Recurrent Congestion
OBD	On-board diagnostics
OPTICS	Ordering Points To Identify the Clustering Structure
RC	Recurrent Congestion
SNN	Shared-Nearest-Neighbor
T-GIS	Transportation Geographic Information System
TMC	Traffic Message Channel
TPI	Traffic Performance Index
Wi-Fi	Wireless Local Area Network (WLAN) based on IEEE 802.11 standards
WLAN	Wireless Local Area Network
xFCD	extended Floating Car Data

1. Introduction

Society in the 2010s has many characteristics influenced by globalization through advanced communication technologies. The development of the internet, since the 1980s, pushes this advancement in many directions. We can communicate via low-cost mobile devices, increase quality of hosting services in the industries and exchange data worldwide. On the other hand, our society is seeking for more information representing dynamics of the world. Examples include the usage of positioning devices within Global Navigation Satellite Systems (GNSS), where dozens of satellites help providing information. These systems give options for detecting individual positions of moving entities on the world's surface. Additionally, there are many Earth observation satellites with optical and non-optical sensors. They help detecting land cover changes over time, which is in this sense also observed movement.

The mentioned example of GNSS technology for monitoring multiple moving entities implies the extensive use of GNSS receivers. Due to the development of decreasing pricing of such devices, everybody is able, without high investments, to produce data of moving entities (Andrienko et al. 2013). Besides this technology, there are as well historical and other alternative sensor technologies that allow movement data acquisition, such as radar, Bluetooth, Wi-Fi positioning, Microwave telemetry or GSM-GPS (Global System for Mobile Communications via Global Positioning System). For the latter, the positioning accuracy will become even higher due to the combination of GNSS, odometry in the background of the installation of the ICARUS module on the International Space Station (ISS), which will allow the use of 5-g GPS/acceleration/magnetometer tags[1]. By observing the terminology of the tracked entity, one focus of this work is less on the movement of fields or surfaces like soil and water movements; it is more on tracking the movement of entities like individual animals, pedestrians, vehicles, and robots. A historical technology and data source connected with the secondly mentioned type of movement is Floating Car Data (FCD), which is data acquisition from tracked individual vehicles, of usually vehicle fleets. Each vehicle is a moving sensor. The records of these sensors result in vehicle movement trajectories, consisting of a number of positions with timestamps. The resulting data has usually the representation of an updated list showing the individual movement positions together with their instantaneous attribute values. Dependent on the tracking device, the acquisition strategy, the number of observed vehicles, the resulting size and the complexity of FCD sets can vary. In case of very large-sized and complex, in the sense of multivariate, data sets, the literature is referring to as massive movement data or massive FCD. The tracked entities of massive FCD are numerous and the positioning accuracies are relatively high, which results in expectable mass information that is difficult to handle. Due to resulting immense data sizes, there is a need of efficient algorithms for inferring reasonable information from the massive data amount (Andrienko et al. 2013).

Andrienko et al. (2015) name the inferred reasonable information as spatial events. As a representation, these spatial events can have abstractions as points in a space-time continuum. In general, handling data streams of multiple moving objects for spatial event detection is challenging, since numerous individual movements are matter of inspection on a global scale (Andrienko et al. 2015). Numerous tracked vehicles may help inferring driving behavior surveys (Schuessler and Axhausen 2009), vehicle-caused air exposures and traffic qualities (Dechenaux et al. 2014). Each tracked vehicle results in movement trajectories, which are sequences of timestamped movement positions for every individual vehicle. For Treiber and Kesting (2013) vehicle movement trajectories are the most accurate representation of traffic.

[1] http://www.icarusinitiative.org/

Nevertheless, trajectories of massive movement data are difficult to handle. There are many proposed methods for the task of analyzing massive movement data, which differ not only in their efficiency but also in the aims of handling the data. Approaches' aims vary from learning periodical movement pattern detection to the derivation of individual behavior and interactions with other entities. There are many differences in applying preprocessing steps. This direction shows that the development is far from standardization or best case examples, especially when respecting positioning accuracy and quality of achievable analysis results. This might be a critical question when inspecting massive movement of tracked vehicles within complex densely built urban environments of cities in the 2010s. Against this background, it appears reasonable to invest effort for finding out the requirements of deriving representative traffic or group movement information. Specifically, the aim is detecting and describing known traffic patterns like traffic congestion events. One possibility of massive movement data collection is the mentioned data acquisition method named Floating Car Data (FCD). In case of FCD, only the vehicle is tracked and not the individual person. In many cases, resulting FCD sets often come from entire taxi fleets that have already installed tracking devices as part of a dispatcher service. The principles of Floating Car Data (FCD) are pictured in a simplified overview in Figure 1, which is motivated by Fastenrath (1997).

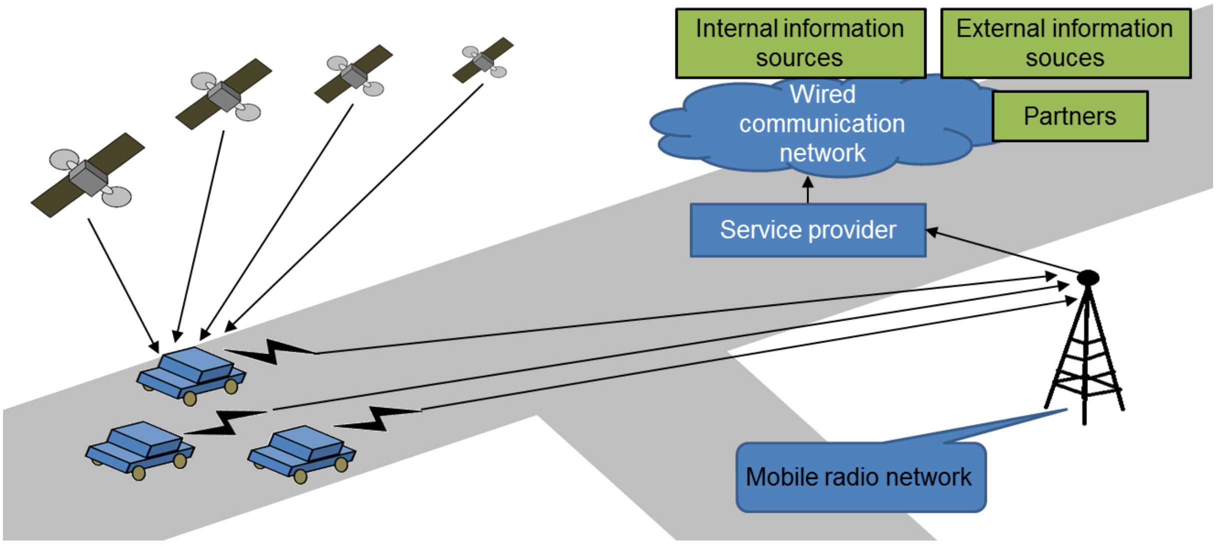


Figure 1: Simplified overview of the FCD technique – positioning via GNSS, data upload via cell towers and data storage.

Detecting traffic patterns from FCD is non-trivial, since tracking all vehicles, participating in the traffic is not feasible, even in the 2010s. The records might only show the trend of movement, in the same way the stream movement of a river system influences a boat; the general traffic flow of a road network system influences FCD acquisition. The system of traffic flow is relatively complex due to a changing number and variety of traffic participants and their individual driving behaviors.

FCD has varying quality. Influences on the data quality are device-dependent temporal sampling intervals (low and high sampling rates), together with the used positioning technology and the suitability of the investigation area for tracking numerous vehicles. This might be the crux for detecting and classifying certain pattern in the data. Another device-dependency is the quality of positioning, which is influencing the resulting spatial accuracy of acquired entity movements.

1.1 Motivation – Focus of the work

The focus of the present work is to develop suitable computational methods for gaining immediate knowledge about general traffic dynamics and congestion events within FCD sets. Besides the detection, also the representation of this information is important, especially when inspecting the data at different scales. The focus of this work is less the prediction of traffic congestion, but more its detection and representation. Therefore, the gained insight might be a starting point for calibrating a data-driven traffic learning and prediction model, in the way of a traffic forecast.

The proposed methods mainly base on computing with timestamped geodata partitioned into time series. This type of representation is nowadays an important element in different applications and object of ongoing research (Shekhar et al. 2015). The subtask of traffic congestion detection and prediction is a prominent task, since it affects millions of people every day with tremendous losses in time, money, and energy. An example, which shows the importance of traffic information, is coming from Los Angeles, where 1,800 interconnected static traffic sensors with a high cost investment and maintenance are used to describe continuously the city's traffic situation[2].

On the other hand, numerous tracking technologies emerge from movement acquisition via mobile sensors. In many cases, movement of an entity has the representation as a trajectory (Treiber and Kesting 2013), which means that a chronologically series of spatial positions with time stamps is produced. The acquired knowledge from trajectory data mining is useful in different application domains (Mazimpaka and Timpf 2016a). Moving object data are prevalent in many domains (Van de Weghe et al. 2006) and have often inherently imprecise location information (Laube 2001).

Since moving objects refer to any kind of moving entity, there are differences between groups of entities, as for example tracked vehicles. The data coming from tracked vehicles is, since the late 1990s, often referred to as probe data or as more specific Floating Car Data (FCD).

When tracked entities are vehicles, literature refers to its acquisition technology and resulting data as Floating Car Data (FCD). This acquisition technology and data has various connections to the domain of traffic engineering. Mobile tracking technology allows different insights into traffic conditions, which was not possible with data coming from static tracking devices such as induction loops. By definition of Cohn and Bischoff (2012), FCD may be produced by different sensors. The most frequently used sensors for creating FCD sets are GNSS receivers (Zhang et al. 2011d, An et al. 2016). Besides these, mobile phones (e.g. A-GPS) and Bluetooth[3] can generate FCD. For the data acquisition, each tracked vehicle of a fleet can have differently built FCD devices, which allow the recording of instantaneous values from on-board vehicle sensors. On-board diagnostics (OBD) allow recording instantaneous speeds, acceleration and other information. BMW terms the outcomes of these acquisitions extended FCD or xFCD (Huber et al. 1999).

Application examples of FCD from the domain of traffic engineering include methods that connect analysis results with the traffic flow theory, traffic dynamics and parameters, and the estimation of different travel times (Treiber and Kesting 2013).

An often-mentioned aim of using this technology is the development of services towards intelligent transportation systems (ITSs), which describe technically more reliable and adapted services compared to classical and historical. Many different research domains are working with FCD. Some of the recent

[2] http://www.forbes.com/sites/jonbruner/2012/01/25/how-los-angeles-keeps-traffic-moving-through-4114-stoplights/
[3] Universidad Politécnica de Madrid. (2015, June 16). Traffic monitoring though detecting Bluetooth devices on vehicles. *ScienceDaily*. Retrieved March 22, 2017 from www.sciencedaily.com/releases/2015/06/150616093457.htm

approaches use classical GIS methods for data analysis. Geoinformatics is for Raper and Livingstone (2001) a connection between physical and human geography. In their, very geographical, point of view, geodata representations are mentioned as conceptualized geographic entities. This allows connecting applications from historically subdivided groups of Geography. Fifteen years later, GIScience is a relatively known domain, which stands between Earth science and computer science, but has its own individual historical background. Additionally, Gudmundsson et al. (2012) introduce the term computational movement analysis, which represents an ongoing research direction of different research domains. One important ongoing topic of this group of the GIScience community is the modelling and representation of moving objects. This comes along with more and more available tracked movement data, as for example from Chinese taxi services (Liu et al. 2012a). This development appeared relatively suddenly and results from increasing positioning accuracies[4] and decreasing pricing of the used sensors. From the end of the 1990 to the early 2010, the situation on geodata availability was ranging from not enough data to massive data, which is difficult to handle. Besides GIScience, there are many working groups from transportation engineering, which work with FCD, especially in the context of connecting this information with classical data coming from static traffic counter devices such as induction loops. In general, it is possible to distinguish between the issues affecting the traffic detection from the direction of data quality (sensor and acquisition geometry specific issues) and from vehicle traffic theoretical models. The first direction includes reasoning on the data quality; in case of inspecting FCD a recorded vehicle trajectory might be affected by two types of errors: measurement error and interpolation error (Ranacher et al. 2016b). For Ranacher et al. (2016a), both types of error influence the calculation of movement parameters. This connects vehicle traffic theoretical models, mainly from the basics of mechanics and in particular traffic physics (Kerner 2004) and the physics of traffic congestion (Takashi 2002). The later implies knowledge of data acquisition methods and therefore makes reconstructions of past vehicle movements possible. There is a connection between the quality of this kind of prediction and the mentioned measurement and interpolation errors. When inspecting interpolation of spatial positions of movement, there is a focus on the sampling interval of the created FCD sets. This means there is a device-dependency when interpolating movement. This factum is a key issue in designing a traffic pattern analysis framework with emphasis on FCD. Especially the spatial interpolation error, which results from trajectory interpolation methods, has no mentioning in the literature, together with available test scenarios or any comparable inspections. Additionally, there are no available research on testing the connection of calculated movement parameters with temporal sampling strategies as stated by Ranacher et al. (2016a).

Concerning the topic of computing traffic parameters, a central aspect is the space of inspected movement. Here the connection to Map Matching (MM) methods, namely those techniques for matching tracked positions of a moving entity to the underlying road network, is clearly visible.

It is difficult to apply those techniques, especially when vehicles are tracked with low sampling rates (Zeng et al. 2016), which might be device-dependent and not only dependent on the positioning technology. Depending on the way of how to connect tracked vehicle positions with traversed road segments, the computed traffic parameter might greatly vary in their values and reliability. Besides using knowledge from the general development in analyzing movement data of moving entities, reviewing the research on FCD in the context of transportation engineering is also important.

The presented approach has its motivation by previous and ongoing research coming from the data mining domain, originating from computer science, and partially from computational geometry, originating from mathematics. The connection to data mining results from the aim of detecting

[4] http://www.gps.gov/systems/gps/modernization/sa/

patterns within the data automatically. In connection to this work, patterns within the data are traffic phenomena within individual vehicle movements. The base for data mining of movement data is the representation of movement as a series of discrete moving objects. The defined patterns for traffic phenomena are usually abstractions of the real world with specific geometrical appearances. These abstractions of movement in geospace are defined as points, lines or areal polygons with connected time components. Galton and Duckham (2006) mention that the area occupied by points can be aggregated or abstracted in many different ways. Additionally to this, the inspected objects in this work that form time-varying geometric abstractions. The temporal variations of these abstractions affect the shapes and other non-geometrical attributes.

The focus of this work is to introduce understandable and reproducible computational methods for detecting traffic situations and patterns, without the knowledge of investigation areas or used FCD. A traffic pattern has to consist of certain quality measure or attribute value at selected locations with temporal information as time of appearance and periodicity. The knowledge of the underlying road networks and its representation as features is not necessarily a precondition of the framework. One further aim is to produce results for further analyses as for example the prediction of traffic situations or the estimation of caused exposures by the vehicles. The latter might be connected with the impact on health effects caused by traffic congestion[5] (Goel and Kumar 2015).

Information representation appears often differently, dependent on the application and the analyzed data type. Consequently, there are numerous different representation possibilities for spatial data and movement. Dependent on the important, to be analyzed, properties of the used data, the used representation can vary in its dimensions, scales and representation philosophies.

The possibly most important space for representing geodata is the two-dimensional Euclidean space. In this space, points, lines and areal polygons represent features of the real world usually in a bird's eye view as in typical GIS software map views. Figure 2a pictures one example for the two-dimensional Euclidean space. Besides, query options based on spatial coordinates, there are two main groups of typically used data types: discrete vector data and continuous raster data, as shown in Figure 2a. In case of representing movement, there are the options of using raster time series or trajectory polylines.

Another movement space includes beside x- and y-axis, also a z-axis representing the temporal component. Space time cubes as in Figure 2b originate from time geography by Torsten Hägerstrand, where the temporal component is part of changes in space. This type of movement space can represent developments of human settlements or habitat developments via space time prisms and other three-dimensional geometries. Additionally, it is possible to represent individual trajectories as spatiotemporal polylines that meet at selected locations and time windows. These are the locations in space time, where moving entities have met. Due to cluttering effects, it is possible to represent only few spatiotemporal trajectories.

Another space of movement is the network space, where arcs and nodes correlate with real world roads and intersections, as pictured in Figure 2c. Arcs and nodes imply topological relations of selected locations and therefore are usable for routing applications. One important precondition for performing these applications is the implementation of realistic connectivity similar to road networks of the real world. Additionally, every appearance at road segments, such as number of lanes, road type, road closure and traffic states are assignable to arcs as weights. In general, the network space focus on representing restricted movement as the movement of a vehicle on a road network. In some case, also

[5] https://www.univadis.co.uk/viewarticle/traffic-jams-are-officially-bad-for-your-health-436822

other types of movement, even in open space, are representable via creating functional networks that do not match any visible network of the real world, but focus more on local knowledge or perceptual aspects. When focusing on road network features as in Figure 2c, there are several connections of arcs and nodes to movement trajectories, especially via map matching (MM) methods and via aggregation of tracked movement positions to road segments with specific identifications. Nodes often represent road intersections and more specifically the points, where directional changes in 2D Euclidean space occur. Usually, these locations imply several options of changing the driving direction.

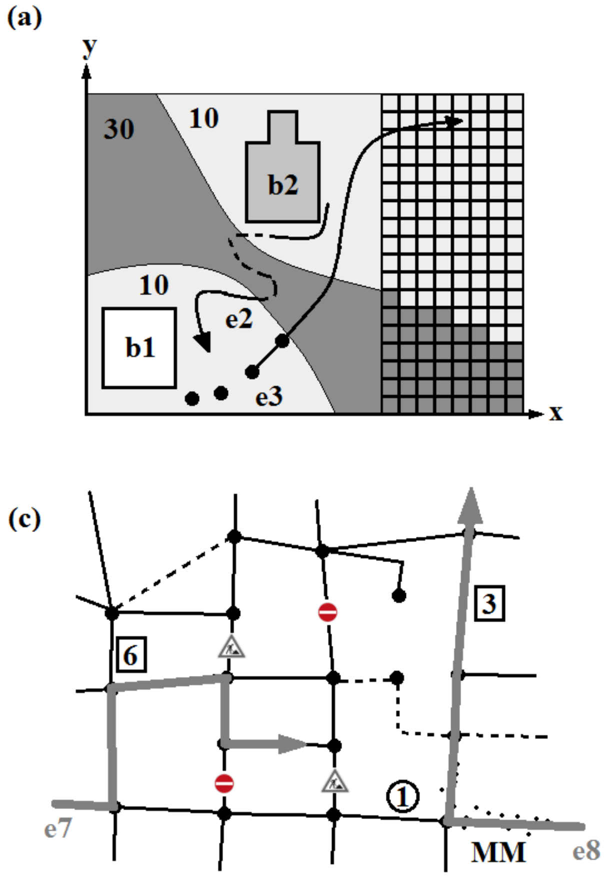


Figure 2: Four different movement spaces, with (a) two-dimensional Euclidean space, (b) three-dimensional space time cube, (c) network space, and (d) irregular tessellation; adapted from Gudmundsson et al. (2012).

The fourth possibility of a movement space is the irregular tessellation, pictured in Figure 2d. The general movement information is difficult to obtain, since the movement representation of one entity only consists of sequences of areal polygons. This change of space at different time windows allows interpreting movement. It refers to the positioning via phone tower cells of mobile phone users. As in the data extracts of the telecommunication providers, call detail records (CDR) imply movement representations as a sequence of visited phone tower cells. These sequences might not represent detailed movement of recent mobile phone users, as pictured in Figure 2d. Additionally, the height component influences the positioning via GSM. Nevertheless, when inspecting CDR on a large scale, for example on the scale of a whole country, inferring representative movement patterns in possible. This is different from the scale level of one city, since movement of entities is difficult to reconstruct.

1.2 Events of moving objects and traffic phenomena

This work has the focus of detecting vehicle traffic phenomena from FCD. In general, there are multiple differing definitions of various vehicle traffic phenomena in the literature. The most prominent examples include traffic congestion event. These types of events are usually occurring in road network space, namely on differently connected road segments of different lengths and widths. Since these events include numerous vehicles that are slowly moving or still standing on parts of the road network, they are called linear events. Events occurring in road network space, which are for Tang et al. (2016a) so-called linear events (LE). Besides the shape and size of these linear events, there are also temporal attributes of the traffic event. Additionally, there are references to the frequency of traffic congestion appearance. There are the terms recurrent congestion (RC) and non-recurrent congestion (NRC) as differentiated by Anbaroglu et al. (2014), Varaiya (2007), Dowling et al. (2004), and Chen et al. (2016). On the other hand traffic accidents can be the causes of traffic congestion events, as defined by Okabe et al. (2009), and by Xie and Yan (2008). The reason is that accidents sites depending on accident time and severity have influences on the surrounding traffic, as road closures and blockades, or cause secondary accidents. Since often assumed these traffic congestion events are caused by unusual driving behavior, they are often part of traffic anomalies.

A traffic anomaly is in general a traffic pattern that is unusual for selected investigation areas, as defined in Lan et al. (2014), Pang et al. (2011) , Pang et al. (2013), Zhang et al. (2011a), and in Chen et al. (2012). The critical question is how to observe something as a typical traffic situation. As in the Google maps vehicle routing service, we can compute usual traffic states for road segments based on averaged travel times of specific working days. In many cases, periodical traffic congestion show similar congestion propagation patterns or has concise association with capacity limits of selected transportation infrastructure elements. Congestion propagation and bottleneck identification are special cases of moving events that reflect, to a certain degree, the quality of the transportation infrastructure, as described by Chawla et al. (2012), Ji et al. (2014), and Long et al. (2008)

When observing the influences of the underlying infrastructure, there are also Phantom jams, which are non-recurrent traffic congestion events without traffic bottlenecks and other influences of the transportation infrastructure, as in Sugiyama et al. (2008), and in Flynn et al. (2009). The model states that small delays of following vehicle drivers can cause phantom jams. Thus, this type of traffic congestion can depend on the driving behavior of individuals.

Besides the difficulty to distinguish between these various traffic phenomena, most of the literature is distinguishing between general infrastructural elements in rural and urban investigation areas. Urban vehicle traffic phenomena are often the focus of ongoing research. One specific phenomenon is urban traffic congestion, which has for Miller et al. (1999) the following characteristics:

- Traffic congestion patterns are complex spatially
- Morning and evening peak periods are declining compared to rural areas
- Non-work trips (spatially and temporally dispersed) are increasing compared to rural areas
- Traffic congestion is a dynamic phenomenon

Due to difficulties in detecting periodical traffic patterns and the frequently changing mobilities of people, traffic congestion in urban road networks still remains a research challenge (Anbaroglu et al. 2014). This statement finds its proof in a questionnaire by Bertini (2005), where hundreds of experts on transportation delivered very differing answers on how to define and to measure urban traffic congestion.

Figure 3 shows examples of real-world traffic congestion in urban environments. Traffic congestion might have various appearances connected with different causes and processes. The challenge here is to find patterns in available data, which pictures only one small selection of traffic participants. Nevertheless, literature defines specific vehicle movement patterns, which allow distinguishing between different types of traffic congestion.

Figure 3: Selected real-world appearances of urban traffic congestions, with (a) cars and motor-assisted bicycles at a light-signal system, (b) vehicles, vans and pedestrians at an urban intersection, (c) congested highway traffic and on-ramps, (d) traffic gridlock due to buses and delivery vans, and (e) evening rush hour on elevated highways. (References in [6])

Keeping this in mind, the introduced framework describes discrete traffic phenomena in connection with the frequently changing mobilities in urban environments. Traffic phenomena have many different appearances with different definitions. Additionally, traffic congestion, as one type of these phenomena, has various definition possibilities. From these numerous definitions of traffic congestion, only those, which have the focus on traffic congestion in urban environments, are the basis of literature review and of practical tests in this work.

After finding rules to distinguish between different traffic event types based on FCD, another unsolved task is to find suitable representations of the detected phenomena. In case of traffic congestion events, the representations can range between linear or punctual features, areal polygons, or parts of a road network for different time components, or as time series. One challenge of the spatiotemporal representation of traffic congestion, is handling the temporal component, since numerous geodata representation approaches base on processing static geodata. Another challenge is inferring vehicle traffic patterns without knowledge about the transportation infrastructure. When inspecting traffic phenomena, it is important to include available road network information into analyses. The accurate representation of connected road segments can serve as a base for evaluating detected traffic phenomena. Several approaches presume even the existence of road network data, as for example, the estimation of traffic densities and the matching of events with road segments.

[6] Image references – URL:
(a) http://riskology-wp.s3.amazonaws.com/wp-content/uploads/2014/06/23221103/bangkok-traffic.jpg
(b) http://www.sheepsheadbites.com/wp-content/uploads/2011/09/sb_rosen.jpg
(c) http://mms.businesswire.com/media/20131105006522/en/390694/5/1652732_Aerial_shot.jpg
(d) https://pacejmiller.files.wordpress.com/2011/03/shanghai-traffic.jpg
(e) http://www.spreephoto.de/wp-content/uploads/2013/07/shanghai-urban-traffic-950x518.jpg

The idea is to use areal polygons for representing traffic events, in particular traffic congestion. Consequently, the traffic congestion representations in this work are not bindingly linear events, but more like areal representation of influenced road segments. The connection from polygons to linear features is feasible via the matching of inferred traffic information (polygons) with road segments (lines, polylines, or arcs).

1.3 Hypothesis, research questions and objectives of the work

After the motivation and the general problem statements in the previous section, one can introduce hypothesis, research questions and general and specific objectives of the work.

Hypothesis

Knowledge gained from vehicle trajectories supports quantitative detection of and differentiation between traffic phenomena (of the real world).

Therefore, data of moving objects needs examination by different approaches in different movement spaces, independent of (data acquisition) quality.

This hypothesis focuses on the possibilities of using vehicle trajectories for detecting those patterns that reveal traffic phenomena of the real world. The challenge here is to use a relatively small number of traffic participants and inspect their properties of movement. Another part of the hypothesis is indicating that examination of vehicle trajectories in Euclidean space is not enough. This assumption requires additional examination of the data in feature spaces that respect also the temporal components of the data together with components of the objects' movement dynamics. The gaining of knowledge about traffic phenomena is independent of the data quality, focusing on the reliability of spatial and temporal components of vehicle trajectories.

Research questions

Resulting from this hypothesis, four research questions appear. The answering of these four questions is one goal of this thesis. The four research questions of this thesis are:

1. How can we detect complex traffic phenomena from massive FCD with varying quality?
2. How much FCD do we need and in which quality for detecting traffic phenomena and its propagation over time? Is there a need for additional information?
3. What are typical characteristics for traffic phenomena, in particular traffic congestion? What are investigation-area-independent properties?
4. Can we benefit from inspecting vehicle movement in different spaces, by means of different views on the movement data, in the way of gaining more detailed knowledge (e.g. for calculating traffic parameters)?

Research question 1 implies a review of techniques that might serve as means for detecting complex traffic phenomena. One precondition is gaining knowledge about the role of data quality and its influence on the results. More specifically, the question includes the search for the thresholds for having the lowest possible quality of FCD for detecting traffic phenomena. In this work, the quality of FCD consists only of its partial properties sampling interval and spatial accuracy.

In the second research question, the focus is on finding the optimal way of observing moving traffic participants for guaranteeing optimal results. During tracking data acquisition, its duration or acquisition time can influence the quality of the outcomes. In a different way, quality of traffic phenomena detection can depend on the number of observed vehicles. The task is to estimate optimal durations and numbers. This question is connected to the device-dependent quality of positioning and acquisition, which is mentioned in question 1. Nevertheless, the bigger focus of question 2 is to clarify the numbers of trajectories of moving entities for defining traffic phenomena. Additionally, the aim is to show when the need of additional information is inevitable. Additional information results from data on the underlying transportation infrastructure or the mode of transports. For this case the type specification of the needed information is important and its dependency on the specific type of analysis.

Question number 3 is about finding the typical characteristics for selected traffic phenomena, or more specific traffic congestion. In case these characteristics have certain patterns, the goal is to define or to find measures or parameters, which are independent of the investigation area.

The fourth question is about reasoning on the way of movement data analysis. Despite analyzing movement in Euclidean space, other spaces might show hidden patterns of movement. The aim of the fourth question is to find extensions in different spaces or in different views on the data for defining additional traffic parameters.

General objective

The general objective of this thesis is to develop a conceptual and methodological framework for detecting and classifying vehicle traffic pattern or so called traffic phenomena based on various FCD.

Since the term traffic phenomenon has a great variety in its interpretation, the focus in this thesis is only on one class of traffic phenomena for further analysis: vehicle traffic congestion. In the following case studies, only vehicle traffic congestion and its propagation are matter of further inspection.

Specific objectives

The specific objectives of this thesis include the following:

1. Calculation of various traffic parameters based on movement representations in different spaces (Euclidean, Spatiotemporal, Network, Feature spaces)
2. Detection and Classification of different traffic phenomena
3. Quality assessment of results – focus on FCD quality (measurement, interpolation)
4. Method development for connecting classified movement patterns (and various other information) with street information for transport networks

5. Defining a technique for FCD sets, which usually come from a small percentage of participating vehicles, which allows estimating the actual (real world) traffic density

Additionally to these specific aims, this work has some follow-ups on ongoing research in the domains of GIScience and Transportation Engineering in the context of FCD. Therefore, the following additional research questions should be answered, which are mainly defined in the outlooks of selected research papers:

- What is an Appropriate Temporal Sampling Rate to Record Floating Car Data with a GPS? (Ranacher et al. 2016b)
- How much GPS data do we need? (Patire et al. 2015)
- How to compare movement? A review of physical movement similarity measures in geographic information science and beyond. (Ranacher and Tzavella 2014)

The first question refers to data quality, and, how proportionate a high sampling interval is for the quality of traffic phenomena detection. The second specific question consists of finding out if the number of observed vehicles is of extraordinary importance for accurately detecting traffic phenomena. In the third question, only those movement similarity measures are selected that deliver useful insights into traffic phenomena. Here the focus is more on indicating, which already known similarity measures are not useful for the traffic pattern analysis framework.

1.4 Approach and methodology

For proving the mentioned hypothesis, one needs to traverse research from various different fields including computational movement analysis, physics (in particular mechanics), traffic engineering (as part of civil engineering) and GIScience. Further connected core disciplines are transport geography and time geography. The used approach bases on previous work on traffic pattern detection and classification within geographic information systems. Following the concept of designing a framework for traffic pattern detection, the proposed methodology makes use of six different methods that benefit the pattern exploration of FCD. After applying the framework methods and inspecting the first outcomes, several aspects contribute to ongoing research.

Approach

The used approach in this thesis for solving the research questions consists of designing a traffic pattern analysis framework for FCD. Therefore, the first part is to find similar ideas from the body of literature. Parts of selected historical and recent research are included into the framework, nearly always in modified and extended form. Other parts of the framework emerge from first practical tests of various techniques for handling this type of data.

Methodology

The terms moving objects, vehicle tracking and traffic patterns appear in many different research domains. This results in a high number of different terminologies. Therefore, the first step for defining a traffic pattern analysis framework is to specify a taxonomy of features and methods to avoid misunderstanding. This taxonomy enables describing the design specifications of the framework precisely.

The methods of the framework rely partially on observing movement of concrete objects in different spaces (Gudmundsson et al. 2008). Besides observing moving objects in two- and three-dimensional Euclidean space, there is also the spatiotemporal component of moving objects, which is the location of the object at a certain time. Since vehicle traffic is bound to transportation infrastructures such as road networks, a modelling of movement in network space is the additional third component of observation. In network space, it is possible to aggregate movement of numerous moving objects and to assign it with connected arcs, which base on street representations of the real world.

After introducing selected preprocessing steps for FCD, the proposed framework continues with the computation of products that rely on preprocessed FCD. The first focus is introducing global indices for respecting the daily traffic situation for the whole investigation area. Dependent on these indices, selected rush hours of the day influenced by local traffic phenomena are selectable.

One central aspect of the framework is detecting and representing traffic congestion events. Therefore, the framework has six techniques that allow not only the traffic congestion detection, but as well the representation of different congestion events. The evaluation of the proposed framework techniques implies testing with specific FCD sets generated by urban taxis in 2007.

1.5 Structure of the thesis

The structure of this thesis consists of an introduction and six following chapters as pictured in Figure 4. Chapter 2 consists of the state of the art on moving objects. The focus of this chapter is to reveal the most important and known methods for analyzing this type of data. Additionally, chapter 2 shows the basics in traffic data acquisition, which consists, besides historical techniques, as well of the options to respect and include road network information into the analysis.

Chapter 3 gives comprehensive definitions of FCD and reviews recent and historical FCD applications. In this chapter, the focus is more on related work that influences the proposed traffic pattern analysis framework in this thesis. Based on the two previous chapters, chapter 4 designs a general framework for analyzing FCD and for detecting vehicle traffic patterns. The last mentioned is the fundamental application area within this thesis. Therefore, chapter 4 lists many definitions of traffic patterns and more specifically of vehicle traffic congestion. Including these definitions into a general holistic definition of traffic congestion is the base for introducing a traffic pattern analysis framework for FCD.

After these definitions, chapter 5 consists of testing the framework with a historical taxi FCD set of 2007 coming from a taxi fleet in Shanghai. The results of the testing and evaluation are part of an extensive discussion and conclusions in chapter 6. Chapter 7 concludes the work with an outlook for unsolved problems. Figure 4 shows the different parts of this thesis with shortly summarized terms for each section.

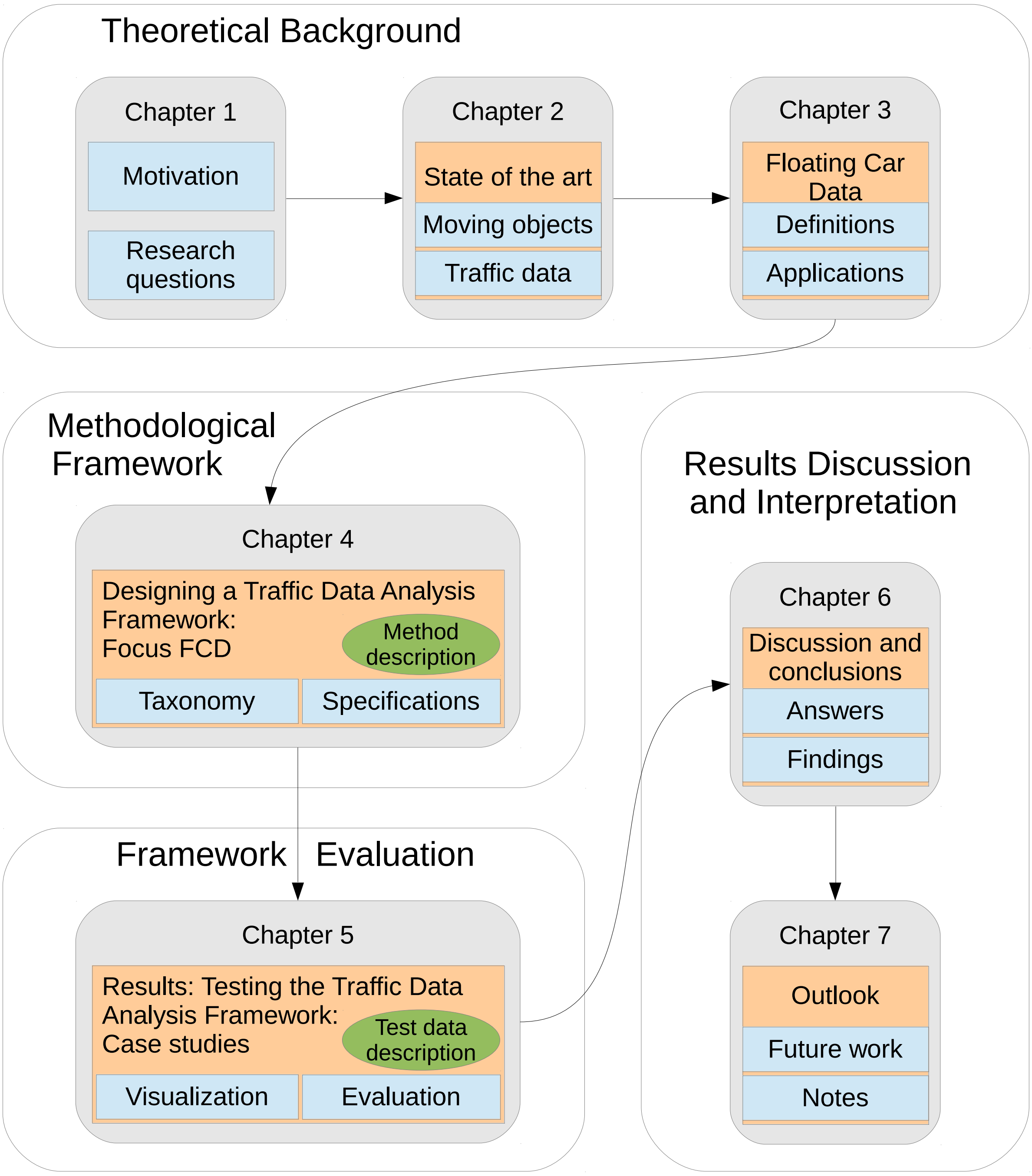


Figure 4: Overview on the structure of this thesis.

2. State of the art on vehicles as moving objects – a review of methods

This part of the thesis presents an introduction on concepts connected with moving objects. The term moving objects refers to observations of various different dynamics in our environment. One common property is that a concrete entity is moving in Euclidean space, be it an organism or a machine. Additionally, and as an extension, it is possible to define other spaces of movement representation. The variety of tracking devices with various data acquisition techniques for acquiring moving objects indicates different qualities of the data. More specifically, movement trajectories acquired with differing tracking devices have different accuracies and resolutions in space and time. In general, a trajectory is the description of individual object movement with a number of records showing spatial positions of the entity with attached time components. Trajectories might be part of data sets with different structure and quality. In many cases, trajectories of moving objects have to be defined first from unsorted and not preprocessed, raw, movement data. Another fact is that every analysis of moving object data is application domain driven. Therefore, the following will show common basics of moving object data analysis but as well domain specific issues that developed historically.

2.1 Computation with data from moving objects

Tracked movement of objects is widely available and used for various applications in our society (Dodge et al. 2016). In case of inspecting individual vehicle trajectories, the movement data has a specification that includes domain knowledge about the moving entity. In this case, the domain knowledge might include the movement restriction on roads or typical values for measurable quantities like speed and acceleration. In the most often cases, there is the, sometimes excessive, usage of tracking devices for collecting the movement of certain entities, as for example people, drivers, robots or animals. These movements result as a series of records with a spatial and a temporal component, namely movement trajectories of entities.

Formal representation of movement trajectories

A movement trajectory consists of a list of records $X_{1:T}$ containing at least two coordinates x, y, and one time component t:

$$X_{1:T} = x_1, \dots, x_t, \dots, x_T \text{ , where } x_t = (x, y, t) \tag{1}$$

x_t is recorded by a positioning device. The simplest way of representing movement trajectories visually within a map view is via polylines, which is a synonym for line graphs. Every point within one single trajectory might then be the base for linear interpolation between consecutive movement positions. FCD representations and their constellation possibilities are part of the following chapter 3.

Progressively varying continuous fields

Another group of movement data is the movement of fields like clouds, wind and water surface movements (Andrienko and Andrienko 2013). The typical representation for these phenomena are progressively varying continuous fields (Goodchild and Glennon 2008), which are continuous fields with additional time components (Keler and Krisp 2015a). Other examples of progressively varying continuous fields are atmospheric movements, ocean circulations, temperature, humidity, or vehicle and pedestrian movements (Keler and Krisp 2015a). With respect to the two last mentioned examples, it is to note that it is generally possible to convert massive moving object data of many entities into raster time series. This is possible via aggregation of movement by aggregating individual movements of entities into streams or flows of moving objects, as it is defined in Andrienko and Andrienko (2013). This conversion uses individual movements of moving entities, usually trajectories, and creates group movements, often represented by movement or position densities.

Movement analysis is possible with tracked animal movement resulting in timestamped movement positions. Extracting patterns for analysis focuses more on individual movement. Usually, traditional analysis methods consist of statistical methods coming from the field of movement ecology (Demšar et al. 2015). One key application in movement ecology is to define animal behavior together with the specific contexts. Besides many similarities between vehicle and animal movement, there are as well many differences resulting from domain-specific issues. One of these issues is the definition of habitats for moving animals and the absence of a road network, which is typical for vehicle movement.

Methods from movement ecology

Animal movement might be an indicator for environmental spatiotemporal changes. The movement is at first sight an indicator for the behavior of individuals. Respecting its long-term patterns may reveal socio-economic processes of the human society, together with indication of the climate and weather conditions. There are historical methods used for acquired animal trajectories by various sensors, in the time even before the appearance of GPS (Demšar et al. 2015). Historical tracking devices are for example handheld radio receivers in combination with installed radio transmitter colliers on the animal's bodies. The different devices are part of wildlife radio telemetry, which has a long development since the 1960s.

There is a great research body from movement ecology on how to analyze moving entities. This can benefit research on moving vehicles in many ways. When observing multiple animals in the same investigation area, the interesting patterns include for example the interactions between the tracked entities. The idea is detecting movement similarity like following and meeting behavior. There is a big variety of different similarity measures for inspecting movement data (Ranacher and Tzavella 2014). Selected research on similarity measures may also benefit applications in the urban environment, such as the different similarity measures can find applications with data coming from observed humans (Ranacher and Tzavella 2014).

Another aspect focuses on connecting animal movement with context. In case of movement ecology, this is often the connection of trajectories or parts of it with specific land use types. Nevertheless, animal movement appears as well in urban environments with massive anthropogenic features. One common approach in movement ecology is the segmentation of trajectories into stops and moves. In a similar way, (Spaccapietra et al. 2008) proposes a conceptual model of movement based on stops and moves. The extracted segments are further observable via visualization methods for looking inside trajectories (Andrienko and Andrienko 2013). The last aspect connects thematic reasoning on the scale of analysis representation, as described by Laube and Purves (2011) and by Gschwend and Laube (2014).

Transport modes of moving humans

Instead of animal movement data, human movement data may consist of different modes of movement, since humans can use vehicles for their transport. The frequent production of these data sets is feasible via numerous mobile phone users and may imply different modes of transportation. In context of moving entities in urban environments, there is an ongoing development towards "smart cities" and using humans as "living sensors" (Souto and Liebig 2016). Movement of animals, as of humans can reveal behavior; human behavior in an urban environment with specific use cases is often more complex than the behavior of tracked animals. Tracking numerous human beings during daily life has a connection with privacy concerns. These concerns arise from the possibilities of inferring behavioral patterns of individuals. Nevertheless, there are various established anonymization possibilities. One example for these anonymization strategies for tracking data includes introducing identification numbers of tracked entities that may hide the identity of the tracked entity. Another example has a connection with data post processing via cutting trajectory segments near respective start and end points. By including anonymization radii at trajectory start and end points, it is possible to post process the data for further scientific and eventually commercial applications.

One exception is the FCD coming from vehicle fleets, as tracked taxis are important for maintaining high service quality in selected investigation areas. The precise spatial and temporal information is crucial for taxi companies. Depending on the availability of tracked information about entering and leaving taxi customers, more or less patterns of the driving behavior, as hunting and waiting for customers are inferable from the taxi trajectory data (FTD).

Technical realization of handling movement data in databases

For Leonardi et al. (2010) a moving object database (MOD) is a specific example of a database, which can handle movement trajectory data and additionally perform spatiotemporal queries on movement data. Motivated from designing application-oriented databases, the focus of MOD is more on providing efficiency in handling various queries. Simple spatiotemporal query operations are helpful for many applications. Prominent examples of moving object databases (MOD) are Hermes (Pelekis and Theodoridis 2006) and Secondo (Almeida et al. 2006). These database types allow higher efficiencies in case of spatiotemporal pattern queries (Sakr and Güting 2011) or efficient indexing of moving objects for k-nearest neighbor search (Güting et al. 2010)

Time and sampling intervals of tracked movement

For Herrera et al. (2010), there are at least two different types of sampling GPS tracks recorded on transportation networks. These two types are temporal and spatial sampling. Laube et al. (2007) focus on spatial sampling and propose for spatial variables the investigation by using local, focal, zonal, or global context operators (DeMers 2002, Tomlin 1990, Worboys and Duckham 2004). These context operators are explicitly two-dimensional (local, focal, zonal, or global context). The subsequent step after spatial sampling is adapting contextual operators into one-dimensional lifelines. These lifelines are resulting from the specific scale of movement analysis level. There are in total four different analysis levels for movement data, which are represented in Figure 5 based on the finding from Long and Nelson (2013). The analysis levels are instantaneous, interval, episodal or total.

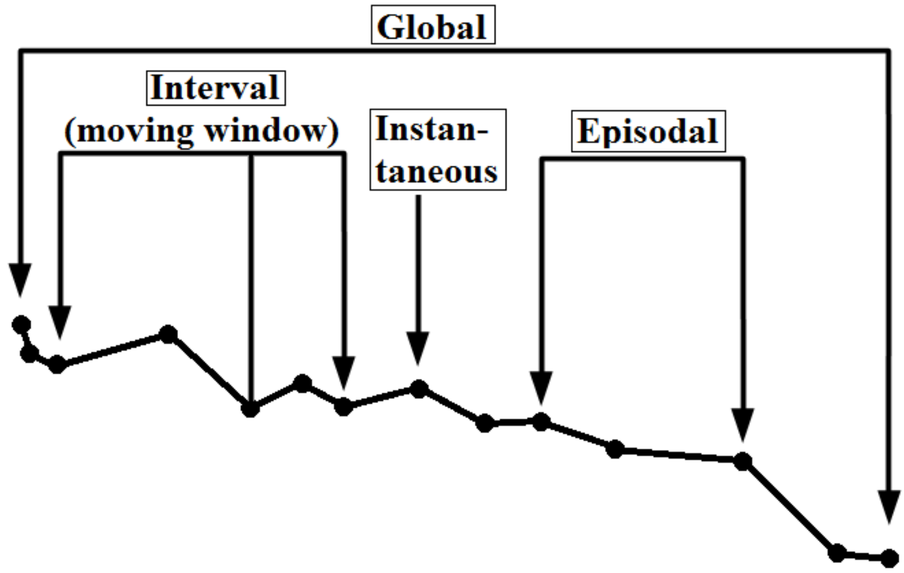


Figure 5: Four analysis levels for movement data, based on Long and Nelson (2013).

The four analysis levels in Figure 5 are heavily dependent on the temporal sampling intervals of the inspected data. The fixes in time might reveal constant or irregular lengths. Benhamou (2004) states that many analytical methods require equal step lengths in time as Figure 5 describes as interval or moving window. In case of irregular step lengths, there are various possibilities of interpolation, both in space and time. Another option is aggregating records with spatiotemporal information into episodes as pictured in Figure 5.

The temporal analysis level of global or total can include the whole data set from the earliest point in time to the very last, or one completely individual trajectory. Figure 5 shows in the lower part a movement trajectory representation with movement position points and connecting polylines. These polylines can result from a linear interpolation. Every FCD record has one specific temporal component. This level of analysis is instantaneous. In real life applications, the instantaneous temporal values of the records can serve for computing temporal intervals or episodes, together with averaging of the associated instantaneous attribute values.

2.2 Movement and context – analyzing first and second order effects

Analyzing movement without the knowledge of contextual information is difficult. The context of movement, when working on movement data without metadata, is difficult to obtain. There are different aspects of context, which are complex and difficult to describe quantitatively. The context of movement of daily commuters is different from the one of daily working taxi drivers. There are different types of behavioral aspect that are present during movement. Within complex urban environments, the context of movement might change dynamically with the recent traffic situation that can influence millions of people at the same time. Therefore, it is to state that the inference of reliable context information needs detailed and accurate data together with technically mature methods for information extraction. Context knowledge is not only helpful for potential analysts of movement data; it facilitates preprocessing steps, since contextual knowledge may reveal helpful insights into data acquisition.

For Purves et al. (2014), recent research on trajectory data mining methodologies for movement pattern identification and the extraction of behavior, does often neglect including the geographical context within the movement data sets (Siła-Nowicka et al. 2016). The meaning of geographical context can be quite unfocused. In general, any additional information besides the acquired movement that reveals spatial and temporal components might count as contextual information. Purves et al. (2014) introduce three groups of possible context information for movement data resulting from different ways of context data acquisition:

1. Context acquisition appears simultaneously with GNSS acquisition of moving entities. One example for this case is an additional video camera device that records the traversed surroundings on the road network during FCD acquisition.
2. Context acquisition by adding traversed space descriptions of individual moving entities. One example for the second option during FCD acquisition is the perceived information by the driver (of the observed vehicle) by moving on a road network.
3. Context acquisition by adding knowledge of physical and biological properties of the movement process. The example with FCD acquisition might include knowledge about vehicle properties together with power range, weight, and accelerations, which allow further insights into vehicle dynamics.

The second and third group requires static information on the transportation infrastructure and metadata on the tracked entities (vehicle properties). One idea of the traffic pattern analysis framework is to generate contextual traffic information solely from FCD. Besides these definitions of context acquisition, which always require additional data and information, there is the context due to interactions of the moving entities. This type of context focuses on the type of interaction between two or more entities at the same time and space. Since this context type is differing from the previously mentioned, there is a classification option that refers to different orders of entity interactions. In general, there is the differentiation between first and second order effects. In general, movement analysis includes the detection of both, first and second order effects (O'Sullivan and Unwin 2010).

Analyzing first and second order effects

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). Consequently, the previously mentioned three definitions by Purves et al. (2014) are first order effects of movement. There is a plethora of research, on how to combine movement data with contextual information. This research area is widely known as research on semantic trajectories. The main contributions in this area are methodologies for efficient, ideally on-the-fly algorithms that allow online functionality.

Research on semantic trajectories is important for web service personalization (Shekhar et al. 2015), navigation applications. Semantic enrichment in the context of smart cities can combine various different data sources (Zhang et al. 2016b).

Second order effects consist of detecting interactions of multiple entities, such as meeting points, convoys and flocks (Gudmundsson et al. 2012). The last mentioned include also the detection of traffic congestion or stoplights in case of analyzing FCD. The methods for detecting second order effects of movement include similarity measures, which can be temporal, spatial or spatiotemporal (Ranacher and Tzavella 2014).

Context-dependent transformation of movement data types

The practical analysis of movement data has in many cases a smooth transition of first order and second order effects of movement. An illustration of this fact in Figure 6 shows the options of transforming different types of spatiotemporal data (Andrienko et al. 2013). Context information is included in the spatial event and can come from spatial time series and trajectories via the extraction step. On the other hand, spatial events can have an aggregation into spatial time series, or an integration into trajectories. This means that in case a spatial event is traffic congestion, it is possible to enrich one single trajectory with this information for representing the affection. Traffic congestion, as a spatial event, implies the representation possibility of time series. Additionally, the description of congestion propagation is feasible via successive time windows on the associated space. The aggregation of trajectories within spatial time series results, in connection with the previous example, as general traffic flows of vehicles, comparable with flock or convoy patterns.

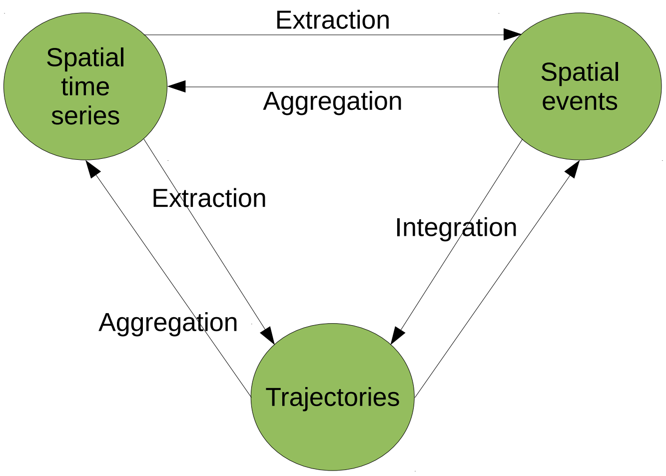


Figure 6: Transformations between different types of spatiotemporal data, adapted from Andrienko et al. (2013).

2.3 Traffic data acquisition

Important domains of vehicle movement analysis are traffic and transportation sciences. Research on the nature of traffic is important for most of the world's population. Understanding traffic, traffic congestion appearance, and propagation in time and space is a task that has high importance for future generations. Besides tremendous emissions caused by traffic (Goel and Kumar 2015) in air[7], land and water, there is a gigantic loss of time and money every day in the world, due to vehicle traffic congestion (Rao and Rao 2012). In general, traffic data acquisition is feasible with static and mobile sensors. The prior are often expensive to establish at numerous sites and usually have a poor spatial coverage. The latter often implies restrictions from the selected positioning technology, as for the cases of using GNSS positioning within urban canyons that might cause signal losses. After data acquisition, one challenging task is the data handling for subsequent traffic classification. Depending on the way of classification, different congestion patterns have different terminologies in the literature.

[7] https://www.univadis.co.uk/viewarticle/risk-of-air-pollution-is-biggest-at-traffic-lights-247048

For the monitoring of traffic conditions or for the collection of traffic volume data, several traditional methods are still in use. One historical technique for car movement acquisition, and considered antecessor of FCD, is the method floating car observer. In 1954, Wardrop and Charlesworth (1954) described this method for the first time. It bases on speed estimation by driving with an individual vehicle speed and flow of the complaisant lanes. Floating car observer is also used nowadays (Kühnel 2012).

The simplest possible acquisition of the traffic flow is simple vehicle counting by human perception on the roadsides of each road lane, additionally supported by stopwatches (Leduc 2008). This is part of a variety of historical static traffic sensors, which are still in use. Counting vehicles is one important principle of traffic engineers for detecting the real traffic flow rate in selected locations. Common devices for accurate traffic flow rate acquisition are induction loops. Induction loops use the fact that vehicles are made of metal and detectable by its electromagnetic appearance. The installation of induction loops within road pavements is, compared to the provision and usage of mobile devices, relatively expensive (Leduc 2008, Klein et al. 2006). Nevertheless, they are helpful in reacting on traffic jams in real-time[8].

Recent development in sensors for movement data acquisition allows information collection with high spatial and temporal scales in real time (Kwan and Neutens 2014, Siła-Nowicka et al. 2016). Others employ newly established remote sensing methods (Yuan et al. 2014). In the 2010s, there is a frequent and ongoing discussion on how to produce and analyze moving object data sets coming from acquisitions with mobile sensor devices such as GNSS, GSM and WLAN. In case, the moving objects are vehicles acquired with GNSS receivers, the resulting data and its acquisition method are then not always referable as Floating Car Data (FCD). Literature has numerous terminologies that are not always understandable. Relatively consistent terms are those connected with taxi trajectories, taxi traces or FTD. Due to its already available communication infrastructure for monitoring the vehicles, taxi fleets of urban environments deliver massive FCD collections.

Analyzing traffic data may consist of various approaches. Based on Chen et al. (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) name three different data input classes for traffic data analysis: static traffic sensors, mobile devices, and merged solutions, which include both classes. The most common form of traffic data is trajectory (Chen et al. 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). Besides traffic and mobility analyses, there is the possibility of inferring air quality and emissions using trajectories acquired via mobile devices (FCD and FTD).

The usage of induction loops allows acquiring the real value of the traffic flow rate, which is besides the traffic density and mean speed one of the macroscopic variables of the traffic flow theory. By definition of Yuan et al. (2014), induction loop systems act as tuned electrical circuits. They have a decade-long history in estimating the velocity and the density of vehicles (Lin and Daganzo 1997, Coifman 2003, Yuan et al. 2014). There are still many attempts, which make use of fixed installed induction sensors, especially on highways in dense populated areas of the world (Klein et al., 2006). One continuously developed technique for vehicle observation uses video cameras. Video processing techniques are frequently used methods for traffic flow analysis, which utilize moving object

[8] Ohio State University. "Software Uses In-road Detectors To Alleviate Traffic Jams." ScienceDaily. www.sciencedaily.com/releases/2003/02/030226073009.htm (accessed March 23, 2017).

recognition and tracking technologies (Daeho and Youngtae 2008, Cucchiara et al. 2000, Michalopoulos 1991). Many post processing steps include image processing techniques for deriving vehicle trajectories from airborne imagery (Kurz et al. 2015), detecting vehicle plate numbers (Wang and Tsapakis 2010) or vehicle colors, as for the example of taxis in New York City[9]. Detecting vehicle plate numbers by image processing methods allows the usage of cameras as traffic counters of individual vehicles, when camera sensors are interconnected. This enables computing travel times of individual vehicles. By using aerial or satellite imagery, it is possible to estimate the traffic density by counting all the vehicles on selected road segments. Yuan et al. (2014) mention that static sensors, as loop detectors and video cameras, usually are only sparsely installed. This makes it difficult to infer detailed information on traffic and daily changes in its dynamics. One relatively cheap and flexible method for detecting traffic flows is the use of pneumatic counters. The idea behind those sensors includes the installment of simple pneumatic tubes across the whole width of roads. Products of these acquisitions are calculated (or averaged) Annual Average Daily Traffic (AADT) and the average distribution of different vehicle types (motorcycles, busses, small vehicles). Patterns of pneumatic counter measurements can indicate the vehicle type. This is realizable via installment of subsequently ordered pneumatic counters.

Traffic data acquisition and traffic sensors are content of Listl (2003), who connects the, by this time, very new technology FCD with traditional traffic data acquisition via static sensors. Moreover, Listl (2003) mentions the problems of how to connect the produced FCD with traffic measures from traditional traffic flow theory. This is tricky, since traffic parameter calculation is strongly dependent on the way of data acquisition of the specific sensor. Based on Chen et al. (2015), there are at least three different classes of working modes of sensors: location-based, activity-based and device-based. When working with GPS data, ones should be familiar with recent research that state that measured distances are always overestimated in length, as discovered by Ranacher et al. (2016a). On the other hand, when computing average speeds from instantaneous velocity of FCD records, there is a constant underestimation compared to the average speed from static sensors.

Traffic theories and models - application and development

In the past, many different traffic theories and models evolved that helped finding insights into principles of daily movements: traffic description. Resulting from these findings, planers may benefit in reorganizing and optimizing traffic. In the 2010s, these theories and models are frequently tested with acquired traffic data, from static and mobile sensors. As Nagel and Herrmann (1993) state, one important theory for understanding the appearance of traffic events, is the Traffic Flow Theory. There is as well the Three-Phase-Traffic theory. Traffic dynamics in empirical probe vehicle data can be studied with three-phase theory by Kerner et al. (2013). An important model for traffic simulation is the Nagel-Schreckenberg model (Nagel and Schreckenberg 1992). It is a theoretical model for the simulation of freeway traffic, and bases on the traffic flow theory (Eisenblätter et al. 1998).

Besides theoretical models, there are as well various data driven traffic models, which address different aspects of daily traffic, coming from various installed, daily operating, sensors. Bliemer et al. (2016), for example, investigate traffic assignment models for transport planning. There are as well approaches for travel time estimation models using FCD, which investigate the impact on the traffic situation in connection with social and meteorological events (Jiang and Jiang 2013).

[9] http://dotsignals.org/

Data fusion techniques and modern applications of Intelligent Transportation Systems (ITS)

The term traffic flow may refer to many different descriptions of movement. Especially when it comes to visualization of traffic flow, the meaning of traffic flow is varying. In some freely accessible traffic maps, it means the density of daily driven road partitions, expressed as an average by the Annual Average Daily Traffic (AADT). In others, it is the general driving direction within a street network or they associate traffic density and average velocity as parts of the traffic flow. In this work, the term traffic flow is not used for a specific quantity or parameter; it is used as a general term for all measurable quantities connected with the dynamics of moving objects, in specific of vehicles.

Recent applications of intelligent transportation systems (ITS) focus on connecting sensors, which operate in real-time, as for example the internet or vehicles (IoV). By definition, the Internet of Vehicles (Bazzani et al. 2011) is an open converged network system supporting human-vehicles-environment cooperation (Yang et al. 2014). For Yang et al. (2016), IoV is a hybrid concept with a fundamental in future directions of achieving cooperative and effective ITSs. One important part of work on vehicle traffic, which is important for ITS applications, is coming from the area of traffic physics (Kerner 2004). Physical laws, especially from mechanics, help to provide the understanding and interpretation of acquired sensor data. The most important part lies mainly in data preprocessing: it is possible to provide reasonability checks on the acquired moving object dynamics. Besides this factum, traffic physics focus also on interactions between numerous traffic participants. Selected work focus also on traffic events, as for example, on the physics of a traffic jam (Takashi 2002).

Data acquisition of transportation infrastructure

Traffic data acquisition based on moving vehicles and their cooperative connection to infrastructural element, as in the mentioned IoV example, requires information about the state of road networks. Additionally, there are more general information on the functioning and the hierarchical classification of selected road types. Marshall (2005) describes the nature of road networks by the three road structure components road type or hierarchy, pattern type and route structure. The differentiation by road type is required for guaranteeing correctness of modelling and high quality results of the analysis. This concludes that detailed information on the transportation infrastructure is important and needed for understanding traffic and its principles. The work of Marshall (2005) by the name of *Streets and Patterns* connects different fields of research with similar objects of investigation: road (networks), transport and urban patterns. Those three investigated objects are inspected in urban planning, architecture, geography and transport engineering (Miller and Shaw 2015).

2.4 Vehicle road networks – theoretical aspects and applications

In the early 2010s, online routing applications for vehicle drivers have numerous daily users. There are newly established concepts like Waze[10] and Uber[11] that allow vehicle drivers to communicate with others and share their services and information. Urban vehicle road networks have often more complex connectivity than highway segments or networks in rural areas. Depending on historical urban

[10] https://www.waze.com
[11] https://www.uber.com

developments connected with population growth, economic development and culture areas, geometric, topologic, and structural arrangements or road networks may vary greatly. When talking about transportation networks used for transporting goods and individuals across certain linear features in Euclidean, the most frequently used models are coming from the network theory. Therefore, the representation of classified real-world transportation infrastructure elements appears in network space. As in most common applications with spatial networks, such as in car navigation software or traffic forecast services, routes and flows have a specification as directed and undirected graphs.

For Li et al. (2015a), edges represent road segments and nodes intersections. Consequently, the edges appear between two intersections. This model is widely used and accepted (Barthélemy 2011). The representation of directed and undirected graphs on a network is applicable in combination with traffic information. For the case of FCD, the Map Matching (MM) step is required for connecting instantaneous parameters together with the spatial position to road segments. In case of massive movement data, coming from hundreds or thousands of tracked vehicles, aggregation techniques are widely used. Andrienko and Andrienko (2007) deliver examples for the aggregation of movement of many individual vehicles, where areas of general movement are defined. This is an example for road network generalization resulting in a spatially abstracted transportation network (Andrienko et al. 2016b).

Zhang et al. (2016b) use another form of road network generalization: Starting from binary road information (e.g. black and white bitmaps) the network is simplified by simple image processing operations. By using data aggregation, it is possible to include traffic or movement models into the network space, as it was done by Li et al. (2015a). In most cases, the urban road network is far more difficult to represent than highway infrastructures.

For Bao et al. (2010) and Shen and Xue (2010), the network of an intersection is complicated, especially when connecting tracked vehicle movement positions with its segments. Consequently, it is an important precondition to guarantee the accuracy of the network connectivity and to represent it in a realistic way with real conditions for its usage. In case of urban intersections, the frequent appearance of traffic lights is assumable. The dynamics of these traffic influencing elements can be derived by using FCD, as it was for example tested by Protschky et al. (2015b) and Protschky et al. (2015a). The MIT group of Carlo Ratti introduced so called slot-based intersections, which allow reasonable intersection regulations without any traffic lights (Tachet et al. 2016). For Tang et al. (2016b), intersection in urban environments are critical parts of the road network where different traffic flows converge and change direction. Consequently, they form bottlenecks and clog points (Tang et al. 2016b). Additionally, the variability of travel times at intersections is observable via segmenting FCD into partitions (Keler et al. 2017b). Travel time might be one indicator for efficiency variations of infrastructural segments. Nevertheless, it is challenging to estimate the efficiency of road segments or of whole road networks.

Marshall (2005) introduces road network properties that are connected with the respective route structure and consider particular movement on network segments. The definition of the route structure is not equal for planners and for actual traffic participants: planners describe it by using the attributes hierarchy, continuity, and connectivity of the route. Participants on the other hand, namely pedestrians and motorists, define route structure by their individual perception, which is influenced by real traffic conditions (Marshall 2005). These perceived real-time conditions influence individual route choices. One of the challenges of detecting individual route choices is to model the traffic conditions. Using FCD or other data from mobile traffic sensors for model definition and evaluation is one possibility. The classical approach for road network planners and engineers to considering these time-varying traffic conditions is to calculate network accessibilities that include predictions of collective travel behavior.

Calculation of movement parameters (from FCD) in network space

For Barthélemy (2011), spatial transportation and mobility networks are examples, where the spatial component is relevant and important in addition to the topology information. Additionally, Barthélemy (2011) states that for representing complex circumstances of the real world within transportations networks, there is a need for topological and spatial information, expressed in Euclidean distances. Important, for a designing a traffic pattern analysis framework, it is important to observe the models behind spatial networks. These imply, especially for vehicle transportation networks, the appearance of state transitions between the different arcs of the network (Barthélemy 2011).

Another prominent example with Euclidean space as the dependable spatial network weight is the internet, since physical cables, which have different lengths and latency times, link a set of routers. Besides the basic knowledge on network theory, which bases on the connection between nodes and arcs, there might be a classification of these connected patterns. In case when arcs connect all nodes, so-called well-connected nodes are detectable: many arcs are connected to one specific node. This node is defined as a hub (Barthélemy 2011). Afterwards, the spatial components to these appearances in network space can be associated: What are the exact locations of the usual nodes and the hubs and what are the exact lengths of the arcs or edges (Barthélemy 2011)?

Problems of common online traffic maps

Online traffic maps can give useful information on the traffic situation on highways. Another case is to describe the situation within complex urban road networks. Since the road network of large cities implies elevation differences, its description is often difficult to model as a topological network and hard to visualize in a two-dimensional map view.

The question here is how the FCD can supply this fact with possible useful information derived from certain patterns. In general, this detection has options of evaluation by freely available information on the urban infrastructure, as for example freely available satellite imagery. This may give a certain view on the urban road network for providing a fast association of photorealistic orientation. Google, Bing or Yahoo offer online map services with a merged solution of multiple different satellite imagery sources. The imagery is in this case an additional layer that may complement the thematic map information. Nevertheless, this information mashup may cause obscurity for the daily traffic map user, because the data is insufficiently combined in space as presented in one example of Google maps in Figure 7a. In this example, the road network is classified by traffic status and is not matching with the imagery data. The additional imagery layer is in this case not meaningful. By unchecking the satellite data information, the map view of Google maps (see Figure 7b) appears where only few road elements have classified traffic states. Another negative example of an online traffic map is pictured in Figure 7c, with the road network of Here.com. Due to not available data on the urban transportation infrastructure, this online map includes only highly generalized information about the highways in Shanghai.

In contrast to Figure 7a, mapquest.com delivers correctly projected imagery and road network data. This online map uses OpenStreetMap (OSM) data for the depiction of the road network, presented in

Figure 7d and Figure 7e. The data appears more detailed than in the other online map solutions, because of the visual annotation of the driving direction on every road surface. Besides this, there is no aggregated information. One big disadvantage of mapquest.com is that traffic information is not available for Shanghai (only for cities in the USA).

Figure 7f shows the traffic map of map.baidu.com with very detailed traffic information for each OSM road element. The problem in this online map is the order of the traffic layers, which is not considering the real elevation level of the road elements. Therefore, some non-elevated road partitions appear elevated and the other way round. This problem was already described by Berglund (2001) with an example of one and the same crossing (see Figure 8), represented by a falsified version (Figure 8a) and its correct representation (Figure 8b).

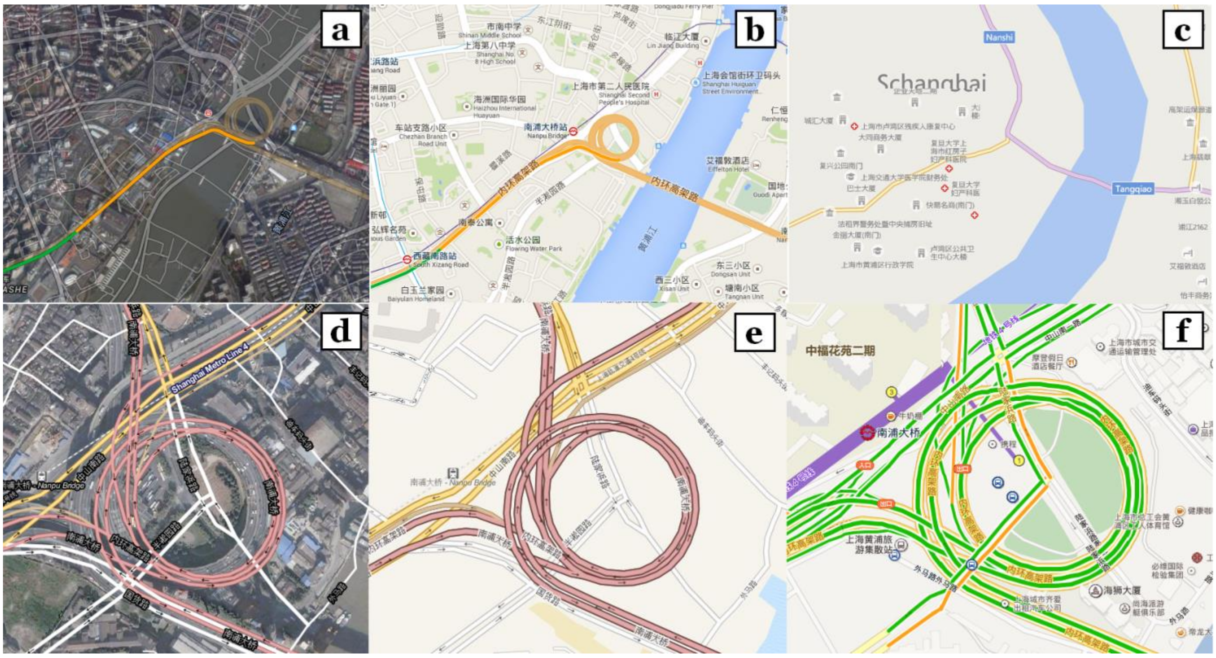


Figure 7: Different examples for traffic map views in Shanghai with (a) Google Maps traffic layer with satellite imagery background and (b) OSM background, (c) Here.com, (d) mapquest.com with satellite imagery and (e) OSM and (f) from map.baidu.com.

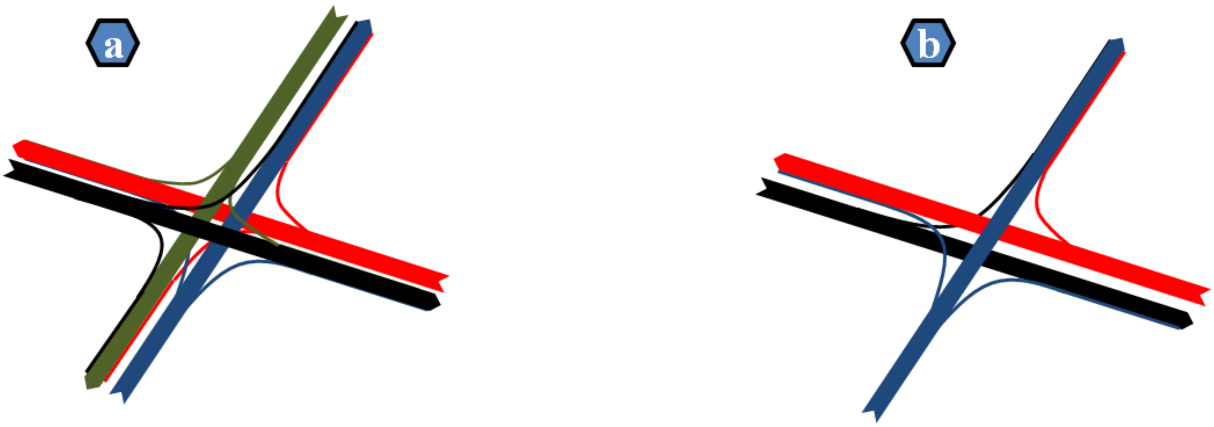


Figure 8: Example representations of the same intersection from (a) standard network from a mapping agency and (b) its correct representation; modified, proposed by Berglund (2001).

All the described disadvantages of common freely available traffic maps will be starting points for designing a more comprehensible traffic map. Against this background, one idea is testing vehicle trajectories on their benefit to plan more understandable traffic maps.

In the 2010s, the ascending requirements of a growing Earth's population of humans and vehicles produce an essential need for providing highly accurate digital information about the transportation infrastructure. This results, not only, from personal views of vehicle drivers (who want confidential information), but as well from the debates of UNO about fulfilling the Kyoto protocol by reducing the worlds exposures into the atmosphere. The last mentioned arises from different statistics with varying percentages caused by private and public vehicles (manned and unmanned). FCD of taxis is usable for detecting the type of infrastructure element, meaning its structure by defining typical traffic flow characteristics and specific vehicle movement patterns. In Keler et al. (2017b), the specific movement patterns are, for example, self-intersections of the same taxi trajectories. Those self-intersections lead to the conclusion of the existence of on-ramps with differing shape (Keler et al. 2017b, Keler et al. 2017a). Fathi and Krumm (2010) show that it is possible to infer road intersections from FCD trajectories.

Dalyot et al. (2013) state that reliable vector geodata is increasing rapidly and often is available as open data. Network data of the same road networks may result in very diverse geodata. There are for example differences in structural, geometric and topological aspects (Dalyot et al. 2013). All there aspects affect the data modelling and further analyses. The example of roads has for Dalyot et al. (2013) the possibility of being represented by areas objects in a cadaster database or by linear objects in a topographic database. The provision of spatial consistency, namely a homogenous unified geodata infrastructure, before the various analyses is needed to guarantee as certain degree of accuracy and reliability of the out coming results (Dalyot et al. 2013).

Based on this problem, Pereira et al. (2009) reason on how to represent road networks for efficient matching FCD on certain part of a road network. One outcome of the matching might be the FCD-based average road link speed v_l^t, from Ji et al. (2014):

$$v_l^t = \sum_{i \in P_{l,t}} s_i \,/\, \sum_{i \in P_{l,t}} h_i \qquad (2)$$

Efficient routing in urban environments can imply different approaches. Some of these approaches find users as daily commuters or people who want to collaborate on avoiding transport time losses. Applications in this sense are coming from the company Waze.

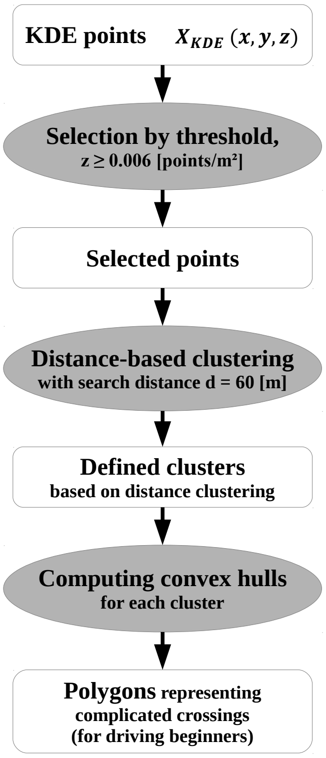
The challenge in this sense is to apply routing on the fly, which means implying awareness of the traffic situation at every position of the route. At every position of the route, it is possible to estimate travel times until the trip destination. One common component of these services is proposing alternative routes for the user, as in the example of Google Maps.

Nevertheless, the average speed information that reveal the traffic conditions for every road segment has varying qualities, possibly resulting from insufficient positioning accuracies of mobile phone users within moving cars. This might result in mismatching to nearby situated road segments and subsequent enriching with falsified road link speeds.

Data-driven detection of perceived complicated crossings

Krisp and Keler (2015) introduce one car navigation solution for avoiding perceived complicated crossings in urban environments. The target group for such services is driving beginners, who are inexperienced and want to avoid confusing crossings.

The method uses the extracted nodes of any input network together with intersection points between the road segments. A perfect connectivity of the segments is not precondition for getting computing results. These extracted nodes are input for a kernel density estimation (KDE). Subsequently, the surface is base for preselection by threshold and polygon creation. Resulting polygons are representatives for complicated crossings and find usage in routing applications as obstacles. Figure 9 pictures a pseudocode of this method together with its workflow.




```
PSEUDOCODE SELECTION BY THRESHOLD & OBSTACLE GENERATION

Input:     Density layer with values z from Kernel density functions f_h (x)
           represented as Points X_KDE (x,y,z) with z as the calculated value
           for the Kernel density function.

Output:    A set of obstacle polygons P (Convex Hulls P)

Function:

Select * from X_KDE (x,y,z)
      where z ≥ 0.006         // unit: points per square meter, [points/ m²]
      X_KDE,thresh  =   S     // Selected Points

d = 60                         // distance d; unit: meter, [m²]
T = Aggregate (S, d)           // Clustering based on distance d
k = 0

for each R in T do             // T = Groups of selected points

pointOnHull = leftmost point in R     // creation of obstacle polygons P
i = 0
 repeat
   C[i] = pointOnHull                 // C contains set of boundary points
   endpoint = R[0]  // initial endpoint for candidate edge on the hull
   for j from 1 to |R|
         if endpoint = pointOnHull
               or R[j] is on left of line from C[i] to endpoint then
            endpoint = R[j]   // found greater left turn, update endpoint
         end if
         i = i+1
         pointOnHull = endpoint
   end for
 until endpoint = C[0]         // reached first hull point
P[k] = createPolygon(C)        // connect points of C by lines
k = k+1                        // update polygon count for next iteration
end for
```

Figure 9: Workflow and pseudo code for detecting complicated crossings.

Besides its use for driving beginners this method shows promising results for connecting perceived street complexity with daily mobility patterns (Keler and Krisp 2016a), and with recurrent traffic congestion (Keler 2017).

2.5 Traffic flow theory

These physical relations in traffic dynamics mainly base on the relations coming from the traffic flow theory. The main components are speed, density and flow rate on the one hand (movement of vehicles) and road infrastructure properties on the other.

In general, the traffic flow theory distinguishes between microscopic and macroscopic traffic flow variables. Microscopic variables refer to individual movement of one entity and are mainly velocity v and acceleration a. Both microscopic variables refer to one trajectory or a partition of it. The formal description of velocity v is as the following:

$$v_a(t) = \frac{dx_a(t)}{dt} \qquad \left[\frac{km}{h}\right] \tag{3}$$

The acceleration a can be either positive (acceleration) or negative (deceleration) with the following formula:

$$a_a(t) = \frac{d^2x_a(t)}{dt^2} \qquad \left[\frac{m}{s^2}\right] \tag{4}$$

Microscopic variables describe often the instantaneous properties within individual movement trajectories. As looking inside a trajectory, it is possible to read for every position point of an individual trajectory instantaneous velocity and acceleration values. Additionally, other instantaneous values of movement are not necessarily connectable with traffic, but with the movement itself. There is for example the driving direction, which is important for distinguishing of the passed road segment direction or lane. This may result in an improved matching of the vehicle positions to the right road segment.

Besides looking inside individual movement trajectories, it is possible to aggregate movement and to observe traffic in a bird's eye by using macroscopic variables of the traffic flow theory.

When observing numerous vehicle trajectories, so called bundles of trajectories, the first idea is to count vehicle positions at selected parts of the road network. Therefore, traffic density k, a macroscopic variable, is used:

$$k = \frac{n}{\Delta X} \qquad \left[\frac{vehicles}{km}\right] \tag{5}$$

n is the number of vehicles, which pass through a location x, within ΔX, which is a certain location area as for example a road section.

Even when not in the formula, k is not only dependent on the location x, but also on time t and the measurement interval S. Subsequently traffic density k might also be expressed in the following way: k(x,t,S) = n / ΔX [vehicles/km].

The maximum traffic density k of an ordinary road varies for Immers and Logghe (2002) within the range of 100 vehicles per kilometer per (road) lane. This definition might appear useful for highways with numerous road lanes. The dependency on certain location interval ΔX concludes that formula 5 might be converted. When assuming time interval dt as infinitesimal, then traffic density k might be converted into the following time interval independent expression:

$$k\,(x,t,S) = \frac{n}{\Delta X} \approx k\,(x,t,S) = \frac{n * dt}{\Delta X * dt} = \frac{\text{total time spent by all vehicles in S}}{\text{area (S)}} \tag{6}$$

Another macroscopic variable is traffic flow rate, which appears consistently as q throughout the literature. Besides this reference many other terms for the traffic flow rate can be found: traffic count, traffic intensity or even traffic volume (Immers and Logghe 2002). Traffic flow rate q is defined by the following formula:

$$\mathrm{q} = \frac{\mathrm{m}}{\Delta \mathrm{T}} \qquad \left[\frac{\mathrm{vehicles}}{\mathrm{h}}\right] \tag{7}$$

Traffic flow rate q is dependent on location x, time t und measurement interval S: $q\ (x,t,S) = \frac{m}{\Delta T}\ \left[\frac{vehicles}{h}\right]$. Within the formula, m is the number of vehicles, which pass through a location x, within a certain time range ΔT. This dependency on time interval ΔT makes a conversion into an infinitesimal location interval dx possible:

$$q\ (x,t,S) = \frac{m}{\Delta T}\ \left[\frac{vehicles}{h}\right] \quad \approx \quad q\ (x,t,S) = \frac{m * dx}{\Delta T * dx} = \frac{total\ distance\ covered\ by\ vehicle\ s\ in\ S}{area\ (S)} \tag{8}$$

The maximum possible traffic flow rate q of a street is referred to as capacity. Consequently, capacity q is a threshold that varies for every different road segment. Capacity q of a highway varies approximately between 1800 and 2400 vehicles per hour per (road) lane (Immers and Logghe 2002).

Taxonomy of features for defining various traffic phenomena

Traffic analysis consists in a simple and general form out of inspecting the parameters of the traffic flow theory. For Qiu et al. (2010), who follow up on this idea state that there are three critical parameters, namely density, speed and flow. All three parameters refer to vehicle, automobile traffic. The calculation with these values depends as well on the way these parameters are measured (or computed), which is connected with the vehicle traffic acquisition devices. First, it is to clarify that speed or velocity has many possible definitions. There are at least three different possibilities for measuring, computing and representing vehicle velocities. Depending on the point of view of vehicle traffic observation, different parameter values are measurable for same areas of investigation. There is the possibility of computing the space-mean-speed, which refers to an observation of a road segment or lane with a certain length within a certain time window.

$$\mathrm{u\ (x,\ t,\ s)} = \frac{1}{n} \sum_n v_i \tag{9}$$

Another possibility is the time-mean-speed, which does not have a spatial component. The third Alternative calculation of the average velocity implies the combination of traffic flow rate q and traffic density k as a ratio. Mean speed u has the following definition:

$$\mathrm{u = q\ /\ k} \tag{10}$$

As reformulation of this equation results in the fundamental relation of traffic flow theory:

$$\mathrm{q = k * u} \tag{11}$$

The knowledge about two of the three variables allows the calculation of the remaining variable (Immers and Logghe 2002). This refers to the fundamental diagram of traffic flow, which describes the theoretical relationship between flow and density with its different shapes (Bliemer et al. 2016).

Capacity and Volume over capacity (VOC)

Capacity of road segments can be dependent on the road type, the number of respective lanes and its dimensions, namely widths and sizes. By definition, the capacity is the maximum traffic flow rate q of a selected road segment or segment lane element (Immers and Logghe 2002). Road capacity is inferable for each road segment of the investigation area. By measuring and estimating capacity, the real road capacity value can serve as an indicator for the traffic state. By comparing those two values, it is possible to define a road congestion index as in Qian et al. (2006):

$$\text{road congestion index} = \frac{\text{indicated road capacity}}{\text{road capacity}} \quad (12)$$

For Qian et al. (2006), the denominator road capacity shows the road supply, whereas the enumerator indicated road capacity is the road demand. By comparing the two capacities in one selected road segment can indicate, if the road is crowded (with vehicles) or not.

Relating trajectories with macroscopic parameters of the Traffic Flow Theory

In a simple and understandable way, Saberi et al. (2014) define network traffic flow variables and road link-based measurements for each road segment. Each segment includes the average network flow, the average network density, and the average network speed. The challenge consists of using individual vehicle movement trajectories for enabling computing with traffic flow variables. Connecting single entity movements with the traffic flow theory variables is feasible via Time-Space diagrams. For understanding this connection, Figure 10 shows a simplified view on how to handle these cases. In Figure 10, there are trajectory-based measurements in the form of curved lines that sometimes cross or traverse a region A.

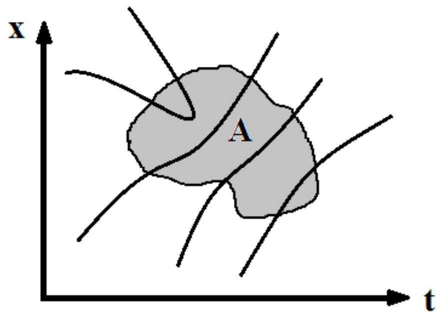


Figure 10: Two-dimensional trajectories within a Time-Space diagram, based on Saberi et al. (2014).

Based on this traversed appearance, q(A), k(A), and v(A) can be defined, which are namely flow, density, and speed for observed vehicles in region A. This shows that these measures are dependent on previous knowledge of A, which is not only a region A, but has also a temporal component. This results that A has the possibility of being a spatiotemporal event.

2.6 Defining traffic congestion – a review

Traffic congestion is a complex traffic phenomenon with various properties. These properties vary in many appearances of daily traffic congestion events worldwide. Along with different perception of traffic congestion, there are different definitions and measurements. The reason for traffic congestion events can have different causes that result in different consequences for the traffic participants. Therefore, this subsection has the aim of reviewing different research connected with this phenomenon. In a first part, there is a discussion on theories and taxonomy behind traffic congestion. The second part of this subsection shows possibilities for detecting, classifying, and representing traffic congestion. In the third part, there is an introduction of the term traffic bottleneck with its connection to an urban road network. This has the idea to reason on the dynamic changes of traffic states in urban vehicle road networks; more precisely, how to connect local traffic flows with the global traffic flow. A short outlook outlines the topic traffic forecast, together with a critical review of recent traffic congestion measures and how urban traffic situations are developing.

2.6.1 Theoretical background and taxonomy

Besides the terms movement data, vehicle trajectory and movement pattern, there is the term traffic congestion. There are various different theoretical backgrounds of traffic congestion. Consequently, parts of the taxonomy differ, especially in domain-specific terminology. This might be confusing, since similar terms as capacity might have different formalization and influence on those events that are actually defined as traffic congestion. In general, traffic congestion can be associated as an event with a specified start and an end. An event is different from a process: the latter changes only its state but is still ongoing (Galton 2008). That is why a congestion event might be a part of a process and the process refers to traffic in general or for a selected area. Besides this temporal component of traffic congestion, there is a more general definition by Qian et al. (2006), who refer more to the built environment that is partially restricted for vehicle drivers. Therefore, Qian et al. (2006) refer to traffic congestion as a result when road demand exceeds road supply. This definition of traffic congestion makes a connection to the capacity of each road and refers to the term traffic bottleneck. In research, there is a plethora of different definitions for traffic congestion (Sofer et al. 2012, Aftabuzzaman 2007). The main reason for this is the assumed existence of many different types of traffic congestion. These events may appear periodically as recurrent traffic congestion or might be caused by accidents (Okabe et al. 2009, Xie and Yan 2008), or by slowly moving heavy goods vehicles.

Tang et al. (2016a) define traffic congestion as a linear event, since its movement is restricted to linear road segments. Besides recurrent congestion (RC) and non-recurrent congestion (NRC) (Anbaroglu et al. 2014, Anbaroğlu et al. 2015, Anbaroglu et al. 2016), there are as well phantom jams that appear suddenly without any detectable previous cause (Sugiyama et al. 2008, Flynn et al. 2009). Anbaroğlu et al. (2015) state that non-recurrent congestions (NRC) cause more frustration than recurrent congestions (RC), since they are not predictable.

Theoretical definition of traffic congestion

The theoretical definition of traffic congestion of Oguchi (2015) bases on the traffic flow theory. Oguchi (2015) distinguishes between indications of a congested and an uncongested traffic state. The congested traffic state occurs in theory, when traffic densities are above critical densities and at the same time, speeds are below critical speeds. The case of an indicated uncongested traffic state is the opposite of the mentioned. The biggest body of research is coming from the transportation sciences, with their various subfields. Another body of literature is coming from urban and transportation planning. Other research is coming from social sciences and implies to focus on the social impact of traffic congestion or in general its influence on daily commuters. Research on traffic congestion is very important, since traffic congestion is perceived differently in different cultures and has a great impact on today's urban societies. Starting with the historically first form of research about traffic congestion origins from the time, where automobiles already influence bigger cities: 1950s with the very first bigger congestion events that are comparable with those of the 2010s. Although, the imagination of vehicle traffic congestion appeared much earlier and resulted from the first ongoing coach and horse-drawn carriage congestions in the early 1900s or even late 1800s. After selected parts of the Worlds' vehicle transportation infrastructures reached its maxima, the problem could be observed by a high partition of citizens, e.g., urban dwellers. There is a big difference between inspecting traffic congestion at a highway level, where the focus is on connections between selected cities, and the traffic congestion inspection within a complex urban environment with many additional modes of transport. For differentiating between travel modes, Qian et al. (2006) apply a system dynamics approach, one methodology for studying and managing complex feedback systems, to understand the complex traffic system of Shanghai. One key component for understanding urban dynamics is to measure impacts of different traffic participants as for example the group of private vehicle owners. Here it is interesting to question the road construction approaches for solving the problems. Sofer et al. (2012) and Aftabuzzaman (2007) mention some historical definitions of traffic congestion that are defined as the following. One often-connected principle in these definitions is the traffic flow theory.

Practical traffic congestion definitions and measurement possibilities

Overviews about traffic congestion definitions and measures are part of the work of Turner et al. (1996), and of Bertini (2005). Bertini (2006) reviews the definition and measurement of traffic congestion by using the results of a survey about urban traffic congestion. The survey received 480 responses from transportation professional and academics. Figure 11 show the responses for the first questions of the survey. These are the fundamental questions of how to define (Figure 11a) and to measure (Figure 11b) urban traffic congestion.

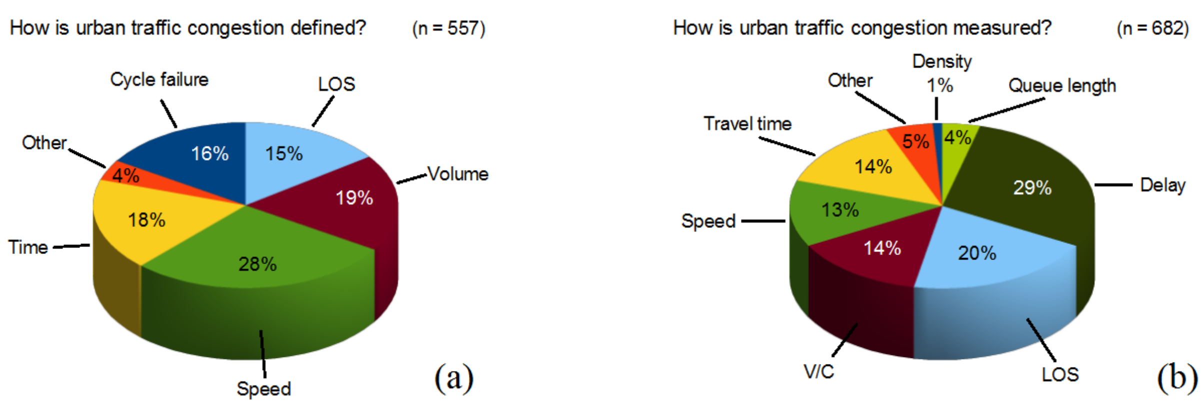


Figure 11: Survey results from experts about (a) defining, and, (b) measuring urban traffic congestion, based on Bertini (2006).

The responses in Figure 11 show that there are higher differences in defining and measuring urban traffic congestion. This is one motivating point of this work to include numerous measures from Figure 11b for introducing alternative definitions of traffic congestions, as part of the traffic pattern analysis framework. Most experts define urban traffic congestion via speed (28%), volume (19%) and time (18%) as pictured in the pie chart in Figure 11a.

There are 16%, who define it via cycle failure, which is difficult to observe. The cycle failure focuses on the signs and operational parameters of light-signal systems. These features, usually installed at road intersections, influence the dynamics of traffic in providing coordination. During rush hours, numerous vehicles occupy many roads to a big proportion. It appears that queued vehicles at traffic lights are not able to leave within one cycle and loose large time windows. In some cases, the traffic congestion reaches even the signalized intersections. One reason for this kind of appearance might be the cycle failure, resulting from unsuitable adjustments of the light-signal system parameters that do not allow lowering the urban traffic flow. Simple indications of cycle failure appearances are the queue lengths of traffic congestion events, which result as longer as the lengths of affected roads. As a result, the traffic congestion events influence numerous connected road segments.

15% of the experts define urban traffic congestion via level of service (LOS). This definition (15%) and measure (20%; see Figure 11b) comes from official manuals about highway capacities in different states of the USA. The key of LOS is its detailed classification of perceived traffic flow rates between the letters A and F. The letter F stands for the worst traffic situation and A for the best. There are specific standards for the definition of each class in specific manuals and books, such as the Highway Capacity Manual (HCM) and AASHTO Geometric Design of Highways and Streets (Green Book), In general, LOS has more focus on the traffic flow rate and on how people perceive different values of the traffic flow. Therefore, LOS has less connection to average speeds or traffic densities.

Besides defining traffic congestion, there are answers of the questionnaire of how to measure urban traffic congestion pictured in Figure 11b. The highest numbers are for the measurement of traffic congestion via delay with 29%. Delay computations consist usually of the difference between usual travel time and travel time during congested traffic. At the second ranking place, LOS has 20%, which summarizes methods that classify traffic flow rate in general.

Rank number three with 14% is the ratio V/C, which is volume over capacity. This ratio shows when volumes of vehicles on road segments are exceeding the value associated with its capacity. This traffic volume value results from the maximum possible traffic flow rate value on a specific road segment. Therefore, lower V/C values indicate free flow. On the same rank as V/C is also travel time (14%), which might indicate the travel delay, when knowledge of usual or typical travel times is available.

Only 13% of the experts believe that speed is a measure for urban traffic congestion. One specific detail of the questionnaire from Bertini (2006) is the 1% of experts, who prefer the density of vehicles or the traffic density as the measure for urban traffic congestion. This shows that vehicle or movement position densities do not deliver to provide higher correlations with traffic congestion events. On the other hand, there are the traffic densities, which are possibly more difficult to obtain in urban environments than at highway networks.

4% of the experts use queue length as the measure for urban traffic congestion. 5% propose another measure that is not connected with the ones mentioned in Figure 11b. The queue lengths of traffic congestion events have also a connection in thematic with cycle failures, since one possible approach to compute cycle failure might be comparing queue lengths and road lengths.

Causes of traffic congestion

The Transportation Research Board defines in the year 2003 seven sources of traffic congestion. The Federal Highway Administration of the U.S. Department of Transportation[12] divides these seven them into the following three categories:

1. Traffic influencing events; with traffic accidents, work zones, and weather
2. Traffic demand; with fluctuations in normal traffic, and special events
3. Physical highway features; with traffic control devices, and physical bottlenecks (capacity)

The three different groups of causes have more focus on causes at highway networks, especially in USA. Nevertheless all three categories are applicable on complex urban environments. In many cases of complex road networks, the three categories are referring to ring highways and other roads that are frequently used. Category 1 refers more to unusual events that cause traffic congestion, such as traffic accidents. Traffic accidents can influence the traffic situation in numerous ways. Additionally, there are different ways of how traffic accidents affect different road types, especially in case complex road segment connections. In general, it is challenging is to estimate the impact of an accident on an urban highway segment that has good connectivity to minor roads. In a similar way planned work zones can cause delays on roads at specific times of the day, as well in a periodical manner. The third example for category 1 is the influence of the weather, which can be dependent on the investigation area.

Category 2 shows the causes that are often very difficult to substantiate, since they connect vehicle traffic with human mobility patterns. By respecting the population densities of the investigation areas, it is possible to distinguish between typical rush hour traffic congestion and these events resulting from special events. The latter can emerge as large social events that attract tourists. This causes variations in the usual traffic patterns, since partially more traffic participants interact on the road networks.

The third category refers to the features that minimize road capacities physically, as closures of lanes or inefficiently built transportation infrastructure. This category refers to static bottlenecks, namely an aspect of a certain feature of the road network that causes periodical traffic congestion events.

Congestion length, factors, and propagation

Congestion length is to differentiate from queue length, since it refers only to traffic congestion events. The length of a traffic congestion event might be frequently varying in time or being relatively static. The measure of congestion length can be very useful in computing the vehicle density. The term congestion propagation refers to the appearance that the spatial dimension of certain traffic congestion events is changing over time. In a figurative example the core point of one congestion event, for example the location with the highest vehicle density and lowest average velocity, is moving along parts of certain road segments.

Specific traffic congestion propagation can be a periodical event on working days in urban road networks, namely recurrent traffic congestion. There is also the possibility of non-recurrent traffic congestion events as for example the unusually long (in time and space) event in Northern China considered being the heaviest traffic congestion event ever[13]. Periodical traffic congestion events

[12] https://ops.fhwa.dot.gov/congestion_report/

[13] "Monster traffic jam ... again" article by He Dan and Wang Qian in the China Daily Updated: 2010-09-04 07:35, accessed September 4, 2010

might indicate static traffic bottlenecks in selected parts of the urban transportation infrastructure. The spatial propagation of traffic congestion might then indicate possible dynamic traffic bottlenecks.

2.6.2 Traffic congestion detection, classification and representation

Definitions of traffic engineering include the use of the traffic flow theory for defining traffic congestions for highway segments. This appears quite detailed by using macroscopic fundamental diagrams, where beginning and ending of congestion is detectable (Daganzo and Geroliminis 2008). Andrienko et al. (2015) define traffic congestion by the ratio between speed and vehicle density. The combination of those two parameters can tell general traffic states, when aggregated in space and time for the later geovisualization. Yuan et al. (2014) mention the internationally recognized problem of traffic congestion on urban expressways. This shows that, even if constructed for massive flows of vehicle, the capacity of urban expressways is reached in usual daily situations. As a consequence, the operational efficiency of various urban transportation systems is decreasing, which has for Yuan et al. (2014) also effects on increasing air pollution or exposures. As an application that would help to decrease traffic congestion, Yuan et al. (2014) propose the use of received on-board real-time traffic information, which has the potential of possible improving of road network efficiency. Traffic congestion detection and information dissemination are key issues of providing traffic information service (Yuan et al. 2014). Yuan et al. (2014) mention also the extended, alternative approaches in detecting traffic congestion. There is the possibility of Vehicular Ad-hoc NETworks (VANETs) that make use of the cooperative vehicle-infrastructure inside selected vehicles (Yuan et al. 2014). This makes it possible to collect and exchange traffic information between vehicles in real time (Yuan et al. 2014). Unfortunately, this method is successful on highways with clear topologies and relatively long road segments. For the case of urban highways with major and auxiliary roadways in dense urban environments VANETs do not perform well for Yuan et al. (2014), since they often depend on a GPS-based vehicular network (Li et al. 2007b).

Movement pattern classification of traffic congestion and its perception

Traffic congestion is a transportation-related problem. For Jun et al. (2006), traffic congestion is directly connected with the frequency of vehicle usage and with individual driver behavior. When drivers get stuck in traffic congestion events, there are only restricted options of movement. This appearance correlates with a certain type of event interpretation, as defined by Lazarus (1966) as undesirable and taxing on personal resources. As a conclusion, experiencing a traffic congestion event as a driver might cause psychological stress (Hennessy and Wiesenthal 1999). Additionally, Jun et al. (2006) state that there is a variety of other problems related to transportation and driving behavior such as crash frequency, energy consumption and vehicle emissions. Based on the previous assumptions of Jun et al. (2006) that traffic congestion is dependent on individual driving behavior and vehicle usage rate, there are already some implemented attempts to solve this problem by introducing driving restrictions as a management strategy. The main aim of these implementations is to force the vehicle drivers to change their usual driving behavior. Besides these, there are historically used traffic performance measures for the qualitative description of traffic.

For Sofer et al. (2012), one focus of governmental and research institutions is finding methods for the reduction of traffic congestion. Therefore, numerous definitions of traffic congestion measures and traffic performance measures appeared since the 1960s. One selection of these measures, which define and measure vehicle traffic congestion are represented in Table 1. This Table 1 has three columns showing in the first column the references of the selected studies. The second column pictures the used measure for traffic state detection and the third column the eventual threshold for traced traffic congestion. Despite all these information, there is no indication for using these measures for highways or urban road networks. One special case, within these definition is those of Robinson (1984), who states that vehicle emissions of hydrocarbons and carbon monoxide are especially high in speeds under 20 km/h. Therefore, the speed value of 20 km/h serves as a threshold for associating traffic congestion.

Governments and researchers focus on ways to alleviate traffic congestion. For reasoning on this attempt, the measures in Table 1 are useable for various different types of traffic analyses. D'Este et al. (1999) introduce in their work the term quantitative congestion measures, that are different from the individual perception of congestion, since everybody has different scales of traffic congestion severity. This means that different cultural aspects, together with the way of constructing transportation infrastructure, influence the traffic congestion perception.

Table 1 shows that besides the measure speed, there are numerous studies, which involve the flow rate or volume of traffic for as measures for traffic congestion. Besides those, there are global indices, which show the general traffic states for selected investigation areas. Another measure is the similarity of traffic states by Qi et al. (2016), who focus on finding typical and comparable patterns for traffic congestion definition.

Table 1: Selection of introduced vehicle traffic congestion measures, mainly influenced by the listing in Sofer et al. (2012).

Study	Measure	Threshold for Congestion
Skabardonis et al. (2003)	Speed	Below 60 mph
Kwon et al. (2006)	Speed	Below 60 mph
Polus (1996)	Occupancy	Above 30%
Cottrell (1993)	v/c	Equal to 1
Lindley (1986)	v/c	Equal to or larger than 0.77
Lorenz and Elefteriadou (2001)	Speed	Average speed in all lanes below 90 km/h
Polus and Pollatschek (2002)	Flow	Flow changes from dense to unstable
Sofer et al. (2012)	Critical Occupancy Point (COP)	Dependent on place, lane and presence of trucks (“moving obstacles”)
Keler et al. (2016), Keler (2015)	Congetion Value c	3 ranges, divided by typical instantaneous velocity distribution on rush hours
Robinson (1984)	Speed	Below 20 km/h
Qi et al. (2016)	Similarity of traffic states	Metaheuristic jam clustering
Qian et al. (2006)	Road congestion index	$\frac{\text{indicated road capacity}}{\text{road capacity}}$
TomTom,Cohn and Bischoff (2012)	Global congestion index	Congestion Level [%]
Wen et al. (2014)	Global, macroscopic and dynamic index	Traffic Performance Index (TPI)
Schrank et al. (1993)	RCI	Equal to or larger than 1.0

D'Este et al. (1999) and Taylor (1992) state that factors reflecting the level of traffic congestion might come from measures like travel delays, travel times, travel costs, queuing at intersections and selected road segments, number of incidents and accidents, the level of energy consumption and the level of pollution.

For Taylor (1992) and D'Este et al. (1999), the more accurate congestion indicators are based on the quantities travel time and speed. Table 2 summarizes nine congestion indicators with the way of calculation and short descriptions of the outcomes, additionally complemented by other sources.

Table 2: Quantitative traffic congestion measures, proposed by D'Este et al. (1999), and complemented by others.

Name	Calculation	Description	Reference
Travel time t_{travel}	$t_{travel} = t_{start} - t_{end}$	Time between beginning and end of journey	D'Este et al. (1999), Taylor (1992)
Average speed $\overline{v}$	$\overline{v} = {d_{travel}}/{t_{travel}}$	Calculated by dividing the total distance travelled by the time taken to travel that distance	D'Este et al. (1999)
Congestion index C	$C_{index} = {(C - C_0)}/{C_0}$	Proportion of state transition between free flow and congestion	D'Este et al. (1999)
Time Moving Cr	Counting and saving times of within trajectories or based on average speeds	Time when travel speed is greater than zero	D'Este et al. (1999)
Proportion Stopped Time F_s	$F_s = {C_s}/{C}$	C_s is the stopped time, which is part of travel time, where vehicle is stationary C is the total travel time	D'Este et al. (1999)
Acceleration Noise	Acceleration of deacceleration	Represents dynamics state transitions	Underwood (1968), Taylor (1992)
Mean Velocity Gradient	Gradient of average velocity	Represents changes in mean velocity	D'Este et al. (1999)
Relative delay rate RDL	$RDL = \frac{[actual\ travel\ rate\ (\frac{min}{mi}) - acceptable\ travel\ rate\ (\frac{min}{mi})}{acceptable\ travel\ rate\ (\frac{min}{mi})}$	Delay in travel time, with travel rate as travel time per segment length or the inverse average speed	Lomax et al. (1997)
Fuzzy inference	1. computing input parameter values 2. classifying input values into groups 3. defining different congestion states 4. determining congestion index	Fuzzy inference system for determining composite congestion index	Hamad and Kikuchi (2002)
Volume-Capacity (V/C) ratio R_{V-C}	$R_{V-C} = {V}/{C}$	Ratio shows the efficiency of each road segment	Dewees (1978)
Congestion Value c	$c = \bar{k} / \bar{v}$	Ratio shows proportion between vehicle density and average velocity for each road segment	Keler (2015), Keler et al. (2016)
Local congestion value iL	$i_L(l) = \begin{cases} 1 & \text{if } \bar{v}(l) < c_t v_0(l) \\ 0 & \text{if } \bar{v}(l) > v_0(l) \\ \frac{v_0(l) - v(l)}{v_0(l) - (1 - c_t)} & \text{else} \end{cases}$	Shows local congestion values of affected roads without looking at global scale; only dependent on NN	Ulm et al. (2015)
LoC (level of congestion)	LoC = TCE (density, average speed)	Shows relation between density and average speed in any traffic congestion estimation (TCE) model	Yuan et al. (2014)
Congestion pockets	Clustering of connected arcs with similar velocity values (estimated via similarity matrix)	Defined as a connected part of the road network which consists of congested segments, for a certain time component	Rempe et al. (2016)
Speed performance index R_v	$R_V = {v}/{V_{max}} * 100$	Ratio between average speed and maximum permissible speed on a road, multiplied by 100	He et al. (2016)

There are the quantitative measures of traffic flow or its performance. There is for example the number of stops, the travel speed or the bandwidth. The last mentioned term includes the maximum quantity of green zones for scheduled movement (Graser et al. 2012). Another group consists of derived values as for example level of service (LOS) or delay (Graser et al. 2012). The LOS consists of six classes, which are named from A to F. For Hamad and Kikuchi (2002), LOS is the traditionally most popular measure of traffic congestion. It consists of the ratios between the supply and the demand of traffic on roads, from which six classes are inferred (Hamad and Kikuchi 2002).

Traffic congestion classification by its appearance

Traffic anomaly detection consists mainly of Data Mining applications for FCD (Chawla et al. 2012). One central aspect of these applications is to distinguish between typical or recurrent traffic congestion and abnormal or non-recurrent traffic congestion (Anbaroğlu et al. 2015). This subsection is about the principles of Recurrent Congestion (RC) and Non-Recurrent Congestion (NRC) influenced by research from Anbaroglu et al. (2014), Varaiya (2007) and Dowling et al. (2004). In general, it is possible to differ between vehicle traffic congestion that is recurrent and non-recurrent. Recurrent traffic congestion (RC) is typical for urban environments. During the so-called peak-hours, commuters travel to working places and achieve the maximum capacity of selected road segments. For D'Este et al. (1999), recurrent traffic congestion is resulting from predictable variations in the level of demand associated with daily commuter patterns. In particular, it is possible to predict morning and evening peaks of commuting and other recurrent events. Recurrent events are of course investigation area dependent. The factors affecting the appearance frequency of recurrent congestion can imply the population density, the built area distribution and the level of motorization.

For Anbaroğlu et al. (2015), the non-recurrent traffic congestion (NRC) is causing unexpected delay for commuters and other participants of the daily traffic. In contrast to the previously mentioned recurrent traffic congestion (RC), NRC is caused by unexpected events. As a result of this unexpected delay the level of frustration for commuters and traffic operators is high (Anbaroğlu et al. 2015). The examples for such events causing NRC are various. Anbaroğlu et al. (2015) mentions traffic accidents, vehicle breakdowns, planned engineering works, public events and (suddenly changing) weather conditions (Kwon et al. 2006). The difficult aspect of these events is the causal connection with congestion, in particular as a component of congestion. This complicates any formal definition of NRC. This makes it difficult to formalize NRC. Besides the difficulty to infer NRC from events (characterized by spatial and temporal components) it is hard to classify NRC events based on duration, timing and location (Anbaroğlu et al. 2015). This makes it difficult to distinguish between RC an NRC. For Anbaroğlu et al. (2015), research on detecting NRC is recent and may provide new important insights for traffic operation centers for understanding the impact of these events on the daily traffic. This research is important as it benefits its measurement and formulization (Hallenbeck et al. 2003). Previous research has shown that NRC leads to longer travel times (Anbaroğlu et al. 2015) and this makes NRC the major source of travel time variabilities (Noland and Polak 2002). Dowling et al. (2004) introduces a methodology for measuring recurrent and non-recurrent traffic congestion. On the other hand non-recurrent traffic congestion appears after unexpected events. These events may be caused by vehicle accidents on roads, extreme weather events, or building sites.

There is research on detecting, so-called, traffic anomalies in selected parts of the road network. Many established methods that are entirely different in their basics, have one aspect in common: using historical traffic information to detect the “usual” (often on weekdays) traffic pattern. After this definition, recent data or data streams can be used to differ between the traffic states. Lan et al. (2011), for example, is using counted vehicles on road segments (after map matching) for calculating vehicle densities and its standard deviations for selected hours of selected days. Later, these standard deviations are used as thresholds for distinguishing between usual and abnormal traffic patterns.

The term phantom jam as defined by Sugiyama et al. (2008) and Flynn et al. (2009) describes traffic congestion that appears without bottlenecks. Consequently, the so-called phantom traffic jams are special cases of non-recurrent traffic congestion. The appearance of such events was proven in experiments explained in 2008, where participants drive in a circle (Sugiyama et al. 2008). After a certain time congestion appears due to small fluctuations of variations in driving behavior[14]. The following formula shows these principles:

$$\rho_L = \frac{\rho_M}{2}\left(1 - \sqrt{1 - \frac{4\beta}{u^2}}\right) \tag{13}$$

Interesting is the measure β, since it is a measure of road condition and driver behavior. In general, β is dependent on the time an individual vehicle needs to slow down its velocity. This is dependent on road segment curvature and weather conditions. If β is small, fever participating vehicles are needed to establish a traffic congestion event. By respecting group dynamics of numerous vehicles it is even possible to describe such complex events as ‘phantom jams’. Those events can be tested on real time conditions (Flynn et al. 2009) and under prepared testing conditions (Sugiyama et al. 2008).

Global traffic congestion indices

For Wen et al. (2014), the cardinal rule for alleviating traffic congestion is to scientifically and quantitatively describe traffic congestion. One possibility is to make use of global traffic indices for describing traffic state changes within a whole day for the whole city. This idea was formulated by numerous previous research (Sun et al. 2009) and resulting in defined technical standards. In 2011, Beijing was officially releasing the technical standard for the dynamic description of global traffic congestion: urban road traffic performance evaluation indices. Those indices are explained as a unity of theoretical methods for measuring dynamic traffic congestion characteristics of the whole road network. For Wen et al. (2014), urban traffic congestion has different typical characteristics and is influenced by different conditions (such as certain days of the week, holidays and weather). One necessity is the setting up of relationships between traffic congestion and the mentioned influencing factors. In their approach, Wen et al. (2014) conduct a microscopic traffic analysis on the causes of traffic congestion. They define TPI patterns that indicate that traffic congestion has inherent characteristics. Their practical case study is carried along with a real-time FCD processing system, which updates the FCD positions every minute, average speed on road segments every 5 minutes and every 15 minutes a global traffic state estimation for the whole city road network. Global indices for traffic flow are vehicle dynamics of a whole city, expressed by the Traffic Performance Index (TPI) (Wen et al. 2014). Traffic congestion performance measures express more the level of congestion and are for example expressed by the Traffic congestion index (Sun et al. 2009). The traffic performance index was developed in 2007 and is a dynamic macroscopic index (Wen et al. 2014).

Besides the mentioned global indices, there are also local traffic congestion parameters as for example, the Traffic Congestability Value (TCV) by Patel and Mukherjee (2015) or the roadway congestion index (RCI) by Schrank et al. (1993). There is a big variety in global and local traffic congestion

[14] http://360.here.com/2016/01/18/the-physics-of-traffic/

indices. The reason is that they should serve helpful tools for potential analysts. The indices have often different input data and show different aspects of traffic congestion. Yang et al. (2015), for example, are using Traffic Flow Prediction Indices for representing the reliability of traffic flow detection. Its assignment is performed via a specific optimization technique. Other indices show the performance of vehicle traffic. Sun et al. (2009) define the macroscopic variables speed, volume, and density as classical parameters for the description of traffic congestion. Their advancement consists of five different traffic congestion performance measures, which have the depiction as symbol as $K_1$to K_5, as pictured in Table 3.

Table 3: The five traffic congestion performance measures by Sun et al. (2009).

Characteristic sort	Code	Index name
Congestion intensity	K1	Traffic congestion index
Congestion spatial distribution	K2	The kilometrage proportion on different congestion grade
Congestion temporal distribution	K3	Congestion Grade duration by time of day
Traffic bottlenecks (congestion frequency)	K4	The number and distribution of recurrent congested nodes and segments
Reliability	K5	Road network reliability index

The K1 is for Sun et al. (2009) the possibly most important index of the five mentioned traffic congestion performance measures in Table 3. It is the relative value of the congestion intensity (traffic load intensity) in a given investigation area within a specified time window. As all indices, K1 has no unit and can have a value from zero to 10. Additionally, there is often a classification with 5 defined classes of traffic congestion (Sun et al. 2009). For the representation of its usefulness, there are plots of the temporal distributions of K1 values for an average working day in June 2006 and June 2007 for the whole area of Beijing. Small variations in congestion intensity are detectable. Sun et al. (2009) introduce traffic congestion indices for whole cities. In particular these global congestion indices found their way into official technical standard, as stated by Wen et al. (2014) for the case of the traffic performance index (TPI) in Beijing in 2011.

Another use case of Sun et al. (2009) is to compare K1 values of Beijing and Shanghai for comparing city specific congestion intensities. In general, Sun et al. (2009) found out that there is a time-dependent trend in Beijing and Shanghai on working days: similar starts and ends of morning and evening rush hours. This time-dependent trend was also proven in Keler and Krisp (2016b), where only travel time on highway segments between 13 prominent was inspected over its variation over the day. This central area of the city is then representative enough to show rush hour dependent travel time variations for the whole city.

2.6.3 Traffic bottlenecks and dynamic traffic state transitions on the road network

Traffic bottlenecks are the locations of a road network that cannot manage road supply. Consequently, traffic congestion results from these traffic bottlenecks, e.g. on selected rush hours on working days.

In general, this definition is clear, but real world traffic congestion appears due to many more reasons that are not necessarily caused by the capacities of transportation infrastructures.

The traffic bottleneck

A traffic bottleneck is a term that indicates the appearance of traffic congestion on selected parts of the transportation infrastructure. In general, bottlenecks indicate recurrent traffic congestion events. The reason for these appearances is in most cases inefficient properties of the built road elements for managing daily vehicle usage, the lack of suitable road capacities or insufficient local traffic regulations on selected parts of the investigation areas. Keler and Krisp (2017c) and Keler (2017) test the correlation between traffic congestion events (inferred from taxi FCD) and the estimated complexity or road intersections (inferred from OSM road network) on the road network of Shanghai. Whereas there are far more traffic congestion events that are not appear on complicated crossings, it is possible to detect those intersections where traffic congestion appears periodical. Those areas have often on-ramps and elevated highways, due to the technique of inferring complicated crossings based on the high densities of nodes and intersection points of vehicle-accessible road segments. Since FCD of many working days is used to derive congestion events, intersections to complicated crossings show very similar areas in the morning and evening rush hours. Nevertheless, this inspection is scale-dependent and the detected traffic bottlenecks show smaller variations in space up to 100 meters. This might be helpful for detecting the influence areas of detected bottlenecks, but less the dynamic propagation of certain traffic congestion events. Detected traffic bottlenecks are useful products for inspecting different scales of the urban traffic flows (Li et al. 2015a). For Li et al. (2015a), transition between free traffic flow and traffic congestion is well studied in highways. The reason for that is that it is not that complex as inspecting the entire traffic of a city within the whole road network. This network scale is essential for Li et al. (2015a) to connect local flows with the global flow in a city.

The detection and classification of traffic phenomena based on tracked entities has a connection with implying the structural properties of the underlying road networks. The term traffic bottleneck indicates such a connection. A bottleneck is by definition a location of the road segment, where there is a capacity change. The reasons for this change might be various. One focus in this context is the connection with the previously defined movement and traffic parameters. It should be noted that a traffic bottleneck is always referring to an element of the given transportation infrastructure. This means that it is a static object or a part of a static object. Nevertheless, traffic bottlenecks might appear as stationary and as moving. Its detection always base on observations of the vehicle dynamics and the group patterns that result from participating vehicles. In general, it is to state that there are many different appearances of traffic bottlenecks with different types of interactions between the built transportation infrastructure and vehicle dynamics.

Percolation and bottleneck identification

For Li et al. (2015a), traffic percolation consists of the collective organization process of local flows in the roads into global city flows. This definition focuses on urban traffic with the road network within administrative city boundaries as the object of investigation. Global traffic consists of dynamically changing clusters of local flows. Those traffic flow clusters are connectable by bottleneck links. By improving the functionality of selected bottleneck roads, Li et al. (2015a) state that the global traffic can be improved in a cost saving way. The connection to this work consists only of the knowledge about percolation. Practical usage of the knowledge about traffic percolation and its possible modelling is possible after gaining the first results of the traffic pattern analysis framework. In particular, traffic percolation modelling needs detected timestamped traffic congestion areas beforehand.

A method for Identifying Bottlenecks is proposed by Li et al. (2015a). It is a method for detecting only the recurrent bottlenecks. Those recurrent bottlenecks affect the global traffic in a city on a bigger scale. This means that The approach by Li et al. (2015a) is based on percolation theory. When talking about bottlenecks, there is a general differentiation between two differently perceived phenomena: the stationary bottleneck and the moving bottleneck. For the graphical and theoretical representation of bottlenecks, ones can make use of the traffic flow theory for modelling and representing bottlenecks. For Sugiyama et al. (2008), it is commonly understandable that traffic congestion appears near bottlenecks. Examples of bottlenecks by Sugiyama et al. (2008) include on-ramps, tunnels and sags.

Structural and dynamic traffic bottlenecks

Structural bottlenecks are defined by Li et al. (2015a) as those bottleneck links, which can be found by traditional network analysis. This means a structural bottleneck is described as a road segment, a lane of a road segment or as an intersection. A road intersection of the real world might then be a node, or in some cases a hub, which affects many different connection roads.

Dynamic bottlenecks vary in time and space. That is why traditional network analysis should be extended for their detection. In case of daily recurring traffic congestion events, the dynamic bottlenecks might show periodical pattern. This means that they affect the same elements of the underlying transportation infrastructure. These appearances might serve as motivation for planning a modification of parts of the infrastructure. For urban and transportation planners measuring the capacity thresholds of these infrastructural elements is challenging, since capacity is traffic density dependent. FCD still might benefit these analyses in the way that periodical congestion events can be detected. The detected periodical events might be used as indication for dynamic bottlenecks.

Queue and queue length

Besides the term bottleneck, there are as well the terms queue and queue spillback. Queues are chains of still standing entities, which appear in the same pattern as queuing pedestrians standing at a bus station and waiting for the bus. The same example in vehicle road networks has its causes by many different phenomena. Since the origin of a traffic jam can vary in many ways, also the appearance of queues has various examples. In general, it is to say that queues can be a result or the appearance of traffic congestion. Additionally, queues are partially predictable in the case of still standing in front of a red traffic light. This occurs often at intersections. As every queueing event includes a certain number of vehicles that varies over time, various individual queue lengths are measurable. This measuring occurs every day on highways with the information coming over the Traffic Message Channel (TMC). This means the queue length has often an aggregated form of length, and often given in km. This is dependent on the sensors that deliver a quite sparse spatial resolution, whereas the temporal resolution of these sensors is relatively high. Another attempt in estimating the queue length is given by Wang and Chen (2013). Wang and Chen (2013) use FCD for inspecting the development of queues at selected intersections. The term queue spillback is referring to the mentioned dynamic properties of queues. The modelling of its appearance is possible via kinematic wave theory. The simplified kinematic wave theory refers to modelling the formation and dissolution of traffic congestion. It was successfully tested in various applications with real traffic data.

Traffic prediction, the accuracy of urban traffic congestion measures and long-term changes of urban traffic congestion

Examples for traffic prediction are coming from researchers of Microsoft, who deliver a small review on traffic prediction (Zhang et al. 2016a). When inspecting this specific body of research, there should be a differentiation between the applications traffic flow prediction (Ko et al. 2016) and alleviation of traffic congestion, as from the approach by the MIT[15] (Wang et al. 2012). Important elements of the latter are distinct and accurate traffic congestion measures. Measuring phenomena, which are difficult to describe, as urban traffic congestion, is challenging, especially when it comes to accurate measures.

By referring to, the previously mentioned, survey of transportation expert, there are also general views on the quality of traffic congestion measures, in particular about the accuracies of these measures. Figure 12a shows this with the responses for questioning the general quality of urban traffic congestion measures. Surprisingly, only 18% of the responses rate these measures as being accurate. Most of the answers (33%) express the appearance of having an intermediate accuracy, which is only partially reliable. 20% say that the accuracy of urban traffic congestion is variable, without mentioning the influences on this variability. 14% of the 525 experts rate urban traffic congestion measures as inaccurate.

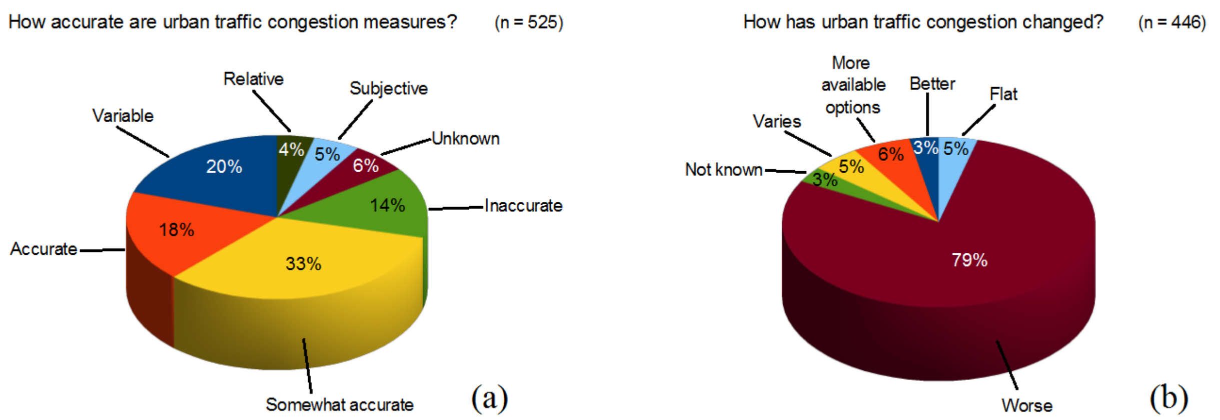


Figure 12: Survey results from experts about (a) accuracy of congestion measures, and, (b) the change of urban traffic congestion, based on Bertini (2006).

Figure 12b reveals the perception of the experts about long-term changes of urban traffic congestion, where 79 % think that urban traffic congestion becomes worse than before. The answers show the problem of urban environments of having increasing traffic problems on their road networks, partially influenced by the growth of the urban population.

15 Massachusetts Institute of Technology. "Traffic congestion can be alleviated throughout a metropolitan area by altering trips in specific neighborhoods, model shows." ScienceDaily. www.sciencedaily.com/releases/2012/12/121220143742.htm (accessed March 24, 2017).

3. Floating Car Data (FCD)

The term Floating Car Data (FCD) appeared since the late 1990s. Since then, this technology emerges in an exponentially growing number of publications, which are coming from various research domains. Besides FCD as a method for data generation, the acquired data product has often the same terminology and is often usable for the same context. By definition FCD refers to the acquisition and resulting data from observed vehicles, be it private vehicles, test vehicles (manned and autonomous), or cars of vehicle fleets as taxis and delivery trucks. For FCD, there is no specification on the used tracking device or on the positioning technology for data acquisition. The only requirements are the presence of at least three attributes: two coordinate values for the spatial position and one time stamp for associating a moment in time (Cohn and Bischoff 2012). For the case of tracked vehicle fleets, there should be an additional vehicle identification attribute for distinguishing the different timestamped vehicle positions of different vehicles. Another not often mentioned part of the definition concerns the installation of the sensor for data acquisition. FCD bases on installed positioning devices within the vehicle without the option of changing to another mode of transport. The installed devices have in many cases the terminology FCD devices and are traditionally installed in taxi vehicles (Lorkowski et al. 2003).

There are also examples of using standardized FCD devices for whole vehicle fleets, which has the reason of guaranteeing similar acquisition requirements, positioning accuracies, and sampling intervals. The determination of underlying sampling intervals is critical for the FCD application, since long temporal intervals are often not useful anymore for detailed traffic information. The size of the sampling interval often indicates the selected data acquisition method. Similar to the GNSS acquisition strategies of tracking applications on mobile phones, there are at least two different acquisition strategies for FCD devices: per constant time interval or per constant distance. This is the way of recording trajectories of moving entities. Both variations imply two different strategies of acquisition.

There are critical ongoing discussions on how to observe numerous or single vehicles for different applications. The variety of applications results from differently built tracking and processing hardware, and different data quality requirements. Data generation and FCD applications often indicate the way of FCD processing and usage. In the domains of transportation sciences, FCD has recurrently usage for optimizing established transportation services. Service optimization is connected with the term intelligent transportation systems or its acronym ITS (Zhang et al. 2011b). Research on technical realizations of the FCD method is mainly coming from electrical and electronics engineering and information technologies, but has as well its interdisciplinary communities. Examples for these interdisciplinary communities are also the different interdisciplinary and successful conference series on ITS, location-based services and mobile technologies in general.

Besides transportation sciences with their various still unsolved problems in the 2010s, there is ongoing research from applied computer science and GIScience. GIScience benefits from the more established, since more mature and interdisciplinary field of ecology. In particular, the connection lies in the ecology of movement of the whole organism, but in part as well with behavioral ecology. The connection between GIScience and movement ecology has the naming computational movement analysis (CMA). Computational movement analysis is the title of Laube (2014) and describes a very general modern domain, which focus on all forms of data from tracked entities.

FCD has a long coexistence with GNSS technology, in particular with GPS. This is the reason why literature refers more frequently to these data sets as GPS trajectories. Nevertheless, the term is confusing, since a GPS trajectory might refer to the real trajectory of a moving GPS satellite (selected orbits in space). Therefore, this work implies only the use of the terms FCD and FCD trajectory. Additionally, there are more the more general terms moving object, moving entity and tracked vehicles for distinguishing the tracked entity. As part of the definition, FCD is moving objects data of tracked vehicles.

3.1 Floating Car Data acquisition – errors, data sampling and device dependencies

In the FCD method, each vehicle acts as a sensor itself for the detection of the coordinates and additional information, such as velocity and the driving direction (Liu et al. 2008). Subsequently, the data-sending step follows to an information center in near real time (Liu et al. 2012).

About FCD device dependencies

GPS receivers are not able to catch GPS satellite signals in tunnels, below bridges or in urban canyons. One typical urban canyon includes road spaces along tall buildings or in densely built areas. Many urban canyons exist in Asian Megacities, especially in Japan, which is introducing an additional Satellite navigation system named QZSS in 2018 to solve this problem.

The mentioned examples indicate signal losses in case less than four satellites are visible (as within tunnels). In urban canyons, a positioning is possible, but often with reduced positioning quality due to the geometry of GPS satellite trajectories. A FCD device create FCD that is usually updated with temporal intervals between 1 second and 10 minutes in the NMEA-0183[16] mode via GPRS sentence (Zhang et al. 2011d).

Another term, namely extended Floating Car Data (xFCD) indicates explicitly the acquisition of on-board sensor and electronics information via connected FCD devices. Examples in the literature that mention this term are Sansone et al. (2002) and Huber et al. (1999). xFCD appears as a solution to model not only movement information, but as well air exposures as in the case of EnviroCar (Bröring et al. 2015).

For Muckell et al. (2014), the trend of producing more and more trajectory data from moving objects has three major groups of problems. There is the time-consuming and expensive transmission of the data, the expensive computational operations for extracting patterns and information from massive trajectory data, and the high amount of redundancy within the data that requires high storage space and efficiency (Muckell et al. 2014). One idea to overcome the high data volume is the compression of trajectory data (Muckell et al. 2014).

For Moreira et al. (2010), only few approaches in movement analysis are inspecting the data quality in a systematic way. This means that the data quality estimation part is not directly included in pattern and model derivation. Besides validating the results of several methods based on data quality, the impact of data quality on traffic pattern detection is huge.

[16] https://www.nmea.org/content/nmea_standards/nmea_0183_v_410.asp

About positioning quality in GNSS and the case of low sampling rates

One example that is not only interesting in GNSS technology is positioning quality. This means how reliable is a derived position on the Earth's surface or not.

One example for implying the positioning quality into movement analysis is presented by Dias et al. (2007). In this specific case, movement of pedestrians is acquired by two devices, which results in a distribution of distances between measured positions that should represent movement of same entities. As a result, it is to mention that no substantial impact on the analysis is detectable.

It is still not clear, if this appearance might occur also in case of tracked vehicles with two GNSS receivers or two FCD devices. This might be interesting, especially in case of tracking vehicles in urban environments.

Low sampling rates appear in many different FCD sets, due to different devices and record modes. Since both parts are not always consistent (or better: nearly always not consistent), an often used method to provide overview on the investigated data set is to represent a temporal sampling interval histogram. By inspecting this kind of histogram, one can immediately tell the distributions of sampling intervals.

In some studies, the low sampling rates, which are long sampling intervals, are excluded from analysis. This may have various reasons. One reason might be the higher complexity in reconstructing the driven vehicle trajectory. For the map matching (MM) case, it is difficult to infer velocities and driving directions based on sparse subsequent vehicle movement positions. So called sparse FCD with long sampling intervals is favorably useful for highway MM, since highway segments are longer, broader and often, besides complicated on-ramps and intersections, easier to represent than urban transportation infrastructure. Additionally, traffic parameters on highways are easier to detect than on the more complex urban road networks.

Another emerging case that is connected with sampling rates and temporal resolution is the loss of signal. The loss of signal is often resulting from the used positioning technologies. In case of GNSS receivers, a certain number of visible satellites is required for allowing positioning and producing FCD records. This means signal losses appear in tunnels, below broad bridges and, which is the most often case in many urban environments, in case of urban canyons. These urban canyons influence the positioning not only by signal losses, which result in missing records for a certain time range, where the vehicle is inside the canyon. The other influences of urban canyons are:

1. Increasing GDOP values[17], and decreasing positioning quality, and
2. Multipath effects

The influence of the temporal sampling interval on the quality of FCD sets

One important term of this work is temporal sampling interval. Sohr et al. (2010) mention this term as the frequency of the transfer interval of the FCD. Depending on the way of data acquisition and creation of intervals between the point records, FCD becomes more or less useful for different applications. For the application of providing traffic information, in real-time or as historical traffic

[17] Computation of GPOP Langley, R. B., 1999. Dilution of precision. *GPS world*, 10(5), 52-59.

parameters associated with road elements, Sohr et al. (2010) mention the precondition of having an interval shorter than 60 seconds. This guarantees an acceptable quality of derived traffic parameters. This factum refers to traffic parameters and not to movement parameters of individual movement. Therefore, there is a need of inspecting each recorded vehicle trajectory within FCD sets separately. For gaining overview on the sampling intervals of inspected FCD sets, it is feasible to introduce histograms on sampling intervals. In those cases, as in Li et al. (2007b), sampling intervals are calculated between all consecutive records of the data sets.

On the other hand, Sohr et al. (2010) mention that the number of individual trajectories resulting from participating vehicles in the FCD set acquisition is less important for gaining higher quality of calculated traffic parameters. More important for guaranteeing acceptable traffic parameter qualities are great mileages of the observed vehicles. This means that a large size of the temporal observation window is more important than a high number of tracked vehicles. Within one selected investigation area, expecting great mileages within one day of data acquisition is a good precondition for computing road segment based traffic parameters with an acceptable quality. Consequently, a long temporal data acquisition window guarantees higher quality of calculated traffic parameters than the number of observed vehicles (Sohr et al. 2010). These assumptions are resulting from various tests of the German Aerospace Center (DLR) in Berlin on computing with FCD sets and products of traffic simulations ((Sohr, 2008 #1116) , (Brockfeld 2009), (Krieg 2008)). The need for comparing FCD and traffic simulation results has a strong connection with the fact that it is not possible to observe every vehicle participating in the daily traffic. This is one object of investigation in selected parts of this work.

One additional aspect about data quality is the acquisition via Floating Phone Data (FPD) or Floating Mobile Data (FMD), which is a matter of investigation of Tettamanti et al. (2014) and of Claudel and Bayen (2008). The examples with mobile phones as sensors are not part of the proposed traffic pattern analysis framework. The reason for this is that Floating Phone Data (FPD) has not the restriction of coming from installed devices within vehicles. Since a human individual can have different modes of transport, the movement is not only restricted to vehicle road networks. Therefore, the framework has not the focus to distinguish between different modes of transport.

Representing FCD accuracy

An often-arising question when using FCD for various analyses is the following: How much FCD do we need? There is no direct answer for this question by Patire et al. (2015), because the size or number of records of each FCD set are always dependent on its specific properties. Another important influence is the favored type of FCD analysis. When talking about accuracy of FCD, there should be a general differentiation between use cases of these terms. The reason is the altering of its meanings together with differences in interpretations. On the one hand, there is positioning accuracy of the used devices. The resulting spatial accuracy of the movement positions can then vary based on the FCD sample size. Additionally, the temporal sampling interval can highly influence the spatial accuracy, especially when inspecting the shapes of individual trajectories. High sampling intervals often indicate unrealistic trajectory estimations when using linear interpolation between FCD positions in Euclidean space. By respecting individual movement in the Euclidean space, it is also possible to inspect the shape of the trajectory by comparing it with possibly traversed road segments. Reasoning on sampling intervals in this way has a thematic connection with the group of algorithms named map matching (MM).

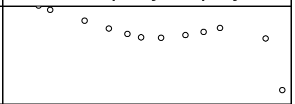
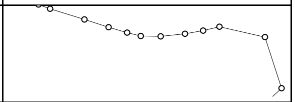
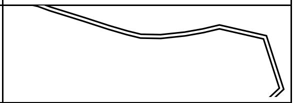
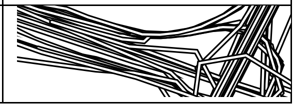
3.2 FCD forms of representation

FCD is a relatively new technology collecting data from numerous observed vehicles equipped with GNSS devices. Based on FCD sets, it is possible to infer road elements (Li et al. 2015a), examine commuting patterns (Dewulf et al. 2015), model typical periodical traffic flow patterns in selected parts of the network (Körner 2011) and giving answer on emissions (Gühnemann et al. 2004). Another important product from FCD is information on traffic congestion. Duan et al. (2009) test FCD for detecting those areas periodically influenced by traffic congestion. In all the mentioned applications, used FCD sets emerge from preprocessing data for preparing further computational steps. Depending on different applications, there are various different FCD representations. The FCD representation can change within specific applications, dependent on the need of various input information for producing intermediate and end results of the analyses. Floating Car Data (FCD) is GNSS-tracked vehicle movement, includes often large data size and is difficult to handle, especially in terms of visualization (Keler and Krisp 2016b). Since many different domains make use of FCD, different applications imply varying forms of FCD representation. This subsection has the aim to mention and to discuss these different forms. Partially important is the thematic connection to the field of information visualization. Findings of the review throughout the different research domains are the backbone for designing FCD visualization approaches of the framework.

Preprocessing possibilities for vehicle movement data

For Protschky et al. (2015b), FCD have a great potential for information retrieval in the vehicular context. The data resulting from applying the FCD technology can improve road maps, detect locations at risks, and provide real-time traffic information (Protschky et al. 2015b). However, different applications need different representation forms of FCD. One possible selection of FCD representation forms, when computing within classical GIS software environments, is pictured in Table 4. This table shows four possibilities for representing various kinds of FCD, during their analysis within vector GIS software environments. These four different options for FCD products in Table 4 are originally proposed by Körner (2011).

Table 4: Possibilities of FCD constellations, proposed by Körner (2011).

Element	Meaning	Exemplary display
Point	Single position	
Vector	Two positions of the same car in the right chronological sequence	
Trajectory	Many positions of the same car in the right chronological sequence	
Bundle of trajectories	Overlay of many single trajectories on the same arc or node of a network	

Traffic data representation in has a variety of forms. Körner (2011) proposes four different levels of analysis for the case of FCD: point, vector, trajectory and bundles of trajectories. FCD representation has a strong connection with its preprocessing. In most cases, visual representations of FCD realize the detection and deviation of further information. Table 4 shows the option of analyzing FCD records as projected points on a map. This might be the most common and simplest form of introducing spatial awareness into the analysis. On the other hand, a trajectory is the most common form of traffic data (Chen et al. 2015). For Treiber and Kesting (2013), a trajectory is even the most accurate form of representing traffic. Partitions of trajectories can be vectors between two consecutive points of one moving entity. Depending on the desired application the connections of points into other FCD representations has different possibilities, and requires in some cases a previous evaluation of the positioning quality.

Moreira et al. (2010) discuss the characteristics of positioning data sets and divide them contextually into the two following properties***:***

1. Sampling rate and accuracy of readings, and
2. Movement representation by vector sets.

The first property is mainly device-dependent and influenced by the data acquisition method. Additionally, there are influencing factors of atmosphere and the built environment. Concerning the second property, there are much more factors that might be very varying, since the vector generation procedure is mainly application-dependent. In general, the step of interpolating the spatial connection between two movement positions can have manifold possibilities.

Estimating FCD-based velocities

In their study on estimating routing velocities based on FCD of taxis, Liu et al. (2008) introduce the Metropolis criteria for estimating realistic average velocities for road segments. Originating from simulating heat balance processes of solid objects, Liu et al. (2008) apply the criteria with the idea of respecting influence areas of coordinated stops at intersections with traffic lights for providing more accurate hour-based average velocities for associated road segments. After a previous MM step, the current car velocities are validated by validation with selected high-speed roads in the city of Shenzhen. For Pan et al. (2011), the average travel speed is one of the most important indices for traffic status identification.

Representing movement by vector sets

Moreira et al. (2010) mention a multidimensional vector representation of GNSS records: every vector includes, besides respectively two spatiotemporal points, also speed and direction of movement. For the movement direction, different calculations with different units of measure are possible. In case of no available instantaneous movement direction values, it is possible to compute movement direction from every two consecutive points.

The formula for calculating speed between consecutive points comes from the basics of Mechanics and can have the following representation. The average velocity $\bar{\upsilon}$ has the definition of:

$$\bar{\upsilon} = \frac{\Delta\, pos'n}{\Delta t} = \frac{displacement}{time} \qquad \text{where } \bar{\upsilon} = \frac{\acute{\upsilon}_i + \acute{\upsilon}_j}{2} \tag{14}$$

Independent of $\bar{\upsilon}$, the average acceleration or deacceleration $\bar{\alpha}$ calculation bases on the difference (Δv) of two instantaneous velocities $\dot{\upsilon}$ in two consecutive points i and j:

$$\bar{\alpha} = \frac{\Delta v}{\Delta t} \qquad \text{where } \Delta \mathrm{v} = v_j - v_i \tag{15}$$

The direction of movement between those points has its description and calculation coming from Geometry or Geodesy. In case of FCD, it is possible to compare instantaneous speed, namely the velocity coming from on-board sensors. In a similar way, the driving direction may have instantaneous values at every point, with their vector-based calculated counterparts (Körner 2011).

Segmenting movement trajectories and selecting trajectory points

Research on Floating Taxi Data (FTD) has often a focus on extracting functional trajectory segments. The semantic information that causes this segmentation might be getting-in and getting-out points of operating taxis. Consequently, it is feasible to segment taxi trajectories into partitions of different operational mode. The other useful information are the positions of taxi pick-up and drop-off points that show interesting spatial distribution patterns for multiple observed taxis within a period of time. For these two point types, points clustering techniques are applicable for detecting spatial hot spots. The spatiotemporal analysis of these hotspots is possible via space time cubes. The idea is to use each timestamp of the raw data for deriving a 3D view with the used or averaged time-interval as a third axis within a space time cube (Kraak 2003). The investigation of Floating Car Data (FCD) within a space time cube is challenging for a high number of observed vehicles due to the cluttering or overlapping effects of multiple trajectories in nearby space time. Based on the origin and destination points of each taxi trip, it is possible to infer patterns like the locations in an urban road network where many taxi customers are getting in or getting out at certain times of the day. Usually, FTD has an attribute of Boolean type that indicates an occupied (1) or non-occupied (0) state of an individual taxi. The inspection of customer pick-up and drop-off points can include point clustering techniques based on similarities derived from their three inspected coordinates longitude, latitude and time, as in the applications of Krisp et al. (2012), Pan et al. (2013), and, Yue et al. (2009).

Visualization of FCD and traffic data visualization

Chen et al. (2015) define the term traffic data as datasets generated by sensors and collected via storage devices, which are connected to moving vehicles. The properties of traffic data sets are in general high-dimensional or spatiotemporal, which is difficult for scientific visualization. This is the reason why information visualization and visual analytics are preferable for traffic data visualization (Chen et al. 2015). Numerous researchers have already examined the visualization of FCD. Stanica et al. (2013), for example, visualize instantaneous velocity values by simple colorized dots that represent the spatial positions of the FCD records. The coloration is altering based on the instantaneous velocity values between high and low velocities. Binning is a popular technique, as it is a simple form of point aggregation. Its practical use enables the creation of heat maps, 2D histograms, and other binned views. There are aggregated views on FCD in the form of spatial grid cells (Andrienko and Andrienko 2007), which work with summation places (Rinzivillo et al. 2008).

Cheng et al. (2013) use derived congestion time based on road elements for providing line-wise extrusion. In Shepherd (2008), the usage of extrusion appears point wise, resulting in spatially narrow two-colored cylinders within a three-dimensional map view. The visualized information is static, describing the proportion of male and female inhabitants by two differing colors with varying ratios within certain cylinders. Depending on the spatial distribution of the cylinders, several cluttering effects appear with the three-dimensional map view. One option for avoiding overlapping features is the introduction of an orthogonal projection of the recent view. Liu and Ban (2013) test the extrusion technique for dynamic data via extruding durations of traffic congestion events on a two-dimensional map view of the road network. The extruded point elements perform the description of congested areas within a certain time ranges. The point associations with partitions of road elements show relatively accurate spatial patterns. By comparing congestion time visualizations of two different time windows (e.g. selected times of one day), it is possible to detect changes of the traffic situation. Another possibility for visualizing FCD bases on interpolation (e.g. density estimation) of the geographic positions of the FCD records. By deriving isolines from traffic or movement data, certain parts of the investigated area are representable as classes of certain interpolated surfaces in a two-dimensional view (Tanaksaranond, 2010).

Cartographic representations of traffic are used in existing real-time traffic maps (traffic layers of Google Maps, Yahoo Maps and Bing Maps) and typically use the stoplight metaphor (green, yellow, red) without any legends (Goldsberry 2008) on road segments. Additionally, Liu et al. (2012a) use coloration of 3 velocity classes (stoplight colors) for both driving directions within the road elements. In case of interactive traffic maps, this appearance is not to handle in higher zoom levels and requires intelligent generalization methods. Traffic maps mainly show the appearance of traffic congestion (Goldsberry 2005, 2008) on two-dimensional road segments, which imply often averaged movement information as average velocities. Due to the classical two-dimensional representation, it is often difficult to visualize complex urban infrastructures in map views.

An important quantity for describing traffic congestion is the average speed for each inspected road element. For instance, Lund and Pack (2010) propose the use of interactive legends in traffic maps, which allow the selection and visualization of the average speed, the reference speed and the recorded speed on road networks. Reinthaler et al. (2007) evaluate taxi FCD data for the derivation of average velocities by using a map matching technique. One example of an interactive traffic map with matched speed values is using the Google traffic layer from Google Maps. The coloration of the road segments follows the stoplight or traffic light metaphor, which means that it is ranging from green over orange to red. Depending on the scale of inspection, the road elements have more or less detailed representation dependent on the selected ways of line generalization. In higher zoom levels, more details are discernible as for example road lanes and elevation differences. The concept of an interactive legend is included in the Google Maps functions via selection possibility between live traffic values and the typical traffic values for each time of the day.

Inspecting massive FCD visually is not always manageable (Andrienko and Andrienko 2011). Therefore, Sun et al. (2009) use single traffic flow parameters and predefined traffic congestion performance measures, like congestion intensity, spatial and temporal distribution, congestion occurring frequency (bottleneck detection) and reliability. These indices describe congestion qualitatively.

It is still challenging to represent these derived indices on a map visually due to the conflicts between the principles of cartographic conventions and the existing real-time traffic map design. These conflicts include for example the typical use of the stoplight metaphor in traffic maps, which allows omitting classes and introducing a continuous color gradient (ranging from green to red) or even removing the whole legend (Goldsberry 2008). Another disagreement of traffic maps with cartographic principles is the representation of unipolar statistics such as vehicle velocity with varying hue instead of varying lightness as proposed by Slocum (2009). All these conflicts are respectable and important issue that one should take into account in case of creating map views.

In many cases, the handling of FCD points has a connection with aggregation and generalization for subsequent visualization. Based on FCD aggregation, it is possible to create Thiessen polygons (Andrienko and Andrienko 2011) or to introduce spatial grid cells (Andrienko and Andrienko 2007). For these features, which are also called summation places (Rinzivillo et al. 2008), colorations based on previous average parameter classifications might be introduced. Other visualization approaches go more into scientific visualization (Ding and Meng 2014) by extending traditional domain-specific graphical representations as for traffic engineering (Helbing and Treiber 2002).

3.3 FCD preprocessing possibilities - a review of methods

There are not always possibilities of visual analytics approaches and EDA for not preprocessed FCD sets. One reason for this is the usually massive size of FCD sets, which makes it difficult to implement interactive data analysis applications.

This is not always possible and depends on the visualization methods and implementations. For Andrienko et al. (2016a), the visual analytics approach has, as all other analytical tasks with movement data, its own specific preprocessing steps.

The following reviews preprocessing steps for the general task of FCD analysis. The focus of the review is less on a certain type of analysis or application, but more on gaining an overview of possible preprocessing procedures for FCD. At selected parts of the review, indications reveal the type of intended analysis of certain preprocessing steps or methods. The preprocessing of the spatial components for example usually focuses on assigning individual movement to road segments. . In this sense FCD with higher spatial accuracy is usable for detecting road intersections, as in Xie et al. (2017) and Liu et al. (2012d). Additionally, it is possible to estimate the travel delay at intersections with FCD (Liu et al. 2013), and to infer statistics of travel distances (Liu et al. 2013).

Overview on possible FCD preprocessing steps

For Ye et al. (2009), the first step in data preprocessing is the inspection of GNSS logs and the GNSS sequence, which is often differing. After ordering every subsequently appearing point, it is possible to reason about the patterns inside trajectories in the way of segmenting trajectories into functional partitions. The simplest possibility for the case of inferring functional parts of trajectories is the detection of stay points and segments of faster movement (Ye et al. 2009).

As previously mentioned, there are processing strategies for not preprocessed raw data coming from GPS receivers without additional information as defined by Schuessler and Axhausen (2009). The focus of the work of Schuessler and Axhausen (2009) is on using GPS trajectories for extracting travel behavior survey. Therefore, the exact traffic situation definition is of less importance. The idea is to use a simple two-step workflow, as in the following:

1. Data cleaning and smoothing
2. Determination of “trips” and “activities” – those correspond to basic operation for detecting “moves” and “stops” of tracked entities (Ye et al. 2009)

Literature on preprocessing FCD has different approaches dependent on different desired tolerances of the outcomes. Castro et al. (2013) propose preprocessing steps that are useful for massive taxi trajectories.

Whereas Andrienko et al. (2016a) propose a more general framework that is applicable to many different massive movement data sets.

FCD preprocessing is also dependent on the available computational resources. Therefore, Pang et al. (2011) mention that a previous estimation of computational complexity is favorable for reasonable preprocessing. Pang et al. (2011) propose a preprocessing approach, where a pruning strategy is introduced in the way of a classical likelihood test statistic. The mentioned pruning strategies are often part of precomputation.

Based on Chen et al. (2015), there are four main steps in FCD preprocessing, mainly resulting from the perspective of traffic data analysis. This means the focus is more on detecting general traffic trends and less behavioral aspects of individual vehicle movement. The four main operations for traffic data preprocessing, which are proposed by Chen et al. (2015) are the following: data cleaning, data matching, data organization and data aggregation.

FCD organization, indexing, and storage

Many applications connected with FCD analyses face frequently appearing problems connected with high computational costs. These problems result from various physical conditions of the data storage and organization. One of these problems is the difficulty of providing efficient spatial data handling, where lexicographic information indexing is not useful. The advantage of R-tree-based index structures is the expelling of point transformations for spatial data storage. This is the reason for Kriegel et al. (2008) that R-tree-based index structures are useful for spatial clustering.

Possibilities for reducing computation times and raising efficiency include specific indexing methods in different types of databases. For the cases of massive geodata, the usage of extended binary trees can often improve the efficiency of geospatial computations (Parlante 2001). Computations in geospace are one reason for the development of the so called tree evolution, especially for the case of R-trees, which evolved since the 1980s (Guttman 1984, Manolopoulos et al. 2010, Balasubramanian and Sugumaran 2012). The background of tree structures is coming from the 1960s and is still continuously developing onto data-specific tasks.

Interesting work, which introduces a quadtree- and octree-based approach for point data selection in 2D or 3D is coming from Peters (2013). Peters (2013) shows that besides spatial indexing of massive point data, trees are also useful for point generalization.

Traditional indexing of geodata includes spatial indexing methods using R-Trees or B-Trees (Guttman 1984). The basis for this is the use of minimum bounding rectangles (MBR) for indexing selected two-dimensional point or line data. There is a plethora of different R-tree variations that consist of at least three different classes (Balasubramanian and Sugumaran 2012). The three main groups are named by the manner of how the methods are modified in comparison to the classical R-tree as it was defined by Guttman (1984).

Besides indexing FCD as a preprocessing step, it is useful as a previous step of FCD clustering. For the example of clustering data in Euclidean space, spatial indexing is useful, since it fastens the computation of distance and density-based clusters. There is even the term cluster index, which is another type of index taking the cluster positions into account, as in the approach by Long et al. (2016). It is important to distinguish between the cluster index, which is originating from data driven clustering in feature space, and the spatial cluster index, which is dependent on the quality of the cluster index and the distribution of the data in Euclidean space. The quality of the cluster index results is measurable via the computational efficiency in applying spatial clustering techniques.

Another challenge, when working with massive FCD, is to handling the additional time component of the data in an efficient way. Within research on the analysis on moving objects, there is always reasoning on how to handle this specific data type. A solution comes with the term moving object database (MOD), which is a special framework of a database, specifically fixed on representing the movement of individual objects. MODs are databases that allow fast and reliable querying and processing functions. One option is the representation of recorded point records for a selected time window. This shows the addition to usual geodatabases: the temporal component. This has substantially different storage properties and makes it more difficult to provide efficient indexing.

Interpolation and density estimation of FCD

In general, when talking about interpolating FCD, there is a need to distinguish between interpolation of time and of space. One exception is the interpolation of spatiotemporal geodata (Li and Revesz 2004). Ranacher et al. (2016b) reason on how to interpolate between consecutive the movement positions of individual moving entities. The information for selecting the most suitable interpolation possibility for specific FCD is not always coming from the properties of the underlying acquisition constellation. The constellation of FCD acquisition includes for example the selected positioning technology and devices. Selecting suitable interpolation methods for specific FCD is also dependent on the specific vehicle dynamics. Besides matching moving object positions on road segments, there is the option of linear spatial interpolation between the subsequent movement positions. Linear spatial interpolation is the easiest and computationally most efficient way of interpolating between moving object positions. Another movement interpolation method refers to as random walk. Random walks have numerous applications in the domain of applied robotics, when for example estimating random indoor movement of robots. Another method is coming from the movement ecology domain and finds mostly usage in sparsely tracked animal movements: brownian bridge movement models (BBMM) find usage in modelling movement between two recorded positions. The Brownian Bridge Movement Model (BBMM) is for Fischer et al. (2013) a probabilistic method in contrast to usual linear interpolation.

One central input information from the, to be interpolated, data for using probabilistic interpolation methods is the time component. Depending on the temporal values and respective Euclidean distances, it is possible to generate additional movement positions between consecutive points with the highest probabilities. The best way to compare the results of linear interpolation with BBMM for massive movement data sets with numerous tracked entities is to provide kernel density estimation for the densified additional points between original positions.

Gudmundsson et al. (2012) mention the usage of many historical GIS operations for the analysis of movement data. These applications include spatial interpolation techniques for creating surfaces (continuous data). In the early 1970s, there were frequent discussions on spatial interpolation techniques, especially considering the reasonable selection of the most suitable interpolation method for specific data sets from different domains. In general, the used spatial data in these examples of spatial interpolation was static. Interpolating spatial point data with numerical values describing one certain attribute has numerous possibilities. The resulting outputs are surfaces. Examples for spatial interpolation are inverse distance weighting (IDW), Kriging, or other spatial interpolation methods. In case of interpolating overlapping points, the mentioned techniques often fail to work, since averaging different attribute values in similar positions is not included. In this context, applying a kernel density estimation (KDE) can solve static overlapping points into surfaces representing the point densities. This concludes a contextual change, since the focus is on interpolating point densities and not on interpolating attribute values.

For estimating the density of movement, there are established advanced KDE techniques for various movement data sets. Krisp and Peters (2011) introduce a directed KDE, where two moments in time are selected for multiple tracked airplanes in Germany. Based on the intersection patterns of the vectors, the direction of movement flows of multiple entities moving in flocks are visually detectable. This proofs again the assumption that movement is definable as progressively varying continuous fields. In case of dense point distributions, it is possible to use kernel density estimation (KDE). Krisp et al. (2011) extend this idea and use adaptive and directed KDE for the visual traffic analysis, which helps detecting movement trends within dynamic point data.

Another form of surface generation consists of estimating the movement densities in raster cells or vector bins, mainly the length and time of the moving entity passing the individual cell or bin. The aim is to find space-time density values as they are defined in Hengl et al. (2008):

$$D_{xyt}(j) = \frac{\hat{d}_j}{\hat{v}_j} \; ; \qquad (16)$$

with $D_{xyt}(j)$ as the space-time density at cell j , $\hat{d}_j$ as the length of the movement path with cell j , and $\hat{v}_j$ as the average velocity of movement within cell j. For a single moving object, the space-time density is, as a simple interpretation, the duration of time the object spends within each cell. Space-time densities are part of spatial field methods (Long and Nelson 2013). The difficult preselection for applying the method is selecting the cell size of j.

Generalization of vehicle movement based on FCD

This subsection of the thesis has the focus to review the possibilities of vehicle movement generalization. Early research on generalization of vehicle movement is coming from transportation engineering (Li et al. 2011, Zhang et al. 2014). Generalization is also helpful for reasoning on general flows of movement. For the case of massive vehicle trajectories, this implies the creation of polygon-based symbols showing origins and destinations of vehicle traffic flows. Andrienko and Andrienko (2011) use a trajectory aggregation technique and visualize those aggregated trajectories via flow symbols. This version of movement generalization varies in degree of abstraction: higher degrees might represent the more general flows within a city between the official cities districts, lower degrees of abstraction are achieved by lowering down the aggregation parameter, which is itself in Euclidean space. The lower degrees of abstraction are useful for detecting specific highway segment structures as selected on-ramps, which only allow movement in one direction (Andrienko and Andrienko 2011).

These examples come from a massive vehicle trajectory data set from private vehicle drivers in Milan (Italy) and consist of an acquisition of several days (Andrienko and Andrienko 2011). The same ideas find application for other case studies as for example tourist taking photos. By inspecting data from photo-sharing social networks, it is possible to visualize the flow movement of certain moving photo tourists. The photo tourist trajectory might imply more abstraction than a tracked vehicle trajectory, since only geo-tagged photos are used to reconstruct the path of entity movement.

Visual analytics approaches for traffic data inspection

With the use of linked views for inspecting traffic data, it is possible to produce different sights on the same data selection. Guo et al. (2011) introduce an interactive visual analytics system, which considers trajectories as entities and includes three linked perspectives: temporal, spatial, and multi-dimensional. Tominski et al. (2012) use selection circles in space, which is referred to as time lens, for selecting movement data partitions. Other approaches include interactive selections based on road segments for deriving traffic congestion durations based on traffic jam propagation graphs (Wang et al. 2013b). These graphs aim to concatenate spatially and temporally related traffic congestion events. This appears problematic, since only two-dimensional views are present with differing colorations, additional symbols varying in size and simple diagrams.

In contrast to Wang et al. (2013b), the visualization approaches in this work tend more to show traffic congestion in a more intuitively recognizable and visually stimulating way for users without detailed knowledge on the visualization system components.

There are many possibilities to group traffic visualization approaches. Chen et al. (2015), for example, propose the grouping into time-based, spatial-property-based, spatio-temporal, and multi-property-based traffic visualization.

Exploratory data analysis (EDA)

In general, it is to say that some of the mentioned visual analysis examples make use of procedure flow described by the term exploratory data analysis (EDA). This term was summarized by Keim et al. (2004) as a sequence of the three steps “Overview”, “Zoom and Filter” and “Details on demand”. The overview represents the inspected data in a summarized view, namely global view in this work. By using the “Zoom and Filter” functions, which are the used data analysis methods, it is possible to detect movement patterns. After detecting some patterns, “Details on demand” refers to the inspection of certain details in the data, which is dependent on the field of interest for analysis. This inspection is a detailed view on the data, namely the local view in this work. The aim of “Details on demand” or the local view is to propose a hypothesis in the end (Keim et al. 2004). Exploratory Data Analysis and the mentioned examples for visual analysis of FCD is the theoretical base for visualizing the results of the, to be presented, test FCD sets.

3.4 The particularity of Floating Taxi Data (FTD)

In comparison with the tracking of private vehicles, tracking taxis has several advantages. One is the often-easier realization of the observation. This advantage results from the already established communication infrastructure with a trunked radio system and a dispatch system (Lorkowski et al. 2003). Other advantages result from knowledge about the taxi service, since the aim of every driver is to earn a high amount of money resulting from many customers. This knowledge allows inferring specific mobility patterns of the investigation areas together with information on the driving behavior of each individual taxi driver. In 2013, Castro et al. (2013) recognize the similarity of Taxi GNSS trajectory data sets and therefore classify the, until the mentioned time, accessible work on the analysis of taxi trajectories. There are mainly three groups, which respectively have subdivisions into areas of similar application. A diagram pictures the whole classification in Figure 13, where each group of taxi data analysis is rectangular and its respective subareas are oval symbols in green. In particular, Figure 13 has the focus on mining Floating Taxi Data (FTD) for gaining further knowledge. Mining FTD is a special case of mining movement data or spatiotemporal data. Therefore, each subarea of research has a connection with the term dynamics and lists its special applications via red rectangles.

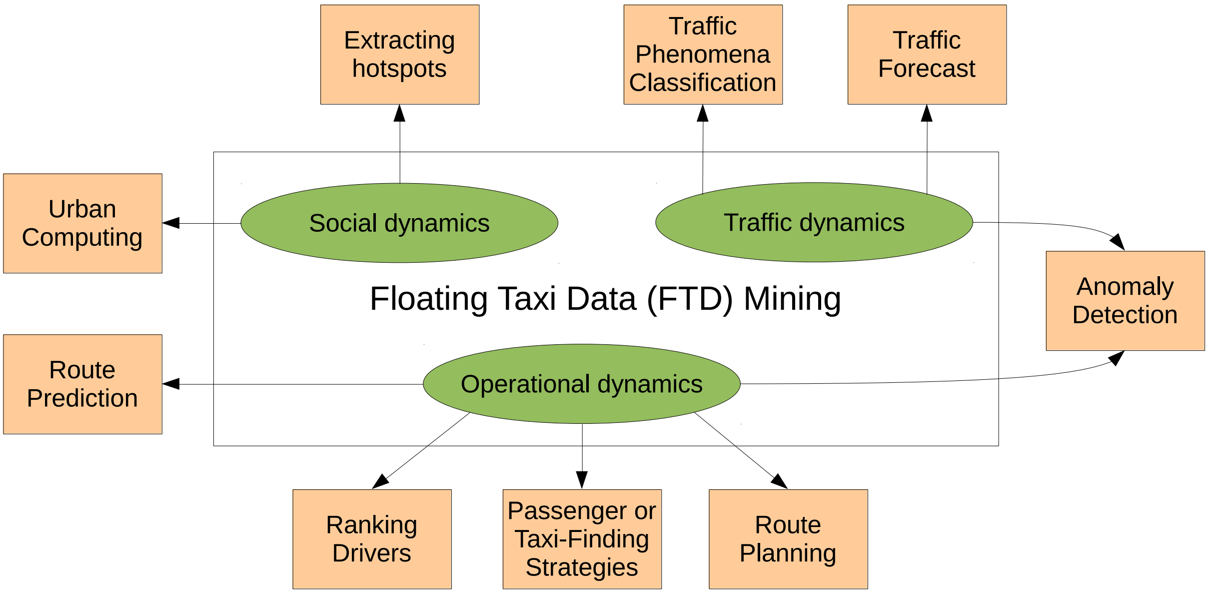


Figure 13: Classification of existing work on the analysis of Floating Taxi Data (FTD) Mining, proposed by Castro et al. (2013).

The three main groups in Figure 13 consist of the fields of interest for different groups of analyst and planners. Nevertheless, connections between the three groups and their applications are very presumable and have numerous examples in the literature. The specific characteristics of taxi FCD in comparison with FCD from vehicles with other functionalities rely mainly in different mobility patterns. Those patterns result from the fact that there is a predetermination of taxi driver destinations by the taxi customers.

FCD from busses, as for example part of the public transport, show more or less very periodical patterns. Exceptions from these patterns can reveal certain events that influenced the bus movements (Mazimpaka and Timpf 2017, 2016b).

In contrast to this, FCD from private vehicles (as for example enviroCar or INRIX) have mostly less periodic movement pattern as busses, but still are in general more periodic as taxi trajectories. The reason is those private vehicle owners usually commute between home and working places. Taxi driver have more a hunting and waiting for passengers pattern. Sometimes many individual taxi drivers share the same vehicle or at least the same identification, as it is observable from simple data analysis tasks. This is the reason why nowadays more and more research on operational dynamics appears.

In the 2000s and 2010s, numerous FCD sets appeared as massive collections of the movement of thousand traffic participants. Especially in Chinese cities, Taxi FCD sets appear as a side product of the usual dispatcher service with already established communication infrastructure for surveying the taxi vehicles (Gühnemann et al. 2004). The data consists in the simplest case of records with coordinates, car identifications and time stamps. Furthermore, many FTD sets have record entries coming from the on-board sensors of the taxi: instantaneous velocity, driving direction and status of passenger load, to mention the most important.

Besides highways between megacities, there are periodical traffic congestion events within the cities (Sun et al. 2009). Successful taxi drivers that avoid traffic congestion locally know these events (Liu et al. 2009). This shows the big advantage of using FTD for the derivation of daily mobility patterns: the knowledge of origins and destinations of taxi trips. Numerous research use this type of data for inferring positions of transport hubs (Ding et al. 2015), detecting mass events, or as an indicator for individual driving behavior (Liu et al. 2009).

Other use cases of FTD in the past years include the estimation of traffic situations, mainly in urban areas (Liu and Ban 2013). Tang et al. (2015) mention that FTD is useful for urban and transportation planning. Ding et al. (2016) use for example the space-time cube as a visualization approach for detecting the connection of pick-up and drop-off points on the one hand, and the taxi states of occupancy and non-occupancy. These patterns are useful for detecting transportation hubs such as airports (Ding et al. 2015). In general, extracting taxi pick-up and drop-off points from raw data is not a problem in case the data has an attribute about the taxi's operational status. The following step might be the subsequent detection of clusters, which has various possibilities. Krisp et al. (2012), for example, are using the k-means algorithm for detecting taxi pick-up and drop-off hotspots.

Precomputing information supplements from the field of behavioral studies

In Pang et al. (2011), taxis are observed 24 hours a day without significant pauses in motion. This assumption mainly comes from local geographic knowledge, because the taxi drivers usually have changing shifts. Keeping this in mind, trajectory segmentations of certain observed vehicles may represent movement of numerous individual taxi drivers. There are many studies on the driving behavior of taxi drivers based on taxi FCD, as the detection of the most successful within selected areas. The detected behavior of successful drivers often show correlations with traffic dynamics, as for example the movement patterns influenced by specific traffic situations throughout the day. It is possible to extract behavior from FCD and to classify it by different movement or traffic patterns.

Visualization of Floating Taxi Data (FTD)

Despite Mustafa et al. (2009), who visualize travel times by coloring bidirectional road links, Keler et al. (2016) use the stoplight metaphor (Goldsberry 2005, 2008) for coloring polygonal road segments. The visualized taxi density on these road segments appears via extruded road polygons in a 2.5D view. Additionally, it is possible to make use of selection circles (Keler and Krisp 2015b) for selecting temporally and spatially restricted (circle in space) data partitions for further inspection (Keler et al. 2016). By using these tools, Keler and Krisp (2016b) identify FTD-based travel times between 13 prominent crossings in Shanghai. Estimated travel times vary in a similar way as in Sun et al. (2009) with their global congestion indices for Shanghai. This shows that the traffic congestion events in the city center have a high influence on the overall traffic situation in a city.

This may come from the similarities of the underlying data sets (2007 and 2009). Comparisons to recent travel times from the Google API, show in most cases similar travel time even if between 2007 and 2015 many change appeared in the urban transportation infrastructure (Keler and Krisp 2016b).

3.5 The map matching problem – state of the art

The previous sections emphasize the particularity of movement data, of which FCD is a special case. In most cases, especially in urban environments, FCD result from vehicles passing over road segments of road networks or other transportation infrastructure elements. Due to the nature of positioning devices for FCD collection, there are several challenges in matching the recorded positions onto network segments. The integration of raw, not preprocessed, FCD trajectories with a road network is for Liu et al. (2016b) often referred to as a map matching (MM) problem. Depending on different investigation areas with different transportation infrastructures, several diverse problems are connected with matching entity movements with network features (Li et al. 2015b).

The use of Map Matching algorithms for vehicle trajectories

Since vehicle movement usually appears on road elements, there is a need for additional transportation infrastructure information when performing movement data analysis. In its practical form, this is often a road network, as for example from crowd-sourced information such as the OpenStreetMap (OSM) project, which has one of the most accurate freely available digitized road networks (Stanica et al. 2013). The association methods of road elements with vehicle routes are widely known as map matching algorithms (Yang and Meng 2015) and appeared in the years from 1989 till 2006 in 35 quite different variations (Quddus et al. 2007).

Positioning technologies like global navigation satellite systems (GNSS) imply certain rules that allow computing a position on the earth surface or atmosphere with a time stamp. For the case of GPS, one needs at least four visible satellites, of which each movement trajectory around the earth's surface is known and accurately describable. It is not always possible to reach four visible satellites, especially in urban environments, due to elevated road segments, tunnels, and urban canyons, due to signal losses or lower GDOP values (Langley 1999).

The idea behind map matching is, not only, connecting or associating movement points (positions of movement) to the digitized transportation infrastructure (road network). It is inferring the actual positions of movement on these segments, which results in many applications in the consecutive change of trajectory position coordinates. Additionally, it is to say that there are different MM algorithms with varying matching qualities. This incorporates the different aims of using MM, for movement data with varying qualities.

Classification of different Map Matching techniques

Brakatsoulas et al. (2005) partition MM techniques into two main groups: incremental MM, which is a "classical" approach and global MM methods. One global MM method consists of curve-graph matching, where the curve is the movement trajectory represented as a polyline (different interpolation types possible).

In general, the number of newly appearing MM methods is growing: in 2012, Zhao et al. (2012) stated a number of 36 known individual MM algorithms. This number is nowadays (2017) far more than 50. As a general classification, Zhao et al. (2012) introduce four groups of MM techniques: geometrical, topological, probabilistic and extended. The left part of Figure 14 pictures this classification. There are at least four different groups for MM, mainly resulting from the different case studies and data sets. Achieving reliable MM results in complex urban transportation infrastructures is difficult due to uncertainties in the two-dimensional FCD and the differentiation between elevated roads and the roads immediate below (Li et al. 2007b). The fact that most FCD sets are available without height components makes MM applications for urban road networks challenging. The supplementary objective of all the existing MM algorithms appears as a tradeoff between the possibility of reading and processing of large data volumes. Large data volumes refer to FCD and to road network data and at the same time. One goal of all MM techniques are reasonable computation times (see Figure 14), which may be the crucial factor for the introduction and managing of MM applications for the daily usage, as traffic forecast and navigation services.

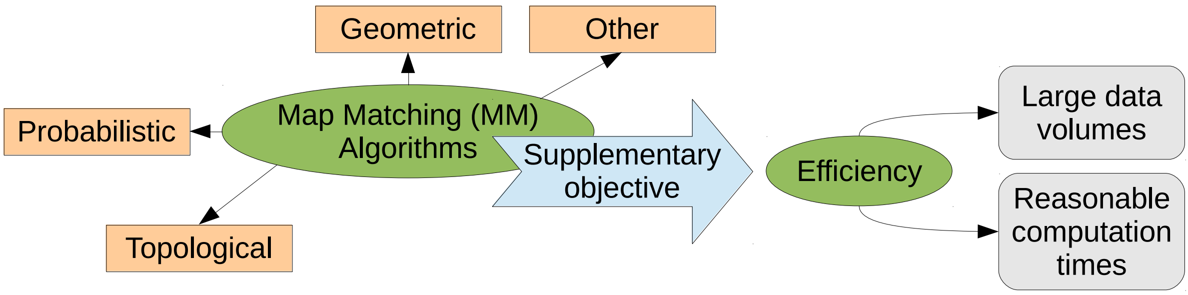


Figure 14: Grouping of available Map Matching algorithms, according to Zhao et al. (2012).

Typical products of MM with FCD are Traffic maps, which can be derived from taxi data (Liu et al. 2012a) and can describe traffic with varying resolution in time and space. The grouping in Figure 14 is extendable with two additional MM classes: naïve MM and Map Inference techniques. Map inference is actually not a MM technique, as it produces a road network with certain topology based on vehicle trajectories. These cases are useful in applications, where no road network information is present and FCD covers many individual road segments.

Naïve MM, the other additional class, describes very simple techniques that appeared in the late 1990s as for example mentioned by White et al. (2000) or Bernstein and Kornhauser (1996). The main idea consists of matching each movement position with the nearest neighboring road segment. A common challenge of these methods was the coping with the absence of instantaneous velocity records. One problem in these methods was the absence of instantaneously recorded velocity values in the used FCD sets. Therefore, computing instantaneous velocity values bases on distances and time differences between consecutive FCD records. Resulting from this absence, there is nearly always an overestimation of velocity at road intersections, when vehicles are turning into other road segments. Ranacher et al. (2016b) discuss the idea of improving velocity estimation by deriving movement parameters from different movement spaces. Observing movement trajectories of vehicles in different spaces can benefit finding more insights on its movement properties. For the case of interpolating the path between two consecutive records, different sampling strategies are possible. With the strategy of matching the movement positions first and subsequent computing of velocities, it is for example possible to falsify velocity values. Therefore, previous testing might benefit the computing of various movement parameters. Due to the important nature of MM techniques to improve different kinds of applications, from routing applications to traffic forecast services, research on this topic is numerous and growing exponentially. Many approaches result in patents, whose major parts are coming from China, as for example the following: Online map matching method based on floating car data on tunnel road section CN 104504898 A[18].

In general, there is a differentiation between online MM and offline MM. Online MM refers more to fast solutions, which include positioning on road segments, together with routing algorithms as in applications of mobile devices. This type of matching focuses more on the accurate matching of the recent positions. Offline MM focus on postprocessing FCD with low sampling rates for reconstructing the detailed movement trajectories of usually numerous drivers. This finds more application in scientific research, for example, when analyzing historical movement trajectories for testing the development of altering qualities of the traffic situation throughout the years.

Geometrical and topological MM methods

Zhao et al. (2012) introduce a MM model that connects FCD and roadway network buffers. Therefore, they are using the three different criteria: distance, direction, and connectivity. The desired result of this method is to represent individual vehicle trajectories by sequences of visited roads. Starting from not preprocessed FCD they focus on removing FCD record outliers based on comparing Euclidean distances (criterion 1) and driving directions (criterion 2) between probe data and road segments. The 3rd criterion is checking for connectivity between different matched road segments (Zhao et al. 2012).

Probabilistic and extended MM methods

The introduction of MM appears differently, depending on the favored product. Liu et al. (2008), for example, want to calculate routing velocities for traffic aware vehicle navigation services. Therefore, they apply matching of massive FCD to road segments of Shanghai's road network. The desired information is the time dependent average velocity in each segment based on historical FCD. For reasons of validation for matching real-time FCD, typical average velocity values are used for comparing the velocity trends. The last mentioned is useful for giving answer on the differently used road lanes and driving directions of road segments (Liu et al. 2008).

There are MM technique, which make use of fuzzy logic (Quddus et al. 2006) or increase matching efficiency by scaling with map reduce (Tiwari et al. 2014). In general, there is a bigger differentiation between high frequency MM and low frequency MM, as mentioned by Quddus and Washington (2015). Low frequency MM appears more difficult in complex urban environments and often requires unconventional approaches. Rohani et al. (2016) show a modern and extended MM approach, where the tracked vehicles communicate via V2V communication technologies or via vehicular ad hoc networks (VANET).

After matching several records with road segments, the information in arcs is usable for vehicle routing applications. Time-dependent average velocity can serve to compute more realistic travel times. With travel time predictions, it is possible to improve vehicle routing applications in the way of including estimated traffic situation qualities into route computation. The question here is how to deal

[18] https://www.google.com/patents/CN104504898A?cl=en

with the temporal component, when the traffic situation is changing during following the proposed route. Solving this question might improve the quality of traffic-aware vehicle routing applications. One remaining challenge is the temporal resolution of averaged values, since traffic situation changes during the traversing the actual (generated) route. Ji et al. (2010) test dynamic travel time prediction with Kalman filtering. Based on partitioned road segments, FCD records are connected and average travel times are computed. Due to the measurements of instantaneous speed from each observed probe vehicle, averaged time is often longer than the prediction with Kalman filtering. The reason for the last mentioned is the Kalman filtering assumption that the observed vehicles are in continuous movement, without short stopping events, where the instantaneous velocity is equal to zero km/h (Ji et al. 2010). Travel times coming from massive FCD sets are often connected to road segments for traffic-aware car navigation services. The base for computing these weights for routing is computing the harmonic mean of velocities from individual vehicle movements.

Construction of routable road networks from FCD – using map inference techniques

Massive FCD can help to reconstruct the road network based on so-called map inference algorithms. Existing literature on map inference consists for Liu et al. (2012c) of three different categories. In general, used FCD sets for map inference should have a relatively high positioning quality and a high spatial distribution. The spatial distribution of FCD records has a connection with the grade of coverage of road segments compared to all available road segments of a whole road network. For inferring the whole road network of a city, it is beneficial, when FCD positions or FCD trajectories have dense large-area coverage of the investigation area. That means that longer trajectories that pass numerous different places of the investigation area are favorable. Additionally, it is beneficial to have long observation periods, since this enables to detect eventual outliers. Many similar trajectories, as how they result from commuting behavior of the driver can benefit the resulting spatial accuracy of the inferred road network.

In general, map inference methods consist of following three steps:

1. Selection of FCD records based on data quality
2. Weighting based on spatial and temporal resolution (not every information is used for derived traffic status or effective road network)
3. Further inspections: e.g. the comparison between a dynamically varying continuous field and movement (described by trajectories)

Updating established road network via map inference techniques

Literature shows crowdsourced solutions for updating established road network geodata via movement trajectories. These trajectories usually result from tracking via smartphone usage. The CrowdAtlas server for example has its own stream processing flow that iteratively infers road segments (Wang et al. 2013a). Uploaded trajectories come from volunteering users that benefit the CrowdAtlas project. The bases for updating are specialized road map inference techniques performed on the server side. The first stage within the road inference process is applying MM of smartphone-generated trajectories on recent OSM road segments. The MM step itself is, in many ways, a preprocessing step as it excludes matching segments and focuses on unmatched traces. This guarantees reasonable computation times and avoids higher server traffic. The final step is inferring the new roads from unmatched trace segments overlaid on the latest OSM base map. The implementation of this procedure has autonomous updating together with a reasonability check within a GIS, which base on using recent satellite or aerial imagery.

These recent applications allow the usage of A-GPS that, in some cases might increase the positioning quality (higher precision). A-GPS allows, in most cases of urban environments, faster positioning and less energy consumption. Using smartphones for positioning allows, in certain cases, even inferring a whole bundle of connected missing roads in an iterative way. Bundles of minor roads at densely-built housing block that are missing in OSM often indicate urban canyon areas (Wang et al. 2013a).

3.6 Detecting FCD anomalies in the network space – research on traffic anomaly detection

In urban environments, the movement acquisition of tracked vehicles via the FCD technology occurs mainly on road networks. The digital map of road networks is for Zhang et al. (2011c) one of the important component for vehicle navigation, which often has the problem of providing not enough details about complex intersections. Besides this, the usage of an additional road network can improve the quality of massive vehicle trajectory data sets (Liu et al. 2012b). This is similar to using FCD traces for improving validation quality of incomplete or inaccurate road network maps (Pereira et al. 2009). The detailed representation of road networks is indisputably helpful for understanding vehicle movement and traffic information in general. It is for example feasible to associated traffic congestion events with selected road intersections or light-signal systems (changing traffic light signal states).

As previously mentioned, traffic congestion has numerous possible definitions. Therefore, it is important to introduce a taxonomy with the two very general terms of processes and events (Galton 2008) and associate all traffic patterns as such. Since traffic appears on linear road segments and lanes, it is possible to define traffic events as linear events (Tang et al. 2016a). This definition allows several analysis options and facilitates understanding when working with vector GIS software. Exceptional cases of linear events might include punctual and surficial events resulting from this definition. A punctual event might be slowing down on a bumper or the spatiotemporal position of a traffic accident. One idea of the traffic pattern analysis framework consists partially of this imagination. By introducing traffic congestion as a surficial event on a very general scale, it is possible to associate bundles of connected road segments.

Traffic congestion appears usually on road networks and its simplest representation includes associating arcs and nodes with such. Additionally, it is feasible to reconstruct traffic congestion propagation graphs and represent the event via associated road segments. These road segments can imply realistic connectivity via directed arcs. A weighted and directed graph might then represent road links (single lines or connected polylines) with associated travel delays. These travel delays result from the difference between longer travel times and expected travel times.

Another idea of the framework is introducing polygons to represent surficial events for showing that traffic congestion might be a barrier for other participants of the urban transportation infrastructure. Therefore, traditional GIS operations may serve for overlapping road segments with areas not accessible for vehicles (Wang and Zlatanova 2013). This association might serve as an indicator for congestion-influenced areas, where areas are only accessible via restricted movement possibilities: cars on roads that are accessible only for vehicle drivers. Many approaches represent traffic congestion by selected road segments that only allow low travel times.

As an extension, Long et al. (2008) make use of the connectivity of road links and detect congestion propagation patterns. But, how to deal with road intersection points or very long road link? Another question arises in distinguishing between traffic light stop patterns of taxis (Protschky et al. 2015b), stop-and go traffic (Ranacher et al. 2016b) and real congestion, where vehicles are still standing. Besides mean travel times for road links for specific times of the day (Sohr et al. 2010), it is possible to detect typical and abnormal traffic as mentioned in Lan et al. (2014) and Long et al. (2008). This procedures focus mainly on the comparison of historical and recent FCD or FTD. This differentiation bases on statistics of instantaneous and average velocities and, resulting from the previous, travel times and congestion durations (traffic delays). Based on matching FCD on road links, designing network models might help to predict travel time in urban environments (Zhang et al. 2016a).

Matching traffic information

After matching FCD to road segments of a network, there are several possible subsequent steps. In case of map matched FCD, it is possible to infer more details of the road infrastructure, as the number of lanes (Dolancic 2016). This is useful for estimating properties of the transportation infrastructure. Another option is inferring mobility dynamics by using FCD. By using the road network characteristic connectivity, when given, it is possible to construct graphs that represent the general dynamics within the vehicle-accessible road network.

Examples for traffic anomaly detection applications are delivered by Lan et al. (2014), Pang et al. (2011), Pang et al. (2013), Zhang et al. (2011a) and Chen et al. (2012). All the mentioned examples imply the analysis in the network space. In particular those analyses imply usage of arcs and nodes of the network that are enriched with FCD (and other types from mobile sensors) or data from static sensors. By estimating vehicle densities or average velocities for selected time windows, it is possible to calculate variances and standard deviations for selected road segments. In case of big differences in the observations for selected times, it is possible to catch certain arcs or nodes with not typical properties. Those detections might indicate non-recurrent traffic congestion, which result from accidents or other unusual events. One important example for traffic anomalies are traffic accidents as explained in Okabe et al. (2009) and Xie and Yan (2008). In general, this information is useful for FCD anomaly detection, since it can then serve as a starting point for further analyses. In most cases, historical traffic accident data is hard to obtain from governmental agencies due to restricted privacy aspects.

Speed profiles

FCD anomalies are detectable via visual inspections within map views. The focus of these geovisualization attempts is to get general information on affected roads by different traffic events. These attempts often do not include the representation of the traffic state variability over time. Nevertheless, this is feasible via creating time series of traffic states with specific time windows. Within interactive visualization systems, changes in traffic states are detectable via interactions as for example the successive changing of the time window. Another option is to create speed profiles based on the instantaneous velocity values of matched points (Gühnemann et al. 2004). Speed profiles usually consist of average velocity values (for a selected day of the week) for partitions of one selected road or road segment, often differentiated into, when applicable, into different driving directions. One example for a computed speed profile is present in Figure 15.

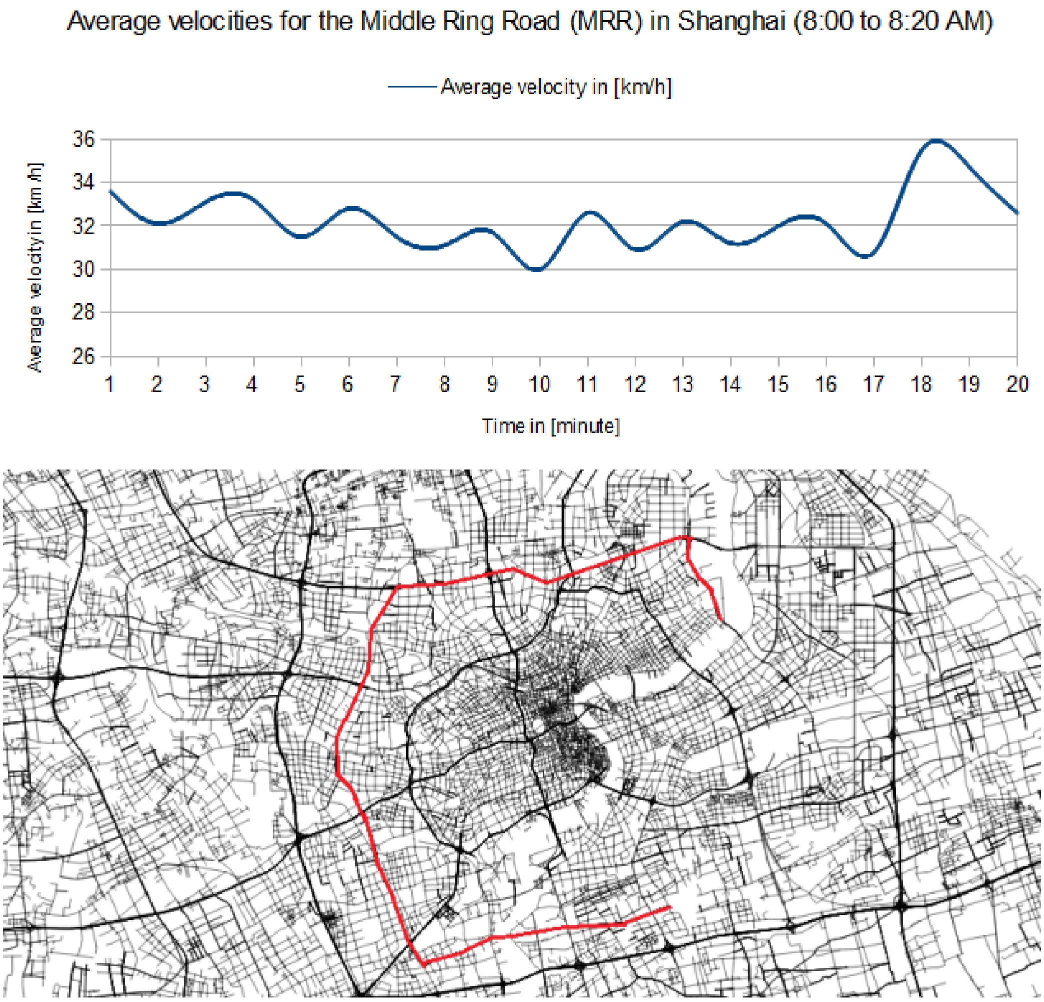


Figure 15: Speed profile for the middle ring road (MRR) in Shanghai for observation time of 20 minutes on one selected Wednesday in 2007.

Figure 15 shows the speed profile for the middle ring road in Shanghai with average velocities of thousands of observed taxis. Useful details are the different average speed minima, which might be associated with intersections and intersections with traffic lights. Another form of speed profile shows the average velocity for selected partitions of the road, which will be explained at a later point.

A complex urban road network might result in the same number of different speed profiles as road segments. For simplifying this step and for increasing computational efficiency there is the possibility of clustering speed profiles. Resulting speed profile clusters might then indicate different types of roads as well as different functionalities of selected road partitions, e.g., connection, on-ramp or intersection (Erdelić et al. 2015).

There are various possibilities for computing travel times. One possibility is weighting segments based on the space-mean speed for computing travel times. Another possibility is matching taxi trajectories with road segments. This can consist of matching instantaneous velocities of every record on each road segment and subsequently computing arithmetic mean speeds (time-mean speed) for the respective segment. There are numerous possibilities for computing travel times. Many options include association of every FTD position to given road segments. Subsequently, it is possible to enrich every road segment with computed average speed values for every observed time window.

Every FCD point has an instantaneous speed values, which is usable for averaging on a selected road segment. After this step, every segment has an average velocity weight. The simplest possibility for travel time computation of various routes on the network is to multiply every average speed value with the length of the road segment. The result is the average travel time for one selected time window. Since instantaneous records are averaged, this average speed is named time-mean speed. Variations in this travel times may indicate changes in the traffic situations. One other option is to compute the average speed with consecutive points of every driver and respective time stamp values. This has one additional condition: the harmonic mean replaces the arithmetic mean of computed velocities. Subsequently, the average speed values on every segment become space-mean speeds.

Road-segment-wise travel times for inspecting congestion propagation and identifying traffic bottlenecks

Besides detecting FCD anomalies via average speed profiles or vehicle densities, also the travel times can serve as indicators. Computation of travel times is possible with matching instantaneous velocity values onto road segments (Ji et al. 2010). By multiplication of average instantaneous velocity with the length of the road segment (in km), travel times for specific time window are inferable. Detected FCD anomalies can also be used for modelling congestion propagation together with the identification of bottlenecks as in Chawla et al. (2012), Ji et al. (2014) and Long et al. (2008). Bottlenecks can be static in case the capacities of selected road segments are exceeding periodically. The case of dynamic bottlenecks might also appear periodical, due to specific traffic congestion situations. Besides temporal aspects as periodical appearance and congestion event duration, there are possibilities to classify congestion by spatial properties. For D'Este et al. (1999), there are two groups of traffic congestion classified by the spatial extent of its appearance. The first one is the so-called point congestion, which affects isolated bottlenecks. This definition indicated the possible representation of point congestion by a point in space. The other variation is network congestion (D'Este et al. 1999). Network congestion is, for D'Este et al. (1999), an area-wide phenomenon and has unstable flow conditions. Network congestion can appear at different parts of the inspected road network. This might be an indication for general lacks of road capacities (D'Este et al. 1999). The inference of road capacities in urban environments are the aim of Liu et al. (2016a), since the knowledge about transport capacity limit can benefit the understanding of complex urban mobility patterns.

Traffic lights and traffic congestion at intersections - from FCD to traffic light dynamics

Protschky et al. (2015b) show one technique how to learn traffic light parameters with FCD. The idea in this approach is to infer traffic light's cycle plans by interpreting the vehicle trajectories within defined road intersection areas. The focus to derive traffic light dynamics is the use of temporally sparse FCD. With a detection accuracy of 99% for cycle times, each intersection needs at least 30 vehicle trajectories. Additionally, traffic lights are for Protschky et al. (2015b) one of the top influencing factors in traffic flow. Despite regulation of traffic, traffic light also have negative effects as they partially lead to traffic jams, increase CO2-emission and decrease driving comfort. As a solution to avoid these traffic light influences Protschky et al. (2015b) propose green light optimal speed advisory and traffic light optimized start-stop control for reducing CO2-emission. For saving travel time, in deed not in respect to traffic congestion events, Protschky et al. (2015b) propose traffic signal adaptive routing. The needed input information for such applications requires future traffic light signal states (as a forecast) and real-time traffic flow information (already given in certain detail by

TMC). Following up the needed information, this information is needed in selected areas of the transportation infrastructure: intersections with traffic lights.

The data problem is again the lack of freely available information, as in Germany where local city administration departments host this information. Therefore, Protschky et al. (2015b) go in the indirect traffic light information retrieval by using FCD. They have an approach of learning traffic lights behavior by interpreting FCD information. Most important attributes of the last mentioned are position, speed and heading of the individual vehicles (Protschky et al. 2015b).

When talking about traffic lights, there are the following continuously appearing fixed terms: Phases, Cycle time, Program and Signal Phase and Timing information (SPaT). In general it is possible to differ between adaptive, dynamic and actuated signals (Protschky et al. 2015b).

For Ramezani and Geroliminis (2014), the queues at signalized intersections are the main cause of traffic delays and travel time variability in urban networks. This shows that besides traffic bottlenecks on selected road segments, also signalization via traffic lights or signs might influence the lowering of capacity. Later, Ramezani and Geroliminis (2014) propose a uniform arrival queue estimation procedure.

Traffic density estimation by using FCD

In Shang et al. (2014), the application is focused on estimating traffic volume (km per minute) from relatively sparse FTD from Beijing. The drawback in this method is that no reliable estimation is guaranteed in case of inaccurate road network information. High vehicle density based on FTD is for Shang et al. (2014) no indication for real high vehicle densities, since no proportion of participating tracked vehicles on the entire traffic is derivable. As methods for FCD-based inference of traffic parameters, Shang et al. (2014) propose interpolation and historical data usage. The last mentioned is useful for finding so-called hot routes, which are for Li et al. (2007a) frequently visited routes in a road network. The knowledge of these routes helps planners and data analysts to better understand the connections between daily traffic flows and the causes of traffic congestion.

Link-based and trajectory-based traffic anomaly detection

In a simpler way, Lan et al. (2014) are introducing a completely data driven framework for detecting traffic anomalies: estimating variances of vehicle position sums for road segments within selected time windows. Afterwards, a linear regression is applied on historical traffic density values and travel times. One bigger opportunity in designing efficient traffic anomaly detection applications is to handle real time data streams (Pang et al. 2011). The difficult task is to gain traffic state knowledge on the fly with the FCD streams, since computation is often extensive. A known algorithm, which can handle this task is iBAT by Zhang et al. (2011a). One compromise as a precondition of anomaly detection is the needed knowledge of entire trajectories (with defined start and end points). Depending on the inner trajectory patterns, Chen et al. (2012) propose a learning algorithm to handle movement outliers. Chen et al. (2012) are using whole trajectories between defined taxi service origin and destination points for detecting anomalous trajectories in real-time. The result consists of trajectory partitions that have the property of being abnormal. In comparison with iBAT, which was defined by Zhang et al. (2011a), iBOAT by Chen et al. (2012) allows an decrease in computation times. In case there is a long-term acquisition of entities, the resulted implies typical daily mobility patterns. For Ye et al. (2009), this case allows data mining of individual life patterns. Starting from the very general trajectory data mining, associations of certain results of stay point detection with predefined life pattern formalizations are possible. When inspecting network space, it should be clear which type of network is used, especially for the traffic analysis framework. Li et al. (2015a), for example, represent the road network of Beijing as a directed road network. Subsequently, when enriched with traffic flow information, clustering results consist of strongly connected components: arcs connected via nodes. Additionally, there is a restriction of movement direction on the network.

Traffic congestion detection by using FCD

Traffic congestion detection by using the tracks of numerous observed vehicles in the same investigation area is often connected with previous map matching to selected road segments. Consequently, numerous FCD points are assigned to segments and deliver the source for aggregated velocity and vehicle density information. This is a common approach for the detection of traffic congestion by using FCD. This is the starting point of Rempe et al. (2016), where connected arcs with similar velocity values are clustered. The similariy of velocity is expressed via similarity matrix. The aim is to find connected parts of the network influenced by traffic congestion for selected time components.

One remaining challenge is to detect moving clusters from moving entities, as the appearance of a traffic congestion propagation with influenced slow moving vehicles. One approach that focus on the discovery of such moving clusters is coming from Kalnis et al. (2005), who introduce three different clustering algorithms. Data clustering can heavily vary in its functioning and is dependent on the, to be clustered, data type and aim of the application.

3.7 Clustering movement - as a data mining task and possible FCD application

Numerous research from different data science domains include the application of clustering techniques. In the context of data mining, clustering is a frequently applied group of methods. Clustering has its origin in the pre computer era and usually consists of a task to find similarities between features based on a distance measures (Jain et al. 1999). In particular, the cluster analysis is the process of organizing a collection of patterns into clusters based on similarity. In case of spatial data, this distance measure is often Euclidean space, which means that spatially nearby-situated features are likely to be in the same cluster. For multivariate data, this might also be a distance measure in feature space, which paraphrases the values of different attributes. In general, Jain et al. (1999) state that patterns within valid clusters are more similar to each other than they are to a pattern belonging to a different cluster.

For Jain and Dubes (1988), the definition of a pattern proximity measure is data domain dependent and thus has to be provided appropriately. Based on the findings of Jain et al. (1999) and Jain and Dubes (1988), it is possible to define at least five general steps or components of a clustering task.

These components have a chronological order, which is shown in the following, (Jain et al. 1999):

1. Pattern representation (optionally including feature extraction and/or selection),
2. Definition of a pattern proximity measure appropriate to the data domain,
3. Clustering or grouping,
4. Data abstraction (if needed), and
5. Assessment of output (if needed)

These five components are summarized parts of a workflow diagram in Figure 16.

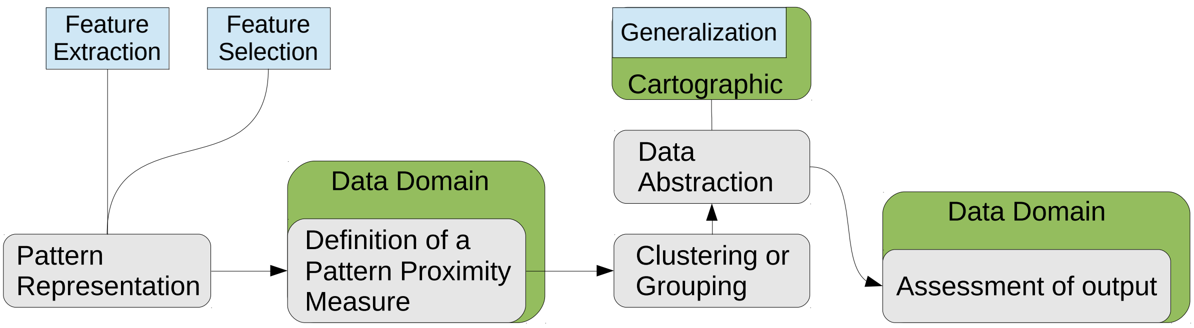


Figure 16: Five general steps of a clustering task, based on Jain et al. (1999).

The mentioned representations of clustering are very general and show only the methodological backbone of nowadays used techniques. In many cases, clustering is a statistical data analysis technique and one task of (exploratory) data mining. Historically, it is also a part of knowledge discovery in databases (KDD). Its usage appears in nearly every specific domain of data science, and it is a helpful component in various decision-making and machine learning applications. Throughout this historical development, clustering techniques developed themselves into groups, often depending on their methodological properties. Shah et al. (2012) classify the established clustering methods into five groups, which are pictured in Figure 17.

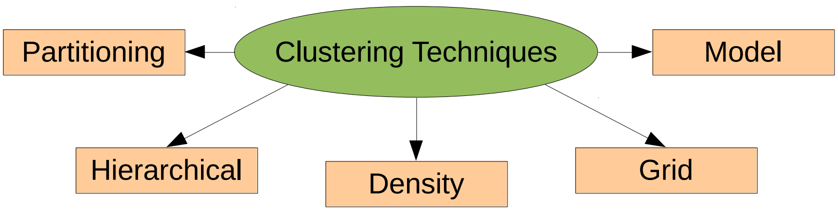


Figure 17: Different groups of clustering techniques, proposed by Shah et al. (2012), modified.

In 1990, only two basic types of clustering algorithms were known: partitioning and hierarchical (Kaufman and Rousseeuw 1990). Besides these grouping, literature states also additional terms as for example graph-based clustering, hard clustering, and soft clustering. Dependent on these groups and the properties of the used algorithms, different clustering tasks are possible for different representations of FCD dependent on the objective of the task. Especially for the groups partitioning and density-based, there are numerous applications for FCD.

Partitioning and hierarchical clustering techniques

Partitioning algorithms construct a partition of a database *D* of *n* objects into a set of *k* clusters. They start with an initial partition of *D* and use an iterative strategy to optimize an objective function (Kaufman and Rousseeuw 1990). The important parameter, which affects the results, is input parameter *k*, namely the number of created clusters. In general, partitioning clustering algorithms make use of two steps. The first step is to choose *k* clusters for minimizing the objective function and the second step is assign each object (e.g. point) to its closest object of a cluster (Ester et al. 1996). Depending on the used technique of partitioning each cluster is represented by its gravity center for the case of k-means algorithm or by one clustered object (e.g. point) next to the gravity center for the case of k-medoid algorithm (Ester et al. 1996). One extended type of the k-medoid method is CLARANS (Clustering Large Applications based on RANdomized Search), as it is defined by Ng and Han (1994).

Partitioning clustering can help in defining functional areas of vehicle fleets. Therefore, a previous extraction of points, where changes of the operational mode occur is the preprocessing step. In this way, Krisp et al. (2012) extract pick-up and drop-off points of taxi passengers by applying k-means point clustering. The aim is detecting the busiest places in Shanghai in terms of frequently visited locations by taxi customers and those locations known as taxi entry points (usually points of interest). The visualization of all detected clusters implies differing coloration within a space-time cube. The idea behind this kind of visualization lies in the efficient differentiation of clusters for facilitating the task of a potential data analyst. Partitioning clustering is not only possible for extracted points of the movement, but for whole vehicle trajectories or regions visited by the observed vehicles. Lee et al. (2007) use a partition-and-group method to first segment trajectories into single line segments and afterwards cluster the trajectory segments by the TRACLUS algorithm. This algorithm is extendable into TraClass (Lee et al. 2008), which consists of two types of clustering: region-based and trajectory-based. Trajectory-based clustering makes use of the trajectories' movement patterns, mainly due to shape and size.

Liu and Ban (2013) use hierarchical clustering for selecting low-speed and stop point of FTD in Wuhan. These extracted stop-and-go movement clusters have a power law distribution. The method for connecting data extracts from different time windows bases on a simple spatial intersection of the clusters in a space-time cube. When no intersection is possible, the certain congestion event is ending. Based on changes in instantaneous speed, Liu and Ban (2013) deliver congestion durations represented as road segment extrusions within a 2.5D view.

Density-based clustering techniques

Another popular group of methods for clustering tracked movement of moving entities is density-based clustering (Shah et al. 2012). The special case of density-based clustering consists not only of assigning clusters based on their distance in feature space or Euclidean space: there is a form of density estimation. The density estimation between points bases on density connectivity. A prominent example of a density-based clustering technique is DBSCAN (Density-Based Spatial Clustering of Applications with Noise), as it is defined by Ester et al. (1996). In DBSCAN, there are two parameters for assigning the clusters: the search distance *epsilon* and the minimum number of points *minPts*. Based on the settings of these two parameters, namely the input information, computed density connectivity of points differs. Depending on the settings, density connected clusters vary in shape and number of points, but not in density. The reason for the latter is the already performed settings of the parameters *minPts* and *epsilon*. There are numerous examples of applying DBSCAN on geodata, especially when estimating hotspots within movement data. Tang et al. (2015) use the density-based point clustering method DBSCAN (Ester et al. 1996) for deriving mobility patterns from Floating Taxi Data (FTD). Whereas OPTICS (Ordering Points To Identify the Clustering Structure) is an extension of the previous in the sense of estimating the most suitable clusters with the most suitable reachability (search distances) for every possible point data set. For the case of massive movement data generated by vehicles, Rinzivillo et al. (2008) are using OPTICS (Ankerst et al. 1999) for detecting traffic congestion of vehicles, which are characterized by their higher vehicle densities. The individual vehicle trajectories are previously aggregated into general flows between bigger areas or whole road segments. When clustering movement data, it is possible to distinguish between clustering single entity movement and group movement. Another question is how to cluster traffic patterns, which in most cases consist of multiple group movements of vehicles. This is challenging, since traffic data, as well as FCD, have different representations.

The Shared Nearest Neighbor (SNN) technique

Another advanced density-based clustering algorithm is Shared Nearest Neighbor (SNN). Originally, SNN originates from Jarvis and Patrick (1973), whose method consists of creating a similarity matrix between inspected features. This similarity matrix includes distances, Euclidean or other specific feature space distances, between every feature of the underlying data set. This idea was later extended by Ertöz et al. (2003), where only the first part of the algorithm was motivated by Jarvis and Patrick (1973). The second part of SNN is density-based clustering similar to Ester et al. (1996) with DBSCAN (Density-Based Spatial Clustering of Applications with Noise).

Based on the Jarvis-Patrick-Scheme (Jarvis and Patrick 1973), a SNN graph is calculated from a similarity matrix by the following equation:

$$\text{similarity}\ (p, q) = \text{size}\ (NN\ (p) \cap NN\ (q)) \tag{17}$$

Ertöz et al. (2003) call this an alternative definition of similarity. In the following, this specific similarity of SNN refers to as SNN-similarity. By means of representing the SNN method by a figurative example, Figure 18 shows the difference between associated next neighbors (a) and SNN (b).

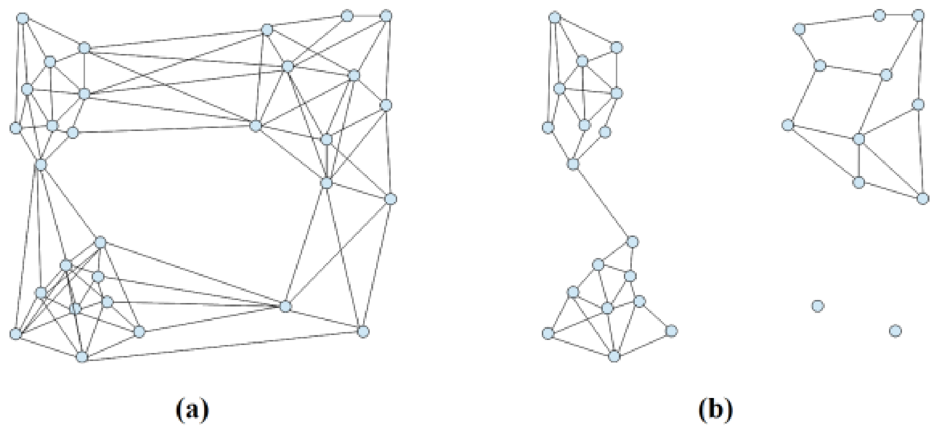


Figure 18: Difference between (a) next neighbors and (b) shared nearest neighbors.

The suitability of SNN for the detection of traffic phenomena is part of this work. The idea behind its use for FCD is coming from the property of the SNN technique to look for similar values in feature space. Similar to Gao et al. (2016), one approach of the framework is to cluster by the spatiotemporal heterogeneity of instantaneous velocity values. The feature spaces might define a combination of Euclidean space and vehicle dynamics values such as instantaneous velocity and driving direction. Nevertheless, many methods for clustering ST data face several problems connected with efficient indexing and, resulting from it, reduction of computation times. The complexity of these methods, density-based clustering techniques, is in general higher than of partitioning clustering techniques. Nevertheless, there are possibilities for increasing computational efficiency.

Creating convex hulls as representation for point clusters

After forming point clusters, it is possible to convert the point groups into area polygons. One possibility for this is computing convex hulls. There are many methods for computing convex hulls or surrounding polygons for a set of points in 2D. Galton and Duckham (2006) mention that the area occupied by points can be aggregated or abstracted in many different ways. These polygons, which result from a certain point distribution in Euclidean space, named as footprint regions of the points. The usual case for mathematicians is, for Galton and Duckham (2006), to create a convex hull. One of its easiest implementations was defined by Jarvis (1973) with the Gift-wrapping algorithm. Many different methods are possible to apply polygon generation based on defined point clusters in Euclidean space. Galton and Duckham (2006) propose nine different criteria for selecting the suitable method for producing the footprint regions.

There are different rules within the algorithms that produce relatively different results. The computational time for executing the algorithms may also vary greatly. The selection of the right technique might surely be dependent on the application. For creating polygons from point clusters in the traffic pattern analysis framework the Gift-wrapping algorithm based on Jarvis (1973) is chosen for creating convex hulls, mainly due to its simplicity and not necessarily due to low computational costs. Table 5 shows also other simple convex hull algorithms that deliver comparable results for two-dimensional point clusters.

Table 5: Overview on simple convex hull algorithms for a set of points in a plane.

Method	Origin	Properties, time complexity
Quickhull	Eddy (1977)	O(n log n)
Gift wrapping	Jarvis (1973)	O(nh)
Monotone chain	Andrew (1979)	O(n)
Graham scan	Graham (1972)	O(n log n)

In the approaches of the framework, there are only performances of the Gift-wrapping algorithm for the detected point clusters. One of the ideas of the framework is representing smaller local congestion events. After applying the gift wrapping algorithm by Jarvis (1973) these smaller congestion events often appear larger in size on a map view. Additionally, it is easier to represent the differing point densities of local traffic congestion clusters via areal polygons. It is to mention that the specific type of convex hull algorithm is not important for the techniques of the framework, since all the algorithms in Table 5 deliver comparable results..

Creating convex hulls as aggregated representations of point clusters

Besides the previous example, it is also possible to created concave hulls. Due to the often fine-grained shape of concave hulls, visual inspection of resulting polygons is difficult on larger scales (for example numerous roads of the network). Since the focus of this work is to observe congestion clusters for selected time windows on a larger scale (in the way of an overview), the convex hull method is preferred. Consequently, concave hulls are more useful for smaller scales, where association with road segments is more exact. Figure 19 pictures this reasoning with Figure 19a as an input of points for creating as convex hull in Figure 19b and a concave hull in Figure 19c.

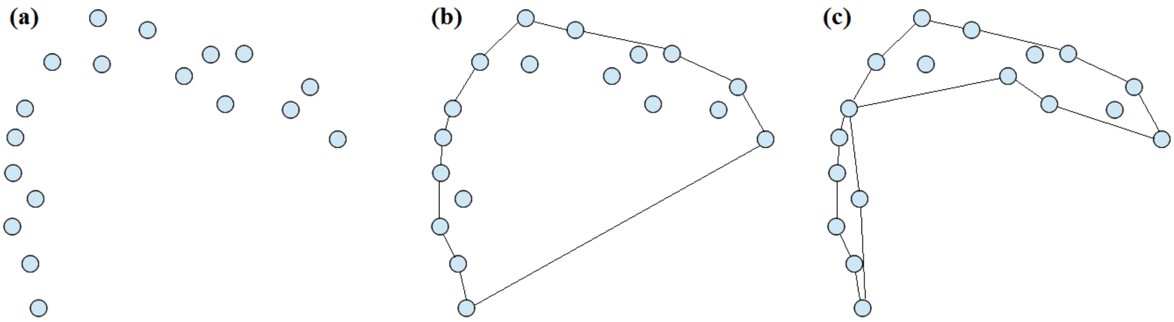


Figure 19: Group of points in space (a), with their convex hull (b) and their concave hull (c), based on Galton (2005).

Galton (2005) mentions that the convex hulls of ‘C’-shaped point clusters, as in Figure 19 dissemble the shape information. In case of generating concave hulls or non-convex hulls, the shape information is usually retained. The disadvantage of concave hulls is for Galton (2005) the absence of a unique way of their definition.

Continuous clustering of movement trajectories and data streams

Besides the handling of static movement data with the previously mentioned clustering techniques, there is extensive ongoing research on how to apply clustering for data streams or dynamic two-dimensional moving data points (Jensen et al. 2007). The group of clustering for spatiotemporal data is named by Costa et al. (2014) as spatiotemporal clustering. Differing from the previous methods is the fact that this mentioned clustering performs continuously for near real-time movement data. This is a far more challenging problem, since the straightforward approach of clustering moving objects periodically is computationally very expensive: for short periods of moving objects data streams previous clustering is not leveraged and long periods may imply parts without any clustering results (Jensen et al. 2007). The main problem is that dynamic data is only handled as usual static object positions without respecting movement trends (Jensen et al. 2007). In general, Nassar et al. (2004) state that designing continuous clustering consists of two options:

1. Developing a new clustering scheme for moving objects, and
2. Extending existent frameworks (for static clustering) by introducing new summary data structures

Jensen et al. (2007) follow the second option and introduce “clustering feature” as the new summary data structures, within which key properties are defined for the incremental detection of moving clusters. These “clustering features” are built as an extension of the Birch algorithm (Zhang et al. 1997).

4. Designing a traffic pattern analysis framework for FCD

This chapter shows the general design specification of a traffic pattern analysis framework. Each step of the framework has its general design specifications in section 4.1. The framework has specific requirements for traffic pattern detection. Therefore, section 4.2 consists of the data preprocessing descriptions. Despite preprocessing vehicle trajectories (FCD) for traffic information derivation, evaluating eventual data on the transportation infrastructure is part of the previous inspection. Section 4.3 focuses on the data handling and the computational steps of the framework, dependent on the inspected traffic phenomena. The contents of this section include six different techniques of the traffic pattern analysis framework. In connection with matching techniques, 4.3.5 shows how to match detected traffic patterns with underlying transportation infrastructure elements. The matched information, which consists of selected road segments together with attribute information, is the input for deriving the macroscopic traffic flow parameter traffic density. After defining computational steps of the framework, its outcomes are matter of visual representation. Therefore, section 4.4 shows the possible visualization options for inferred vehicle traffic patterns, which help evaluating and interpreting the results. Besides establishing a macroscopic view on traffic, this work focuses on inspecting urban traffic congestion in a microscopic view. There are the restrictions for the other vehicle drivers in the same area, which participate in the traffic. It is difficult to assign any thresholds for selected traffic parameters for indicating traffic congestion events. Therefore, most of the inferred products of the framework are areal polygons that represent traffic congestion events for selected time windows. This implies that individual taxi drivers are detectable within specific areas for specific time windows.

The case study for testing the proposed framework bases on Floating Taxi Data (FTD) from thousands of observed vehicles in Shanghai in the year 2007. The aim is to focus more on the relations between the vehicle trajectories (Brakatsoulas et al. 2004) and less on the relations between vehicle trajectories and their environment. The definition of vehicle traffic congestion bases on typical rush hours of the investigation area (Sun et al. 2009) as vehicle-generated barriers for other vehicles. This indicates inspecting a large portion of recurrent traffic congestion events, not leaving the fact out that detection of non-recurrent events appears in a similar way. Additionally, the creation of a semantically enriched network is another goal of the framework. A functional or semantically enriched network might serve as practical information sources for vehicle drivers. By assigning segments of the road network with spatiotemporal information about traffic congestion, it is possible to avoid the affected road segments. In a similar way, the outcomes of the framework might facilitate already established traffic forecast approaches. In this way, it might serve as a traffic forecast service, where the probability of traffic congestion at selected locations of the road network has a representation as discrete moving obstacles for potential traffic participants. Its discrete representation appears as a congestion propagation trajectory that serves as a movement trajectory of various congestion events, depending on computed core points and their movements for selected time windows.

4.1 Design specifications

The framework in this thesis consists of three main components: 1. Data Preprocessing, 2. Data Handling and Computational Steps, and, 3. Result Representation and Visualization. The certain techniques and properties within the components are pictured in Figure 20.

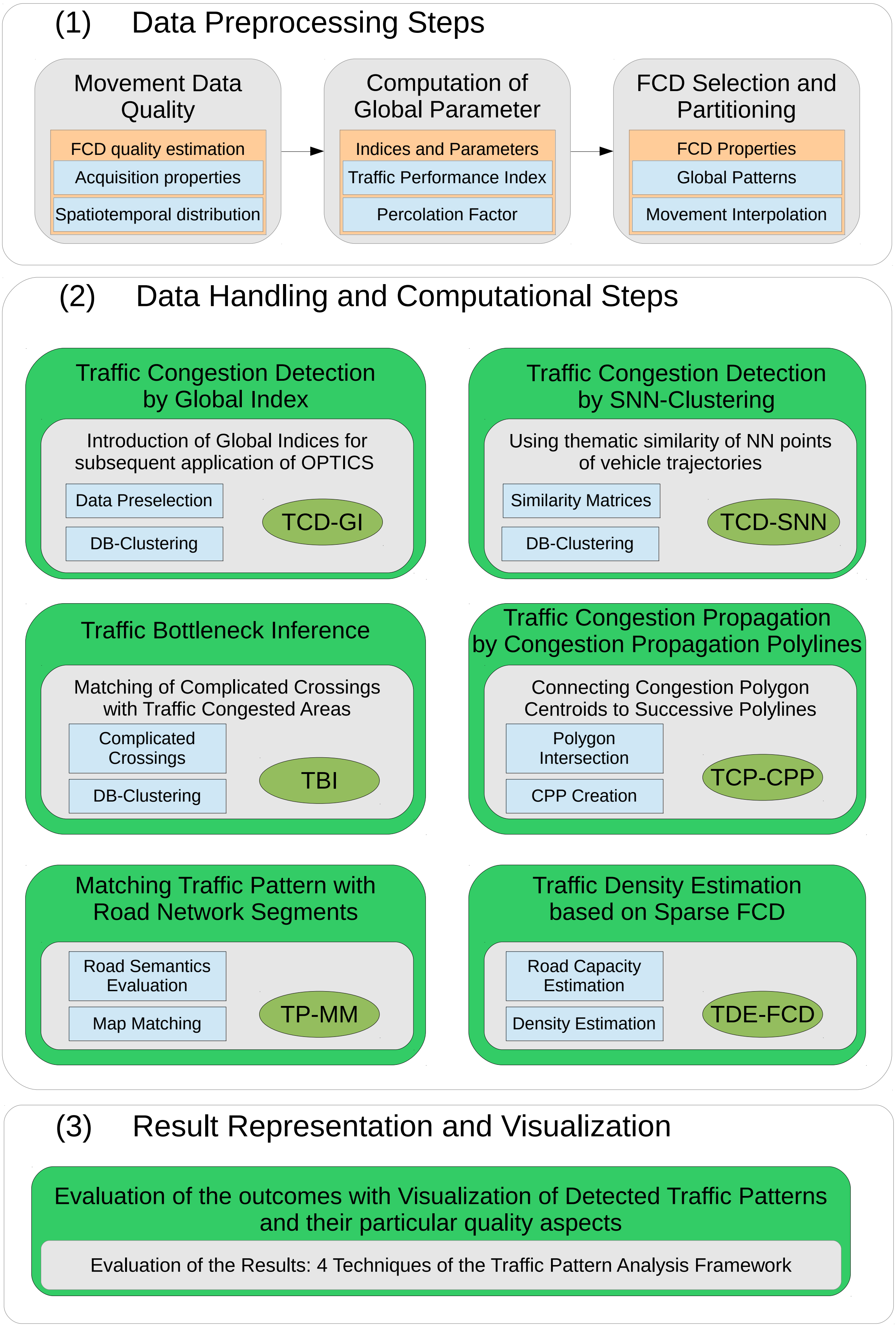


Figure 20: Components of the traffic pattern analysis framework for FCD.

4.2 Data preprocessing steps

Floating Car Data (FCD) sets need in many cases post-processing steps for complying with specific applications. For performing the six different methods of the framework, it is necessary to exclude eventual outliers from realistic records within the data. Consequently, an outlier has attribute values that are different from the larger part of the data.

Excluding unrealistic attribute values

Depending on the spatial distributions of the FCD sets, it is possible to introduce tolerance regions of the investigation area by querying the longitude and latitude values. The second query option is excluding unrealistic values of instantaneous velocity, as for example values over 200 km/h at rush hours. In numerous cases, errors in the temporal component appear in connection with falsified date time format. Before applying data partitioning, the framework proposes selecting and excluding FCD records as the first preprocessing step. The performing of the first query bases on information about the administrative city boundary of Shanghai. For the further analysis, selections consist only of those FCD records with longitude and latitude values inside the city boundary. Additionally, longitude and latitude values have to be within the borders of the investigation area and on neighboring areas with 150 km tolerance distances. These preselection values result from testing different FCD sets of different investigation areas. The second step of data selecting is querying unrealistic FCD record values. This includes selecting FCD records that have realistic instantaneous velocity values. This means all tracked vehicles within the investigation area imply the assumption to drive less than 150 km/h. After testing numerous not preprocessed FCD sets, these outliers, namely errors in attribute values, which result mainly from the data acquisition, imply usually a portion between 1% and up to more than 10% of the entire given dataset. Nevertheless, previous data inspection of individual data sets influences the individual construction of FCD preselection queries. Outliers in the data might not only include falsified attribute values, but as well loss of signal and imprecise positions. In urban environments, these outliers occur often at urban canyons, where GDOP values are relatively high. One additional third option for excluding unrealistic attribute values is coming from the research of Andrienko et al. (2016a), where the same vehicle identifications are used for individual vehicles. There are numerous options to exclude and to repair this appearance in the data. The simplest method for detecting duplicate IDs is the visualization of individual trajectories by polylines, since those polylines result in a characteristic zig-zag-pattern (Andrienko et al. 2016a).

Excluding FCD trajectory partitions with jumps in space and time

The exclusion of spatiotemporal outliers bases on detected jumps of FCD trajectories in space and time. Therefore, the framework proposes connecting FCD points with the same vehicle identification into polylines. The ascending time component of every tracked entity serves for connecting every consecutive point. The results of these constructions often depend of the inspected data partition, and if there is an application of a previous partitioning step. Each resulting trajectory record or polyline element gets additional attributes, which are created supplementary to the process of connecting points to polylines, in the form of < length, start_time, end_time, time_diff, x_start, y_start, x_end, y_end >. Important for the following steps are the trajectory attribute values length (of the trajectory) and time_diff, which is the time difference between starting and end point. Afterwards, it is possible to visualize the trajectory polylines by classified lengths. Unrealistically long trajectories are colored in red, as pictured in Figure 21a. This type of visualization facilitates subsequent exclusion of affected trajectories (Keler et al. 2017a). It is feasible to exclude whole trajectories or only selected trajectory partitions, before applying the data-partitioning step.

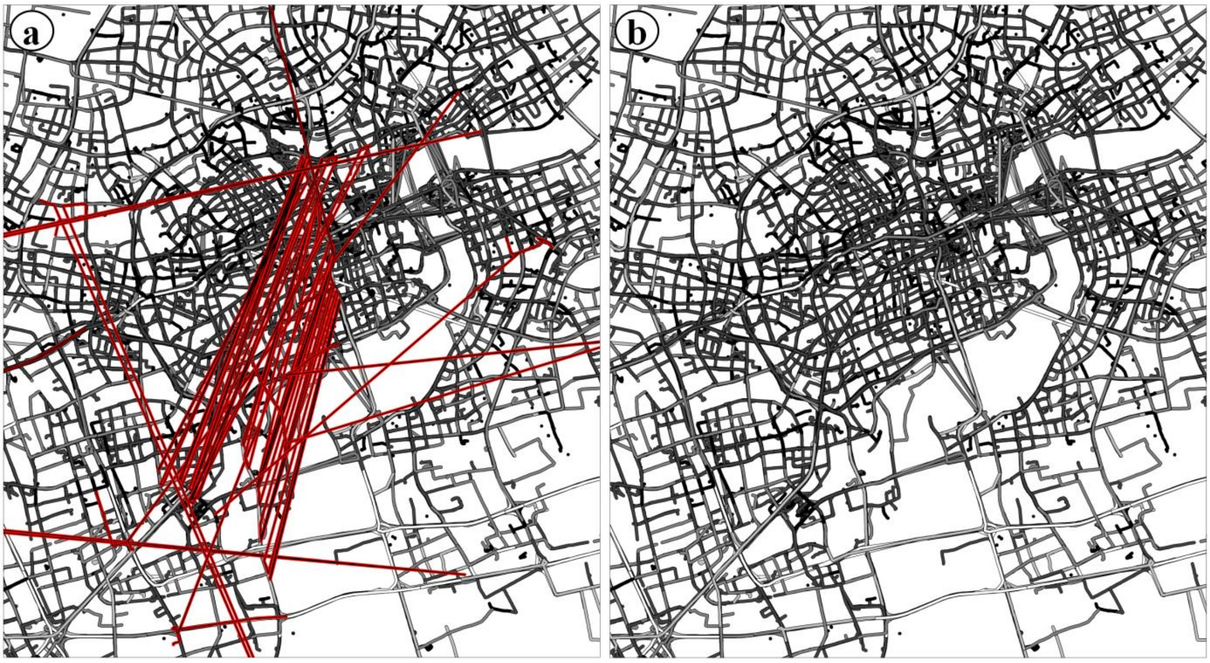


Figure 21: Comparison of the unfiltered (a) and filtered (b) trajectory velocity visualization.

Since numerous trajectories have jumps in space and time, especially in the inner city as pictured in Figure 21a, it is more reasonable to select the second option for exclusion. This means excluding only unrealistic trajectory partitions, as lines or polyline segments, is preferable due to avoiding information loss. Figure 21b shows the filtered result with excluded jumps in space and time.

Exclusion of vehicle trajectories based on sampling intervals

A sampling interval over one minute is not useful for selected FCD-based applications (Sohr et al. 2010). The distances of successive points of each individual trajectory together with the respective time difference are important for further processing steps. Therefore, it is possible to compute the sampling intervals between every successive movement positions of every trajectory, by the time difference of every successive time stamp. This helps estimating the usefulness of the data to perform different kinds of analyses. In case of the traffic pattern analysis framework, these computations are only useful for estimating the movement data quality. It is not an essential part of the framework, but can serve for further evaluation of the results. Especially, when comparing average velocities averaged by the arithmetic mean with instantaneous velocity values in every point and those averaged by the harmonic mean for every road segment. Computing velocities based on vectors from FCD implies more computational steps, as for example the projection of WGS84 coordinates to UTM coordinates. This transformation step allows the calculation of vector velocity values in km/h. Usually, there is a big difference when inspecting vector mean speed values and time-mean speed values for selected working days. Falsified vector mean speed values results from imprecise positions and successive distances between the points. A previous map matching (MM) step does often not benefit an improvement in the resulting values. By computing the time-mean speed, there is often a systematical underestimation of the real average velocities, since many zero values in the FCD records have a high influence.

Preprocessing and data smoothing

In the presented approach, various preprocessing techniques are used. In particular, movement data aggregation techniques are used for taxi movements in Shanghai on selected rush hours, which do not aggregate into moving object streams. Even if this step is not a typical preprocessing technique, it helps to get an overview on the data values via data smoothing. The subsequent step is to create clusters of taxi trajectory partitions that have similar velocity and driving direction values.

Road network information

One optional data type for the traffic pattern analysis framework is road network information. Independent of the source, the network has to represent road segment in an accurate way concerning lengths and shapes. In the OSM project, it is possible to reconstruct routing graphs in post processing. In some cases, the routable networks are inferable via automated approaches. This information is often not available throughout the database and there is no guarantee of realistic road segment connectivity. A more important issue is the data storage of historical road network information. In many cases, Asian cities have only one fixed state of road networks, without available data extracts from former years. By using map inference techniques, it is sometimes possible to reconstruct changes in the transportation infrastructure by simple polyline matching techniques. OpenStreetMap (OSM) was founded in 2004 by Steve Coast as a means to create an open geodatabase of the world by volunteered mappers. The common term in the literature for this type of data is Volunteered Geographic Information (VGI). According to Stanica et al. (2013), the OSM project has one of the most accurate freely available digitized road networks. This makes even vehicle routing applications possible as it was tested by Graser et al. (2015). Nevertheless, there are sometimes remaining problems in computing reasonable routes, due to specific geodata issues such as missing connectivity of digitized road segments. Graser et al. (2015) evaluated the quality of the OSM road network in a case study for its suitability for vehicle routing.

Data partitioning

Ehmke et al. (2010) aggregate FTD into one hour partitions for calculating average velocities for whole road partitions. The same partitions are used for computing the global index of the framework, which is the first technique. The other techniques of the traffic pattern analysis framework use smaller time windows, since the aim is to detect the propagations of traffic congestion in a more detailed way. Therefore, 10-minute partitions serve for FTD rush hour selections.

Approaches within the traffic pattern analysis framework do not bindingly include the detection of phantom traffic jams. The temporal resolution in this approach is 10 minutes and consequently this indicates that phantom jams might be difficult to detect. The reason for this statement results from the usual short durations of phantom jams. Nevertheless, such events are partially detectable via extended clustering techniques for the case of numerous associated taxi drivers with similar driving directions within the phantom jam. Nevertheless, in many cases, these events might appear similar to stopping at traffic lights. One idea of excluding this kind of misclassification is using additional information coming from volunteered geographic information (VGI), as from the OSM project. This is possible, since traffic congestion events are definable via underlying spatial and temporal extents.

4.3 Data handling and computational steps of the framework

Based on the previous findings, this subsection proposes simple and reproducible methods for detecting vehicle traffic congestion based on Floating Car Data (FCD), in particular Floating Taxi Data (FTD). The central framework component of data handling and computational steps consists of six different techniques that are applicable consecutive or intermittent on preprocessed FCD sets.

4.3.1 Traffic congestion detection by global indices

The main aim, of applying traffic congestion detection by global indices, is selecting those data partitions that represent the rush hours of traffic within specific investigation areas. In a practical sense, it is the selection of rush hours and its subsequent extraction as resulting FCD partitions. The motivation behind this method originates from the congestion indices of previous studies in the literature. The preselection step implies a practical simplification of the data inspection task of FTD sets, since only those partitions are matter of further analysis, where recurrent traffic congestion appears most likely. The usage of global traffic indices aim finding out the highest or most influential times of traffic congestion in Shanghai. In general, the computation of global indices has no strict requirements in the framework, and can be a part of the previous data preprocessing steps. Figure 20 shows this optional setting as the middle part of the data preprocessing steps of the framework. Besides computing the traffic performance, also inspecting traffic percolation is possible and is included as an optional component. Another question when representing traffic patterns is the choice of the data partition. Ehmke et al. (2010) aggregate FTD into one hour partitions for calculating average velocities for individual road segments. These semantically enriched road segments include 24 states per individual day: as average velocity and as travel time, since there is the possibility to infer knowledge of road lengths and numbers. Instead of this approach, subsequent k-means clustering is not part of the present procedure. For this method, only the one-hour FCD partitions are used for extracting hour-wise average velocities. In particular, averaging consists of computing the time-mean speed via arithmetic mean of instantaneous velocities for the whole investigation area. For Ehmke et al. (2010), one-hour partitions are small enough for detecting typical traffic states such as all the daily rush hours. Therefore, plotting average velocities together with the number of records is usable for revealing the rush hours of inspected days.

The global traffic index values I_{hour}^{cong} for each hour of the day in this work consist of temporal distributions of the number of records $n_{records}$, the number of records with zero instantaneous velocity $n_{zero}^{\bar{v}}$, and the time-mean speeds $\bar{v}$. All these measures are input information for the following formula for computing the hour-wise global traffic index values I_{hour}^{cong}:

$$I_{hour}^{cong} = \cos\left(\frac{n_{zero}^{\bar{v}}}{n_{records}} * \bar{v}\right) \tag{18}$$

The global traffic index values base on the cosine of a ratio between zero velocity records and all records, multiplied by the time-mean speed. This formula (18) results for previous testing of different FCD sets. The variation of values shows many similarities with the changing traffic situations of selected investigation areas. Additionally, there are options to extend the presented formula. Optional properties of data partitions are applicable for restricting global traffic index values in time and space. These optional properties might imply the size of inspected area, temporal sampling intervals, possible acquisition windows or the absolute number of points for a certain vehicle. For the framework, the computations consist only of the number of points and the average velocities for each hour. This has the reason of implying only measured values into average velocities, without previous estimation of extended global indices. The number of points is the representation of reliability, in the sense that the number of records should not be significantly lower than in other time windows.

4.3.2 Traffic congestion detection by shared nearest neighbor (SNN) clustering

The second technique of the framework consists mainly out of the shared nearest neighbor (SNN) method, which was, in its underlying form, introduced by Ertöz et al. (2003). Using the shared nearest neighbor (SNN) algorithm for detecting traffic congestion originates from the approach of Gao et al. (2016). The idea of Gao et al. (2016) is to cluster the spatiotemporal heterogeneity of instantaneous velocity values. In a similar way the similarity of instantaneous velocity is measured, with the expectation that traffic congestion results in similar low values at specific location for selected times of the day. The 10-minute FTD partitions that result from the previous data partitioning step are data input for the shared nearest neighbor (SNN) algorithm, which was, in this form, defined by Ertöz et al. (2003). From multiple available attributes, only instantaneous velocity and driving direction are useful for computing the similarity matrix in this approach. Those two attributes are selected for defining similarity with the aim to be able to distinguish between different road lanes and elevation levels. The following step is to sparsify the matrix with k = 10 for only keeping the 10 most similar neighbors in every FTD point. A SNN graph is then constructed by using the Jarvis-Patrick algorithm from Jarvis and Patrick (1973).

The second part of SNN algorithm is identical to DBSCAN (Ertöz et al. 2003). Therefore, the first part of SNN serves more as a preprocessing step for excluding points with less similar velocity and driving direction values for further cluster generation. The second part of the SNN approach is applying the popular variant of density-based clustering DBSCAN introduced by Ester et al. (1996). The input parameters for creating density-based clusters are the minimum number of points *minPts* for calculating density and a search distance measure *epsilon*, which might be set in Euclidean space. For the following testing *MinPts* has the value of 3 points and the search distance *epsilon* is 50 meters in Euclidean space. After creating SNN clusters for 10 minute FTD partitions the Gift-Wrapping algorithm by Jarvis (1973) is used to create convex hulls. The method of applying the SNN algorithm on FTD and its subsequent conversion into polygons is pictured in Figure 22.

This approach (TCD-SNN) includes SNN with the similarity values of instantaneous speed and driving direction. The addition of the driving direction aims to benefit the differentiation of clusters with similar velocity values on same or adjacent road lanes with the same driving direction. In general, it should serve as an indicator for similar road lanes. Another idea of this combined similarity value is to distinguish between elevated highway segments and roads below these segments. This differentiation is even more difficult since similar velocities might appear on both elevated segment with similar driving directions. The idea behind instantaneous velocity as similarity measure results from detecting taxis with similar speed that are less than 50 meters away from each other within a time window of 10 minutes. These detected appearances may indicate traffic congestion events.

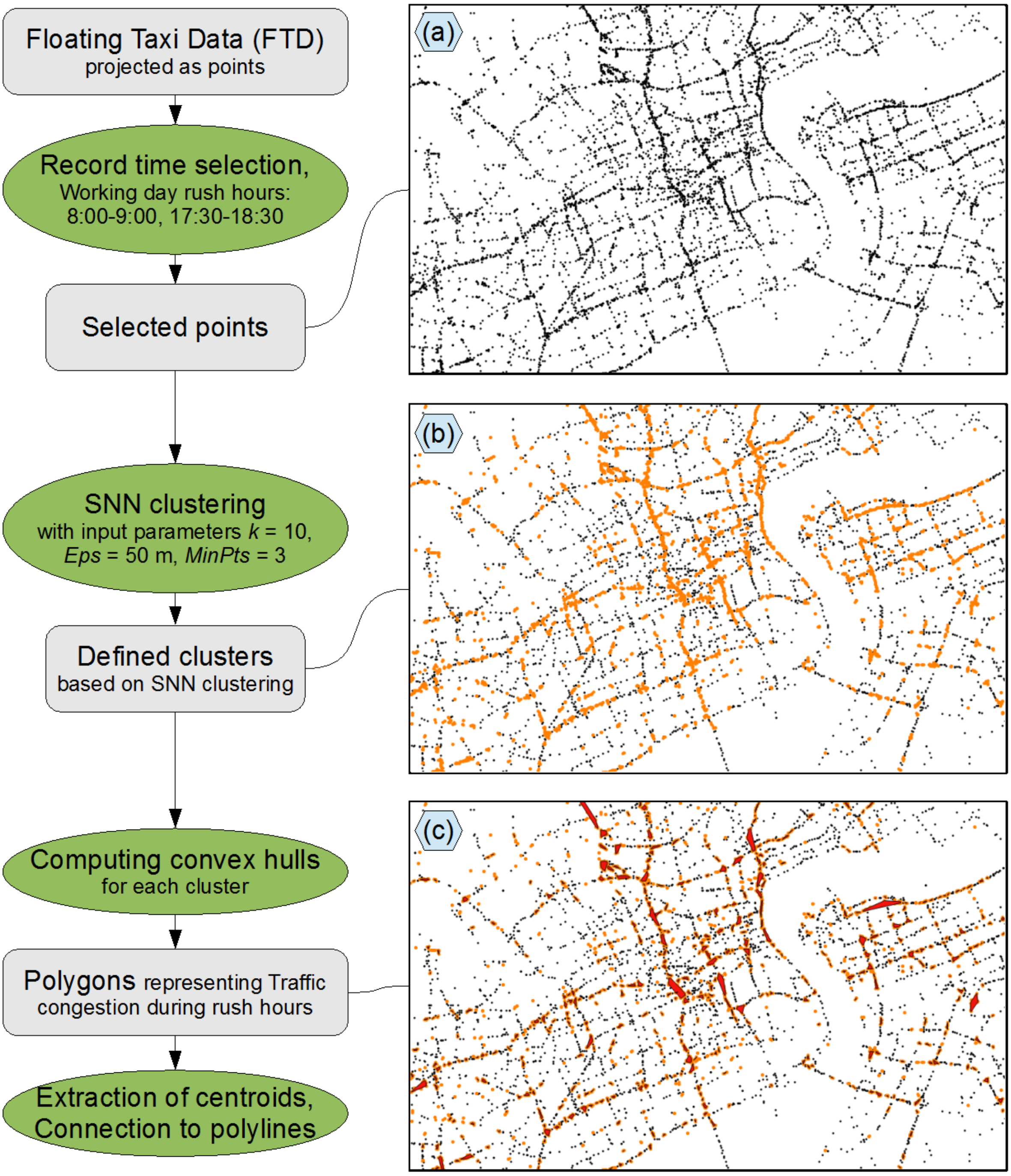


Figure 22: Creating congestion polygons with (a) selecting taxi movement points of rush hours, (b) point clustering based on similarity by SNN and (c) creating congestion polygons.

Figure 22 shows that only quite few SNN clusters are detectable. One reason for this appearance is the way of constructing the similarity matrix with k = 10. When focusing on the similarity of two different attribute values, it is assumable that only few selected points show similarity in both attributes. After focusing on rush hours, the second step of the SNN techniques implies defining SNN clusters, which is challenging. Creating similarity matrices of instantaneous velocities serve as preselection for subsequent similarity estimations of driving directions. This procedure shows more reasonable results than the one of creating single similarity matrices from both attributes. Subsequently, resulting clusters appear small and often narrow in their spatial distribution. Additionally, there is a need for manual inspections, since directional similarities require testing due to numerous outliers in the direction attribute values within the test data sets. Other specific types of outcomes of the SNN technique are similarities in higher velocities at selected parts of the highways, which also comply with similarity in driving direction. These polygons are extractable via querying the SNN polygon outcomes. Therefore, several outcomes of the SNN-technique are questionable, even though many similarity clusters appear reasonable.

4.3.3 Traffic bottleneck inference

Literature distinguishes between static and dynamic traffic bottlenecks. This differentiation is dependent on the way of how traffic congestion interacts with the underlying transportation infrastructure. Therefore, (Keler et al. 2017c) introduce a novel and yet simple technique to infer traffic bottlenecks in urban environments. The traffic bottleneck inference bases on a study on inferring complicated crossings and distinguishing between congestion events and traffic bottlenecks. This technique consists of handling two data sources with two methods. The approach begins with computing with two input data sets, more detailed FCD and road network information, and results in the detected traffic bottleneck locations within the investigation area. Figure 23 pictures the workflow of this approach mentioned as TBI (traffic bottleneck inference) throughout this work. One aim of the TBI technique is to simplify the representation of traffic bottlenecks via area polygons that overlay the affected road segments. By creating traffic bottleneck polygons for different time windows of different days and weeks, it is feasible to reason on distinguishing between static and dynamic traffic bottlenecks.

The blue colored background in Figure 23 summarizes the method of Krisp and Keler (2015), whereas the yellow background shows the elements of detecting FTD-based traffic congestion. The resulting product of Technique 3 (TBI) bases on the matching and intersection of the outcomes of the two mentioned methods.

The derivation of complicated crossings within the road network is very difficult, since no survey of locals’ opinions is available. Instead, Volunteered Geographic Information (VGI) of the vehicle road network from the OSM project is matter of data extraction. The extracts serve for applying node extractions and subsequent polygon creation and extraction for classifying complicated crossings in Shanghai. The detection and classification of complicated crossings is based on the approach proposed by Krisp and Keler (2015), where complicated crossings in Munich were defined for inexperienced drivers. After numerous evaluations, the results for Munich, mainly the inferred polygons, appear reasonable as they often express the perceived complicated crossings in reality. Another case study on the road network in New York City (NYC) shows less reasonable results, mainly due to numerous misclassifications of park areas as complicated crossings. This shows also the disadvantages of the method (Keler and Krisp 2016a). As a conclusion, it is to note that node density of road segments alone is not enough to estimate the complexity of road partitions.

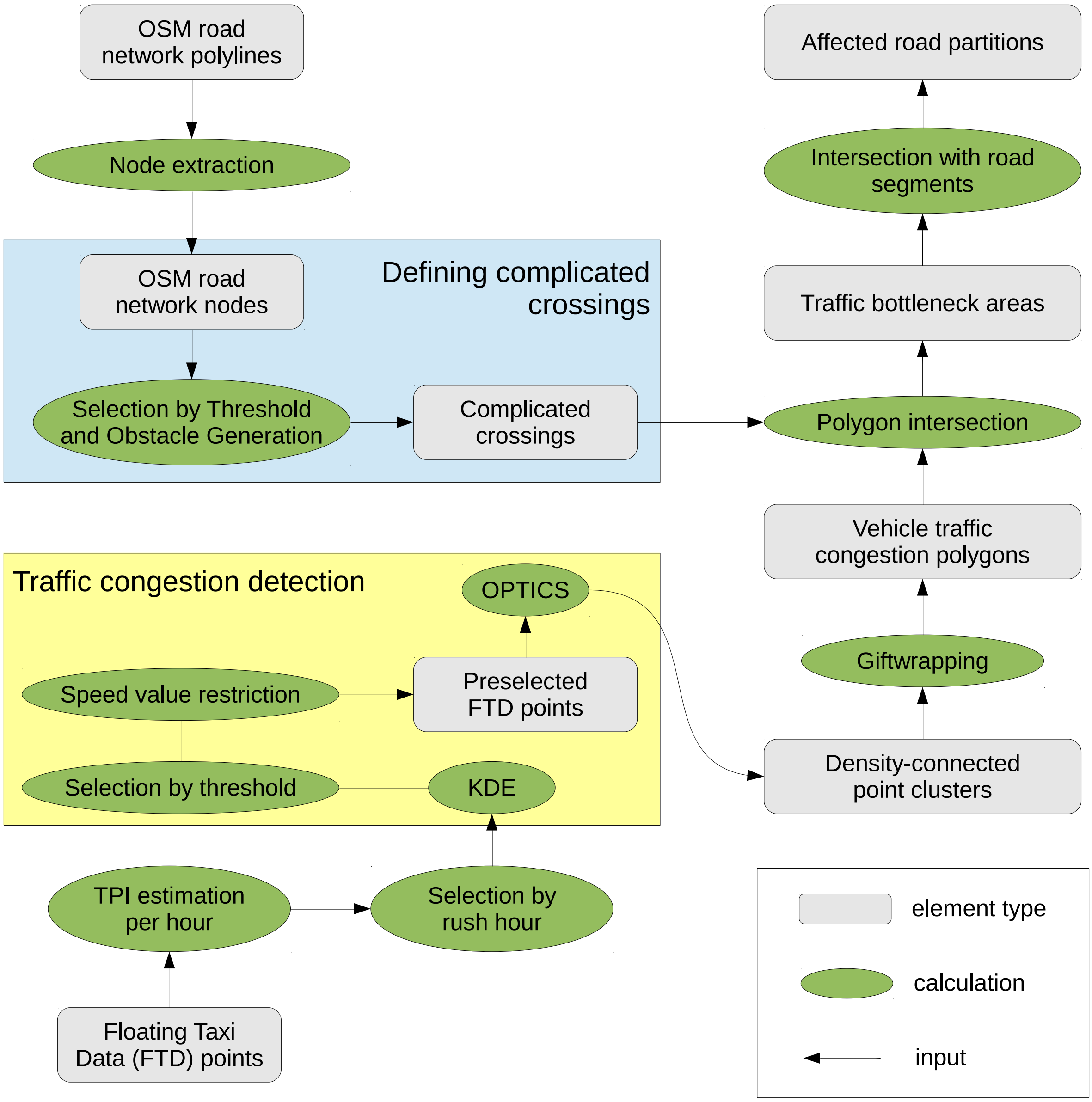


Figure 23: Workflow of the approach ("TBI") for defining complicated crossings (blue), detecting traffic congestion (beige) and defining traffic bottleneck areas.

The aim within the traffic bottleneck identification method is to test the method of Krisp and Keler (2015) for the road network data of Shanghai. As provided in Krisp and Keler (2015), the average distance of roundabouts (in this case 60 m) can serve as a selected threshold for the creation of obstacle polygons. These obstacle polygons are detected complicated crossings and base on two-dimensional Kernel Density Estimation of extracted street nodes. After inspecting the transportation infrastructure in Shanghai, the search radius remains 60 meters. In case of the road network of Shanghai, which consists of 83,797 polylines, has after the node extraction 777,377 nodes. These nodes are input for detecting complicated crossings based on the method by Krisp and Keler (2015). The mentioned number of nodes transform into 3,080 complicated crossings in Shanghai.

The idea of FTD-driven bottleneck identification bases on preselecting rush hours and lower vehicle speeds on the one hand and on the other hand density-based clustering. The results of density-based clustering are density-connected point clusters that represent areas of congestion influence on selected road segments. This indicates that there is no performance of a previous matching of traffic information (FCD positions) on road segments. One claim is that road segments are not spatially accurate enough to represent traffic congestion. Therefore, this approach is classifiable as an alternative method for detecting and representation of vehicle traffic congestion.

The first step of computing traffic congestion locations and bottlenecks of the road network is the introduction of a global congestion value for the whole network of one day of FTD. Keler et al. (2016) define a congestion value c, with a subsequent classification into 10 classes of equal range:

$$c = \frac{k}{v} \quad \text{with vehicle density k} \left[\frac{vehicles}{km}\right] \text{ and average velocity v} \left[\frac{km}{h}\right] \tag{15}$$

By using this index, which is influenced by the traffic performance index (TPI) as explained in Wen et al. (2014), it is possible to detect the rush hours of selected working days within selected investigation areas. Only data partitions of morning and afternoon rush hours on working days are selected. Afterwards, these selected partitions are matter of further segmentation into smaller subsets of 10-minute time windows.

The subsequent step consists of a Kernel Density Estimation (KDE) for FTD points. The used kernel for KDE is a quartic kernel. By applying KDE with quartic kernels, surfaces result for every 10-minute partition of selected rush hours. The following step is to extract kernel densities above a specified threshold. There is no indication for specific density thresholds, since its selection is data dependent. In this work, for example, a massive FCD set with relatively high sampling rates is used. The easiest way to obtain a suitable threshold is by visual data inspection. One can make use of instantaneous velocity value in every point together with observing possible visual clusters of the FCD points. Consequently, the step of density threshold selection comes after the preselection based on the global index, the subdivision into 10-minute steps, and its visual inspection within a GIS application. Applying spatial and visual analysis steps with GIS software can help the potential analyst to select optimal thresholds.

After the setting of density thresholds for surfaces (based on 10-minute partitions of FCD), only parts of the surfaces are selected. The respective FCD points inside these areas serve for further analyses. This is possible after introducing the thresholds for extracting dense FTD point concentrations. Since the results of KDEs are continuous fields, it is feasible to define polygons with characteristic rounded edges. Only the records inside these areas are matter of extraction. One additional restriction for extracting FTD points is the introduction of a threshold for the attribute instantaneous velocity. Here, 20 km/h is set as the maximum speed for an ongoing traffic congestion event. This is motivated by a previous study of Robinson (1984), who states that vehicle emissions of hydrocarbons and carbon monoxide are especially high in speeds under 20 km/h.

In a subsequent step, density-based clustering is applied: OPTICS algorithm, which was introduced by Ankerst et al. (1999). The input parameters for creating density-based clusters are the following: for *minPts* 3 points are used and the search distance *epsilon* is 50 meters in Euclidean space. Derived traffic congestion clusters are subsequently converted into convex hulls by using the gift wrapping algorithm (Jarvis 1973).

4.3.4 Traffic congestion propagation by congestion propagation polylines (CPP)

One approach of the framework (TCP-CPP) makes space-time connections between the respectively defined spatial FCD clusters of certain time windows in a similar way to Liu and Ban (2013). The extension of this idea is to match records with same taxi identification in different time windows and by incorporating a tolerance search distance, even if consecutive congestion clusters are not matching. The cluster results for selected consecutive time windows are connectable via its size, shape, and location, if applicable. This should provide a possibility to detect long duration congestion events. In the latter case, the idea is to reason on the possibility to describe the movement of the congestion event or the bottleneck itself. Nevertheless, only abstractions of whole events represent the traffic phenomena by vector elements. Therefore, the aim is to extract density core points from SNN-clusters.

Subsequently, these core points are connected for selected rush hours into polylines. These polylines are input information for the recurrent traffic congestion movements. By taking the underlying road network structure into account, these congestion movements are transformed into traffic bottleneck movements. When measuring polyline lengths of the computed phenomena, one needs to respect additional computation due to coordinate transformations, for expressing the lengths in meters or kilometers.

The aim is to detect traffic bottleneck movement by introducing congestion propagation polylines (CPP). After the generation of congestion cluster polygons for respectively six time windows of each daily rush hours (morning and evening), spatiotemporal connections of congestion clusters of consecutive time windows are tested. The first idea is to evaluate different methods for this connection coming partially from previous research. After successfully matching subsequently appearing congestion polygons based on the Euclidean distance, the second idea is to detect changes in topological relationships, similar to the approach of Salamat and Zahzah (2010). After the first tests, another option is selected, which might appear more reliable for detecting bottleneck movements: querying identical taxi identifications in matching subsequently appearing polygons.

For the detection of congestion propagation, the centroids of created clusters for every 10 minutes of the FTD sets are matched in Euclidean space. Depending on the results of this overlay or matching, the following possibilities, among others, might be expected:

1. One cluster polygon of one time window is intersecting another polygon of the subsequent time window (1:1)
2. One cluster polygon of one time window is intersecting many polygons of the subsequent time window (1:n)
3. Many cluster polygons of one time window are intersecting one polygon of the subsequent time window (n:1)
4. Clusters of subsequent time window are not intersecting, but feature has certain distances to other clusters.

By respecting these four possibilities, certain overlapping polygons might be associated as an ongoing traffic congestion event. Additionally, indications, as lengths and durations of CPPs can serve for classifying traffic congestion event, as for example if a formation of a traffic jam is still occurring with heavy intensity or if the event is dissolving. After detecting several polygon intersections, only the matched consecutive polygons with an additional search distance of 50 meters are selected. Another condition is to respect only the shortest distances between the matching polygon centroids. A preselection assures that there are matching polygons in all six time windows of each rush hour. A final selection pics only these polygons that include the same taxi identifications in all six time windows within connected polygons. Afterwards, a congestion propagation polyline is defined by connecting all centroids of matching polygons of the rush hours. In case of multiple taxi identification in the same polygons, it is feasible to introduce weightings for extending the created polylines. The more taxi identifications within matching consecutive polygons, the higher are the assigned weights for each connecting line.

4.3.5 Matching traffic phenomena on road networks

For the transformation into network space a so called naïve Map Matching Algorithm was used based on Bernstein and Kornhauser (1996) and White et al. (2000). In contrast to matching recorded positions of individual movement, this case intends to match congestion polygon centroids to their closest situated OSM road segments. In this case, point to line matching makes only sense, when the aim is on the association of traffic congestion propagation. When having connected centroids of traffic congestion polygons, it is possible to associate the traffic congestion propagation with the most influenced road segments. The congestion polyline is then defined as a specific traffic congestion event.

The matching of whole taxi trajectories within thaffic congestion events on underlying road segments is far more complex. The reason for this is first the complex and densely built transportation infrastructure in Shanghai, and, second the insufficient positioning accuracies of the used positioning devices. The possibly most reasonable approach is to match vehicle trajectory partitions within congestion propagation polylines, in particular the taxi movement inside matching polygons.

The matching approach consists of matching only clustered subsets of FCD with the underlying transportation infrastructure. The first matching of clusters with the road network originates from created SNN clusters. Here the idea of distinguishing between different elevation levels is helpful for matching with elevated and non-elevated road segments.

The condition, if selected clusters are associated with elevated and non-elevated road segments implies at least two preconditions. The first is the intersection pattern. The idea is to compare the lengths of associated road segments with the actually matched parts of the segments in space. The second precondition is to check connectivity of road segments. In many cases, as in parts of the OSM road network information, there is an indication of elevation levels via specific attribute values. In case of no available elevation assignment information, a connectivity check is required for the influenced road segments. Subsequently, it is possible to distinguish the different elevation levels, which might be a difficult task, since the road network in urban areas is very densely situated.

The connectivity check consists of spatial queries that give insights on how road segments of different types are connected. When highway segments match with minor roads at or nearby situated intersections, it is in general an indication for the existence of on-ramp segments. In case, there are connections of segments without on-ramps, there should be a guarantee of implying a reasonable connection. In most cases, these are connections with segments of similar road type. The connectivity check and the respective query outputs can serve for finding out more about complex connections of the urban transportation infrastructure.

Since the input information is mostly FCD from observed taxis, there is a need to respect certain quality aspects. This includes especially the varying positioning qualities in urban areas (urban canyons), where only lower GDOP values or less than four GNSS satellites are visible.

4.3.6 Traffic density estimation based on sparse FCD

One difficult task in using FCD for traffic detection and forecast is the computation of vehicle or traffic densities for selected parts of the road network. Traffic density is one of the three basic quantities of the traffic flow theory. Due to the fact that it is still not possible to acquire the movement of all participating vehicles (but maybe in the near future), the real traffic densities can only be estimated. Gühnemann et al. (2004) propose the estimation of the traffic flow rate by including average velocity profiles with fundamental diagrams. Fundamental diagrams might have various appearances, since individual fundamental diagrams have to be defined for street segments with different road types, speed limits and influences by signals and traffic lights (Gühnemann et al. 2004). By applying the fundamental traffic flow relations, it should be noted that the first step is to infer a fundamental diagram between traffic flow rate q and space mean speed u with a proportional travel time (Gühnemann et al. 2004). The second step is then to provide a function between speed and resulting density by inversion of the previous function between speed and traffic flow rate. Gühnemann et al. (2004) state that this is often connected with over- and underestimation of traffic densities, especially for low-density roads with usually constant speeds.

The idea of deriving traffic volume from FCD is already described by Shang et al. (2014). In this approach, sparse FTD is first map matched to road links for calculating average travel speeds for one-minute time windows. The road links that are not intersecting taxi trajectories are then enriched with estimated travel speeds by using the so called Travel Speed Estimation model (Weng et al. 2016). Since the idea of Shang et al. (2014) is to estimate emissions and fuel consumption of all road partitions in the investigation area, the traffic volume is needed. Shang et al. (2014) use an unsupervised Bayesian Network for estimating traffic volume per minute for each road segment. The input information consists of multiple factors: travel speed, weather conditions, and geographical features of a road.

The proposed method does not follow a probabilistic approach, but uses more the configurational information of the network and their segments. The main idea consists of saving average velocity differences at every road segment with given or estimated speed limits. Differences indicate in general variations in traffic densities.

Based on averaged measure, as the average velocities, it is possible to estimate typical or usual traffic density values for every road. Only if selected instantaneous velocities are averaged to non-typical speed values, non-typical or unusual road densities are calculated. Unusual road densities are not incorporating recurrent traffic congestion. This is only provided by detected traffic congestion, otherwise traffic density is calculated via attribute information of every specific segment. The number of lanes given as an attribute increases the estimated capacity by multiplication with number of lanes. Additionally, the attribute width, together with the estimation of average lane widths and total widths of road segments are introduced, as pictured in Figure 24.

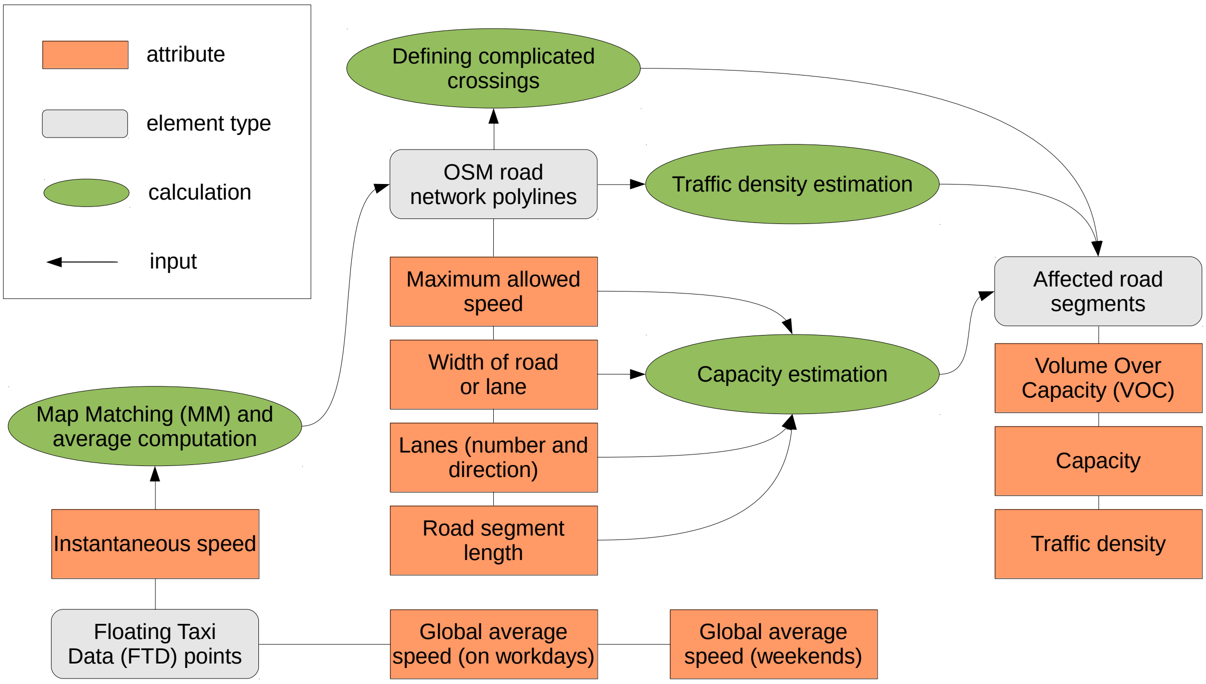


Figure 24: Workflow for traffic density estimation based on sparse FCD.

The proposed method in Figure 24 consists of three different computations. All three computational steps base only on those road network, which are influenced by detected congestion events. The workflow of this method starts with enriching FTD points with global average speeds of workdays or weekends respectively, resulting from data partitions. Together with the instantaneous speed values in every point, global average speed is assigned to road segments via naïve MM as in White et al. (2000). The enriched OSM road segments are polylines with several additional attributes: “, “, “, “. These are used to estimate the road capacity of each road segment. Additionally, the traffic density is estimated based on interpolating vehicle numbers, which base on previously matched FCD and road type. By combining traffic density and road capacity, road segments are extractable, which are affected by exceeding the respective road capacity. As an addition, the method by (krisp and keler, 2015) is included for providing a reliability check for estimating the node densities of the road network. As an outcome, volume over capacity (VOC), capacity, traffic density are the computed attributes of each road segment. All the basic steps of this traffic density estimation technique are tested with a FTD set from Shanghai. This step facilitates also the selection of input parameters of the previously described SNN method, mainly in that sense that a reasoning step might be included. Another aim is to reason on the idea of integrating real time data (e.g. from taxi companies) into the traffic pattern analysis framework. The questions here are how to modify the clustering method and if historical traffic information (like historical FCD) can improve efficiency of clustering near real time data. Similar to Shang et al. (2014), a method for estimating the traffic volume is introduced. The focus of estimating traffic volume is not only providing traffic information on road links, but as well on estimating the traffic volumes in congestion polygons.

Additionally, the macroscopic fundamental diagram can support the derivation of traffic densities in the urban road network. The relation between average velocity and traffic density can result in estimating the traffic flow rate. This technique includes, as most other FCD-based traffic density estimation methods, a high degree of prediction ambiguity. One addition for increasing precision of the method is to include the previously mentioned forms of traffic congestion representations. In its simplest case, this would be spatiotemporal matching of detected cluster polygons and propagation polyline with the corresponding road segments. This is the reason, why this technique is more a tool for evaluating the results of the other methods of the framework.

An optional method for comparing similarity of detected CPP is using fréchet distances. This method consists of comparing different polylines by its similarity in shape and is also used in Brakatsoulas et al. (2005). After estimating traffic densities, it is possible to connect and to relate them with selected vehicle trajectories. Another option is observe traffic density estimation with the computed congestion propagation polylines. This is a complex task, since density core points are often (spatially) outside the road segments. The movements of these core points are representable via CPP in two-dimensional Euclidean space.

4.4 Result representations of the framework outcomes

There are numerous visualization methods for large FCD. In particular, the initial inspection of massive FCD, before the data preprocessing, includes often data visualization, which might help to discover new insights into various traffic patterns. Aiming to have an insight into spatial and temporal properties of massive FCD, is the base of selected geovisualization techniques. Visualization is used in parts of the traffic pattern analysis framework. This aims to provide more understanding of traffic phenomena like traffic congestion, especially in connection with its change in shape and size at different time windows.

Another focus of result visualization is the visualization of differently acquired instantaneous attribute of FCD records. This is useful, when reasoning on density-connected point clusters that result from shared nearest neighbor similarity matrices. In particular, it is possible to observe colorized instantaneous velocities with averaged velocities of similarity cluster polygons. In a similar sense, evaluating similar driving directions can imply coloration based on radiant or degree.

The visual analysis process itself is often dynamic and implies testing of different visualization techniques in separate windows. In a similar way, this is part of the traffic pattern analysis framework: results of selected methods are intermediate results and in the same way input for following approaches. This idea goes into the basics of visual analytics with the projection of geospace. Another motive of the interactive visual analysis is the idea of exploratory data analysis (EDA). In addition to this fact, there are established visual analytics methods for movement data (Andrienko et al. 2013).

Starting from both ideas, it will be discovered, if analysis on FCD benefits from personal inspection (interactivity) of individual mobility and travel times. The idea behind this visualization design is not to inform daily commuters but to give deep insights into traffic congestion dynamics for planners. The most reasonable visualization for daily commuters might be the simple coloration of road segments by using the stoplight metaphor within an interactive 2-D web map.

This approach focuses on the information from historical massive FCD, which connectable to road segments and additionally is observable in the Euclidean space.

Visual representation of traffic congestion and traffic bottlenecks

The detected traffic congestion polygons and complicated crossings are the inputs for detecting traffic bottlenecks. In this work, vehicle traffic bottlenecks are defined as those locations that are complicated crossing and that are affected by traffic congestion. When starting from FCD positions, every point is clustered by a density-based clustering technique.

The cluster results are pictured in Figure 25a, where each cluster has a different point density, colorized from lower densities in green to higher densities in red. Afterwards convex hulls representations are established for represent every point cluster in a visually more, which is picture in Figure 25b.

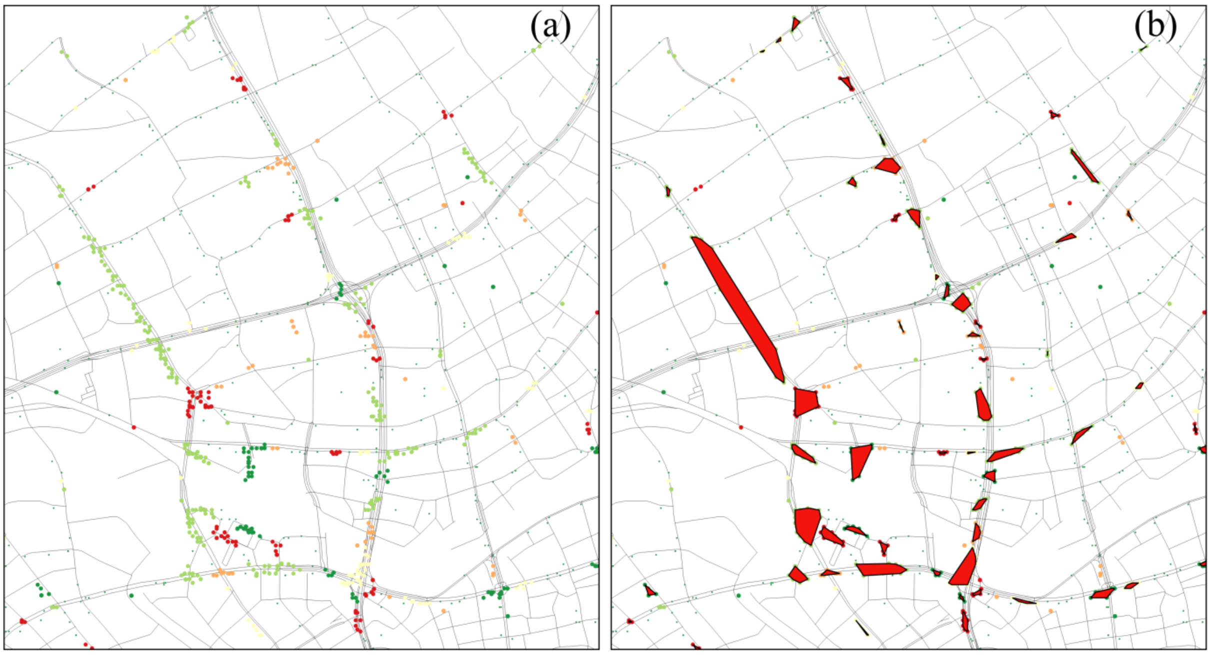


Figure 25: Comparison of clustered FCD points (a) and, resulting convex hulls (b), representing traffic congestion events.

The convex hulls in Figure 25b represent the detected traffic congestion events. Dependent on the distributions of detected congestion events, only portions of the events are connectable with complicated crossings. Since there are individual results of traffic bottlenecks for each time window, one further step might consist of a visual representation of its spatiotemporal variations. For avoiding visual overload, the 2-D polygonal representations imply 50 % opacity of the colorations. The second option is the 2.5D-view with extruded bottleneck polygons. This extrusion bases on FTD point density of selected 10-minute partitions within each polygon.

Visual representation of traffic congestion propagation

Besides animation, it is possible to show selected traffic congestion propagation patterns in static map views. The question here is how to include the time component, e.g., together with some other attribute value for visualizing density. One part of the traffic pattern analysis framework consists of detecting visited places by inspected moving vehicles. The detected places are following connected with the analyzed trajectories. Therefore, techniques are applied for semantic trajectory generation for annotating raw vehicle trajectories with spatiotemporal information.

The bigger aim of this step of semantic enrichment is to provide a connection between static infrastructural properties and dynamic traffic parameters. This connection can correlate for example traffic congestion with quality values of road networks. The practical steps consist of investigating the intersections of polygons with network topology. This procedure is sequentially, since the polygons are stored as a series of timestamped features. A series of timestamped feature is a time series that allows enriching the road network with averaged traffic information, which has a graph representation.

The second step introduces abstractions of parts of the polygon time series, which are polylines with timestamped nodes, namely congestion propagation polylines (CPP). These CPP are representations of the spatial positions of the highest vehicle densities or the lowest speeds within detected congestion events. There are as well at least two different possibilities to connect points to polylines: connection of core points and connection of recorded taxi positions. Figure 26a shows a polyline of spatially connected congestion events for successive time windows. This is feasible by connecting the centroids of polygons that are intersecting each other. The other option for congestion propagation polylines is pictured in Figure 26b, where individual trajectory partitions inside matching congestion polygons are connected to polylines. This shows the tracked movement of individual drivers influenced by traffic congestion.

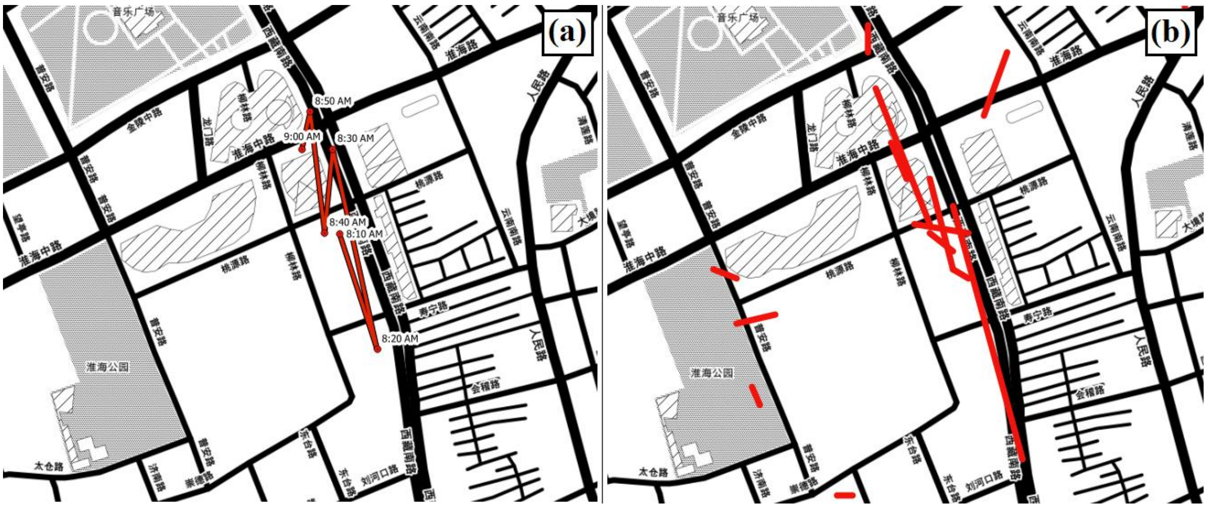


Figure 26: Comparison of the congestion propagation polyline (CPP) of moving polygon centroids (a) and, of individual taxi positions (b).

Figure 26a shows the possibility of additional labeling on parts of the propagation polylines. This labeling is possible for two possibilities of propagation polylines in Figure 26. Another possibility is to show every consecutive line segment or core point by varying color that for example show the starting of a traffic congestion by green and the end by red. The bigger aim of working with CPP is, besides measuring the similarities between CPP for the same time windows, but different working days of the weak, to match CPP with correspondent road segments. This allows connecting a very general abstraction of congestion propagation with real world road segments that are actually affected by traffic congestion events.

The idea behind the last step is also to connect traffic congestion with the topology of the used space and to reason on the interaction between connected road segments. An important term and measure of the last mentioned analysis step is phase transition.

5. Testing the traffic pattern analysis framework for traffic congestion in urban environments

After discussing the detailed elements of the traffic pattern analysis framework for FCD in the previous chapter, the flexible nature of every proposed technique is observable. This indicates the possible adaptation of different data types and sets. The six methods of the framework can adapt different movement trajectories by previous processing steps. In particular, the traffic pattern analysis framework can find application for various traffic data of different formats. As proposed by the flowchart in Figure 20, the framework relies strongly on previously calculated global traffic indices, which classify the rush hours of the data and influences data partitioning (number of time windows). This factum shows that the subsequent detection and classification of traffic patterns is strongly dependent on the underlying FCD and the respectively selected investigation area. In the following, the traffic pattern analysis framework is tested with real data.

The resultant insights are subject of evaluation. This evaluation serves for finding answers for the research questions in this thesis. Subsequently, each research question receives an answer based on empirical results from this chapter in the following chapter 6. After presenting the properties of the test data sets and the investigation area, each method of the framework has a testing step applied by using the FCD set from Shanghai. The results of the specific methods of the, previously presented, framework include time series of polygons and subsequently created polylines. By using the two products, the last two methods of the framework connect inferred traffic information with road network information. As an illustration, the most representative results of selected partitions of the entire data set are matter of discussion in the following sections.

5.1 Validation of vehicle traffic congestion measures with FCD and OSM of Shanghai

In the following, the general design specifications of the framework and the computational steps for inferring traffic information are tested with specific FCD sets and road network data. The investigation area for the case study is the city of Shanghai (PRC). Shanghai has an average population density of 3,631 people per km² and an absolute number of 23 million inhabitants (year 2000), thereof 30 % are temporary inhabitants. Shanghai has a highly complex transport infrastructure with elevated roads, tunnels and inner-city highways (Qian et al. 2006). This high complexity results from excessive governmental investments to improve the previous severe traffic problems. For Qian et al. (2006), the traffic condition have improved in Shanghai since the early 1990s after building additional roads, widening of older roads and including elevated segments. Nevertheless, the improvement has decreased by a high increase in private vehicle owners in Shanghai (Qian et al. 2006). This causes traffic flow rates that are not comparable to the early 1990s, since they are immensely higher. Shanghai has nowadays (2017) one of the best-managed taxi services in China[19]. There are around four different bigger taxi companies, which differentiate themselves by having typical vehicle colorations. As the used FTD in this case study has anonymous taxi identifications, it is difficult to guess the source of data, namely the taxi company from where the taxi vehicles are coming. The taxi service in Shanghai has fixed pricing that has not changed since 2007. Table 6 shows the system of taxi fares in Shanghai.

---

[19] http://www.meet-in-shanghai.net/taxi.php

Table 6: Shanghai taxi fare as of the year 2007[20].

Shanghai Taxi Fare	0 to 3 km	3 to 10 km	Above 10 km
Day time: 5:00 to 23:00	RMB 14	RMB 2.4 /km	RMB 3.6 /km
Night time: 23:00 to 5:00	RMB 18	RMB 3.1 /km	RMB 4.7 /km
Waiting	5 minutes waiting equals to one kilometer in charging		

In Table 6, three classes of trips are defined: from 0 to 3 km, from 3 to 10 km, and above 10 km. Dependent on these trip lengths and the time of the day the taxi fares in Shanghai are consistently calculated. Additionally, waiting costs for five minutes are the same as for one km of trip length. This is important for the case of still standing within traffic congestion events. There are as well some facts of local knowledge about the taxi service in Shanghai. It is known that most taxi driver have expertise in negotiating busy traffic in Shanghai. In the rush hours, it is difficult to hail taxis and on rainy days almost impossible[21]. Another fact is the ban for taxi drivers to pick up passengers within the radius of 30 meters from an intersection. This might result in specific distribution patterns of pick up points.

5.1.1 Tracked taxis in Shanghai from 2007 – the SUVnet-Trace Data

The central data set for testing the traffic pattern analysis framework is the SUVnet-Trace Data[22]. This FTD set was obtained from the Wireless and Sensor networks Lab (WnSN) at Shanghai Jiao Tong University. It represents records of thousands of tracked taxis in Shanghai in the year 2007. Specifically, it includes FTD from an acquisition between 31st of January 2007 until the 1st of March 2007. Every daily partition includes around 12 to 13 million records of four to seven thousands frequently observed vehicles. Within this month, there are around 10,000 different taxi identifications, which is detectable via inspection of individual taxi ID for every hour of the day. Depending on the time of the day, there are more or less taxi drivers with operating tracking devices. Table 7 shows the data structure of the SUVnet-Trace Data via the columns name, type, length and description for every attribute.

Table 7: Data format of the SUVnet-Trace Data from February 2007.

Attribute name	Type	Length	Description
ID	NUMBER	10	Primary key
TAXIID	NUMBER	7	ID of taxi
LONGITUDE	NUMBER	9	Longitude (6)
LATITUDE	NUMBER	8	Latitude (6)
SPEED	NUMBER	3	Speed
ANGLE	NUMBER	3	Angle
DATETIME	TIMESTAPMP	6	Time that GPS record was sent
STATUS	NUMBER	1	1: taxi is occupied, 0: taxi is vacant
EXTENDSTATUS	NUMBER	1	unknown (not documented)
REVERSED	NUMBER	1	1: reserved taxi, 0: taxi without reservation

20 http://www.meet-in-shanghai.net/taxi.php
21 http://www.meet-in-shanghai.net/travel-tips/taxi
22 http://wirelesslab.sjtu.edu.cn/taxi_trace_data.html.

Most of the inspected original 10 attributes find no usage in the study. One preprocessing step is the preselection of certain attributes of the data. These selected attributes for further analyses are the car ID, longitude, latitude, time, and instantaneous velocity. By previous inspection of the SUVnet-Trace Data, it is detectable that the sampling interval in time is differing. This means the jumps in time between consecutive points of the same vehicle are not constant. These time jumps vary between 1 second and 30 seconds and have an average of around 12 seconds for each inspected hour of the data set. In some cases, the sampling interval becomes extremely high, due to specific elements of the transportation infrastructure as GNSS signal losses at tunnels detectable via jumps of successive taxi positions. Several of these jumps also occur at specific urban canyons in the city center, which causes missing taxi trajectory partitions of numerous taxi trips across the densely built environment.

Previous findings from analyzing the SUVnet-Trace Data

This taxi probe data is inspected by Keler and Krisp (2016b) in the way of using time window selections. Previous work reveals interesting findings that indicate specific mobility patterns in Shanghai. Keler and Krisp (2016b) use time-stamped and interconnected area selection for observing travel time differences within working days. On selected rush hours, it is possible to show that calculated travel times, from the SUVnet-Trace Data and via shortest path routing in Google Maps, in 13 prominent crossings in Figure 27 are comparable, even after 10 years of difference.

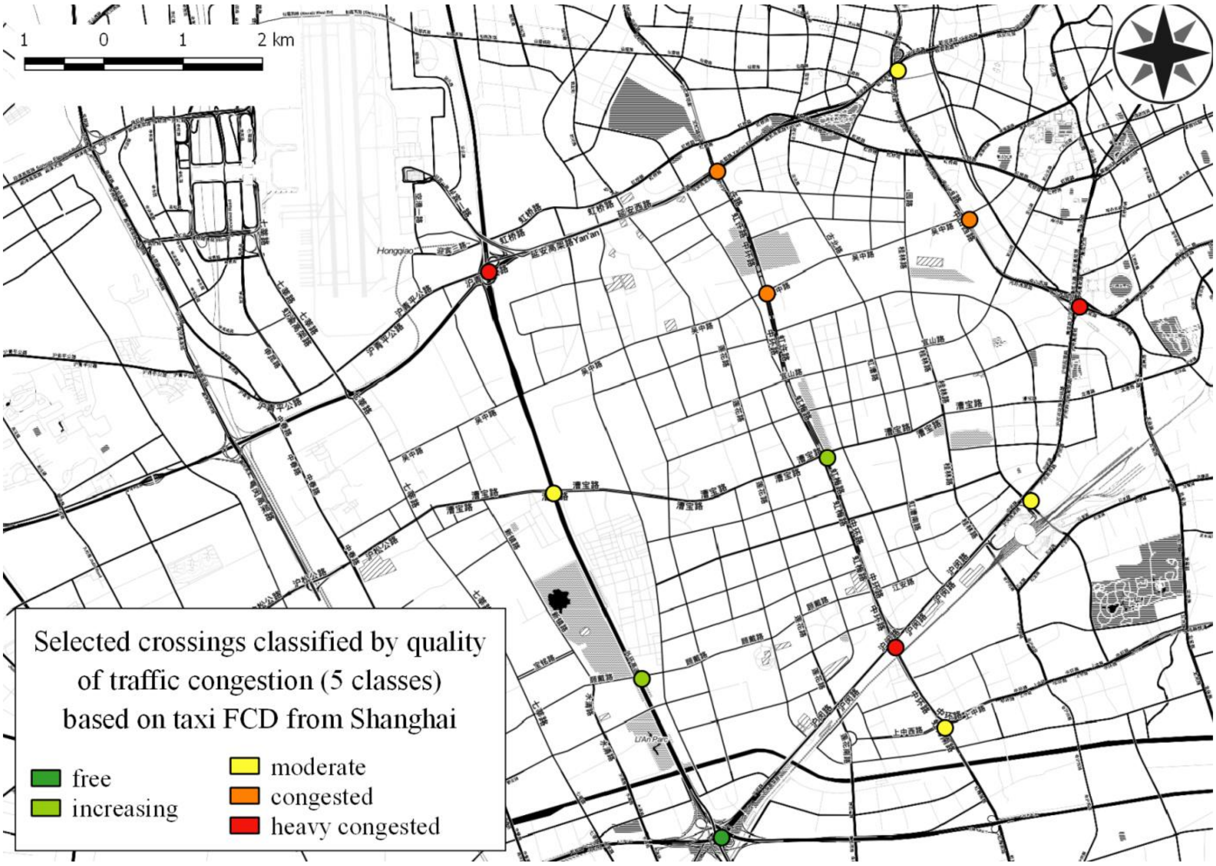


Figure 27: Selected crossings in Shanghai classified by quality of traffic congestion (5 classes) based on taxi FCD.

Within a working day, there are comparable travel time variations between the crossings in Table 8, resulting with the daily morning and afternoon rush hours.

Based on the findings in Keler and Krisp (2016b), it is assumable that traffic patterns (mainly average speed or travel time) between 13 prominent crossings in Table 8 can indicate the periodic rush hours of the city. This means that the estimated traffic flow on the highway segments may show the general situation within the city, or at least its center. As a compromise between data amount and short time windows, 10-minute partitions serve for further traffic analyses. It is assumable that six states within one hour (e.g., selected rush hours on working days) are small enough for detecting differences in travel time variations

Table 8: Selected 13 prominent crossings in Shanghai, with naming, coordinates and selected area identification.

ID	Sel_area	Sel_lon	Sel_lat	Sel_ID
388	Wai Huan Hu Qing Ping Li Jiao Qiao	121.358094	31.183053	301
389	849 Hong Xu Lu	121.388660	31.193598	302
390	827 Zhong Shan Xi Lu	121.412212	31.205220	303
391	1885 Cao Bao Lu	121.367646	31.158105	304
392	1619 Gu Dai Lu	121.379107	31.136764	305
393	Xin Zhuang Li Jiao	121.389511	31.118375	306
394	708 Wu Zhong Lu	121.394401	31.180858	307
395	2 Wu Zhong Lu	121.421279	31.189230	308
396	71 Cao Xi Lu	121.436061	31.179283	309
397	688 Cao Bao Lu	121.403223	31.162119	310
398	Hu Min Lu Fu Lu	121.419189	31.148112	311
399	76 Liu Zhou Lu	121.429660	31.156339	312
400	710 Shang Zhong Xi Lu	121.418159	31.131703	313

5.1.2 Road network information from the OSM project as part of the transportation infrastructure

Additionally, this case study utilizes the OpenStreetMap (OSM) road network for aggregating individual FTD points for the calculation of the traffic parameters in network space. This requires previous matching of FTD-based congestion polygons or polylines with the underlying road network. Additionally, OSM data find usage for the later visualization of the calculated results. It has the function as a base map to display the traffic congestion information. The usage of road segments is not necessarily part of the traffic pattern analysis framework. It is optional for the first four techniques, but required for the last two techniques of the framework.

For the practical traffic analysis in Shanghai based on FTD, OSM data is extracted for the administrative borders of Shanghai. The main information consists of the road network represented as polyline elements. Additionally, it is to mention that digitizing road segments or lanes has no standard procedure within the OSM project. Therefore, static information of the transportation infrastructure requires further inspection. Another step is inspecting the reasonability of this information, mainly in road types, driving directions and restrictions. All these attributes of the OSM road segments are crucial for achieving reasonable routing results, as well as performing reasonable map matching (MM) approaches. In terms of connectivity of different road segments, Shanghai’s road network needs a testing step with computed test routes. The first results indicate a more or less realistic connection between the roads, with some small mistakes in driving directions and at complicated crossings (Krisp and Keler 2015). For comparing the reliability of the computed OSM network routes, the routing service of Google Maps is used. For the city of Shanghai, all available road networks are extracts that are accessible by car. As an illustration, only the highway segments in Shanghai are subject of testing reasonable connectivity. Besides the road network data quality, there is one connection with different visualization possibilities connected with the OSM project. Therefore, there is a need for representing abstracted road segment information (centerlines, lanes) within understandable and realistic map representations. Li et al. (2014) take this as a starting point to extend polyline road information into extended polygonal street information that is realistic by introducing an automated method. Realistic means that estimated street widths comply with the ones of the real world.

In comparison to Liu et al. (2016b), the road network of Shanghai in OSM is more detailed. Whereas Liu et al. (2016b) have a network with 109,986 vertices and 138,978 segments, the used OSM network for testing the framework has 212,110 vertices resulting from 29,857 polyline road segments. This means that especially highway segments are relatively long polyline elements, which results from the OSM data acquisition.

OSM road network information of Shanghai for the detection of complicated crossings

One additional step of the previously presented technique has the focus on extracted nodes from road segment polylines. This has the reason of applying the method of Krisp and Keler (2015) for detecting and extracting complicated crossings. This step is beneficial to the previous procedures of testing reasonable connectivity, because it helps estimating the node densities of the OSM road segments, especially at intersections. The extraction of complicated crossings is feasible with any possible road types that are accessible via vehicle. The different road types include besides highways, also major and minor arterial roads, and connection roads between the mentioned. Additionally, there is an implication of any type of local roads, together with cycle ways and pedestrian paths. Especially the elevated highways and expressways[23] have multiple specific quality problems. More precisely, the OSM of Shanghais’ Inner Ring Road at Puxi has some junctions, exits, and bridges with unrealistic appearance. Additionally, the Inner Ring Road at Pudong has wrongly mapped junctions at Zhangjiang at Longyang Road and Luoshan Road. Some bridges and exits in this area need additional modifications. Shanghais’ Middle Ring Road proceeds from Zhoujiazui Road to Shangzhong Road and includes unrealistic connections to bridges, exits and tunnels. The mapping of Yixian Elevated Road and the North-South Elevated Road are not complete. Yan'an Elevated Road has wrongly mapped bridges and exits. The Humin Elevated Road exists in OSM, but has wrongly mapped exits. The road below (ground level) Humin elevated road has additional mapping errors. The mapping of the Huaxia Elevated Road needs additional modifications, since partially not complete. Additionally there are elevated roads around Hongqiao Railway Station, which are missing.

In general, it is to note that OSM road segments are in many cases not reliable. Especially in Shanghai and other big cities in China, the recent state of digitized road networks is questionable. First inspections show besides numerous missing segments, also newly established that were constructed after 2007. Since the often-modified transportation infrastructure has multiple changes, it is difficult to get historical states of the road network.

[23] Xylem 23:02, 30 January 2010 (UTC), http://wiki.openstreetmap.org/wiki/Shanghai

OSM road network adaptation into the state of 2007

A relatively reliable method for creating a Shanghai road network similar to real conditions consists of a combination of automated and manual methods. The most suitable technique is to perform an automated matching of the segments. The used procedure consists of performing a combination of automated and manual inspections of the road network. One useful source of validations, are satellite imageries from the Landsat missions. The resolution for Landsat satellite imagery is in most cases high enough for detecting changes in shape and position of selected road segments. In many of those cases, where changes appeared between 2007 and 2017, creating taxi trajectory polylines can provide more insights. Matching taxi trajectory polylines with the recent road network in an automated way delivers general results about the places of changes. Nevertheless, it is hard to achieve automatic derivation of the conditions of 2007, since the historical state of OSM is often not given (data storage). Especially the conditions from the year 2007, where the OSM project had not the same popularity as recently, are hard to obtain. Another option to get historical road network conditions is to obtain governmental road network data, as from the cities' agencies.

For the further analyses, one road network data extract of Shanghai from the OSM project is used with various manual modifications. Those modifications result from previous preprocessing and data analysis steps. There are in total a number of 58 segments, which are modified. The ground truth for validating the recent OSM road network is road network of Shanghai from 2007 of unknown source, with less detailed road descriptions and attributes as the one from the OSM project. Many segments of the two data sets match perfectly. Introducing matching tolerance areas via creating polyline buffers based on the recent road type can benefit the manual detection of differing or missing segments.

5.2 Representation of traffic congestion and its propagation with Floating Taxi Data (FTD) from Shanghai

After reviewing the test data sets, the methods of the traffic pattern analysis framework are tested with selected FTD partitions. Those data partitions result mainly from rush hours on working days, where heavy traffic congestion events are present. The reason for testing the framework on these hours of the day is detecting traffic congestion propagation patterns. The first steps in testing the FTD of Shanghai include extraction and preprocessing. Table 9 shows an example of the selected days for the further analyses.

Table 9: Example of selected days for estimating traffic congestion events.

Weekday	Wednesday	Tuesday	Tuesday	Tuesday	Tuesday
Date	31.01.2007	6.02.2007	13.02.2007	20.02.2007	27.02.2007
Number of taxis	5,123	6,003	4,954	4,344	4,896
Number of records	12,425,395	12,764,883	11,877,341	11,464,943	11,146,926

Within the selected days, there is a variation in number of taxis and records. The first preprocessing step is data partitioning into segments (episodes) of one hour respectively 10 minutes for all days of acquired FTD, as the Tuesdays in Table 9. The further analysis includes extracts of morning rush hours on working days. The reason for selecting working days is the usually greater significance of working day rush hours, compared to rush hours of weekends. The morning rush hours of working days are, compared to the evening rush hours, more intense and easier to detect.

5.2.1 Traffic congestion detection by global index (TCD-GI)

The test results for the first technique of the traffic pattern analysis framework show significantly differing patterns for the morning and evening rush hours on working days in Shanghai. Figure 28 shows these appearances with one selected day in Shanghai.

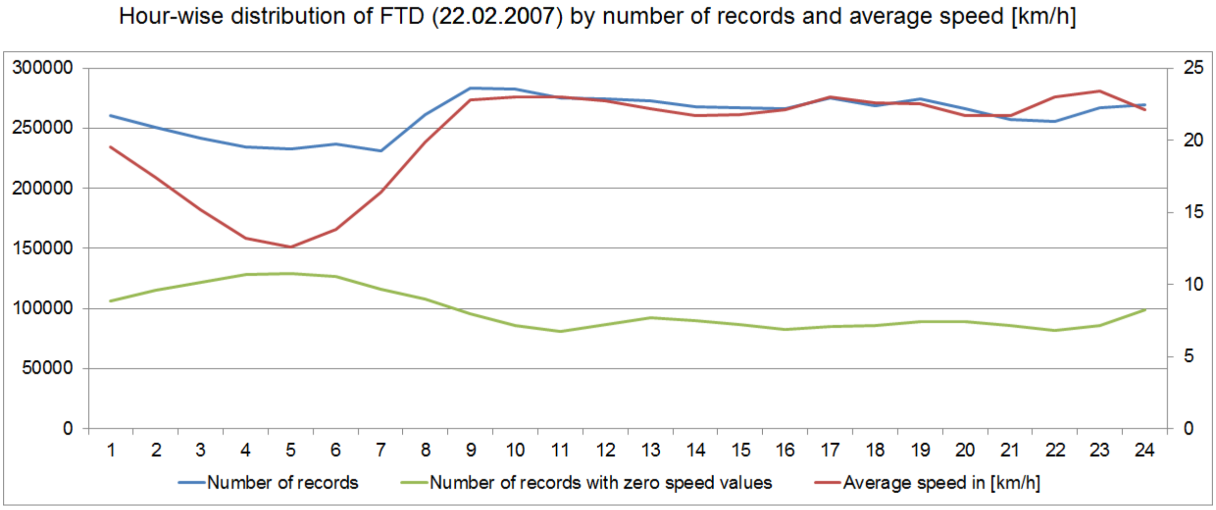


Figure 28: Example for distribution plot of input information for defining global traffic congestion index values for 24 hours of FTD in Shanghai on Feb 22, 2007.

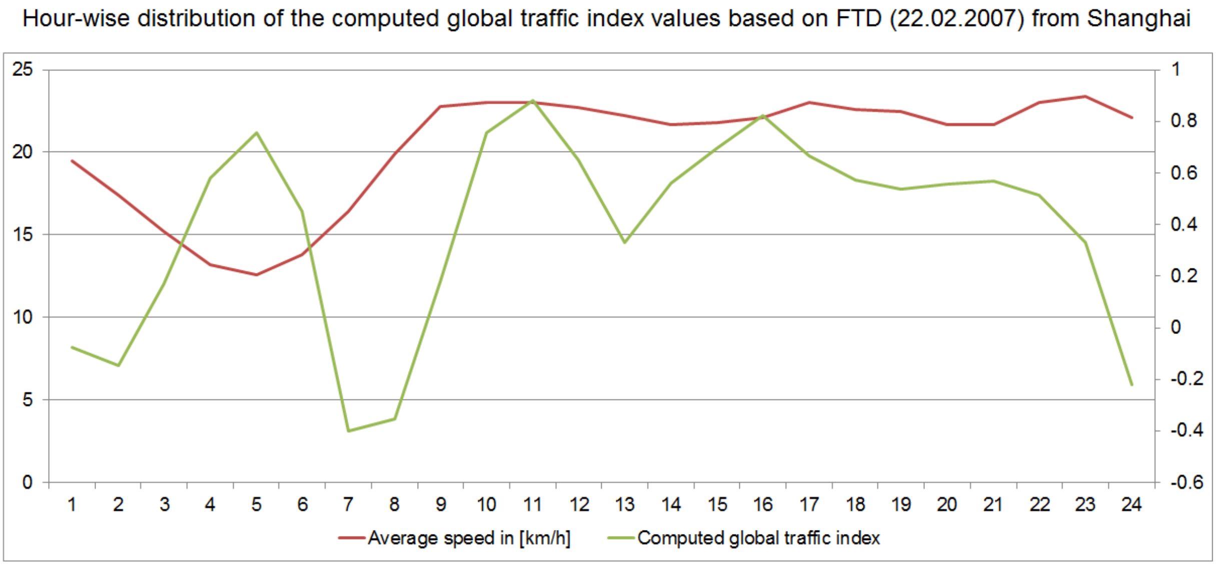


Figure 29: Example for distribution plot of the global traffic congestion index values for 24 hours of FTD in Shanghai on Feb 22, 2007.

The blue colored curve in Figure 28 shows the number of records per hour and the green curve shows only those records, which have zero velocity as the attribute value for instantaneous velocity. The values of both curves refer to the left vertical axis. The averaged instantaneous velocity is plotted in red and refers to the vertical axis on the right.

The distribution of average velocities on this working day in Shanghai is not typical. Lower average velocities appear earlier than on other days of the FTD acquisition period. The evening rush hour is not even detectable. By applying the formula for the global traffic index I_{hour}^{cong}, it is possible to infer further information. Figure 29 shows the distribution of the computed global traffic index $\mathrm{I}_{\mathrm{hour}}^{\mathrm{cong}}$ for the values, which are pictured in Figure 29 with the green curve. Additionally, the plot includes the average velocity curve with the associated left vertical axis for the values. The right vertical axis in Figure 29 refers to the values of the global traffic index. After numerous tests, it is to say that negative global traffic index values often indicate rush hours. Due to the high numbers of records with zero velocities, negative index values appear also in the late night hours, which falsify the selection of rush hours. Nevertheless, the minima of the global traffic index are between 7AM and 8 AM and indicate the morning rush hours in a very distinct way. As a completion for an insight into the computed global traffic index values, Table 10 shows the results for the selected data partition, together with all values of the input measures.

Table 10: Distribution of number of records and average velocities of one day (Feb 22, 2007) of FTD from Shanghai.

Time [h]	$n_{records}$	$n_{zero}^{\bar{v}}$	$\bar{v}$ [km/h]	I_{hour}^{cong}
0 to 1	260438	105913	19,5	-0,0760606414
1 to 2	251009	115397	17,4	-0,1448528043
2 to 3	241540	122103	15,2	0,1692775195
3 to 4	234489	128505	13,2	0,5811154018
4 to 5	232325	129041	12,6	0,7549148681
5 to 6	236489	126578	13,8	0,4508267724
6 to 7	230861	116359	16,4	-0,4004221871
7 to 8	261210	107832	19,9	-0,3532867681
8 to 9	283141	95305	22,8	0,1785598846
9 to 10	282637	85957	23,0	0,7572582059
10 to 11	275487	81110	23,0	0,8830058957
11 to 12	274614	86454	22,7	0,6499859419
12 to 13	272865	92401	22,2	0,3300312168
13 to 14	267611	89544	21,7	0,5588939638
14 to 15	267358	86867	21,8	0,6968292038
15 to 16	265918	82884	22,1	0,8224081059
16 to 17	275033	85199	23,0	0,6662002548
17 to 18	268278	85996	22,6	0,5725353255
18 to 19	274226	88792	22,5	0,5385175837
19 to 20	266584	89234	21,7	0,556621202
20 to 21	257226	85912	21,7	0,5698364579
21 to 22	255947	81419	23,0	0,5119717398
22 to 23	267199	85834	23,4	0,3307081061
23 to 24	269700	98568	22,1	-0,2211232544

5.2.2 Traffic congestion detection by SNN-Clustering (TCD-SNN)

The TCD-SNN technique implies SNN clustering of all taxi movement positions and subsequent convex hull generation. Performing SNN clustering consists of creating a similarity matrix for instantaneous speed values and driving direction values. The outcomes of the first tests deliver many insights into urban dynamics during working day rush hours; more precisely, it represents detected traffic congestion in Shanghai by time series of polygons. These polygons imply the additional similarity of the driving direction, which is challenging to provide. The reason for this is the lack of insights into reasonable driving direction values. This TCD-SNN technique aims at defining congestion events for selected time windows that affect vehicle drivers moving in the same direction. For the case of one selected rush hour, there is a partitioning step for the preprocessed FCD into 10-minute partitions. These partitions represent episodes of the traffic situation. Figure 30 shows one example for February 19, 2007 during a morning rush hour (8 to 9 AM).

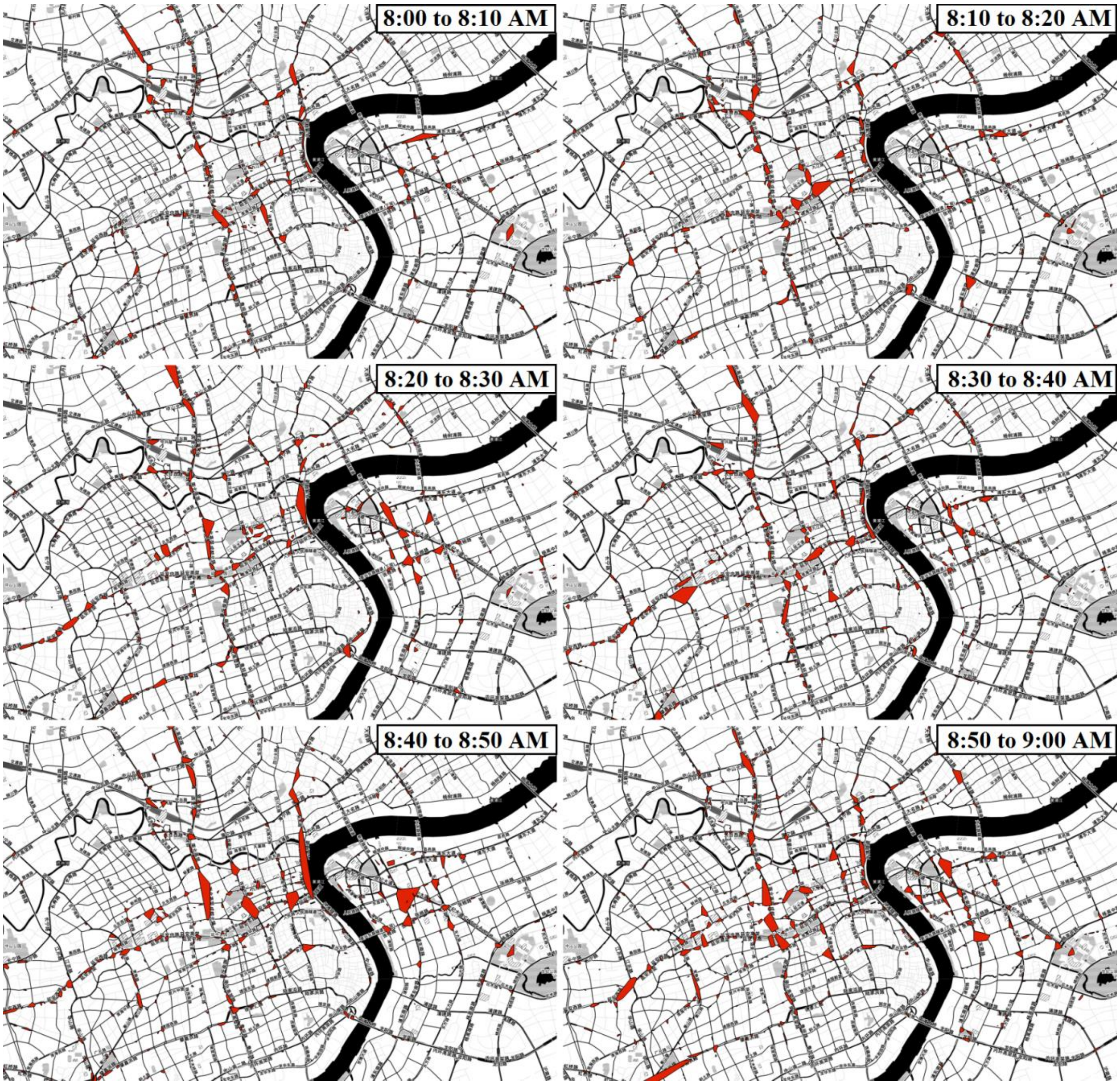


Figure 30: Example of defined congestion events in the Center for Shanghai for a morning rush hour (8 to 9 AM) on the 19th of February with 6 proposed episodes.

Figure 30 shows that the distribution of congestion polygons is frequently changing every 10 minutes. This represents the highly dynamic manner of urban traffic congestion. By first visual inspection, changes in polygonal shapes and sizes at the same road intersections and partitions are detectable. These appearances might indicate traffic bottleneck movements. Resulting from the idea of detecting changes in topological relationships, matching traffic congestion events with available road segments is possible. This is partially difficult to perform, since elevated and non-elevated roads appear at the same areas, as pictured in the lower left part of Figure 31. In some cases, the appearance of traffic congestion events is hard to associate with certain road segments. One helpful procedure for visual analysis is computing density ranges for the resulting SNN clusters with varying densities. In Figure 31, the color scheme is ranging from dark green as low-density clusters, over yellow as mediocre density, to red as SNN clusters with high density of points.

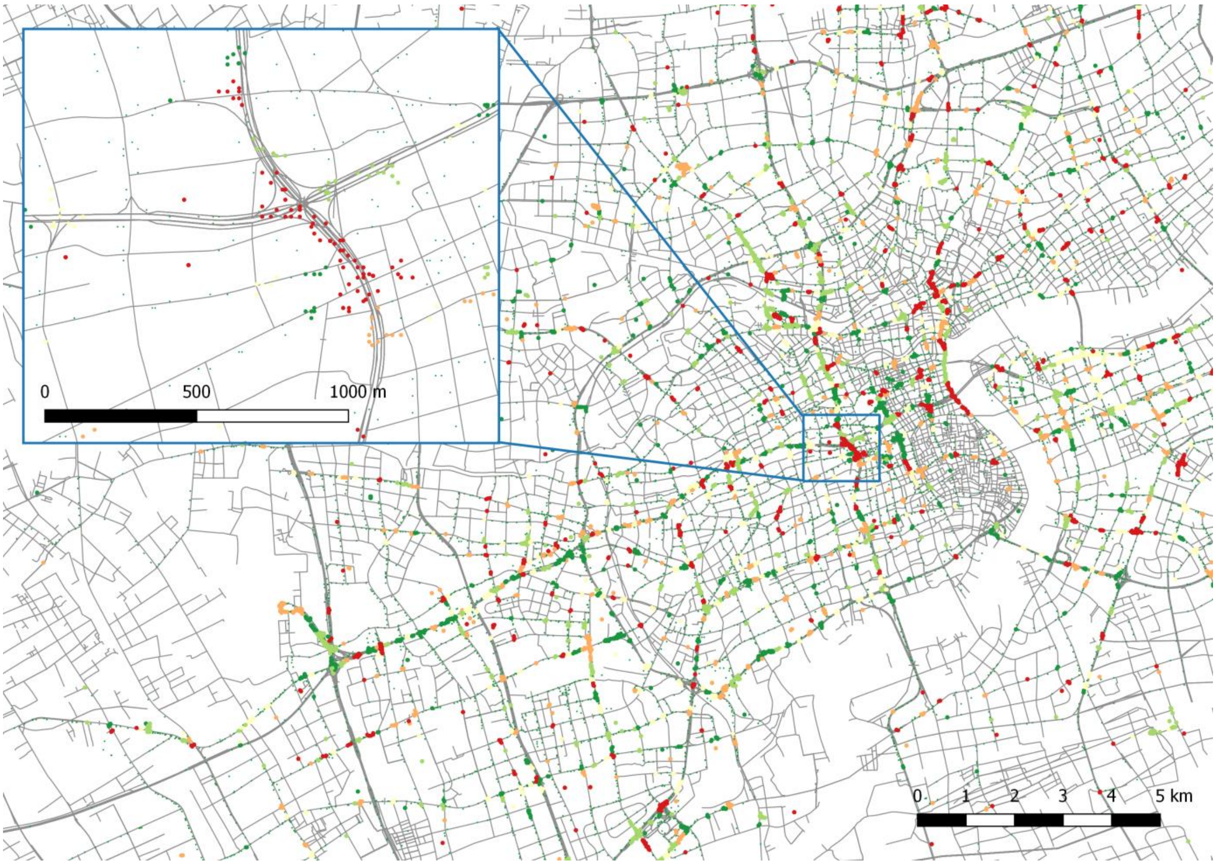


Figure 31: Original SNN clusters for a selected 10 minute FTD partition colorized by variation in density.

Figure 31 shows different detected densities for different similarities of taxi velocity and driving direction values. Up to 90 percent of all derived SNN density core points are located within a 50 meter distance to certain road segments.

One simple possibility to match the SNN clusters with road segments is to provide Naïve MM, one of the very first MM approaches, with computed congestion core points and its nearest road segments. Based on the appearance of certain taxi identifications in the FCD sets, it is possible, up to a certain level, even to distinguish between elevated and non-elevated road segments. An interesting appearance is also detectable at certain parts of the elevated highways: on-ramps between minor roads and highways. Those transportation infrastructure elements can sometimes imply similarity with higher velocity values. Those clusters are spatially nearly always smaller than congestion polygons. In Figure 31, there is no differentiation between slow moving traffic with lower similarities (dark green color) and higher velocity values with lower similarities, which represent in most cases on-ramps.

For distinguishing between recurrent and non-recurrent congestion (Chen et al. 2016), the focus of the framework is to inspect aggregated congestion duration times by the previous findings with matching defined congestion polygons. Consequently, the presented congestion detection model has a resolution of 10 minutes. Based on these findings, there are detectable congestion events with different durations. The central aspect for congestion classification is the matching into (1D) network space. In a subsequent step, it is possible to analyze the periodical appearances of congestion events for longer periods of time.

5.2.3 Traffic bottleneck inference (TBI)

Besides the usefulness of the previous, there are remaining clusters that are showing similar movement that is actually not traffic congestion. TBI consists of using OPTICS for defining clusters with different densities of preselected FCD. The TBI method focus on those FCD points that have instantaneous velocities below 20 km per h, which are associated with less effective vehicle driving with greater exposures and fuel consumption. The second extension is the use of OSM road networks for applying the method by Krisp and Keler (2015).

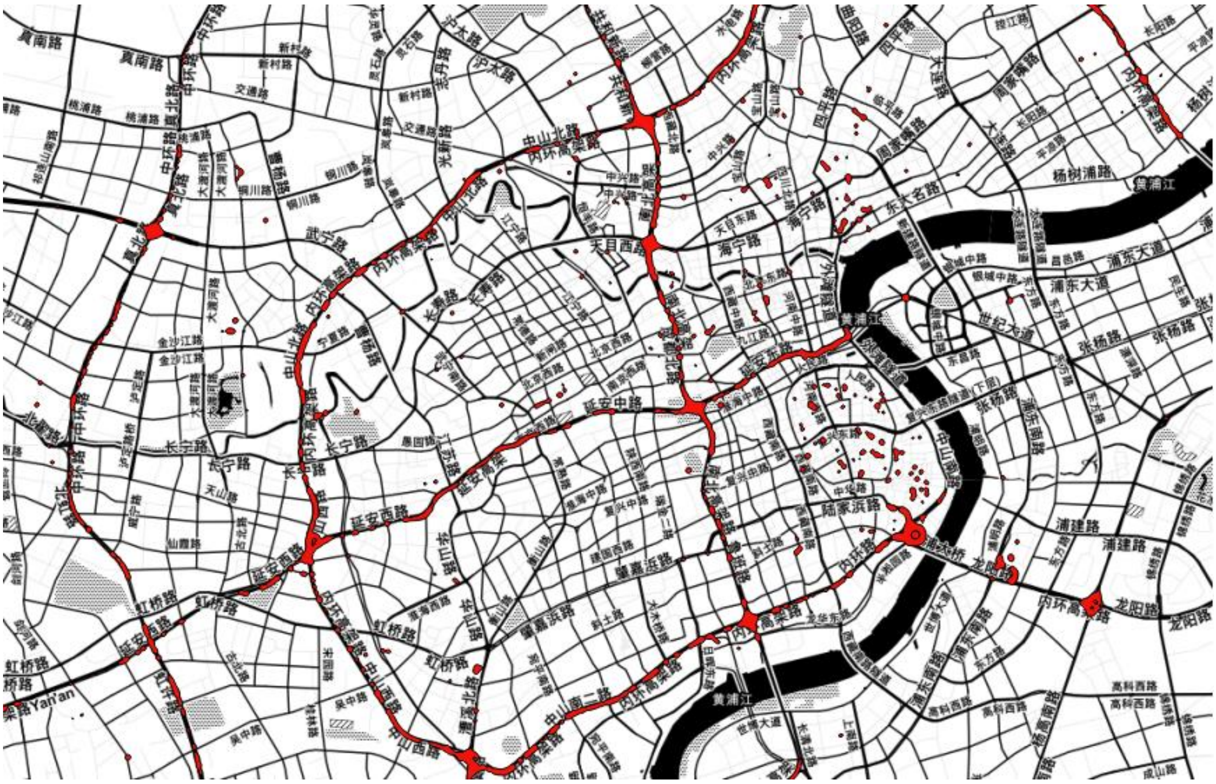


Figure 32: Detected complicated crossings in the center of Shanghai, by applying the method by Krisp and Keler (2015) for OSM road network of Shanghai.

By combining the results from the two approaches, it is possible to provide an alternative definition for moving traffic bottlenecks. By applying the method of Krisp and Keler (2015), a high number of complicated crossings is detectable, mainly resulting from multiple intersections of different road segments. Figure 32 pictures a cutout of detected complicated crossings in the city center of Shanghai.

Focusing on the central part in Figure 32, Figure 33b shows some large-area polygons, as a cutout of detected complicated crossings in the same area of Shanghai. This appearance might appear clearer by some explanation from the literature: Shanghai's overall traffic load consists of around 30% resulting from its elevated road network (Li and Zuo 2004, Zhang 2004, Xu et al. 2012). Furthermore, Xu et al. (2012) state that this appearance is difficult for identifying elevation levels of vehicle trajectories. For deriving the traffic states, it is also difficult to say if traffic congestion is on or below the elevated highway segment. By means of demonstration selected 10 minute partitions of the overall FTD set are used to represent results. Detected congested areas of 10 minutes in the morning rush hours of one selected working day (Monday the 12th of February 2007, from 8:30 to 8:40 AM) are pictured in Figure 33a.

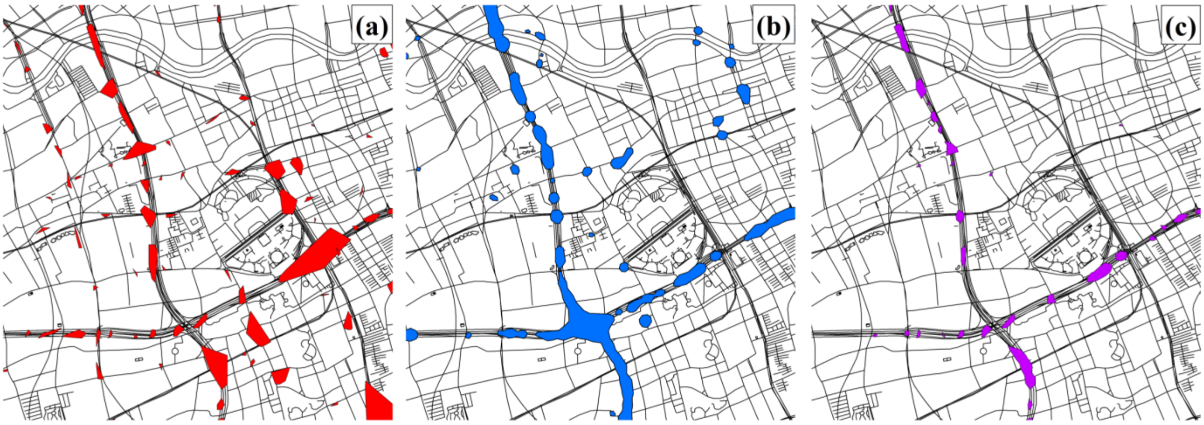


Figure 33: Congested areas (a), complicated crossings (b), and detected bottlenecks (c) in Shanghai.

By matching the spatiotemporally varying congestion polygons with detected complicated crossings, inference of potential bottlenecks locations is feasible for the same time windows as the traffic congestion polygons. Figure 33c pictures the result of detected bottleneck locations. Via visual inspection of the results, there are several aspects on the shapes and sizes of the resulting polygons. More precisely, Figure 33a includes relatively angular polygons due to the giftwrapping algorithm. Figure 33b has more rounded polygons that mainly appear at more complex intersections. The matched areas between these two mentioned polygon types are traffic bottleneck polygons on parts of the underlying road network (Figure 33c).

This shows that no previous Map Matching (MM) technique is needed for connecting traffic information with road segments. Dependent on the input data quality and quantity various results are possible. This is also connected with variations in the level of aggregations, namely the variation is spatial and temporal accuracy of detected dynamic traffic bottlenecks. One possible analysis options for detected traffic bottlenecks is to compare the results for different time windows. Especially the size variation of traffic bottleneck areas from subsequent time windows is interesting, since changes of the traffic situation can be identified (e.g. for selected hours of working days).

There are numerous possibilities of inspecting variations of traffic bottlenecks in time and space. One option is shown in Figure 34, where detected bottleneck polygons are presented for the three time windows between 8:30 and 9:00 AM (19th of February 2007) within a cartographic view. The traffic bottleneck polygons in this view are pictured with 50 % opacity of the colors, which serves for showing the polygon boundaries of intersecting polygons.

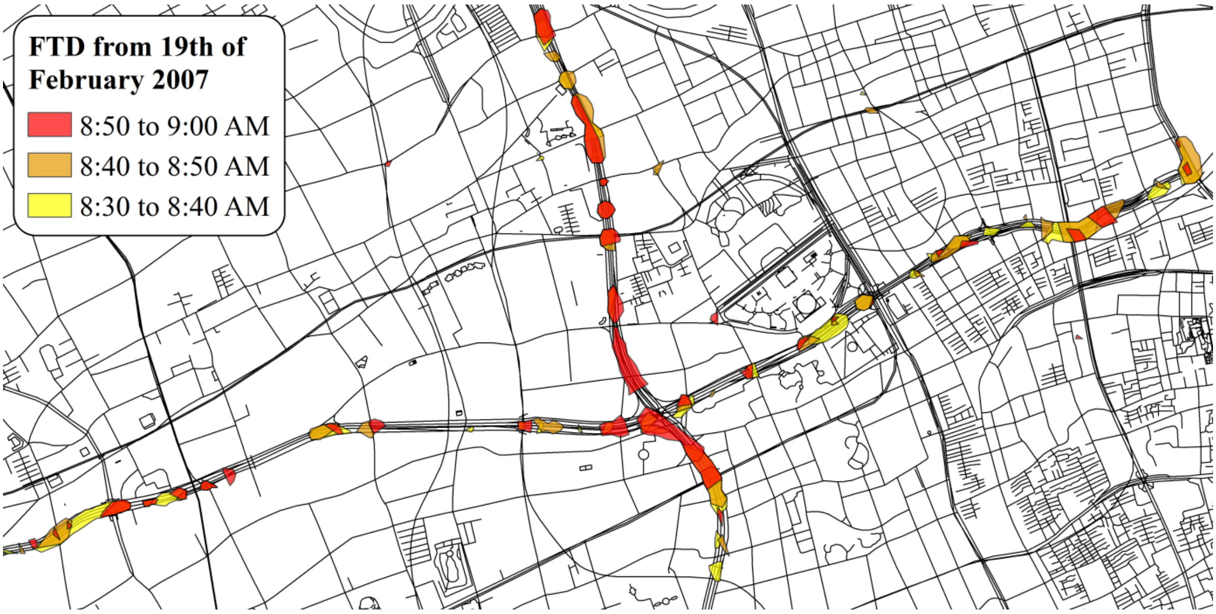


Figure 34: Geovisualization possibility for three consecutive traffic bottleneck polygons within a two-dimensional view.

Figure 35: Geovisualization possibility for three consecutive traffic bottleneck polygons within an extruded three-dimensional view.

The second option for cartographic representation is the extrusion of the polygons within a 2.5D view. By means of representing bottleneck propagation between consecutive time windows, each polygon is enriched by the number of FTD points with instantaneous speed lower than 20 km/h, motivated by the traffic congestion definition of Robinson (1984). The following step is to extrude congestion polygons based on the FTD point density of selected 10 minutes, as it is pictured in Figure 35. Figure 35 shows that there are high variations in vehicle densities of the detected traffic bottlenecks. The bottlenecks in the eastern part of Shanghai in Figure 35 show a decrease in vehicle density, whereas the western part bottlenecks increase in vehicle density (Keler et al. 2017c).

5.2.4 Traffic congestion propagation by congestion propagation polylines (TCP-CPP)

The fourth technique of the traffic data analysis framework delivers results that are difficult to interpret. The reason is that the results do not represent individual movement of vehicles, but the movement of one complex traffic congestion event or numerous events. The possibly best way for interpreting resulting traffic congestion propagation polylines (CPP) is via visual inspection of the polyline properties. Based on the extracted data partitions, respective series of six congestion polygon ensembles are used for creating CPP.

For reasons of clear representations, one example exemplifies the generation of congestion propagation polylines (CPP). A first test consists of selecting one morning rush hour in the central area in Shanghai (12^{th} of February, 2007). For the six data partitions, six polygon ensembles are extracted via SNN clustering. The subsequent step consists of pairwise matching of the polygons of subsequent partitions. For the case of matching numerous polygons of one time partition, the following condition is introduced: selection appears only for the polygon closest to the centroid of the previous polygon (of the previous time window) or of the successive polygon (of the successive time window). The application of this polygon matching has one extension for getting more polygon intersections: addition of a search distance of 50 meters, by including a spatial buffer around each polygon.

Figure 36a shows polylines of different lengths, but with the same number of nodes (six). It is to mention that the centroids of the congestion clusters are not always matching the road segments. In many cases, it is difficult to detect the right road segment, which is influenced by congestion, as the two polylines in the Northwestern part of Figure 36. In this area of Shanghai (see Figure 36), there are around 30 to 60 congestion polygons per 10 minute partition. After querying the consecutive overlapping polygons for matching taxi identification in all six subsequent time windows, nine results for congestion propagation polylines are received. These successfully connected CPP are represented in Figure 36a. As an addition for further visual inspection of the propagation patterns, it is possible to add labels onto the polylines, more precisely at the position of the congestion polygons centroids, as pictured in Figure 36b.

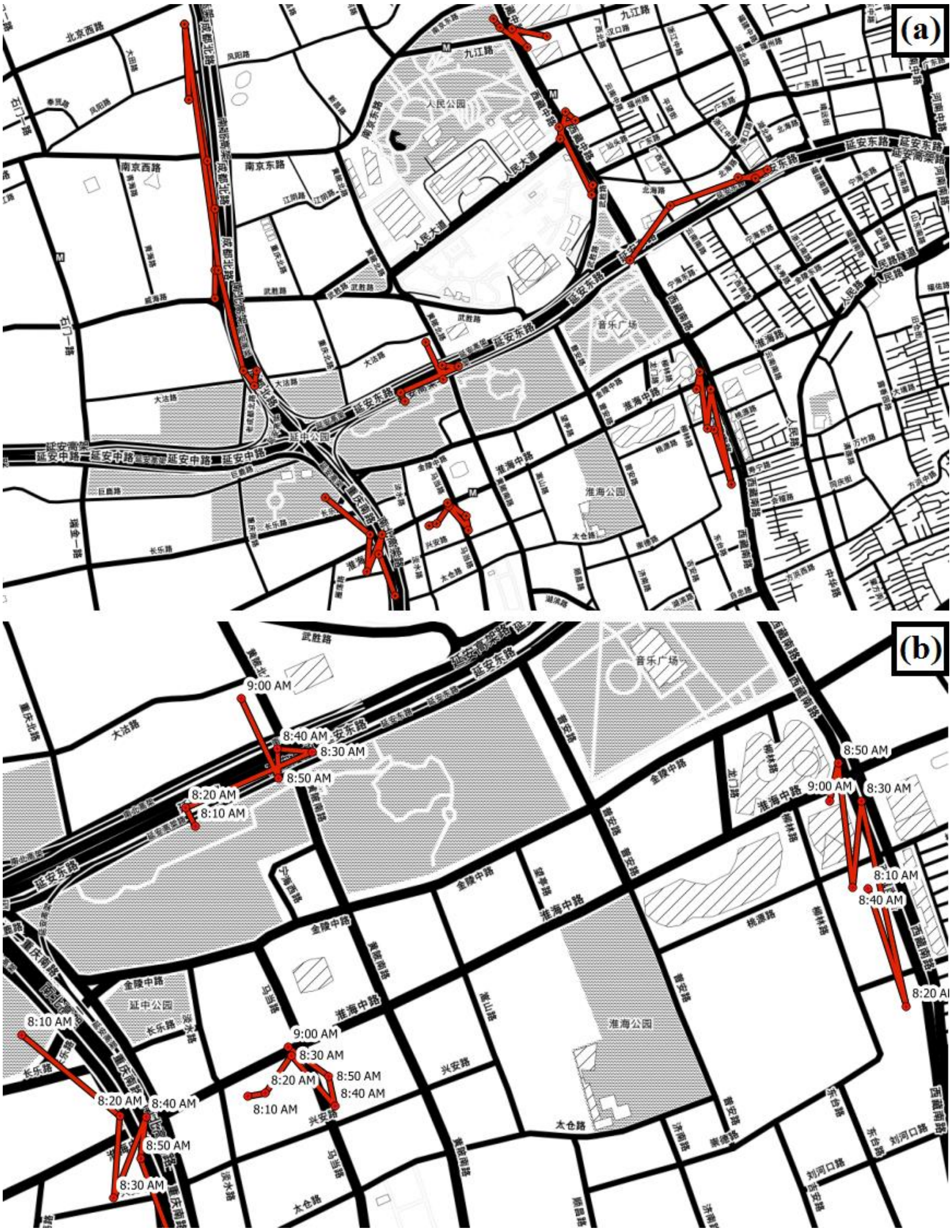


Figure 36: Congestion propagation polylines (CPP) for the morning rush hour on February 12, 2007 with (a) nine individual taxi drivers getting stuck in traffic congestion, and (b) labeled CPP for the same time window with 4 individual taxi drivers getting stuck in traffic congestion.

Remarkable details of the resulting propagation polylines are the directions of polygon centroid movements. In many cases, congestion propagation movement is in the opposite direction, when comparing to the closest situated road, namely against the driving direction. Nearly all detected congestion propagations appear at locations with high road densities and in many cases at elevated highways. Another frequent appearance is the detection of ongoing traffic congestion events on on-ramp segments.

The second option for CPP is the connection of taxi trajectory partitions influence by the detected traffic congestion polygons. Figure 37a shows an overview for detected CPP of individual drivers for the morning rush hour (8 to 9 AM) of the 12th of February 2007. There are in total 16,443 polylines with higher coverage in the city center of Shanghai.

Figure 37: Overview over created test CPP with taxi trajectories from the morning rush hour (8 to 9 AM) on February 19 with (a) distribution of CPP in Shanghai, and (b) the center of Shanghai with CPP showing long durations.

Figure 37b shows a more detailed view on these polylines. Most of the detected CPP consist of only two to three taxi positions, which might result from the fact that taxi drivers leave the defined congested zones within one hour. Another pattern in Figure 37b is that many CPP appear at on-ramps or intersections with elevated highway segments.

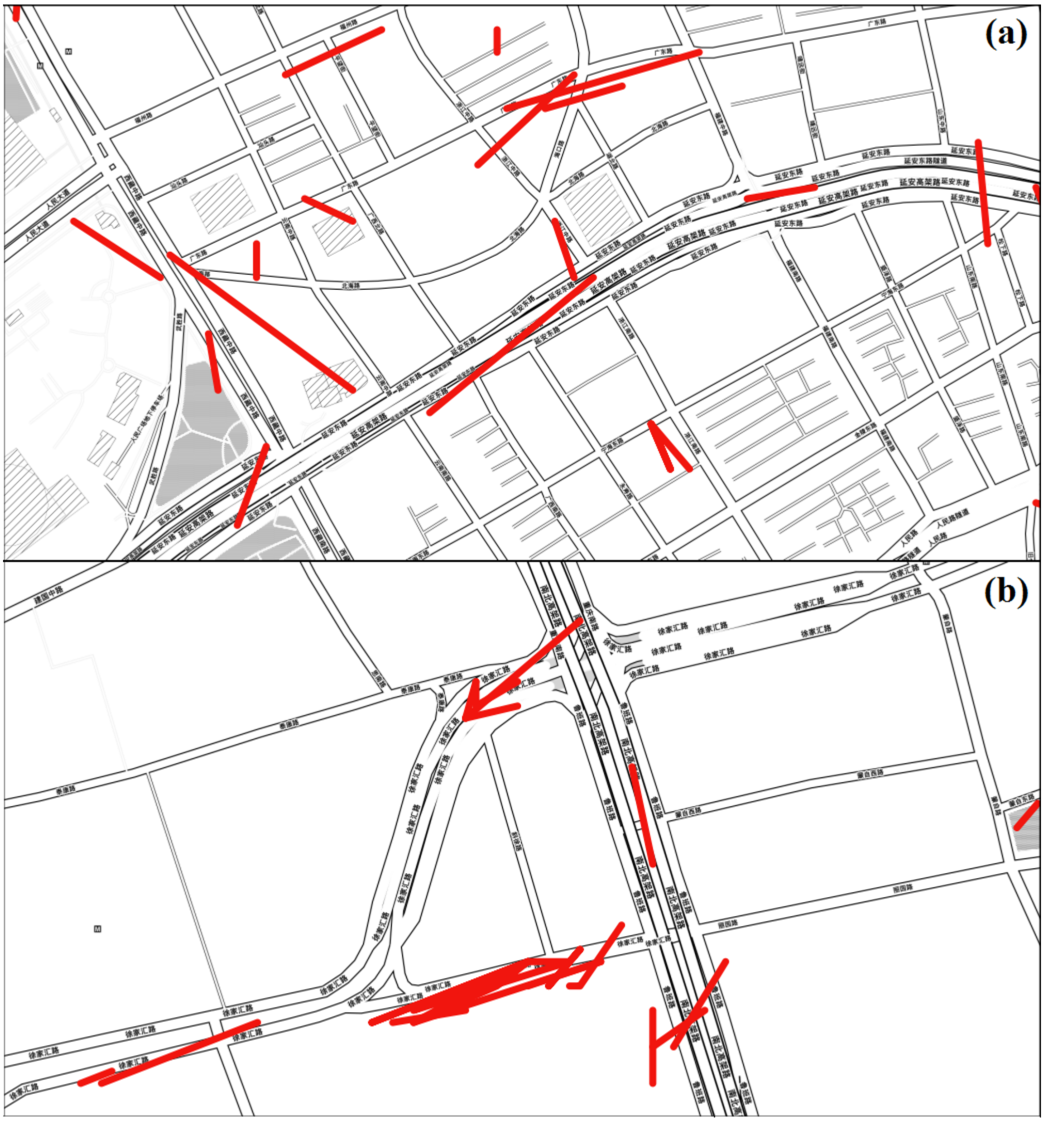


Figure 38: Selected example in congested areas of Shanghai for the morning rush hour with (a) traffic congestion propagation below elevated segments, and (b) congestion propagation at on-ramps.

Figure 38 shows these details on a different scale in the center of Shanghai. Figure 38a shows the high inaccuracy in the measured taxi positions. Nevertheless, it is assumable, which road segments of the network are affected by the detected traffic congestion. Figure 38b shows one example with congestion at on-ramps, where taxi drivers spend more than 30 minutes within the connected congestion polygons. In the southern part of the view in Figure 38b, are three individual taxi drivers, who move within 40 minutes on small on-ramp segment.

5.2.5 Matching traffic phenomena on road network segments (TP-MM)

The results of matching the detected traffic congestion events with road segments, deliver mostly not very accurate results. The reason for this is that many congestion events occur at those locations, where the density or transportation infrastructure elements is very high. Figure 39 presents one test result of the matching, where congestion polygons base on 10 minutes of data. The matching consists of assigning polygons with the respective locations of the OSM road network. Figure 39a shows these congestion polygons in red with opacity and Figure 39b the partitions of affected road segments.

Figure 39: Road network in the center of Shanghai influenced by traffic congestion events with (a) overlapping of traffic congestion polygons with road segments, and (b) extracted influenced road segments.

The subsequent step is to test the connectivity of affected road segments. First tests show that extracted influenced road segments as in Figure 39b, have connected network partitions on two different elevation levels. This might be useful for differentiating between elevated and non-elevated segments. Therefore, the following step is to extract affected road segment and to test their connectivity.

After visually inspecting both types of data, the first impressions result in the following three statements:

1. Massive FCD records have often-inaccurate positions in densely built centers of urban environments due to urban canyons (signal losses), artificial surfaces (multi-path effect), and lower GDOP values.
2. Massive FCD records have sometimes positions not matching the road network due to differences in correctness of data: e.g., road network has changed after the FCD acquisition.
3. Matching inferred FCD trajectories and road network connectivity can deliver reasonable and unrealistic results, since the height component in the used FCD sets is missing.

5.2.6 Traffic density estimation based on sparse FCD (TDE-FCD)

The method for estimating the traffic density based on sparse FCD, delivers results that are difficult to interpret. The most challenging step is estimating capacities of road segments, and differentiating between different road types. The attribute information in OSM is not accurate enough to provide realistic estimation of traffic densities. In many cases, available OSM attributes have missing values, which makes it difficult to estimate specific capacity values. Nevertheless, counted positions of taxis are useful for relating to traffic congestion events. The reason, why the computed traffic density information has less usefulness, is resulting from unrealistic values. Certain parts of the network deliver values that are physically not possible, especially when comparing to the detected traffic congestion events.

6. Discussion and conclusions

In general, the different techniques of the proposed framework allow gaining more insights into traffic situations. The results of applying the framework methods show the possibility of detecting and representing specific patterns of traffic congestion via taxi FCD. Specific patterns imply the differentiation between different types or classes of traffic congestion in urban environments. This classification, together with the pattern detection is dependent on data quality. In this thesis, several aspects of data quality are not part of the content. Only sampling interval and positioning accuracy of the tracking devices are part of data quality reasoning. Besides these two aspects, several other data quality properties are essential for reasoning on the quality of the method outcomes.

One focus of this work is distinguishing between different congestion events based on massive FCD. The part of distinguishing different traffic events presumes the previous detection of these events. The six framework methods serve as an attempt for first detect traffic patterns from FCD and then classify them by several gained insights. Therefore, the framework has specific preprocessing steps, which might imply fore or less computations, as data preselection and partitioning. After excluding most of the unrealistic FCD records, the six steps allow the potential analyst to observe FCD sets in global views by comparing average velocities for different times of the day. Afterwards, the views on the data appear with more details via various map views on detected traffic patterns. These local views enable to observe specific spatiotemporal propagation patterns, which might imply periodicity when observing FCD from different days, weeks or months. Therefore, differentiation of periodical and unusual traffic congestion events do not necessarily include data quality aspects.

There are at least two common principles of the proposed information extraction methodology. These principles include (1) finding speed similarities and variations, and (2) estimating the density of movement positions and vehicles. Optionally and as part of the SNN technique, there is distinguishing between different driving directions. Detecting speed value variations and density estimation of movement positions are the bases for modelling traffic congestion propagation. Furthermore, the daily mobility patterns are connectable with (static) transportation infrastructure. This is a task, which can have numerous solutions, depending on expected outcomes. Reasoning on this connection is part of this thesis and is an extendible approach.

First results of the presented traffic bottleneck detection and visualization techniques show promising insights into features of the urban transportation infrastructure affected by traffic congestion. These features or elements, in many cases, correlate with periodical traffic congestion and imply a structure of lower efficiency. The practical use of the presented results might also benefit the understanding of traffic occurrence and propagation. Using the computed products in practice is the best way to evaluate the results, since this alternative method is heavily data dependent and scalable. Scalable means that, it is possible to set different thresholds for distinguishing between slow moving traffic and actual traffic congestion, which might have different perceptions. In general, it is to state that the presented alternative technique in this work has simple computational steps and is easy to perform, which is beneficial for potential analysist, since there is less input of time and effort into data preprocessing. One benefit of the method is to observe the connection between extracted road network segment densities and detected traffic congestion event. The reasonability of defining intersection areas as traffic bottlenecks needs further evaluation, since selected polygon results might be questionable.

Even when mentioned as massive, the inspected FCD sets are small portions of the real movements within daily traffic. Therefore, one idea is to introduce a model for traffic volume estimation, as a useful addition. All methods for detecting traffic congestion base on relatively sparse data of tracked vehicles. As it is mentioned by Shang et al. (2014), it is still not possible reconstruct the real participants on selected road segment for certain time periods based on few tracked entities participating in traffic. Nevertheless, estimated traffic volumes are not only dependent on averaged recorded travel speeds of the individual vehicle trajectories. Traffic volumes also base on road types and (data-driven) road popularities. Therefore, the attempts for estimating traffic density deliver further insights into FCD sets.

Revisiting the hypothesis and answering the research questions

The hypothesis of this work states that applying the presented framework methods on vehicle movement trajectories enables the detection of traffic phenomena of the real world. More specifically, using the framework makes a differentiation possible between detected traffic phenomena. The presented examples in this thesis involve different forms of urban traffic congestion. There are at least two distinguishable forms of urban traffic congestion, which are namely recurrent and non-recurrent congestion events. This is manageable by traffic pattern detection in different time windows of the FCD sets. By focusing on the same times of the day on, for example working days, recurrent events imply appearance of similar spatiotemporal patterns. The six techniques of the framework enable this differentiation, especially by providing global and local views on FCD for detecting traffic congestion.

Consequently, the hypothesis is true in the sense that data of moving objects needs examination by different approaches. These approaches are applicable as the proposed methodological sequence in this thesis. Helpful insights result from investigations in different movement spaces, such as the two-dimensional Euclidean space and the network space. The statement on the independency of the approaches of data quality is difficult to proof. It is to state that the two inspected aspects of data quality, namely sampling rate and spatial positioning accuracy, have a high influence on traffic pattern detection. Therefore, this statement does not apply the real outcomes and has no proof in the sense of an evaluation.

This concludes to answering the first research question of how to detect complex traffic phenomena from massive FCD with varying quality. Numerous research mentions the fact that traffic dynamics are detectable with FCD that implies sampling intervals smaller than 60 seconds. This facilitates computing accurate traffic delays for every road partition. Additionally a sampling interval smaller than 60 seconds allows detection of traffic signals, such as traffic light signaling the resulting traffic dynamics. When the focus is on detecting complex traffic phenomena, the more important component is not the short sampling interval, but the number of tracked traffic participants that imply eventual interactions. Similarities in movement position densities and instantaneous velocities are more beneficial for the analysis than having only few tracked moving vehicles with shorter sampling intervals. Nevertheless, longer sampling intervals make is more difficult to reconstruct movement trajectories in a complex urban environment. Another more important property of data quality is positioning quality. With the circumstances of having very low positioning quality, it is nearly impossible to reconstruct detailed traffic events. Therefore, the first research question has the answer that it is generally possible to reconstruct complex traffic phenomena with FCD of varying quality. When applying various detection and representation methods, there is a need for taking into account the estimated FCD quality. One example is the creation of a dependency of density-based clustering parameters (search distance, minimum number of points) on FCD quality properties. This might allow introducing a certain area of tolerance within density-connected point clusters resulting from adapted point clustering. In case of long sampling intervals or low spatial accuracy of the used FCD, the resulting point clusters and thereof dependent polygons usually become spatially larger.

The second research question focuses on finding optimal properties of FCD for gaining optimal results. In many cases, this question is difficult to answer, since number of FCD records and numbers of trajectories are hard to associate with traffic events and their propagation over a period. The latter is only inferable via relatively long time windows of FCD of numerous days. On the other hand, it is challenging to define traffic events with single trajectories, due to the missing similarities to other movement trajectories within the same time window (episode). Therefore, there should be a detectable density of movement position points coming from at least two entities moving in the same temporal partition. The question concerning the needed amount of FCD for enabling traffic pattern detection can be answered in the following way: on the one hand, a high number of tracked entities is beneficial for detecting traffic phenomena as traffic congestion or standing in a queue of an intersection with traffic lights. The distinguishing between congested traffic and free flowing appears more reliable with a high temporal acquisition frequency (high temporal sampling rate), respectively a short sampling interval. A high number of tracked entities with long sampling intervals for example make it harder to classify or to distinguish between different traffic phenomena, as different types of traffic congestion events. Relatively high numbers of tracked entities often deliver more spatial distribution of the records within urban investigation areas. This gives the potential analyst more options of detecting congestion propagation patterns and their spatiotemporal connection. Additionally, in the last mentioned case, the calculated global traffic indices might be more reliable, when the traversed spaces of the investigation areas are relatively small. On the other hand, high sampling rates allow a more accurate classification of traffic phenomena, as they include more frequent changes in instantaneous attribute values. Stop-and-Go patterns as mentioned in Ranacher et al. (2016b) imply a precondition of used FCD to have relatively high sampling rates. This allows different ways to compute velocities, which enables the modeling of temporal variabilities in a more detailed manner. Examples include daily average speed profiles for selected road lanes or road segments based different speed computing methods. Besides similarities in speed, Flocks and convoys, which indicate group movement of entities, are hard to detect in FCD due to the nature of moving vehicles on restricted space of a network. These patterns are more typical for moving animals or pedestrians in less restricted spaces and can indicate social connections between the entities. This is not the case for spatiotemporal taxi movement patterns in urban environments.

The third question has the focus on finding typical characteristics for specific types of traffic congestion other phenomena independent of the investigation area. The investigation area independent parameters are resulting from the appearance of still-standing vehicles at a certain area for period of time. This means that this appearance should be differentiable from standing at a road intersection with traffic lights, which might also imply the appearance of traffic congestion, especially during rush hours in an urban environment. The independent parameters include as well the number of records and instantaneous velocity. Additionally, computing global traffic indices is independent of the investigation area. The vehicle position densities can depend on the investigation area, since the complexity of road networks might influence the clustering parameter for density-based point clustering.

Comments on the general and specific objectives

The discussing of the insights gained from the research questions, enables commenting the general objective of this thesis. There are numerous possibilities of obtaining traffic phenomena from FCD.

When focusing on traffic congestion, the six presented methods can deliver specific results that can show different aspects of these patterns. The consideration of introducing these six methods with the specific preprocessing steps is to gain knowledge on the general vehicle movements and specific traffic congestion propagation patterns. Yet, it is not possible to detect all aspects of the data, which motivates to extend each method in adaptation or to advance the framework with additional methods.

Additionally, there are five specific objectives of this thesis. The first is testing the calculation of various traffic parameters based on movement representations in different spaces (Euclidean, spatiotemporal, network and feature spaces). There are highly beneficial aspects in introducing the network space into analysis. The reason is that appearances as percolation are only detectable in these spaces. Another benefit from network space is the reasoning on congestion propagation and on how connected streets are affected. Here it is also possible to assign traffic bottlenecks with selected parts of the transportation infrastructure. This thesis introduces a novel method of representing vehicle traffic congestion: congestion propagation polylines (CPP). By extracting only the core points of congestion clusters of a previously adjusted traffic congestion detection (point clustering and convex hull generation), nearby situated events are connectable for consecutive time steps. The connection of the core points is non-trivial. This is due to the querying of vehicle identification of intersection consecutive congestion polygons, which is dependent on the choice of the data time windows and the probability that vehicle drivers are inside successive congestion polygons. The connection of these core points consists of finding the same vehicle IDs in every two consecutive polygons. Otherwise, CPP vary in their length, in case of not finding the same vehicle identifications. This CPP generation is base for various possible modifications. Besides querying identical vehicle identifications, there are numerous other possibilities of connecting or associating congestion polygons. The framework is relatively open to new established rules and restriction in creating congestion propagation polylines. This indicates that different results in matching subsequent congestion polygons will result in differently modelled propagation patterns. In this work, only one CPP strategy has find usage for the first test results. This is the reason why only very few affected taxi drivers are detectable within the propagation patterns. The detected trajectory partitions inside propagation polygons are in many cases useful representations, since they represent real taxi drivers, who get stuck in ongoing traffic congestion events. In this work, different possibilities for inferring and modeling traffic congestion from massive FCD sets are matter of investigation. As a conclusion, several insights are derivable. Within the analysis of this work, there are numerous detectable and interpretable traffic phenomena. A precondition of this analysis and for gaining reasonable results, preprocessing is a large part of the proposed framework. More precisely, the most time-consuming part of the analysis is the preprocessing of the data input. An unsolved problem, in case of guaranteeing higher spatial accuracy is the choice of suitable MM techniques. The previous derivation of congestion events and its subsequent matching delivers results that are in some cases not very reliable. Mismatching, especially at road intersections and elevated roads in urban environments makes it difficult to find the most suitable matching technique. As an additional difficulty, the complex built structures of urban environments cause often multipath effects and signal losses, while positioning in urban canyons. This makes the association of road segments with vehicle trajectories often difficult and in many cases impossible. One benefit for more insights on FCD from urban environments would be to compare the results of the framework application with the results that base on a previous MM step, coming from any MM technique. This is made more complicated due to the frequent appearance of urban canyons in the city center areas. Nevertheless, these appearances are relatively simple to detect, within the preprocessing steps. This information is not only useful for planners, but as well for traffic forecast. The traffic phenomena in urban environments are still difficult to describe and to formalize. Here, the proposed framework might help in understanding selected portions of these phenomena that are dependent on the numerous individual traffic participants.

The second specific objective is very general and consists of detecting and classifying different traffic phenomena. This work shows possibilities of inferring multiple complex vehicle traffic phenomena from massive FCD. In particular, results show views on different forms of traffic congestion. Even though the detection of vehicle traffic congestion implies many difficulties, its differentiation is more complicated. This thesis proposes a selection of established and adapted methods that serve as partial processes of the data analysis workflows. The combination of these methods allow for example distinguishing between recurrent and non-recurrent congestion. The presented traffic pattern analysis framework is applicable with various kinds of FCD, specifically coming from different types of tracked vehicles of different vehicle fleets. Besides gained knowledge on temporal and spatial distribution of detected FCD-based traffic congestion and bottlenecks, it is also important to estimate the detection (prediction) quality of the framework methods and especially of the framework measures. One missing aspect in this work is the inclusion of taxi driving behavior into analysis, as unoccupied taxis reveal specific customer searching patterns. These patterns are possibly not present in tracked vehicles of car sharing or transport vehicle fleets. The typical patterns for all traffic phenomena base on measurable similarity. For the microscopic view on movement this might be, besides the similarity of trajectory shape and direction, also the values inside the trajectory as velocities and acceleration. For macroscopic views on movement, those could be for each segment with temporally varying values of traffic flow rate, traffic density and average velocity. The similarities and dissimilarities of the three mentioned parameters are indicators for traffic phenomena, especially in case of traffic congestion or bottlenecks. All the mentioned are also investigation-area-independent parameters. The exception would be selected global indices that require the total lengths of the road network or capacities that rely on road type and its certain properties.

The third specific objective is reasoning on the quality assessment of results. In general, this has a connection with the quantity and quality of FCD for detecting traffic phenomena and its propagation over time and eventually needed additional information. For detecting recurrent traffic congestion events, the FCD should be coming from a long-term acquisition of at least two weeks. This is the case in nearly all recent scientific approaches. Another precondition is the density of tracked entities. The higher the numbers of tracked participants the more different traffic patterns are inferable. This indicates the fact that not all the traffic participants are observable in selected investigation areas. Therefore, estimating the traffic density is a challenging and complicated task. Consequently, with varying probabilities and dependent on individual driving behaviors, only few patterns are inferable, without exact knowledge on numerous local congestion events. The most important condition is the availability of different attribute information. The more details the data includes, the more differentiations of traffic patterns are possible. One important example of such information is the recorded instantaneous driving direction. This information is helpful for MM, but as well for distinguishing between different events that are in the same time window but spatially on different elevation levels. The needed additional information is the road network and its connectivity. This information would allow calculating accurate traffic congestion propagation patterns based on comparing the connection of affected road segments.

The connection of classified movement patterns with street information is in many cases feasible. One challenge is the matching of these two components in densely built urban environments, because guaranteeing a relatively high spatial accuracy is often not possible due to various reasons. One reason is general property of complex urban road networks of having different nearby situated traffic phenomena, especially at roads covered with additional elevated road networks. Other reasons include properties of FCD sets as signal losses in urban canyons or multipath effects from positioning, which indicate missing FCD for detecting traffic events. All these aspects make it difficult to establish connections between classified movement patterns and road segment information.

Another specific objective concerns the design of a FCD-based technique for estimating the actual real world traffic density. Method six of the framework proposes this kind of design. After testing the technique, it is to mention that no reasonable results are obtainable. This shows that this type of approach is challenging. Nevertheless, other approach could possibly achieve results that are more reasonable. The missing component for delivering this statement is an extensive testing of every component of the approach. One attempt for improving the results of method six is to introduce a machine learning approach for gaining more knowledge of the traffic dynamics. The used test data set is with one month of acquisition a useful source for further investigation.

Contribution

The contribution of this work consists of a traffic pattern analysis framework for FCD that allows, dependent on the quality of the underlying data source, extracting representations of real-world traffic events. The respected quality aspects in this work consist only of the temporal sampling interval of the data and the spatial positioning accuracy. There are six presented methods enabling the detection of traffic congestion events with vehicle trajectories. The three first methods allow the traffic congestion detection without additional road network information. In general, the six methods allow analysts to first detect events where vehicles are participating in a traffic jam or participating in not congested traffic. Extracting movement patterns of taxis might be beneficial to understand the nature of urban traffic.

Applying the proposed methods on vehicle movement data set from different urban investigation areas can be a starting point for extracting local knowledge. This knowledge can include the origins in space of periodically appearing traffic congestion. It might include areas or places with high congestion affection on drivers, together with the classification into static and dynamic bottlenecks. This can help identify specific elements of the urban transportation infrastructure, which influence or are responsible for traffic congestion events.

Other contributions of the framework are alternative definitions and measures for urban traffic congestion that combine definitions and measurement techniques from previous research. One example for such a new definition is the bottleneck identification technique (Keler et al. 2017). As a part of this technique, the proposed traffic congestion detection step consists of the three approaches clustering (Rinzivillo et al. 2008), global indices (Wen et al. 2014) and pre-selections via velocity values (Robinson 1984). This new definition aims at avoiding misclassification of parking and still standing vehicles at road intersections. Nevertheless, this avoidance is still challenging, especially when no information on traffic light positions is available.

Concerning the differentiation between recurrent and non-recurrent traffic congestion, the framework proposes a simple possibility via intersecting congestion polygons. This is an alternative way of distinguishing and its result reliability heavily depends on the size of the selected time windows.

After extracting traffic congestion of selected rush hours, it is possible to estimate the underlying traffic density to a certain degree of reliability. Due to the questionable results, there is less contribution in estimating traffic densities from sparse FCD in this work. This information matching method allows further analyses in connection with state transitions between the traffic flow information on connected road segments.

This work shows a new methodology in matching traffic information on road segments. Comparing to previous research in the field of map matching, there is a reversed workflow of traffic phenomena detection. Usually, the first step is to match FCD records onto the road network and afterwards reason on traffic phenomena detection. This work first defines the traffic phenomena and then does the matching onto road segments, resulting in time-varying information of the road network features. This is challenging, especially in urban environments with complex urban transport systems, as in the case study of Shanghai with its numerous elevated road segments. Consequently, this work contributes not only to the detection of different traffic phenomena in urban environments, but also to the challenge of matching these phenomena to the actually affected road segments.

Limitation of the approach

The proposed framework implies several different limitations concerning data analysis. One limitation is the missing assigning of not preprocessed movement position to the associated real world road partition, namely missing map matching (MM) techniques within the framework. Another disability is finding and extracting individual vehicle driver behaviors via automated methods. The aiming to find general traffic pattern, allows the extraction of selected traffic event details as traffic congestion propagations, but misses information on behavioral patterns of individual movement trajectories. The possibly most important restriction of the framework methods is the dependency on point clustering parameters and any other parameter values for extracting certain point densities. This is important, when extracting traffic congestion events, but as well for distinguishing between standing vehicles at signalized road intersections and reals traffic congestion events. This is still challenging and difficult to evaluate, since a combination of both is also possible, namely traffic congestion events at and around signalized road intersections. When focusing on traffic congestion propagations, there are numerous options for connecting traffic congestion polygons. The used possibility with querying successively appearing overlapping congestion polygons by same vehicle identifications captures only a relatively small portion of ongoing traffic events. Therefore, it is possible that eventually less severe traffic congestion event that propagate faster are not detectable with the present options. Nevertheless, traffic congestion propagation enables to extract unusually severe traffic congestion events depending on the inspected data partitions and time windows. This concludes to the last important cutback of the approach, which consists of the lacking of real-time capability of the proposed techniques. Since all techniques, the clustering as well as estimating global indices and connecting to road segments, have the precondition of having FCD partitions as input for detecting traffic patterns of one specific time window, no real-time traffic information is extractable. Nevertheless, aggregated information can serve as a valuable base for further reasoning, since time windows smaller than 10 minutes might hinder the differentiation of certain vehicle traffic patterns, as standing at road intersections and propagating traffic congestion.

7. Outlook

One direction of future work goes into the evaluation of the created results with local knowledge, in particular from taxi drivers, on the urban dynamics. This can come from questionnaires or any other kind of qualitative data. Another option might be the usage of other sources of traffic information, including data from static sensors as induction loops.

When observing an ever-changing environment of a complex city, it is obvious to detect differences in periodical traffic patterns via data of other periods. There are numerous changes in the road infrastructure of Shanghai from 2007 to 2017, which results in differences in the general traffic flows and the general traffic dynamics. One question in connection to this fact is how local knowledge is changing and which drivers are affected by this change. Another type of evaluation is the rethinking process of the proposed techniques for different data sets of other modes of transport.

Additionally, the traffic pattern analysis framework might help understanding the sequences that are leading to specific traffic congestion events as traffic gridlocks, which describe entire traffic breakdowns at large parts of the whole road network. This information will not only be useful for planners and associated stakeholders; it will help millions of commuters to plan their daily route.

The more general question here is if these products are useful for traffic forecast. Machine learning methods like artificial neural networks might benefit from training congestion propagation polylines (CPP). In the field of data-driven traffic forecast, usually road links are base for associating traffic states. One challenge would be to define different representation forms of urban traffic congestion, such as point clusters, polygons and propagation polylines for training a forecast system. The option for connecting these representations to road links would be representing the affected segments together with the specific road connectivity information.

The presented visualization possibilities for traffic congestion and traffic bottlenecks might be included in mobile applications, for showing the users places and times with high influences by recurrent traffic congestion events.

Another direction of future work is to develop methods for novel geovisualization of traffic. The specific question here is, how to represent the temporal component, which is crucial for everything thematically connected with traffic, in 2D, 2.5D and 3D views. One attempt for further exploration of visualization possibilities is to find novel visualization approaches for processes. These approaches could be fundamentally different from previous, and include attempts of avoiding animation and map view series. The challenge is to define novel geovisualization approaches of processes and events in a static view without animation.

One addition could be the provision of statistically adapted traffic congestion appearance probabilities, together with selected fuzzy boundaries of the affected areas (e.g. via fuzzy polygons). FCD derivatives as congestion polygons and congestion propagation polylines deliver useful information. Its usefulness might be evaluated in future studies and approaches with massive FCD. As an abstraction of different traffic events with different traffic flow parameters, congestion polygons can indicate the influenced road segments. Since the spatial accuracy of these polygons is relatively low, MM might be provided in a naïve way. Further research might compare different MM methods for matching polygons onto road segments. Several options arise for associating detected traffic events with overlapping road segments. Nevertheless, it is challenging to incorporate and to associate elevated transportation infrastructure with traffic events. The proposed techniques might find usage in different potential applications. Applying the six techniques might be an additional preprocessing step before map matching (MM).

In the field of planning, various parts of the applications can benefit from the presented framework, such as reasoning on the abrasion of selected road segments. Another example might benefit transportation planning in the way of observing differences in capacities for different road types and popularity or road segments in connection with centrality measures. The layers of the resulting polygon features might benefit various tasks of spatial analyses and planning.

In future work, the mentioned aspects might be subject of investigation in analyses that make use of the framework. Additionally, further inspections can go into the direction of finding correlations between the framework products, results from other data sources and simulated results. Other data sets might include tracked animal movement trajectories. Applying the framework methods on these data sets might benefit understanding of the framework values. Since, for example, migrating birds meet at selected parts of the world for hunting or breeding, traffic congestion polygons or propagation polylines for these data sets might result as meeting points and specific parts of the habitats. The meeting point propagation patterns might then, eventually, facilitate understanding group behavior, especially the migration patterns, from a different perspective.

The results of the framework method outcomes might be useful information for traffic forecast approaches, as the frequency of congestion appearance is representable. These periodically appearing traffic patterns might be part of the visualization component of future car navigation services. The extracted layers, areal polygons and polylines, might serve as additional layers in vector maps of vehicle navigation software graphical interfaces. In this way, local knowledge on periodically appearing or typical traffic congestion events is representable for tourists without any local knowledge. Following up on this idea, also research on vehicle driving behavior can benefit from the presented framework in the way that it can be included in GIS analyses.

References

Aftabuzzaman, M., 2007. Measuring traffic congestion: a critical review. *In: 30th Australasian Transport Research Forum (ATRF 2007)*, 09 Melbourne, Victoria, Australia, 1-16.

Almeida, V. T. d., Guting, R. H. and Behr, T., 2006. Querying Moving Objects in SECONDO. *In: 7th International Conference on Mobile Data Management (MDM'06)*, 10-12 May 2006, 47-47.

An, S., Yang, H., Wang, J., Cui, N. and Cui, J., 2016. Mining urban recurrent congestion evolution patterns from GPS-equipped vehicle mobility data. *Information Sciences,* 373, 515-526.

Anbaroğlu, B., Cheng, T. and Heydecker, B., 2015. Non-recurrent traffic congestion detection on heterogeneous urban road networks. *Transportmetrica A: Transport Science,* 11(9), 754-771.

Anbaroglu, B., Heydecker, B. and Cheng, T., 2014. Spatio-temporal clustering for non-recurrent traffic congestion detection on urban road networks. *Transportation Research Part C: Emerging Technologies,* 48, 47-65.

Anbaroglu, B., Heydecker, B. and Cheng, T., 2016. How travel demand affects detection of non-recurrent traffic congestion on urban road networks. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLI-B2, 159-164.

Andrew, A. M., 1979. Another Efficient Algorithm for Convex Hulls in Two Dimensions. *Information Processing Letters,* 9, 216-219.

Andrienko, G., Andrienko, N., Bak, P., Keim, D. and Wrobel, S., 2013. *Visual analytics of movement.* Springer Science & Business Media.

Andrienko, G., Andrienko, N. and Fuchs, G., 2016a. Understanding movement data quality. *Journal of Location Based Services,* 10(1), 31-46.

Andrienko, N. and Andrienko, G., 2007. Designing Visual Analytics Methods for Massive Collections of Movement Data. *Cartographica,* 42(2), 117-138.

Andrienko, N. and Andrienko, G., 2011. Spatial Generalization and Aggregation of Massive Movement Data. *IEEE Transactions on Visualization and Computer Graphics,* 17(2), 205-219.

Andrienko, N. and Andrienko, G., 2013. Visual Analytics of Movement: an Overview of Methods, Tools, and Procedures. *Information Visualization,* 12(1), 3-24.

Andrienko, N., Andrienko, G. and Rinzivillo, S., 2015. Exploiting Spatial Abstraction in Predictive Analytics of Vehicle Traffic. *ISPRS International Journal of Geo-Information*, 4(2), 591-606.

Andrienko, N., Andrienko, G. and Rinzivillo, S., 2016b. Leveraging spatial abstraction in traffic analysis and forecasting with visual analytics. *Information Systems*, 57, 172-194.

Ankerst, M., Breunig, M. M., Kriegel, H.-P. and Sander, J., 1999. OPTICS: ordering points to identify the clustering structure. *In: Proceedings of the 1999 ACM SIGMOD international conference on Management of data*, Philadelphia, Pennsylvania, USA, 49-60.

Balasubramanian, L. and Sugumaran, M., 2012. A State-of-Art in R-Tree Variants for Spatial Indexing. *International Journal of Computer Applications,* 42(20), 35-41.

Bao, J., Chow, C. Y., Mokbel, M. F. and Ku, W. S., 2010. Efficient Evaluation of k-Range Nearest Neighbor Queries in Road Networks. *In: 2010 Eleventh International Conference on Mobile Data Management*, 23-26 May 2010, 115-124.

Barthélemy, M., 2011. Spatial networks. *Physics Reports*, 499(1–3), 1-101.

Bazzani, A., Giorgini, B., Gallotti, R., Giovannini, L., Marchioni, M. and Rambaldi, S., 2011. Towards Congestion Detection in Transportation Networks Using GPS Data. *In: 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom)*, 9-11 Oct. 2011, 1455-1459.

Benhamou, S., 2004. On the expected net displacement of animals' random movements. *Ecological Modelling*, 171(1–2), 207-208.

Berglund, S., 2001. *GIS in Transport Modelling.* Dissertation, Royal Institute of Technology Stockholm (Sweden).

Bernstein, D. and Kornhauser, A., 1996. *An introduction to map matching for personal navigation assistants.* TIDE Center, New Jersey.

Bertini, R. L., 2005. Congestion and its extent. *In: Access to destinations.* Emerald Group Publishing Limited, 11-38.

Bertini, R. L., 2006. You Are the Traffic Jam: Examination of Congestion Measures. *In: Transportation Research Board 85th Annual Meeting*.

Bliemer, M. C. J., Raadsen, M. P. H., Brederode, L. J. N., Bell, M. G. H., Wismans, L. J. J. and Smith, M. J., 2016. Genetics of traffic assignment models for strategic transport planning. *Transport Reviews*, 1-23.

Brakatsoulas, S., Pfoser, D., Salas, R. and Wenk, C., 2005. On map-matching vehicle tracking data. *In: Proceedings of the 31st international conference on Very large data bases*, Trondheim, Norway, 853-864.

Brakatsoulas, S., Pfoser, D. and Tryfona, N., 2004. Modeling, Storing, and Mining Moving Object Databases. *In: Proceedings of the International Database Engineering and Applications Symposium*, 68-77.

Brockfeld, E., 2009. *ORINOKO (Operative regionale integrierte und optimierte Korridorsteuerung) : Projektabschlussbericht DLR.* Berlin, Germany: Deutsche Forschungsanstalt für Luft- und Raumfahrt.

Bröring, A., Remke, A., Stasch, C., Autermann, C., Rieke, M. and Möllers, J., 2015. enviroCar: A Citizen Science Platform for Analyzing and Mapping Crowd-Sourced Car Sensor Data. *Transactions in GIS,* 19(3), 362-376.

Castro, P. S., Zhang, D., Chen, C., Li, S. and Pan, G., 2013. From taxi GPS traces to social and community dynamics: A survey. *ACM Computing Surveys (CSUR),* 46(2), 1-34.

Chawla, S., Zheng, Y. and Hu, J., 2012. Inferring the Root Cause in Road Traffic Anomalies. *In: Proceedings of the 2012 IEEE 12th International Conference on Data Mining*, 141-150.

Chen, C., Zhang, D., Samuel Castro, P., Li, N., Sun, L. and Li, S., 2012. Real-Time Detection of Anomalous Taxi Trajectories from GPS Traces. *In:* Puiatti, A. and Gu, T. eds. *Mobile and Ubiquitous Systems: Computing, Networking, and Services: 8th International ICST Conference, MobiQuitous 2011, Copenhagen, Denmark, December 6-9, 2011, Revised Selected Papers.* Berlin: Springer, 63-74.

Chen, W., Guo, F. and Wang, F. Y., 2015. A Survey of Traffic Data Visualization. *IEEE Transactions on Intelligent Transportation Systems,* 16(6), 2970-2984.

Chen, Z., Liu, X. C. and Zhang, G., 2016. Non-recurrent congestion analysis using data-driven spatiotemporal approach for information construction. *Transportation Research Part C: Emerging Technologies,* 71, 19-31.

Cheng, T., Tanaksaranond, G., Brunsdon, C. and Haworth, J., 2013. Exploratory visualisation of congestion evolutions on urban transport networks. *Transportation Research Part C: Emerging Technologies,* 36, 296-306.

Claudel, C. G. and Bayen, A. M., 2008. Guaranteed bounds for traffic flow parameters estimation using mixed Lagrangian-Eulerian sensing. *In: 2008 46th Annual Allerton Conference on Communication, Control, and Computing*, 23-26 Sept. 2008, 636-645.

Cohn, N. and Bischoff, H., 2012. Floating car data for transportation planning. Explorative study to technique and applications and sample properties of GPS data. *In: NATMEC 2012,* Dallas.

Coifman, B., 2003. Identifying the onset of congestion rapidly with existing traffic detectors. *Transportation Research Part A: Policy and Practice*, 37(3), 277-291.

Costa, G., Manco, G. and Masciari, E., 2014. Dealing with trajectory streams by clustering and mathematical transforms. *Journal of Intelligent Information Systems,* 42(1), 155-177.

Cottrell, W., 1993. *The Growth of Recurring Roadway Congestion.* FHWA, U. S. Department of Transportation.

Cucchiara, R., Piccardi, M. and Mello, P., 2000. Image analysis and rule-based reasoning for a traffic monitoring system. *IEEE Transactions on Intelligent Transportation Systems,* 1(2), 119-130.

D'Este, G. M., Zito, R. and Taylor, M. A. P., 1999. Using GPS to Measure Traffic System Performance. *Computer-Aided Civil and Infrastructure Engineering,* 14(4), 255-265.

Daeho, L. and Youngtae, P., 2008. Measurement of traffic parameters in image sequence using spatio-temporal information. *Measurement Science and Technology,* 19(11), 115503.

Daganzo, C. F. and Geroliminis, N., 2008. An analytical approximation for the macroscopic fundamental diagram of urban traffic. *Transportation Research Part B: Methodological,* 42(9), 771-781.

Dalyot, S., Dahinden, T., Schulze, M. J., Boljen, J. and Sester, M., 2013. Integrating network structures of different geometric representations. *Survey Review,* 45(333), 428-440.

Dechenaux, E., Mago, S. D. and Razzolini, L., 2014. Traffic congestion: an experimental study of the Downs-Thomson paradox. *Experimental Economics,* 17(3), 461-487.

DeMers, M. N., 2002. *GIS modeling in raster.* John Wiley & Sons Incorporated.

Demšar, U., Buchin, K., Cagnacci, F., Safi, K., Speckmann, B., Van de Weghe, N., Weiskopf, D. and Weibel, R., 2015. Analysis and visualisation of movement: an interdisciplinary review. *Movement Ecology,* 3(1), 5.

Dewees, D. N., 1978. Simulations of traffic congestion in Toronto. *Transportation Research,* 12(3), 153-161.

Dias, E., Edwardes, A. and Purves, R., 2007. Analysing and aggregating visitor tracks in a protected area. *In: 5th International Symposium on Spatial Data Quality proceedings. Electronic Proceedings. Enschede, ISPRS,* 15-16.

Ding, L., Krisp, J., Meng, L., Xiao, G. and Keler, A., 2016. Visual exploration of multivariate movement events in space-time cube. *In: The 19th AGILE International Conference on Geographic Information Science,* Helsinki, Finland.

Ding, L. and Meng, L., 2014. A comparative study of thematic mapping and scientific visualization. *Annals of GIS,* 20(1), 23-37.

Ding, L., Yang, J. and Meng, L., 2015. Visual Analytics for Understanding Traffic Flows of Transport Hubs from Movement Data. *In: International Cartographic Conference*, Rio de Janeiro, Brazil, 23-28.

Dodge, S., Weibel, R., Ahearn, S. C., Buchin, M. and Miller, J. A., 2016. Analysis of movement data. *International Journal of Geographical Information Science,* 30(5), 825-834.

Dolancic, M., 2016. Automatic lane-level road network graph-generation from Floating Car Data Page. *GI_Forum 2016, Volume 1*, 1, 231-242.

Dowling, R., Skabardonis, A., Carroll, M. and Wang, Z., 2004. Methodology for Measuring Recurrent and Nonrecurrent Traffic Congestion. *Transportation Research Record: Journal of the Transportation Research Board,* 1867, 60-68.

Eddy, W. F., 1977. A New Convex Hull Algorithm for Planar Sets. *ACM Transactions on Mathematical Software (TOMS),* 3(4), 398-403.

Ehmke, J. F., Meisel, S. and Mattfeld, D. C., 2010. Floating Car Data Based Analysis of Urban Travel Times for the Provision of Traffic Quality. *In:* Barceló, J. and Kuwahara, M. eds. *Traffic Data Collection and its Standardization.* New York, NY: Springer, 129-149.

Eisenblätter, B., Santen, L., Schadschneider, A. and Schreckenberg, M., 1998. Jamming transition in a cellular automaton model for traffic flow. *Physical Review E,* 57(2), 1309-1314.

Erdelić, T., Vrbančić, S. and Rošić, L., 2015. A model of speed profiles for urban road networks using G-means clustering. *In: 2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO),* 25-29 May 2015, 1081-1086.

Ertöz, L., Steinbach, M. and Kumar, V., 2003. Finding Clusters of Different Sizes, Shapes, and Densities in Noisy, High Dimensional Data. *In: Third SIAM International Conference on Data Mining (SIAM 2003)*, 47-58.

Ester, M., Kriegel, H.-P., Sander, J. and Xu, X., 1996. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. *In: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD'96),* 226-231.

Fastenrath, U., 1997. Floating Car Data on a larger Scale. *In: ITS-World Congress 1997*, Berlin, Germany.

Fathi, A. and Krumm, J., 2010. Detecting Road Intersections from GPS Traces. *In:* Fabrikant, S. I., Reichenbacher, T., Van Kreveld, M. and Schlieder, C. eds. *Geographic Information Science: 6th International Conference, GIScience 2010, Zurich, Switzerland, September 14-17, 2010. Proceedings.* Berlin: Springer, 56-69.

Fischer, J. W., Walter, W. D. and Avery, M. L., 2013. Brownian Bridge Movement Models to Characterize Birds' Home Ranges. *The Condor,* 115(2), 298-305.

Flynn, M., Kasimov, A., Nave, J.-C., Rosales, R. and Seibold, B., 2009. Self-sustained nonlinear waves in traffic flow. *Physical Review E,* 79(5), 056113.

Galton, A., 2005. Dynamic Collectives and Their Collective Dynamics. *In:* Cohn, A. G. and Mark, D. M. eds. *Spatial Information Theory: International Conference, COSIT 2005, Ellicottville, NY, USA, September 14-18, 2005. Proceedings.* Berlin: Springer, 300-315.

Galton, A., 2008. Processes and Events. *In: Encyclopedia of GIS.* Boston, MA: Springer US, 913-917.

Galton, A. and Duckham, M., 2006. What Is the Region Occupied by a Set of Points? *In:* Raubal, M., Miller, H. J., Frank, A. U. and Goodchild, M. F. eds. *Geographic Information Science: 4th International Conference, GIScience 2006, Münster, Germany, September 20-23, 2006. Proceedings.* Berlin: Springer, 81-98.

Gao, P., Liu, Z., Tian, K. and Liu, G., 2016. Characterizing Traffic Conditions from the Perspective of Spatial-Temporal Heterogeneity. *ISPRS International Journal of Geo-Information,* 5(3), 34.

Goel, A. and Kumar, P., 2015. Characterisation of nanoparticle emissions and exposure at traffic intersections through fast–response mobile and sequential measurements. *Atmospheric Environment,* 107, 374-390.

Goldsberry, K., 2005. Limitations and Potentials of Real-Time Traffic Visualization for Wayfinding. *In: 22nd ICA/ACI International Cartographic Conference,* July 9–16, 2005 A Coruna, Spain.

Goldsberry, K., 2008. GeoVisualization of Automobile Congestion. *In: AGILE 2008 Conference*.

Goodchild, M. F. and Glennon, J. A., 2008. Representation and computation of geographic dynamics. *In:* Hornsby, K. S. and Yuan, M. eds. *Understanding Dynamics of Geographic Domains.* Boca Raton: CRC Press, 13-30.

Graham, R. L., 1972. An efficient algorith for determining the convex hull of a finite planar set. *Information Processing Letters,* 1(4), 132-133.

Graser, A., Ponweiser, W., Dragaschnig, M., N. Brändle and Widhalm, P., 2012. Assessing traffic performance using position density of sparse FCD. *In: 2012 15th International IEEE Conference on Intelligent Transportation Systems*, 16-19 Sept. 2012, 1001-1005.

Graser, A., Straub, M. and Dragaschnig, M., 2015. Is OSM Good Enough for Vehicle Routing? A Study Comparing Street Networks in Vienna. *In:* Gartner, G. and Huang, H. eds. *Progress in Location-Based Services 2014.* Cham: Springer International Publishing, 3-17.

Gschwend, C. and Laube, P., 2012. Challenges of context-aware movement analysis—Lessons learned about crucial data requirements and pre-processing. *In:* Rowlingson, B. and Whyatt, D. eds. *GIS Research UK 20th Annual Conference*, Lancaster, UK, 241-246.

Gschwend, C. and Laube, P., 2014. Scale Effects in Relating Movement to Geographic Context. *In:* Stewart, K., Pebesma, E., Navratil, G., Fogliaroni, P. and Duckham, M. eds. *Extended Abstract Proceedings of the GIScience 2014*, Vienna, Austria, 51-56.

Gudmundsson, J., Laube, P. and Wolle, T., 2012. Computational Movement Analysis. *In:* Kresse, W. and Danko, D. M. eds. *Springer Handbook of Geographic Information.* Berlin, Heidelberg: Springer Berlin Heidelberg, 423-438.

Gühnemann, A., Schäfer, R.-P., Thiessenhusen, K.-U. and Wagner, P., 2004. Monitoring Traffic and Emissions by Floating Car Data. *Institute of Transport Studies Working Paper,* (ITS-WP-04-07).

Guo, H., Wang, Z., Yu, B., Zhao, H. and Yuan, X., 2011. TripVista: Triple Perspective Visual Trajectory Analytics and its application on microscopic traffic data at a road intersection. *In: 2011 IEEE Pacific Visualization Symposium*, 1-4 March 2011, 163-170.

Güting, R. H., Behr, T. and Xu, J., 2010. Efficient k-nearest neighbor search on moving object trajectories. *The VLDB Journal,* 19(5), 687-714.

Guttman, A., 1984. R-trees: a dynamic index structure for spatial searching. *ACM SIGMOD Record,* 14(2), 47-57.

Hallenbeck, M. E., Ishimaru, J. and Nee, J., 2003. *Measurement of recurring versus non-recurring congestion.* Washington State Department of Transportation.

Hamad, K. and Kikuchi, S., 2002. Developing a measure of traffic congestion: fuzzy inference approach. *Transportation Research Record: Journal of the Transportation Research Board,* (1802), 77-85.

He, F., Yan, X., Liu, Y. and Ma, L., 2016. A Traffic Congestion Assessment Method for Urban Road Networks Based on Speed Performance Index. *Procedia Engineering,* 137, 425-433.

Helbing, D. and Treiber, M., 2002. Critical discussion of "synchronized flow". *Cooper@tive Tr@nsport@tion Dyn@mics,* 1, 2.1-2.24.

Hengl, T., van Loon, E. E., Shamoun-Baranes, J. and Bouten, W., 2008. Geostatistical analysis of GPS trajectory data: space-time densities. *In: Proceedings of the 8th international symposium on spatial accuracy assessment in natural resources and environmental sciences*, 17-24.

Hennessy, D. A. and Wiesenthal, D. L., 1999. Traffic congestion, driver stress, and driver aggression. *Aggressive Behavior,* 25(6), 409-423.

Herrera, J. C., Work, D. B., Herring, R., Ban, X., Jacobson, Q. and Bayen, A. M., 2010. Evaluation of traffic data obtained via GPS-enabled mobile phones: The Mobile Century field experiment. *Transportation Research Part C: Emerging Technologies,* 18(4), 568-583.

Huber, W., Ladke, M. and Ogger, R., 1999. Extended floating-car data for the acquisition of traffic information. *In: 6th World Congress on Intelligent Transport Systems (ITS),* November 8-12, 1999 Toronto, Canada.

Immers, L. and Logghe, S., 2002. *Traffic Flow Theory.* Leuven: Katholieke Universiteit Leuven.

Jain, A. K. and Dubes, R. C., 1988. *Algorithms for clustering data.* Prentice-Hall, Inc.

Jain, A. K., Murty, M. N. and Flynn, P. J., 1999. Data clustering: a review. *ACM Computing Surveys,* 31(3), 264-323.

Jarvis, R. A., 1973. On the identification of the convex hull of a finite set of points in the plane. *Information Processing Letters,* 2(1), 18-21.

Jarvis, R. A. and Patrick, E. A., 1973. Clustering using a similarity measure based on shared near neighbors. *IEEE Transactions on computers,* 100(11), 1025-1034.

Jensen, C. S., Lin, D. and Ooi, B. C., 2007. Continuous Clustering of Moving Objects. *IEEE Transactions on Knowledge and Data Engineering,* 19(9), 1161-1174.

Ji, H., Xu, A., Sui, X. and Li, L., 2010. The applied research of Kalman in the dynamic travel time prediction. *In: 2010 18th International Conference on Geoinformatics*, 18-20 June 2010, 1-5.

Ji, Y., Luo, J. and Geroliminis, N., 2014. Empirical Observations of Congestion Propagation and Dynamic Partitioning with Probe Data for Large-Scale Systems. *Transportation Research Record: Journal of the Transportation Research Board,* 2422, 1-11.

Jiang, X. and Jiang, S., 2013. Estimation of Normal Traffic and Traffic Impact Based on Float Car Data. *Journal of Traffic and Logistics Engineering,* 1(1), 1-4.

Jun, J., Guensler, R. and Ogle, J. H., 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.

Kalnis, P., Mamoulis, N. and Bakiras, S., 2005. On Discovering Moving Clusters in Spatio-temporal Data. *In:* Bauzer Medeiros, C., Egenhofer, M. J. and Bertino, E. eds. *Advances in Spatial and Temporal Databases: 9th International Symposium, SSTD 2005, Angra dos Reis, Brazil, August 22-24, 2005. Proceedings.* Berlin: Springer, 364-381.

Kaufman, L. and Rousseeuw, P. J., 1990. *Finding Groups in Data: An Introduction to Cluster Analysis.* John Wiley.

Keim, D. A., Panse, C., Sips, M. and North, S. C., 2004. Visual data mining in large geospatial point sets. *IEEE Computer Graphics and Applications,* 24(5), 36-44.

Keler, A., 2015. Cartographic representations for visual traffic analysis based on floating car data (FCD) of taxis. *gis.Science - Die Zeitschrift fur Geoinformatik,* 2, 51-61.

Keler, A., 2017. Ableitung von Verkehrsengpässen anhand Staudetektion mit Fahrzeugtrajektorien. *In:* Kolbe, T., Bill, R. and Donaubauer, A. eds. *Geoinformationssysteme 2017 – Beiträge zur 4. Münchner GI-Runde.* Berlin/Offenbach: Wichmann, 172-184.

Keler, A., Ding, L. and Krisp, J. M., 2016. Visualization of traffic congestion based on Floating taxi data. *Kartographische Nachrichten,* 66(1), 7-13.

Keler, A., Krisp, J. and Ding, L., 2017a. Detecting vehicle traffic patterns in urban environments using taxi trajectory intersection points. *Geo-spatial Information Science, (in press).*

Keler, A. and Krisp, J. M., 2015a. Spatio-temporal Visualization of Interpolated Particulate Matter (PM2.5) in Beijing. *GI_Forum 2015*, 2015, 464-474.

Keler, A. and Krisp, J. M., 2015b. Visual Analysis of Floating Taxi Data based on selection areas. *In:* Gartner, G. and Huang, H. eds. *1st ICA European Symposium on Cartography (EuroCarto 2015),* Vienna, Austria, 58-59.

Keler, A. and Krisp, J. M., 2016a. Is there a relationship between complicated crossings and frequently visited locations? – A case study with boro taxis and OSM in NYC. *In:* Gartner, G. and Huang, H. eds. *13th International Symposium on Location-Based Services,* Vienna, Austria, 186-190.

Keler, A. and Krisp, J. M., 2016b. Visual Analysis of Floating Taxi Data Based on Interconnected and Timestamped Area Selections. *In:* Gartner, G., Jobst, M. and Huang, H. eds. *Progress in Cartography: EuroCarto 2015.* Cham: Springer International Publishing, 115-131.

Keler, A., Krisp, J. M. and Ding, L., 2017b. Detecting Travel Time Variations in Urban Road Networks by Taxi Trajectory Intersections. *In: GISRUK 2017*, 18-21 April 2017 Manchester, UK.

Keler, A., Krisp, J. M. and Ding, L., 2017c. Visualization of Traffic Bottlenecks: Combining Traffic Congestion with Complicated Crossings. *In:* Peterson, M. P. ed. *Advances in Cartography and GIScience: Selections from the International Cartographic Conference 2017.* Cham: Springer International Publishing, 493-505.

Kerner, B. S., 2004. Introduction. *In: The Physics of Traffic: Empirical Freeway Pattern Features, Engineering Applications, and Theory.* Berlin: Springer, 1-10.

Kerner, B. S., Rehborn, H., Schäfer, R.-P., Klenov, S. L., Palmer, J., Lorkowski, S. and Witte, N., 2013. Traffic dynamics in empirical probe vehicle data studied with three-phase theory: Spatiotemporal reconstruction of traffic phases and generation of jam warning messages. *Physica A: Statistical Mechanics and its Applications,* 392(1), 221-251.

Klein, L., Mills, M. and Gibson, D., 2006. *Traffic Detector Handbook: Third Edition—Volume I.* U.S. Department of Transportation (Federal Highway Administration).

Ko, E., Ahn, J. and Kim, E., 2016. 3D Markov Process for Traffic Flow Prediction in Real-Time. *Sensors,* 16(2), 147.

Körner, M., 2011. Nutzungsmöglichkeiten von Floating Car Data zur Verkehrsflussoptimierung. *In:* Strobl, J., Blaschke, T. and Griesebner, G. eds. *Angewandte Geoinformatik 2011. Beiträge zum 23. AGIT-Symposium Salzburg.*, Berlin/Offenbach: Wichmann, 381-386.

Krieg, S., 2008. *Simulation und Verkehrserfassung: Evaluierung von Floating Car Daten.* Diploma Thesis, Brandenburg University of Applied Sciences.

Kriegel, H.-P., Kunath, P. and Renz, M., 2008. R*-Tree. *In:* Shekhar, S. and Xiong, H. eds. *Encyclopedia of GIS.* Boston, MA: Springer US, 987-992.

Krisp, J. M. and Keler, A., 2015. Car navigation – computing routes that avoid complicated crossings. *International Journal of Geographical Information Science,* 29(11), 1988-2000.

Krisp, J. M. and Peters, S., 2011. Directed kernel density estimation (DKDE) for time series visualization. *Annals of GIS*, 17(3), 155-162.

Krisp, J. M., Peters, S. and Mustafa, M., 2011. Application of an adaptive and directed kernel density estimation (AD-KDE) for the visual analysis of traffic data. *In: GeoViz2011*, Hamburg, Germany, 9-11.

Krisp, J. M., Polous, K., Peters, S., Fan, H. and Meng, L., 2012. Getting in and out of a taxi: spatio-temporal hotspot analysis for floating taxi data in Shanghai. *In: 6th International Symposium "Networks for Mobility"*, September 27/28, 2012.

Kühnel, C., 2012. *Verkehrsdatenerfassung mittels Floating Car Observer auf zweistreifigen Landstraßen.* kassel university press GmbH.

Kurz, F., Rosenbaum, D., Runge, H. and Reinartz, P., 2015. Validation of advanced driver assistance systems by airborne optical imagery. *In: Mobil.TUM 2015 - International Scientific Conference on Mobility and Transport*, 30. Juni - 1. Juli 2015 Munich, Germany, 1-11.

Kwan, M.-P. and Neutens, T., 2014. Space-time research in GIScience. *International Journal of Geographical Information Science,* 28(5), 851-854.

Kwon, J., Mauch, M. and Varaiya, P., 2006. Components of congestion: Delay from incidents, special events, lane closures, weather, potential ramp metering gain, and excess demand. *Transportation Research Record: Journal of the Transportation Research Board,* (1959), 84-91.

Lan, J., Long, C., Wong, R. C.-W., Chen, Y., Fu, Y., Guo, D., Liu, S., Ge, Y., Zhou, Y. and Li, J., 2014. A New Framework for Traffic Anomaly Detection. *In: Proceedings of the 2014 SIAM International Conference on Data Mining.* 875-883.

Langley, R. B., 1999. Dilution of precision. *GPS world*, 10(5), 52-59.

Laube, P., 2001. A classification of analysis methods for dynamic point objects in environmental GIS. *In:* Konecny, M. ed. *4th AGILE Conference,* Brno, Czech Republic, 121–134.

Laube, P., 2014. *Computational movement analysis.* Berlin: Springer International Publishing.

Laube, P., Dennis, T., Forer, P. and Walker, M., 2007. Movement beyond the snapshot–Dynamic analysis of geospatial lifelines. *Computers, Environment and Urban Systems,* 31(5), 481-501.

Laube, P. and Purves, R. S., 2011. How fast is a cow? Cross-Scale Analysis of Movement Data. *Transactions in GIS,* 15(3), 401-418.

Lazarus, R. S., 1966. *Psychological stress and the coping process.* New York, NY, US: McGraw-Hill.

Leduc, G., 2008. Road Traffic Data: Collection Methods and Applications. *In: Working Papers on Energy, Transport and Climate Change N.1*.

Lee, J.-G., Han, J., Li, X. and Gonzalez, H., 2008. TraClass: trajectory classification using hierarchical region-based and trajectory-based clustering. *Proceedings of the VLDB Endowment,* 1(1), 1081-1094.

Lee, J.-G., Han, J. and Whang, K.-Y., 2007. Trajectory clustering: a partition-and-group framework. *In: Proceedings of the 2007 ACM SIGMOD international conference on Management of data*, 593-604.

Leonardi, L., Marketos, G., Frentzos, E., Giatrakos, N., Orlando, S., Pelekis, N., Raffaet, A., Roncato, A., Silvestri, C. and Theodoridis, Y., 2010. T-Warehouse: Visual OLAP analysis on trajectory data. *In: 2010 IEEE 26th International Conference on Data Engineering (ICDE 2010)*, 1-6 March 2010, 1141-1144.

Li, B., Zhang, D., Sun, L., Chen, C., Li, S., Qi, G. and Yang, Q., 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)*, 21-25 March 2011, 63-68.

Li, D., Fu, B., Wang, Y., Lu, G., Berezin, Y., Stanley, H. E. and Havlin, S., 2015a. Percolation transition in dynamical traffic network with evolving critical bottlenecks. *Proceedings of the National Academy of Sciences*, 112(3), 669-672.

Li, H., Kulik, L. and Ramamohanarao, K., 2015b. Robust inferences of travel paths from GPS trajectories. *International Journal of Geographical Information Science*, 29(12), 2194-2222.

Li, J. and Zuo, L., 2004. Shanghai urban elevated roads and metropolis traffic. *Urban Roads Bridges & Flood Control,* 21(6), 14-17.

Li, L. and Revesz, P., 2004. Interpolation methods for spatio-temporal geographic data. *Computers, Environment and Urban Systems,* 28(3), 201-227.

Li, Q., Fan, H., Luan, X., Yang, B. and Liu, L., 2014. Polygon-based approach for extracting multilane roads from OpenStreetMap urban road networks. *International Journal of Geographical Information Science,* 28(11), 2200-2219.

Li, X., Han, J., Lee, J.-G. and Gonzalez, H., 2007a. Traffic Density-Based Discovery of Hot Routes in Road Networks. *In:* Papadias, D., Zhang, D. and Kollios, G. eds. *Advances in Spatial and Temporal Databases: 10th International Symposium, SSTD 2007, Boston, MA, USA, July 16-18, 2007. Proceedings.* Berlin: Springer, 441-459.

Li, X., Li, M., Shu, W. and Wu, M., 2007b. A practical map-matching algorithm for GPS-based vehicular networks in Shanghai urban area. *In: IET Conference on Wireless, Mobile and Sensor Networks 2007 (CCWMSN'07)*, 12-14 Dec. 2007, 454-457.

Lin, W.-H. and Daganzo, C. F., 1997. A simple detection scheme for delay-inducing freeway incidents. *Transportation Research Part A: Policy and Practice,* 31(2), 141-155.

Lindley, J. A., 1986. *Quantification of urban freeway congestion and analysis of remedial measures.* U.S. Department of Transportation, United States Federal Highway Administration, Traffic Systems Division.

Listl, G., 2003. *Anwendung neuer Technologien zur Erfassung des Verkehrsablaufs.* Kasseler Dissertation PhD, Universität Kassel.

Liu, C., Li, N., Huang, M. and Wu, H., 2012a. City Routing Velocity Estimation Model under the Environment of Lack of Floating Car Data. *Journal of Geographic Information System,* 4(1), 55-61.

Liu, C., Meng, X. and Fan, Y., 2008. Determination of Routing Velocity with GPS Floating Car Data and WebGIS-Based Instantaneous Traffic Information Dissemination. *The Journal of Navigation,* 61(02), 337-353.

Liu, G., Gao, P. and Li, Y., 2016a. Transport Capacity Limit of Urban Street Networks. *Transactions in GIS,* 21(3), 575–590.

Liu, L., Andris, C., Biderman, A. and Ratti, C., 2009. Uncovering Taxi Driver's Mobility Intelligence through His Trace. *IEEE Pervasive Computing,* 160, 1-17.

Liu, S., Liu, C., Luo, Q., Ni, L. M. and Krishnan, R., 2012b. Calibrating Large Scale Vehicle Trajectory Data. *In: Proceedings of the 2012 IEEE 13th International Conference on Mobile Data Management (mdm 2012),* 222-231.

Liu, X. and Ban, Y., 2013. Uncovering Spatio-Temporal Cluster Patterns Using Massive Floating Car Data. *ISPRS International Journal of Geo-Information,* 2(2), 371.

Liu, X., Biagioni, J., Eriksson, J., Wang, Y., Forman, G. and Zhu, Y., 2012c. Mining large-scale, sparse GPS traces for map inference: comparison of approaches. *In: Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining,* Beijing, China, 669-677.

Liu, X., Liu, K., Li, M. and Lu, F., 2016b. A ST-CRF Map-Matching Method for Low-Frequency Floating Car Data. *IEEE Transactions on Intelligent Transportation Systems,* PP(99), 1-14.

Liu, X., Lu, F., Zhang, H. and Qiu, P., 2012d. Estimating Beijing's travel delays at intersections with floating car data. *In: Proceedings of the 5th ACM SIGSPATIAL*

International Workshop on Computational Transportation Science, Redondo Beach, California, 14-19.

Liu, X., Lu, F., Zhang, H. and Qiu, P., 2013. Intersection delay estimation from floating car data via principal curves: a case study on Beijing's road network. *Frontiers of Earth Science,* 7(2), 206-216.

Lomax, T., Turner, S., Shunk, G., Levinson, H., Pratt, R., Bay, P. and Douglas, G., 1997. *Quantifying Congestion. Volume 2: User's Guide.* TRB, National Research Council, Washington, D.C., 0309060710.

Long, J. A. and Nelson, T. A., 2013. A review of quantitative methods for movement data. *International Journal of Geographical Information Science,* 27(2), 292-318.

Long, J. C., Gao, Z. Y., Ren, H. L. and Lian, A. P., 2008. Urban traffic congestion propagation and bottleneck identification. *Science in China Series F: Information Sciences,* 51(7), 948-964.

Long, Z., Duckham, M., Li, S. and Schockaert, S., 2016. Indexing large geographic datasets with compact qualitative representation. *International Journal of Geographical Information Science,* 30(6), 1072-1094.

Lorenz, M. and Elefteriadou, L., 2001. Defining freeway capacity as function of breakdown probability. *Transportation Research Record: Journal of the Transportation Research Board,* (1776), 43-51.

Lorkowski, S., Brockfeld, E., Mieth, P., Passfeld, B., Thiessenhusen, K.-U. and Schäfer, R.-P., 2003. Erste Mobilitätsdienste auf Basis von "Floating Car Data". *In: Tagungsband zum 4. Aachener Kollouium "Mobilität und Stadt"*, 93-100.

Lund, A. and Pack, M., 2010. Dynamic wide-area congestion and incident monitoring using probe data. *Transportation Research Record: Journal of the Transportation Research Board,* (2174), 1-9.

Manolopoulos, Y., Nanopoulos, A., Papadopoulos, A. N. and Theodoridis, Y., 2010. *R-trees: Theory and Applications.* Springer Science & Business Media.

Marshall, S., 2005. *Streets and patterns.* New York: Spon.

Mazimpaka, J. and Timpf, S., 2017. How They Move Reveals What Is Happening: Understanding the Dynamics of Big Events from Human Mobility Pattern. *ISPRS International Journal of Geo-Information,* 6(1), 15.

Mazimpaka, J. D. and Timpf, S., 2016a. Trajectory data mining: A review of methods and applications. *Journal of Spatial Information Science,* 2016(13), 61-99.

Mazimpaka, J. D. and Timpf, S., 2016b. A visual and computational analysis approach for exploring significant locations and time periods along a bus route. *In: Proceedings of the 9th ACM SIGSPATIAL International Workshop on Computational Transportation Science,* Burlingame, California, 43-48.

Michalopoulos, P. G., 1991. Vehicle detection video through image processing: the Autoscope system. *IEEE Transactions on Vehicular Technology,* 40(1), 21-29.

Miller, H. J. and Shaw, S.-L., 2015. Geographic Information Systems for Transportation in the 21st Century. *Geography Compass*, 9(4), 180-189.

Miller, H. J., Wu, Y.-H. and Hung, M.-C., 1999. GIS-based dynamic traffic congestion modeling to support time-critical logistics. *In: 32nd Annual Hawaii International Conference on Systems Sciences, (HICSS-32, 1999),* p. 9.

Moreira, A., Santos, M. Y., Wachowicz, M. and Orellana, D., 2010. The Impact of Data Quality in the Context of Pedestrian Movement Analysis. *In:* Painho, M., Santos, Y. M. and Pundt, H. eds. *Geospatial Thinking.* Berlin: Springer, 61-78.

Muckell, J., Olsen, P. W., Hwang, J.-H., Lawson, C. T. and Ravi, S. S., 2014. Compression of trajectory data: a comprehensive evaluation and new approach. *GeoInformatica,* 18(3), 435-460.

Nagel, K. and Herrmann, H. J., 1993. Deterministic models for traffic jams. *Physica A: Statistical Mechanics and its Applications,* 199(2), 254-269.

Nagel, K. and Schreckenberg, M., 1992. A cellular automaton model for freeway traffic. *Journal de physique I,* 2(12), 2221-2229.

Nassar, S., Sander, J. and Cheng, C., 2004. Incremental and effective data summarization for dynamic hierarchical clustering. *In: Proceedings of the 2004 ACM SIGMOD international conference on Management of data*, Paris, France, 467-478.

Ng, R. T. and Han, J., 1994. Efficient and Effective Clustering Methods for Spatial Data Mining. *In: Proceedings of the 20th International Conference on Very Large Data Bases*, 144-155.

Noland, R. B. and Polak, J. W., 2002. Travel time variability: A review of theoretical and empirical issues. *Transport Reviews,* 22(1), 39-54.

O'Sullivan, D. and Unwin, D. J., 2010. Geographic Information Analysis and Spatial Data. *In: Geographic Information Analysis.* John Wiley & Sons, Inc., 1-32.

Oguchi, T., 2015. Traffic engineering. *In: Traffic and Safety Sciences—Interdisciplinary Wisdom of IATSS.* International Association of Traffic and Safety Sciences (IATSS), 40-48.

Okabe, A., Satoh, T. and Sugihara, K., 2009. A kernel density estimation method for networks, its computational method and a GIS-based tool. *International Journal of Geographical Information Science,* 23(1), 7-32.

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

Pan, S., Jiang, B., Zou, N. and Jia, L., 2011. Average travel speed estimation using multi-type floating car data. *In: 2011 IEEE International Conference on Information and Automation*, 6-8 June 2011, 192-197.

Pang, L. X., Chawla, S., Liu, W. and Zheng, Y., 2011. On mining anomalous patterns in road traffic streams. *In: Proceedings of the 7th international conference on Advanced Data Mining and Applications - Volume Part II,* Beijing, China, 237-251.

Pang, L. X., Chawla, S., Liu, W. and Zheng, Y., 2013. On detection of emerging anomalous traffic patterns using GPS data. *Data & Knowledge Engineering,* 87, 357-373.

Parlante, N., 2001. *Binary Trees.* Stanford University: CS Education Library.

Patel, N. and Mukherjee, A. B., 2015. Assessment of network traffic congestion through Traffic Congestability Value (TCV): a new index. *Bulletin of Geography. Socio–economic Series,* 30(30), 123-134.

Patire, A. D., Wright, M., Prodhomme, B. and Bayen, A. M., 2015. How much GPS data do we need? *Transportation Research Part C: Emerging Technologies,* 58, Part B, 325-342.

Pelekis, N. and Theodoridis, Y., 2006. Boosting location-based services with a moving object database engine. *In: Proceedings of the 5th ACM international workshop on Data engineering for wireless and mobile access,* Chicago, Illinois, USA, 3-10.

Pereira, F. C., Costa, H. and Pereira, N. M., 2009. An off-line map-matching algorithm for incomplete map databases. *European Transport Research Review,* 1(3), 107-124.

Peters, S., 2013. Quadtree- and octree-based approach for point data selection in 2D or 3D. *Annals of GIS,* 19(1), 37-44.

Polus, A., 1996. Evaluation of congestion trends on chicago expressways. *Journal of Advanced Transportation,* 30(3), 79-100.

Polus, A. and Pollatschek, M. A., 2002. Stochastic nature of freeway capacity and its estimation. *Canadian Journal of Civil Engineering,* 29(6), 842-852.

Protschky, V., Feit, S. and Linnhoff-Popien, C., 2015a. On the Potential of Floating Car Data for Traffic Light Signal Reconstruction. *In: 2015 IEEE 81st Vehicular Technology Conference (VTC Spring)*, 11-14 May 2015, 1-5.

Protschky, V., Ruhhammer, C. and Feit, S., 2015b. Learning Traffic Light Parameters with Floating Car Data. *In: 2015 IEEE 18th International Conference on Intelligent Transportation Systems*, 15-18 Sept. 2015, 2438-2443.

Purves, R. S., Laube, P., Buchin, M. and Speckmann, B., 2014. Moving beyond the point: An agenda for research in movement analysis with real data. *Computers, Environment and Urban Systems,* 47, 1-4.

Qi, H., Liu, M., Wang, D. and Chen, M., 2016. Spatial-Temporal Congestion Identification Based on Time Series Similarity Considering Missing Data. *PLoS ONE*, 11(9), e0162043.

Qian, Y., Zhong, Y. and Yuan, J., 2006. Investigating traffic congestion in Shanghai: a system dynamics approach. *International Journal of Services Operations and Informatics,* 1(4), 363-379.

Qiu, T., Lu, X.-Y., Chow, A. and Shladover, S., 2010. Estimation of Freeway Traffic Density with Loop Detector and Probe Vehicle Data. *Transportation Research Record: Journal of the Transportation Research Board,* 2178, 21-29.

Quddus, M. and Washington, S., 2015. Shortest path and vehicle trajectory aided map-matching for low frequency GPS data. *Transportation Research Part C: Emerging Technologies,* 55, 328-339.

Quddus, M. A., Noland, R. B. and Ochieng, W. Y., 2006. A High Accuracy Fuzzy Logic Based Map Matching Algorithm for Road Transport. *Journal of Intelligent Transportation Systems,* 10(3), 103-115.

Quddus, M. A., Ochieng, W. Y. and Noland, R. B., 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.

Ramezani, M. and Geroliminis, N., 2014. Queue profile estimation in congested urban networks with probe data. *Computer-Aided Civil and Infrastructure Engineering,* 30(5), 414-432.

Ranacher, P., Brunauer, R., Trutschnig, W., Van der Spek, S. and Reich, S., 2016a. Why GPS makes distances bigger than they are. *International Journal of Geographical Information Science,* 30(2), 316-333.

Ranacher, P., Brunauer, R., van der Spek, S. and Reich, S., 2016b. What is an Appropriate Temporal Sampling Rate to Record Floating Car Data with a GPS? *ISPRS International Journal of Geo-Information,* 5(1), 1.

Ranacher, P. and Tzavella, K., 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.

Rao, A. M. and Rao, K. R., 2012. Measuring Urban Traffic Congestion-A Review. *International Journal for Traffic and Transport Engineering,* 2(4), 286-305.

Raper, J. and Livingstone, D., 2001. Let's get real: spatio-temporal identity and geographic entities. *Transactions of the Institute of British Geographers*, 26(2), 237-242.

Reinthaler, M., Nowotny, B., Hildebrandt, R. and Weichenmeier, F., 2007. Evaluation of speed estimation by floating car data within the research project Dmotion. ITS America, 1336–1342.

Rempe, F., Huber, G. and Bogenberger, K., 2016. Spatio-Temporal Congestion Patterns in Urban Traffic Networks. *Transportation Research Procedia,* 15, 513-524.

Rinzivillo, S., Pedreschi, D., Nanni, M., Giannotti, F., Andrienko, N. and Andrienko, G., 2008. Visually driven analysis of movement data by progressive clustering. *Information Visualization*, 7(3-4), 225-239.

Robinson, R., 1984. *Problems in the urban environment: traffic congestion and its effects.* Department of Geography, University of Wollongong, NSW, Australia.

Rohani, M., Gingras, D. and Gruyer, D., 2016. A Novel Approach for Improved Vehicular Positioning Using Cooperative Map Matching and Dynamic Base Station DGPS Concept. *IEEE Transactions on Intelligent Transportation Systems,* 17(1), 230-239.

Sakr, M. A. and Güting, R. H., 2011. Spatiotemporal pattern queries. *GeoInformatica,* 15(3), 497-540.

Salamat, N. and Zahzah, E. h., 2010. Fuzzy Spatio-temporal Relations Analysis. *In: 2010 Seventh International Conference on Information Technology: New Generations*, 12-14 April 2010, 301-306.

Sansone, F., Ti Taranto, C., Kramer, F., Van Essen, R. and Zijderhand, F., 2002. Using extended floating car data to improve safety on the road. *In: e-safety Congress and exhibition proceedings: IT solutions for safety and security in intelligent transport*, 16-18 September 2002 Lyon, France, 10p.

Schrank, D., Turner, S. M. and Lomax, T., 1993. *Estimates of Urban Roadway Congestion—1990.* Texas Transportation Institute, College Station.

Schuessler, N. and Axhausen, K. W., 2009. Processing Raw Data from Global Positioning Systems Without Additional Information. *Transportation Research Record: Journal of the Transportation Research Board,* 2105(2105), 28-36.

Shah, G. H., Bhensdadia, C. K. and Ganatra, A. P., 2012. An Empirical Evaluation of Density-Based Clustering Techniques. *International Journal of Soft Computing and Engineering (IJSCE),* 2(1), 216-223.

Shang, J., Zheng, Y., Tong, W., Chang, E. and Yu, Y., 2014. Inferring gas consumption and pollution emission of vehicles throughout a city. *In: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining,* New York, USA, 1027-1036.

Shekhar, S., Feiner, S. K. and Aref, W. G., 2015. Spatial computing. *Communications of the ACM,* 59(1), 72-81.

Shen, J. and Xue, G., 2010. Trace Interpolation Algorithm Based on Intersection Vehicle Movement Modeling. *Wireless Sensor Network*, 2(11), 823-827.

Shepherd, I. D. H., 2008. Travails in the Third Dimension: A Critical Evaluation of Three-Dimensional Geographical Visualization. *In: Geographic Visualization.* John Wiley & Sons, Ltd, 199-222.

Siła-Nowicka, K., Vandrol, J., Oshan, T., Long, J. A., Demšar, U. and Fotheringham, A. S., 2016. Analysis of human mobility patterns from GPS trajectories and contextual information. *International Journal of Geographical Information Science,* 30(5), 881-906.

Skabardonis, A., Varaiya, P. and Petty, K., 2003. Measuring recurrent and nonrecurrent traffic congestion. *Transportation Research Record: Journal of the Transportation Research Board,* (1856), 118-124.

Slocum, T. A., 2009. *Thematic Cartography and Geovisualization.* Pearson Prentice Hall.

Sofer, T., Polus, A. and Bekhor, S., 2012. A new, comprehensive, congestion measure for freeways. *Advances in Transportation Studies - An International Journal,* 2012(Special Issue 1), 45-58.

Sohr, A., Brockfeld, E. and Krieg, S., 2010. Quality of Floating Car Data. *In: 12th World Conference on Transport Research (WCTR),* Lissabon, Portugal, paper nr 02392.

Souto, G. and Liebig, T., 2016. On Event Detection from Spatial Time Series for Urban Traffic Applications. *In:* Michaelis, S., Piatkowski, N. and Stolpe, M. eds. *Solving Large Scale Learning Tasks. Challenges and Algorithms: Essays Dedicated to Katharina Morik on the Occasion of Her 60th Birthday.* Cham: Springer International Publishing, 221-233.

Spaccapietra, S., Parent, C., Damiani, M. L., de Macedo, J. A., Porto, F. and Vangenot, C., 2008. A conceptual view on trajectories. *Data & Knowledge Engineering,* 65(1), 126-146.

Stanica, R., Fiore, M. and Malandrino, F., 2013. Offloading Floating Car Data. *In: 14th International Symposium and Workshops on a World of Wireless, Mobile and Multimedia Networks (WoWMoM),* 1-9.

Sugiyama, Y., Fukui, M., Kikuchi, M., Hasebe, K., Nakayama, A., Nishinari, K., Tadaki, S.-i. and Yukawa, S., 2008. Traffic jams without bottlenecks—experimental evidence for the physical mechanism of the formation of a jam. *New Journal of Physics,* 10(3), 033001.

Sun, J., Wen, H., Gao, Y. and Hu, Z., 2009. Metropolitan Congestion Performance Measures Based on Mass Floating Car Data. *In:* Yu, L., Lai, K. and Mishra, S. eds. *Second International Joint Conference on Computational Sciences and Optimization (CSO 2009),* 109-113.

Tachet, R., Santi, P., Sobolevsky, S., Reyes-Castro, L. I., Frazzoli, E., Helbing, D. and Ratti, C., 2016. Revisiting Street Intersections Using Slot-Based Systems. *PLoS ONE,* 11(3), e0149607.

Takashi, N., 2002. The physics of traffic jams. *Reports on Progress in Physics,* 65(9), 1331.

Tang, J., Liu, F., Wang, Y. and Wang, H., 2015. Uncovering urban human mobility from large scale taxi GPS data. *Physica A: Statistical Mechanics and its Applications,* 438, 140-153.

Tang, L., Kan, Z., Zhang, X., Sun, F., Yang, X. and Li, Q., 2016a. A network Kernel Density Estimation for linear features in space–time analysis of big trace data. *International Journal of Geographical Information Science,* 30(9), 1717-1737.

Tang, L., Kan, Z., Zhang, X., Yang, X., Huang, F. and Li, Q., 2016b. Travel time estimation at intersections based on low-frequency spatial-temporal GPS trajectory big data. *Cartography and Geographic Information Science,* 43(5), 417-426.

Taylor, M., 1992. Exploring the nature of urban traffic congestion: concepts, parameters, theories and models. *In: Proceedings of the 16th ARRB Conference*, 9-13 November 1992 Perth, Western Australia.

Tettamanti, T., Horváth Márton, T. and Varga, I., 2014. Road Traffic Measurement and Related Data Fusion Methodology for Traffic Estimation. *Transport and Telecommunication Journal,* 15(4), 269-279.

Tiwari, V. S., Arya, A. and Chaturvedi, S., 2014. Framework for Horizontal Scaling of Map Matching: Using Map-Reduce. *In: 2014 International Conference on Information Technology*, 22-24 Dec. 2014, 30-34.

Tominski, C., Schumann, H., Andrienko, G. and Andrienko, N., 2012. Stacking-Based Visualization of Trajectory Attribute Data. *IEEE Transactions on Visualization and Computer Graphics (Proceedings IEEE Information Visualization 2012),* 18(12), 2565-2674.

Tomlin, D. C., 1990. *Geographic information systems and cartographic modeling.* Englewood Cliffs: Prentice Hall.

Treiber, M. and Kesting, A., 2013. Trajectory and Floating-Car Data. *In: Traffic Flow Dynamics: Data, Models and Simulation.* Berlin: Springer, 7-12.

Turner, S., Lomax, T. and Levinson, H., 1996. Measuring and estimating congestion using travel time-based procedures. *Transportation Research Record: Journal of the Transportation Research Board,* (1564), 11-19.

Ulm, M., Heilmann, B., Asamer, J., Graser, A. and Ponweiser, W., 2015. Identifying Congestion Patterns in Urban Road Networks Using Floating Car Data. *In: Transportation Research Board 94th Annual Meeting*.

Underwood, R. T., 1968. Acceleration noise and traffic congestion. *Traffic Engineering & Control,* 9(7), 120-130.

Van de Weghe, N., Cohn, A., De Tré, G. and De Maeyer, P., 2006. A qualitative trajectory calculus as a basis for representing moving objects in Geographical Information Systems. *Control and Cybernetics,* 35 (1), 97-119.

Varaiya, P., 2007. Finding and Analyzing True Effect of Non-Recurrent Congestion on Mobility and Safety. *PATH research report*, 2007(TO-5321), 1-23.

Wang, D.-z. and Chen, Y.-y., 2013. Floating Car Data-Based Algorithm for Calculating Vehicle Queue Length at Intersections. *Journal of Highway and Transportation Research and Development (English Edition),* 7(3), 84-89.

Wang, J. and Tsapakis, I., 2010. *Impact Analysis of Speeds and Space-Time Analysis for the Transportation Network in London*, online available: http://standard.cege.ucl.ac.uk/workshops/jiaqiu.html, accessed: 16.01.2014.

Wang, P., Hunter, T., Bayen, A. M., Schechtner, K. and González, M. C., 2012. Understanding Road Usage Patterns in Urban Areas. *Scientific Reports,* 2, 1001.

Wang, Y., Liu, X., Wei, H., Forman, G. and Zhu, Y., 2013a. CrowdAtlas: self-updating maps for cloud and personal use. *In: Proceeding of the 11th annual international conference on Mobile systems, applications, and services*, Taipei, Taiwan, 469-470.

Wang, Z., Lu, M., Yuan, X., Zhang, J. and Wetering, H. v. d., 2013b. Visual Traffic Jam Analysis Based on Trajectory Data. *IEEE Transactions on Visualization and Computer Graphics,* 19(12), 2159-2168.

Wang, Z., Ye, T., Lu, M., Yuan, X., Qu, H., Yuan, J. and Wu, Q., 2014. Visual Exploration of Sparse Traffic Trajectory Data. *IEEE Transactions on Visualization and Computer Graphics,* 20(12), 1813-1822.

Wang, Z. and Zlatanova, S., 2013. An A*-Based Search Approach for Navigation Among Moving Obstacles. *In:* Zlatanova, S., Peters, R., Dilo, A. and Scholten, H. eds.

Intelligent Systems for Crisis Management: Geo-information for Disaster Management (Gi4DM) 2012. Berlin: Springer, 17-30.

Wardrop, J. G. and Charlesworth, G., 1954. A method of estimating speed and flow of traffic from a moving vehicle. *Proceedings of the Institution of Civil Engineers,* 3(1), 158-171.

Wen, H., Sun, J. and Zhang, X., 2014. Study on Traffic Congestion Patterns of Large City in China Taking Beijing as an Example. *Procedia - Social and Behavioral Sciences,* 138, 482-491.

Weng, J., Wang, C., Huang, H., Wang, Y. and Zhang, L., 2016. Real-time bus travel speed estimation model based on bus GPS data. *Advances in Mechanical Engineering,* 8(11), 1-10.

White, C., Bernstein, D. and Kornhauser, A. L., 2000. Some map matching algorithms for personal navigation assistants. *Transportation Research Part C: Emerging Technologies,* 8(1-6), 91-108.

Worboys, M. F. and Duckham, M., 2004. *GIS: a computing perspective.* New York: CRC press.

Xie, X., Liao, W., Aghajan, H., Veelaert, P. and Philips, W., 2017. Detecting Road Intersections from GPS Traces Using Longest Common Subsequence Algorithm. *ISPRS International Journal of Geo-Information,* 6(1), 1.

Xie, Z. and Yan, J., 2008. Kernel Density Estimation of traffic accidents in a network space. *Computers, Environment and Urban Systems,* 32(5), 396-406.

Xu, X., Li, X., Hu, Y. and Peng, Z., 2012. A novel algorithm to identifying vehicle travel path in elevated road area based on GPS trajectory data. *Frontiers of Earth Science,* 6(4), 354-363.

Yang, F., Wang, S., Li, J., Liu, Z. and Sun, Q., 2014. An overview of Internet of Vehicles. *China Communications,* 11(10), 1-15.

Yang, J. and Meng, L., 2015. Feature Selection in Conditional Random Fields for Map Matching of GPS Trajectories. *In:* Gartner, G. and Huang, H. eds. *Progress in Location-Based Services 2014.* Cham: Springer International Publishing, 121-135.

Yang, Q., Wang, J., Song, X., Kong, X., Xu, Z. and Zhang, B., 2015. Urban Traffic Congestion Prediction Using Floating Car Trajectory Data. *In:* Wang, G., Zomaya, A., Martinez Perez, G. and Li, K. eds. *Algorithms and Architectures for Parallel Processing: 15th International Conference, ICA3PP 2015, Zhangjiajie, China, November 18-20, 2015, Proceedings, Part II.* Cham: Springer International Publishing, 18-30.

Yang, S., Liu, Z., Li, J., Wang, S. and Yang, F., 2016. Anomaly Detection for Internet of Vehicles: A Trust Management Scheme with Affinity Propagation. *Mobile Information Systems*, 2016(5254141), p. 10.

Ye, Y., Zheng, Y., Chen, Y., Feng, J. and Xie, X., 2009. Mining Individual Life Pattern Based on Location History. *In: 2009 Tenth International Conference on Mobile Data Management: Systems, Services and Middleware*, 18-20 May 2009, 1-10.

Yuan, Q., Liu, Z., Li, J., Zhang, J. and Yang, F., 2014. A traffic congestion detection and information dissemination scheme for urban expressways using vehicular networks. *Transportation Research Part C: Emerging Technologies,* 47, Part 2, 114-127.

Yue, Y., Zhuang, Y., Li, Q. and Mao, Q., 2009. Mining time-dependent attractive areas and movement patterns from taxi trajectory data. *In: 2009 17th International Conference on Geoinformatics,* 12-14 Aug. 2009, 1-6.

Zeng, Z., Zhang, T., Li, Q., Wu, Z., Zou, H. and Gao, C., 2016. Curvedness feature constrained map matching for low-frequency probe vehicle data. *International Journal of Geographical Information Science,* 30(4), 660-690.

Zhang, D.-Z., Sun, D. J. and Peng, Z.-R., 2014. A Comprehensive Taxi Assessment Index Using Floating Car Data. *Journal of Harbin Institute Of Technology (New Seriers),* 21(1), 7-16.

Zhang, D., Li, N., Zhou, Z.-H., Chen, C., Sun, L. and Li, S., 2011a. iBAT: detecting anomalous taxi trajectories from GPS traces. *In: Proceedings of the 13th international conference on Ubiquitous computing*, Beijing, China, 99-108.

Zhang, F., Zhu, X., Hu, T., Guo, W., Chen, C. and Liu, L., 2016a. Urban Link Travel Time Prediction Based on a Gradient Boosting Method Considering Spatiotemporal Correlations. *ISPRS International Journal of Geo-Information,* 5(11), 201.

Zhang, J., Wang, F. Y., Wang, K., Lin, W. H., Xu, X. and Chen, C., 2011b. Data-Driven Intelligent Transportation Systems: A Survey. *IEEE Transactions on Intelligent Transportation Systems,* 12(4), 1624-1639.

Zhang, N., Chen, H., Chen, X. and Chen, J., 2016b. Semantic Framework of Internet of Things for Smart Cities: Case Studies. *Sensors,* 16(9), 1501.

Zhang, T., Ramakrishnan, R. and Livny, M., 1997. BIRCH: A New Data Clustering Algorithm and Its Applications. *Data Mining and Knowledge Discovery,* 1(2), 141-182.

Zhang, T., Yang, D., Li, T., Li, K. and Lian, X., 2011c. An improved virtual intersection model for vehicle navigation at intersections. *Transportation Research Part C: Emerging Technologies,* 19(3), 413-423.

Zhang, Y.-C., Zuo, X.-Q., Zhang, L.-T. and Chen, Z.-T., 2011d. Traffic Congestion Detection based on GPS Floating-Car Data. *Procedia Engineering*, 15, 5541-5546.

Zhang, Y. M., 2004. The thinking of relieving the congestion of elevated road system. *Traffic and Transportation*, 5, 37-38.

Zhao, Y., Qin, Q., Li, J., Xie, C. and Chen, R., 2012. Highway map matching algorithm based on floating car data. *In: 2012 IEEE International Geoscience and Remote Sensing Symposium*, 22-27 July 2012, 5982-5985.

List of publications relevant to the thesis

- **Keler, A.**, Krisp, J. M., Ding, L., 2017. Detecting vehicle traffic patterns in urban environments using taxi trajectory intersection points. Geo-spatial Information Science, in press.

- **Keler, A.**, Krisp, J. M., Ding, L., 2017. Visualization of Traffic Bottlenecks: Combining Traffic Congestion with Complicated Crossings. In: Peterson, M. P., (ed) Advances in Cartography and GIScience, Selections from the International Cartographic Conference 2017. Cham: Springer International Publishing, pp. 493-505.

- **Keler, A.**, Krisp, J. M., Ding, L., 2017. Detecting Travel Time Variations in Urban Road Networks by Taxi Trajectory Intersections. GISRUK 2017 - 25th GIS Research UK Conference. Manchester, United Kingdom.

- **Keler, A.,** 2017. Ableitung von Verkehrsengpässen anhand Staudetektion mit Fahrzeugtrajektorien. In: Kolbe, T. H., Bill, R., Donaubauer, A. (eds) Geoinformationssysteme 2017 – Beiträge zur 4. Münchner GI-Runde. Berlin/Offenbach: Wichmann, pp. 172-184.

- **Keler, A.**, Krisp, J. M., 2016. Is there a relationship between complicated crossings and frequently visited locations? – A case study with boro taxis and OSM in NYC. In: Gartner, G., Huang, H. (eds) Proceedings of the 13th International Symposium on Location-Based Services, pp. 186-190. Vienna, Austria.

- Ding, L., Krisp, J. M., Meng, L., Xiao, G., **Keler, A.**, 2016. Visual exploration of multivariate movement events in space-time cube. The 19th AGILE International Conference on Geographic Information Science. Helsinki, Finland.

- **Keler, A.**, Krisp, J. M., 2016. Visual Analysis of Floating Taxi Data Based on Interconnected and Timestamped Area Selections. In: Gartner, G., Jobst, M., Huang, H. (eds) Progress in Cartography, EuroCarto 2015. Cham: Springer International Publishing, pp. 115-131.

- **Keler, A.**, Ding, L., Krisp, J. M., 2016. Visualization of traffic congestion based on Floating Taxi Data. Kartographische Nachrichten, 66(1): 7-13.

- **Keler, A.**, Krisp, J. M., 2015. Visual Analysis of Floating Taxi Data based on selection areas. In: Gartner, G., Huang, H. (eds) Proceedings of the 1st ICA European Symposium on Cartography (EuroCarto 2015), pp. 58-59. Vienna, Austria.

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

- **Keler, A.,** 2015. Kartographische Darstellungen für die visuelle Verkehrsanalyse auf Basis von Floating Car Data (FCD). gis.SCIENCE, 28(2): 51-61.