



MASTER'S THESIS

of Zhichun Liu

Date of Issue: 2020-07-15

Date of Submission: 2021-01-15

Topic: Exploring the Evolution Mechanism of Traffic Congestions and Its Impact on Normal Traffic Based on Data-driven Method

Properly understanding the urban traffic congestion evolution in network scale is crucial to effectively mitigate congestion and its impact proactively. Even though many studies focus on this hotspot, it has not been fully and clearly revealed. This study is therefore to explore the network traffic congestion evolution patterns based on a large-scale dataset of naturalistic vehicle trajectories collected by a swarm of drones.

With advances in technology, a variety of methods for traffic surveillance and monitoring are now available, including traffic cameras, loop detectors, ramp metering systems, and GPS data on probe vehicles (taxi fleets or mobile devices). Nevertheless, these methods above have their limitations on the network-scale research for traffic congestion formation, propagation and impact on urban context with multi-modal, multi-lane environment. As a case study, the central district of the city of Athens, Greece was selected that can allow a variety of transportation phenomena to be examined.

Utilizing a swarm of drones could overcome a significant number of limitations of the abovementioned methods, and pragmatizing an actual one for massive data collection in a busy, multimodal urban environment had not been conducted before. Recently, a field experiment called pNEUMA (New Era of Urban traffic Monitoring with Aerial footage), has collected traffic streams data over an urban setting using the Unmanned Aerial Systems (UAS) and shown preliminary results of congestion characteristics. pNEUMA is a large-scale dataset of naturalistic trajectories of half a million vehicles that have been collected by a one-of-a-kind experiment by a swarm of drones in the congested downtown area of Athens, Greece.

Based on their big dataset, therefore, this study will propose a data-driven method based on machine learning to effectively analyze such tremendous amount of data, which is involved with data cleaning, data processing and machine learning to explore the formation and impact of traffic congestion in urban context. The potential models for this research include deep learning (e.g., Recurrent Neural Network), Supported Vector Machine and reinforcement learning. The traffic data will be used to train and validate the models and their capability will be compared in terms of computation duration, accuracy and other possible factors.

Traffic stream data contains continuous trajectory of vehicles, which is able to track the exact routes, origins and destinations in the scale of network. Besides, with high resolution camera, the microscopic behaviors, such as car following and lane changing of vehicles are available for further investigation. Therefore, the sub-tasks will be derived as following:

- Compare the factors of congestion for both scenarios, namely signalized intersections and non-signalized intersections.
- Explore the relationship between the formation of congestion and the layout of different intersections (4-arm, 5-arm etc.)
- Explore car-following behaviors of different vehicle types and their contribution to the traffic jam
- Analyze the impact of congestion on some important factors, such as the O-D distributions, local route choices, headways of vehicles, and mean travel time.

The completed methodology of this thesis could be applied to other traffic data analysis and shed a light on the following relevant research. The results of this thesis would also be able to give a clearer understanding of the nature of urban congestions with a series of traffic parameters and reveal its impact on normal traffic flow in micro- and macroscopic levels, ranging from behaviors of specific type of vehicle to the congestion propagation throughout the studied network.

The master student will present intermediate results to the mentor Kui Yang, a PostDoc Research Associate (TUM) in the 6th, 12th, 18th, 24th weeks. In the meanwhile, there will be regular discussions and feedbacks every two or three weeks.



The student must hold a 20-minute presentation with a subsequent discussion at the most two months after the submission of the thesis. The presentation will be considered in the final grade in cases where the thesis itself cannot be clearly evaluated.

A handwritten signature in black ink, appearing to read 'K. Antoniou'.

Prof. Dr. Constantinos Antoniou

Acknowledgement

Foremost, I would like to express my deep and sincere gratitude to my academic advisor, Dr. Kui Yang, for supporting me to successfully complete my thesis, and for his kind assistance, insightful comments, and knowledge which added considerably to my academic experience. I sincerely appreciate the the Data source: pNEUMA – open-traffic.epfl., who enables such valuable and extremely novel data available for everyone. My research would not have succeeded without them.

I would like to express my sincere gratitude my friends who helped and supported me through this Master study period. Thank you, Fanli Zeng, Yumeng Fang, Mingda Xu, Shupe Chen, Ruoyu Chen, Lu Han for encouraging and supporting me.

It is a really great experience to spent a part of my lifetime in Singapore and Munich, Germany. My deep gratitude and love to each one of you. In addition, a special thank you goes to my parent and my sister for their constant encouragement and financial support and without the help of whom I could not have pursued this Master's degree.

Last but not least, thank you Technical University of Munich and TUM Asia.

Abstract

Nowadays, traffic congestion becomes a big concern in many metropolitan world-widely. Traffic congestion causes numerous detrimental impacts, such as prolonged travel time, air pollution and emissions of green gas. Therefore, it is important to solve this issue so that reliability in road maintenance can be accomplished. However, the mechanism of evolution of traffic congestion in an urban area is somehow unclear. This is crucial to traffic operates and participants since the congestion pattern can reveal the keys to alleviate the negative impacts on the society and environment.

The main objective of the study is to explore the congestion patterns in an urban context with the help of trajectory dataset from a first-of-its-kind experiment pNEUMA. This field experiment provided a high-resolution GPS traces data which contains all vehicles attributes including type of vehicle, location, speed associated with timestamps and etc., in the populated zone of Athens City, Greece.

The study adopts a series of methods to fulfill the goal. Firstly, the trajectory data is processed to discard the static records and prepared for map-matching. Next, a Hidden Markov Method algorithm is employed to link the traces of vehicles to the road network. Inspired by the concept of virtual detector loops, the traffic characteristics on every travelled link will be calculated and analyzed. The time series method will be applied to find out the temporal trend of traffic streams. Finally, a proposed algorithm will distinguish the traffic states and define the congestion on every detector.

The result reveals the spatio-temporal distribution of traffic congestion in Athens during 8:00 to 9:30. Focusing on the representing street, the conclusions are: traffic congestion intend to form at the merging sections of a primary and secondary roads. And traffic congestion propagates along the traffic travel direction in the very beginning. The longest duration of congestion state on the research corridor reached up to ten minutes.

Keywords: Spatio-temporal analysis, Traffic congestion, Trajectory data, Urban network

Table of Contents

| | | |
|----------|---|-----------|
| 1 | Introduction | 1 |
| 1.1 | Background | 1 |
| 1.2 | Aim and research question | 1 |
| 1.3 | Contributions | 2 |
| 1.4 | Thesis structure | 2 |
| 2 | Literature Review | 4 |
| 2.1 | Traffic flow theory | 4 |
| 2.2 | Research on traffic congestion..... | 4 |
| 2.2.1 | Causes of congestion | 4 |
| 2.2.2 | Measuring traffic congestion..... | 5 |
| 2.3 | Trajectory-based traffic estimation | 6 |
| 2.4 | Networkwide traffic congestion | 7 |
| 2.5 | Summary | 8 |
| 3 | Data Processing | 9 |
| 3.1 | Source of data: pNEUMA experiment..... | 9 |
| 3.2 | Application Tooling | 11 |
| 3.3 | Pre-processing of Trajectory Data..... | 11 |
| 3.4 | Establishing of road network | 14 |
| 3.4.1 | Retrieving the network of research area..... | 14 |
| 3.4.2 | Extract specified road for further research..... | 17 |
| 3.5 | Map-matching of trajectory metadata | 18 |
| 3.5.1 | Map-matching theory and methodology | 19 |
| 3.5.2 | Implementation and results..... | 20 |
| 3.6 | Extraction of traffic characteristics | 22 |
| 3.6.1 | Virtual loop detector | 22 |
| 3.6.2 | Traffic parameters extraction | 23 |
| 3.6.3 | Results analysis..... | 25 |
| 3.7 | Summary | 27 |
| 4 | Congestion Evolution in spatial temporal spectrum..... | 29 |
| 4.1 | Metrics for congestion identification | 29 |

| | |
|---|-----------|
| 4.1.1 Theory background..... | 29 |
| 4.1.2 Traffic state distinguished..... | 29 |
| 4.1.3 Establishing of Thresholds..... | 30 |
| 4.2 Congestion identification..... | 31 |
| 4.2.1 A closer look at time session 8:00 – 8:30 am..... | 31 |
| 4.3 Congestion evolution..... | 34 |
| 4.3.1 Spatial distribution..... | 34 |
| 4.3.2 Temporal distribution..... | 37 |
| 4.4 Summary..... | 40 |
| 5 Conclusion and future work..... | 41 |
| 5.1 Conclusion..... | 41 |
| 5.2 Limitations and future work..... | 41 |
| List of References..... | 42 |
| List of Abbreviations..... | 46 |
| List of Symbols..... | 47 |
| List of Figures..... | 48 |
| List of Tables..... | 49 |
| Appendix A:..... | 50 |
| Appendix B: Survey Data..... | 51 |
| Declaration concerning the Master’s Thesis..... | 53 |

1 Introduction

1.1 Background

Congestion is not new. It existed before the industrial revolution, the motor vehicle, and the modern city. In general, traffic congestion stems from an imbalance between the availability of transport services and the need for them [FALCOCCHIO ET AL., 2015]. Nowadays, traffic saturation contributes to chronic congestion with many detrimental effects for in terms of aggravation, stress, lost time and environmental nuisance [DERBEL ET AL., 2020, p. 23-36]. In Metropolitan areas, urban traffic congestion is one of the most severe problems of everyday life, which affects all facets of the quality of life in cities adversely and dramatically.

Concerning on abovementioned negative impacts by traffic congestion, it is essential to understand the evolution of urban traffic congestion on a network scale correctly, so that the corresponded counter measurements can be taken into action. It is believed that identification of congestion characteristics is the first step for such efforts since it is an essential guidance for selecting appropriate measures [MOHAN RAO ET AL., 2012, p. 286-305]. Therefore, many efforts were made in order to alleviate congestion. More relevant research will be discussed in Chapter 2 Literature Review. Even though there have been several studies concentrated on this hotspot, it has not been exposed thoroughly and explicitly.

In the new era of Big Data, the enormous quantities of information are foreseen to be beneficial to the transportation profession and research community. With advances in technology, a variety of methods for traffic surveillance and monitoring are now available, including traffic cameras, loop detectors, ramp metering systems, and GPS data on probe vehicles (taxi fleets or mobile devices). Faced with such high quantity and complexity of data, the traditional approaches to model and predict traffic characteristics have reached their limits. In this research the trajectory data sourced from pNEUMA experiment [BARMPOUNAKIS ET AL., 2020, p. 50-71] recorded a huge amount of vehicle trace with high sampling rate from the congested district of Athens, with over 100km-lanes of road network. More details about the data will be discussed in Chapter 3 Data Processing.

1.2 Aim and research question

The proposal was concerned about exploring the evolution mechanism of traffic congestions in the urban context based on the open dataset from the pNEUMA field experiment. With respect to enormous information from traffic data, the data-driven methods are expected to be the promising way to carry out the analysis. Existing literature reveals that current data sources, such as on-board GPS trajectory [RAMEZANI ET AL., 2015, p. 414-432] [BHASKAR, A. ET AL., 2015, p. 113-122], navigation applications via smartphones, loop detectors [BHASKAR,

Ashish ET AL., 2011, p. 433-450] or fixed surveillance camera in Next Generation Simulation [ET AL., 2006] may not capture important traffic phenomena properly due to the limited level of information. The main aim is to reveal the formation and development of traffic congestion. Moreover, the topic will be examined and studied in different scale of time and space. Therefore, it is expected to choose and utilize appropriate algorithms and models to solve different sub-problems, then to evaluate and compare each algorithm performance.

Main Research questions:

Given the aforementioned need the following two main research questions are formulated:

How the congestion developed and formed in an urban area temporally and spatially?

To address the above research question, a wide spectrum of sub-problems will be derived as following:

1. How to pre-process the trajectory dataset and extract of traffic characteristics?
2. How to match the GPS points to the road network and extract traffic characteristics at a given location and time?
3. What kind of traffic parameters should be considered to identify and measure congestion?
4. How to capture the spatial and temporal patterns of congestion?

1.3 Contributions

The main achievements, including contributions to the field can be summarised as follows:

1. Proposed congestion identification methods based on virtual loop detectors network-wide
2. Provided road classification-related algorithm to distinguish traffic state
3. Explore the congestion evolution spatially and temporally through network

1.4 Thesis structure

This thesis is organized as follows:

Chapter 1 introduces the background of the traffic congestion, pNEUMA traffic data source and raise the objective of research questions and research framework.

Chapter 2 conducts a wide spectrum of literature review on the most relevant subjects for this thesis, including the definition and causes of traffic congestion, common traffic characteristics and their usage, relevant big data algorithms and applications in traffic prediction.

Chapter 3 performs pre-processing of trajectory data and map-matching to the road network. The traffic characteristics are then extracted and processed according the need to next analysis.

Chapter 4 outlines the methods and the set of algorithms which are expected to be applied for congestion identification and conducts spatial-temporal analysis on the congestion evolution in the network.

Chapter 5 summaries the results from previous chapters and concludes the limitations of the work and points out the research work in the future.

2 Literature Review

In this chapter, related work is reviewed and analyzed in order to acquire a deeper understanding of the problem in hand, and find the most appropriate methodologies to identify gaps in the literature and guide towards the development of methods that answer the research questions. The research topics are reviewed thematically as following.

2.1 Traffic flow theory

Traffic flow is an important metric for evaluating road congestion.

For two states of traffic flow, EDIE ET AL. [1965, p. 139-154] suggested his linear models: one is chosen to describe the relationship between density and the velocity logarithm above the optimum velocity within uncongested traffic flow; and the other is the relationship between velocity and the spacing logarithm, the opposite of density, under the condition of congestion. Discontinuities have always occurred in flow-concentration data at around maximum flow in other experiments, and some researchers have attempted to use several curves to model the "discontinuities" variable. Kerner defined traffic flows in three categories: free flow, coordinated flow, and stop-and-go flow.

Relevant studies have focused on the interactions among various traffic participators, such as vehicles, drivers, pedestrians, and bicyclists and infrastructure, including highways, signal control devices, aiming to reveal the relationship between individual traffic participants and the resulting traffic flow phenomena [LI, Li ET AL., 2020, p. 225-240]. As a result, traffic flow studies are empirical studies that heavily rely on high-quality measurements of real data. In the past 50 years, the scientific community has proposed various traffic models in its attempt to understand vehicular traffic flow. There are continuous fluid dynamical methods such as the Lighthill-Whitham theory in the 1950s, and Navier-Stokes-like momentum equation; discrete models such as the follow-the-leader model; as well as stochastic discrete models such as the Nagel-Schreckenberg Cellular automata models. Regardless continuous or discrete, deterministic or stochastic, the purpose of these models is to shed light on the physics of traffic flow and to provide predictions of various traffic phenomes [QUEK ET AL., 2014, p. 289-298].

2.2 Research on traffic congestion

2.2.1 Causes of congestion

Studies on the causes of congestion are well documented, it is also well acknowledged that the disparity in supply and demand for transportation infrastructure results in traffic congestion [FALCOCCHIO ET AL., 2015, p. 35-37]. On one hand, supply is limited by history and geography, the administration and working procedures of transport, and the extent of expenditure in roads and highways. The demand, on the other hand, benefits from the concentration of travel in

both space and time. Practically, the causes of traffic congestion are more specifically concluded as (1) concentrations of trips in time and space—including temporal surges in travel demand on roads of generally constant capacity physical, operational, and design deficiencies that create bottlenecks, (2) traffic demand that exceeds roadway capacity, and (3) physical and operational bottlenecks.

If all travel demand were evenly distributed among the various sections of the urban area, the traffic congestion problem would be a rare event. Similarly, if all travel were evenly distributed to each hour of the day there would be little congestion. But travel demand patterns reflect the concentration in time and space of daily activities: where and when people work, shop, recreate, move goods and provide services. It is the peaking of these spatial and temporal travel patterns that contributes to the recurring traffic congestion problem. As observed in many metropolitan cities, growth in population, employment, and car use vehicle kilometres of travel (VKT) increase congestion on urban roads and highways where capacity growth has not kept pace with growth in VKT. Furthermore, bottlenecks are perhaps the most common cause of congestion. They result from the convergence of a greater number of lanes in the upstream roadways than are available in the downstream roadways. Bottlenecks delay is typically found in hours of peak flow where the number of lanes converging on a roadway, bridge or a tunnel exceeds the number of lanes these facilities have. Bottlenecks are also created by roadway incidents that reduce block travel lanes and restrict traffic flow, or they are created by bad weather conditions a work zone, poorly timed traffic signals, or driver behaviour. Based on the causes of congestion, the analysis of these three points will be developed in the following sections

Several researches have been proposed to classify congestion into recurrent congestion (RC) and non-recurring congestion (NRC). The interruption that travellers frequently encounter or anticipate at known travel hours, such as the morning and evening peaks, is chronic congestion. One method employed by [AN S ET AL., 2018] is mining historical taxi trajectory data to investigate recurrent congestion patterns, which is involved with grid-based congestion detection, a customized cluster algorithm and indicators to reflect RC evolution patterns. The non-recurring delay in congestion is caused by unexpected or random incidents that impede the flow of traffic. This include incidents such as car breakdowns or crashes; road maintenance and poor weather; special events that cause unexpected market spikes, such as the conclusion of a sporting event; and natural or man-made disasters. In the off-peak hours, non-recurring congestion can either produce new congestion or can improve the pause encountered during periods of recurring congestion. [CHEN, Zhuo ET AL., 2016, p. 19-31] employed a data-driven method to dynamically determine the spatiotemporal extent of individual incidents, so that Incident-Induced Delay (IID) can be quantified

2.2.2 Measuring traffic congestion

Congestion in transportation occurs when the occupancy of spaces (roadways, sidewalks, transit lines and terminals) by vehicles or people reaches unacceptable levels of discomfort

and delay [FALCOCCHIO ET AL., 2015, p. 93-110]. Over time, an extensive literature has developed on measuring traffic congestion. In 1925, Mc Clintock explains street traffic congestion as a condition resulting from a delay in movement below that required for contemporary street users [MCCLINTOCK ET AL., 1925, p. xi, 233 p.]. Researchers [ALTSHULER ET AL., 1979] indicates that the term congestion denotes any condition in which demand for a facility exceeds free-flow capacity at maximum design speed. Hereby the measurement is quantified by speed and capacity. Concluded from several studies [SIGUA ET AL., 2008], [MEYER ET AL., 1997],[LOMAX ET AL., 1997], traffic congestion reflects the difference between the travel time at busy traffic periods and when the road is lightly traveled. It is also expressed as the ratio of actual travel time and uncongested travel time or the ratio of actual versus uncongested travel time rates. The three basic components of traffic congestion include intensity or amount), extent area or network coverage, and duration of the jam.

Duration refers to the amount of time the road or the network is congested, which is usually represented by the hours when facility operates below acceptable speed. Extent means the number of people or vehicle affected or geographic distribution, quantified by VKT or kilometers of congested road. Intensity reflects level or total amount of congestion. Travel rate, delay rate and average speed can be utilized to describe intensity. Reliability measures the variation in the amount or duration of congestion [LOMAX ET AL., 1997].

2.3 Trajectory-based traffic estimation

GPS-probe vehicle has recently become a promising technique for traffic states monitoring, with advantages of cost-effectiveness and dynamic traffic conditions surveillance. GPS-probe based traffic surveillance systems enable to obtain traffic states according to GPS data periodically collected from GPS-probe vehicles on road sections. Over the last few decades, the estimation of traffic conditions has gained increasing attention in the academic community [STATHOPOULOS ET AL., 2003, p. 121-135, ANTONIOU ET AL., 2006, p. 103-111, BERTHELIN ET AL., 2008, p. 185-220, D'ANDREA ET AL., 2017, p. 43-56]. Early works on traffic estimation have studied traffic states on highways using on-site detectors and surveillance such as loop detectors and video cameras. Previous methods are limited to road segments with lengths of a few kilometers. Recent years, increased emphasis has been compensated to the effective collecting and retrieval of GPS trace data by data mining, intelligent transport networks, repositories, and smart cities communities. GPS data are usually collected by probe vehicles or specific fleets such as buses or taxies[TANG ET AL., 2019, p. 137-174]. GPS traces and advanced models have been applied to traffic state estimation across a range of settings, such as traffic speed estimation [YU ET AL., 2020, p. 136-152] [WANG ET AL., 2016, p. 499-508], queue profile estimation [RAMEZANI ET AL., 2015, p. 414-432], density (Papadopoulou et al., 2018), congestion [KAN ET AL., 2019, p. 229-243], and travel times [LI, W. ET AL., 2017, p. 100-113].

From the relevant works by [ZHENG ET AL., 2008, D'ANDREA ET AL., 2017, p. 43-56, YU ET AL., 2020, p. 136-152, LI, W. ET AL., 2017, p. 100-113], the general working flow can therefore be concluded as following, when conducting a traffic study based on trajectory data:

Firstly, GPS trajectories data is pre-processed to parse the inherent noise and extract the desired attributes and parameters such as longitude, latitude, time stamps and etc. The second step is map-matching, which addresses the problem of mapping off-the-road GPS points onto a road network and identifies the true traversed path between consecutive GPS points. In other words, the results of map-matching consist of two components, linked traces on road and the ground-true travelled paths of road network. Normally, there are a wide spectrum of methodologies and algorithms for map-matching, including geometric, topological, probabilistic and advanced map-matching algorithm. In particular, Hidden Markov Model (HMM) - based algorithms and their variants [GOH ET AL., 2012, p. 776-781] have been adopted for their abilities to concurrently evaluate multiple hypotheses of the actual mapping in order to find the eventual maximum likelihood solution. These techniques [KAFFASH ET AL., 2021]. have been shown to be tolerant of extremely noisy measurements.

The rest of the map-matching procedure is varied from the final purpose of research objects. As for this study, the trajectory data is derived from aerial video images. The quality and sampling rate are high enough for further traffic characteristic extraction. Thus, the data recovery or missing rate evaluation for the matching results is not necessary and will not be performed in this paper.

2.4 Networkwide traffic congestion

The scale of traffic congestion studies ranges from microscopic (vehicle level), to mesoscopic (link level) and to macroscopic (network level). These classifications regard to traffic representation and operating scales. Besides, different level of resolution of data source will result in different domain of traffic research.

In this literature [AMELIA ET AL., 2014, p. 1-5], the researchers attempted to investigate the association between road segments by using GPS trajectories of traffic congestion. To get the congestion connection in each of the road segments, they use a clustering algorithm and conducted spatio-temporal analysis. But in this paper, the dataset only contains taxis as a mode of transport and the scope is limited within a link-level congestion. Cell- or grid-map [YU ET AL., 2020, p. 136-152, LIU ET AL., 2017] are popular simplification methods to perform a network-wide traffic estimation. By determining the cell congestion state, the traffic state of a specific zone can be estimated. Using large traffic data-sets obtained from traffic sensors, visualization is an efficient method for analyzing traffic congestion. The paper of [CHENG ET AL., 2013, p. 296-306] proposed three 3D exploratory visualization techniques to explore the evolution of traffic congestion on the entire road network, especially on dense urban networks. The identification of congestion is based on link travel time from local authorities. Another research [ZHANG ET AL., 2020] raised a representation framework for traffic congestion data in urban road traffic networks. In comparison to other grid-map based research, they employed a pooling operation to calculate the maximum value in each grid, which is cost efficient in terms of computational resources for traffic congestion prediction. Roads in networks are spatially and

temporally linked. The travel times of its neighbors are influenced by road volumes. An apparent importance is bare upstream and remote roads are negligible. To boost the predictive performance of the volume of traffic, a variety of models have been tested. In a study Zhang (2012) uses a method of traffic clustering to group road points that are spatially and time related. Scholars [XU ET AL., 2013, p. 2421-2427] also monitored traffic states over large road networks, which has become a major service in the entire transportation system. As they stated, the stationary equipment, such as loop detectors and ultrasonic detectors, can collect various traffic data, but only limited to specific road sections, and require a large amount of expenditures.

Previous studies in the network scales are mainly focused on simplified the representation of road networks. Some of them use the data from probe cars of mobile devices with GPS function to instigate the whole performance of the traffic. In this thesis, with the help of full trajectory data in the observed network, such simplification should be conducted in a new way. Inspired by the work (source: https://pNEUMA_mastersproject.readthedocs.io/) from Joachim Landtmeters, the traffic characteristics can be extracted on any location in the network by installing virtual loop detectors. Thus, the link-level congestion can be identified by designing a threshold for key parameters. In traffic congestion context, an area that has traffic congestion is an area where there are many links in the jam status.

In summary, traffic needs to be analysed at a level of spatial and temporal detail in order to better understand the causes and impacts of congestion and to be able to find potential remedies against it since it mostly affects particular links in the network at specific times of the day.

2.5 Summary

This literature review has reviewed the state-of-the-art, organized evidence around four key themes of relevance to the traffic flow theory, congestion identification, trajectory data processing and popular used methods for network-scaled traffic state analysis. However, in this study, a brand-new dataset is available for exploring the congestion evolution, which contains high accuracy in position and high sampling rate trajectory data of all vehicles in a dense network. Therefore, the study will employ and combine series of methodologies to investigate the dataset and mining the mechanism of congestion evolution.

3 Data Processing

3.1 Source of data: pNEUMA experiment

pNEUMA [BARMPOUNAKIS ET AL., 2020, p. 50-71] is a first-of-its-kind experiment, which aims to create the most complete urban dataset to study congestion. In October 2018, it collected traffic streams data over an urban setting using the Unmanned Aerial Systems (UAS) to establish a large-scale video-image dataset during 2.5 hours in the morning peak (8:00 - 10:30 or 8:30 - 11:30) for all weekdays in a week (24/10, 29/10, 20/10 and 01/11) for the city of Athens, Greece.

The team employed a swarm of ten drones hovering over the central business district of Athens (see Figure 3.1) over multiple days to record traffic streams in a congested area of a 1.3 km² area with more than 100 km-lanes of road network, around 100 busy intersections, many bus stops and the dataset includes around half a million trajectories. In order to maximize both the area covered and the number of important points of interest, the number of drones were determined as ten and the hovering point for each of the drone were selected. The flight plans were planned for each drone, including the path to and from the hovering point, ensuring that no intercepting routes between the drones were present.

The action sequence of drone operating is illustrated in Figure 3.2. As dotted line shows, the swarm would take off and each drone would move to its specific hovering point. Then the recording of the traffic stream would begin simultaneously when all the drones were in place. The actual recording action is shown as green line, where starts from the confirmation of all drones being in position. Due to the battery life and extra time for routing, take-off and landing, the duration of each deployment is set to 30 minutes. After recording is accomplished, all drones return home and change batteries for the next session. This cycle was repeated five times and last for 2.5 hours until the end of the morning peak.



Figure 3.1 Researched area in Athens divided into 10 zones

(source: pNEUMA – open-traffic.epfl.ch)

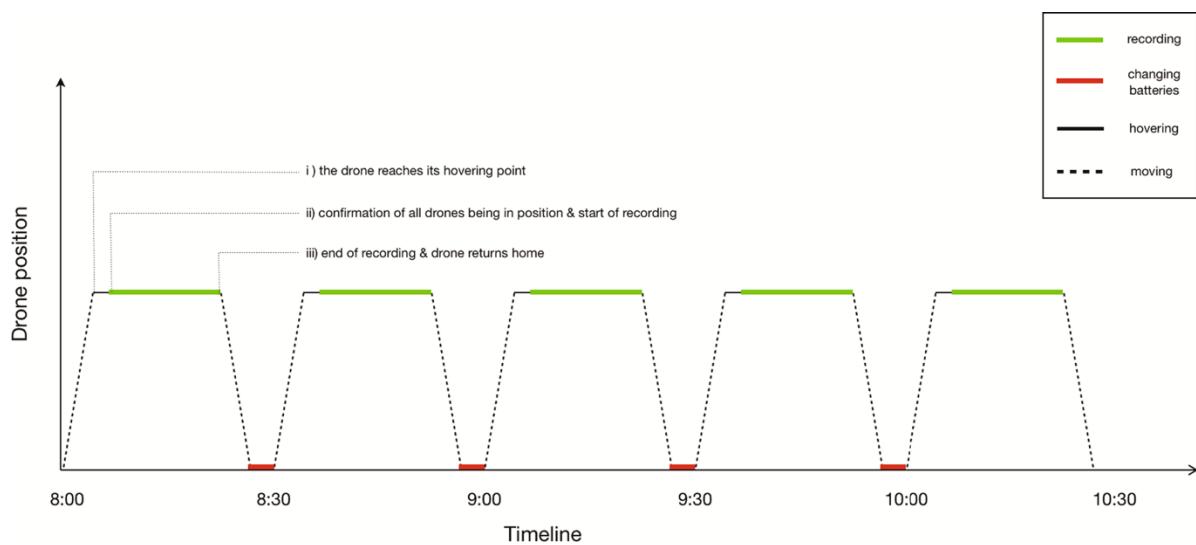


Figure 3.2 Time phases of UAV data collection

(source: pNEUMA – open-traffic.epfl.ch)

With a series of computer-vision techniques, the aerial video footages are converted into detailed trajectories of the vehicles tracked, calibrated in the WGS-84 system. The featured parameters can therefore be produced using the position information, for example speed (first derivative of position), acceleration (second derivative of position), distance traveled etc. The type of each vehicle (car, taxi, motorcycle, bus, heavy vehicle) is also available thanks to the high resolution of image and vehicle identification methods. To make sure the accuracy of the data, real-world coordinates are assigned to the pixels of video image. An advanced Kalman filter is applied to filter out the noise in the measurements up to a level of 3.3 cm, which is equivalent to 2.97 km/h in terms of speed error. Re-identification process was conducted to

synchronize and align up the vehicles which are tracked throughout the period it remains in the study area.

Such a unique experiment provides distinctive research opportunities, allowing the deep investigation of critical traffic phenomena. A complete trajectory dataset contains an enormous amount of information for the microscopic and macroscopic level since vehicles' movements can accurately be analysed and less assumptions are needed to measure different characteristics. Based on their big dataset, this study will utilize data-driven methods to handle tremendous amount of data, which is involved with data cleaning, data processing and machine learning to explore the formation and impact of traffic congestion in urban context. The traffic data will be used to train and validate the models and their capability will be compared in terms of computation duration, accuracy and other possible factors.

3.2 Application Tooling

To address about problems, the application tools need to be chosen properly. In this work there are a number of aspects that needed to be considered. It involved with a lot of data preprocessing, map-matching algorithms, spatial data mining, visualization, time series analysis. After a series of researching, it is concluded that Python provides many of the tools necessary for data and scientific processing. The main packages are introduced in Table 3.1.

Table 3.1 Main Python Modules

| Package | Description |
|--------------------|--|
| Pandas | data analysis and manipulation tool |
| Geopandas | geospatial data operations on geometric types |
| OSMnx | retrieve spatial data from OpenStreetMap and construct road network |
| LeuvenMap-Matching | align traces of coordinates to a map of road segments based on a Hidden Markov Model |

3.3 Pre-processing of Trajectory Data

The collected dataset is distributed and stored in form of comma-separated values (CSV). For each CSV file the data are organized as following: each row representing the data of a single vehicle, the first ten columns in the first row containing the columns' names, the first four columns represent the track_ID, the type of vehicle, the distance traveled and the average speed of the vehicle, respectively. The later six columns repeating every six columns based on the sampling frequency. In the dataset, there are six types of vehicles, which are car, taxi, bus, medium vehicle, heavy vehicle and motorcycle. The columns can be categorized into two groups, fixed tags and variable tags, with respect to the relation with time.

| Fixed Parameters | Variable Parameters |
|-----------------------|---------------------|
| Track_id | Latitude |
| Type | Longitude |
| traveled_distance [m] | Speed [km/h] |
| average_speed [km/h] | Lon_acceleration |
| | Lat_acceleration |
| | Time [s] |

Table 2 Columns of pNEUMA trajectory dataset

It is worth pointing out that these track points are derived from video image recognition. The longitude and latitude of them are also calculated from them, which means that these vehicle points exist independently of the road network data. For the purpose of map-matching, the geometry column needs to create from the coordinate pairs. Then the new matrix with geometry column will be converted to a GeoDataFrame object, which is 2-dimensional labeled data structure of the python package geopandas. The coordinate reference system of the geometry objects is set as WGS: 84 to keep consistence through the whole procedure of data processing.

The main steps for preprocessing the metadata are shown in Figure 3.3. Firstly, read the csv files into pandas and extract every tracked vehicle into single files. Next, the travel direction is calculated for every vehicle using compass bearing. This information indicates the direction from the point at the first time step to the point at the next time step in degrees, clockwise from north to east. With bearing column, the static trajectories can be filtered out. Besides, travel direction will be used to align the trajectory to the edge of road segments in the map-matching step. Finally, all of the GeoDataFrames will be stored in a list and serialized in form of pickle files on the local disks.

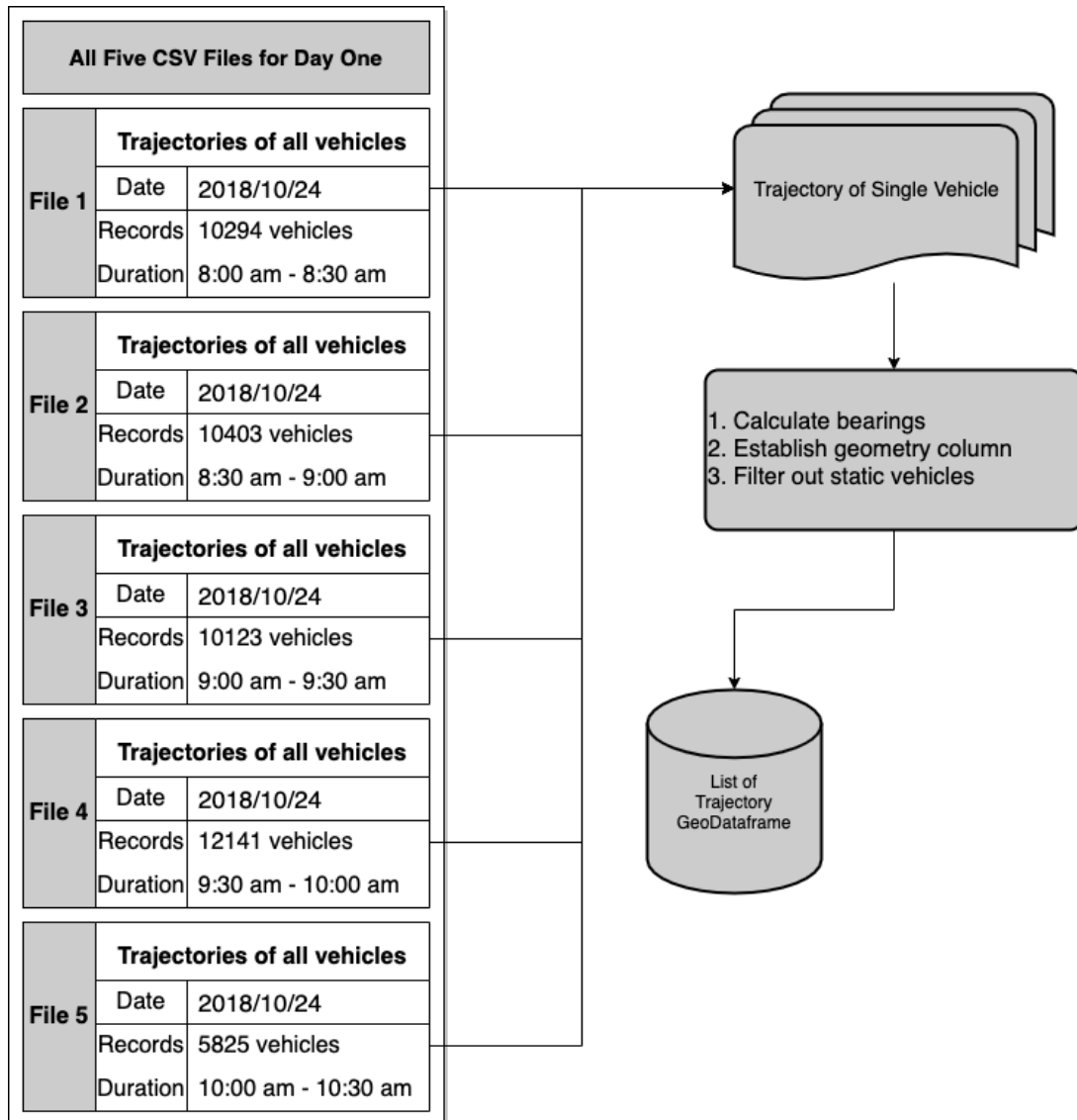


Figure 3.3 Procedure of preprocessing of metadata

In summary, the result of preprocessing is to obtain a list of geodataframes. Each geodataframe contains all information of a tracked vehicle with newly added bearing column and geometry column. The connection between road network and trajectories will be established on the basis of these two columns. Table 2 shows an example of one resulting trajectory. The column tracked vehicle contains the track id from the metadata. This vehicle is a car and traveled 101.4 seconds with a distance of 299.73m during the recording phase of one drone in the same surveillance zone. The coordinates and instant speed are tracked at each time step. Geometry column gives this trajectory a geometric attribute.

Table 3 Example of one resulting trajectory matrix

| Tracked Vehicle | Type | traveled_dist | Avg.Speed[km / h] | Lat | Lon | speed | tan_accel | lat_accel | time | geometry | bearing |
|-----------------|------|---------------|-------------------|-----------|-----------|---------|-----------|-----------|----------|---------------------------|------------|
| 34.0 | Car | 299.73 | 10.641184 | 37.983928 | 23.728834 | 1.3605 | 0.0364 | 0.0000 | 0.0 | POINT (23.72883 37.98393) | 270.000000 |
| 34.0 | NaN | NaN | NaN | 37.983928 | 23.728833 | 1.3625 | -0.0087 | 0.0000 | 40.0 | POINT (23.72883 37.98393) | 270.000000 |
| 34.0 | NaN | NaN | NaN | 37.983928 | 23.728833 | 1.3592 | -0.0364 | -0.0000 | 80.0 | POINT (23.72883 37.98393) | 270.000000 |
| 34.0 | NaN | NaN | NaN | 37.983928 | 23.728833 | 1.3520 | -0.0642 | -0.0000 | 120.0 | POINT (23.72883 37.98393) | 270.000000 |
| 34.0 | NaN | NaN | NaN | 37.983928 | 23.728833 | 1.3389 | -0.1179 | 0.0000 | 160.0 | POINT (23.72883 37.98393) | 270.000000 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 34.0 | NaN | NaN | NaN | 37.983483 | 23.726545 | 17.3402 | 0.1227 | 0.7343 | 101240.0 | POINT (23.72655 37.98348) | 201.509108 |
| 34.0 | NaN | NaN | NaN | 37.983481 | 23.726544 | 17.3612 | 0.0817 | 0.6855 | 101280.0 | POINT (23.72654 37.98348) | 237.610338 |
| 34.0 | NaN | NaN | NaN | 37.983480 | 23.726542 | 17.3789 | 0.0556 | 0.6375 | 101320.0 | POINT (23.72654 37.98348) | 218.244763 |
| 34.0 | NaN | NaN | NaN | 37.983479 | 23.726541 | 17.3930 | 0.0118 | 0.5901 | 101360.0 | POINT (23.72654 37.98348) | 218.244764 |
| 34.0 | NaN | NaN | NaN | 37.983477 | 23.726539 | 17.4008 | -0.0536 | 0.5430 | 101400.0 | POINT (23.72654 37.98348) | 218.244764 |

3.4 Establishing of road network

In order to investigate the traffic characteristics of the trajectory data, it is necessary to obtain information about the road network in the study area. The trajectories of observed vehicles are matched to the road network and various traffic parameters are extracted with the help of virtual loops.

In this study, the Open Street Map (OSM) was utilized as the source of the road network data. It is a well-known collaborative map, where users can easily access and retrieve streets data with customized configuration. With increasing quality of collected geospatial data, OSM becomes a reliable source of data for transportation research. By accessing keys and tags of the OSM, different types of roads can be extracted for further usage, e.g. drivable roads or arterials, etc. Since the following map-matching will adopt Python package “Leuven Map-Matching”, the network data will therefore also be established in a Python environment. The OSMnx [BOEING ET AL., 2017, p. 126-139] package offers the opportunities to collect data from the OSM and create and analyse street networks in a simple, consistent, automatable way. Even more valuable to this research is that it is also sound from the perspectives of graph theory, transportation, and urban design.

3.4.1 Retrieving the network of research area

The boundaries of research area are defined by the maximum and minimum latitudes and longitudes among all tracked records. The retrieved graph from OSMnx is in form of Multidigraph, a data structure storing directed graphs with self-loops and parallel edges. Parallel edges will occur when vehicles are travelling in bi-directions on the same street. In order to add customized functions, the multidigraph is transformed into two dataframes with geometry column, one storing the edges and another for the nodes. For edge matrix, it was assigned with new columns to include dedicated bus lanes and bearings of consecutive links. Then the edge matrix and nodes matrix can be merged and joint to generate the final road network matrix with needed attributes and columns. The working flow is illustrated in Figure 3.4. And

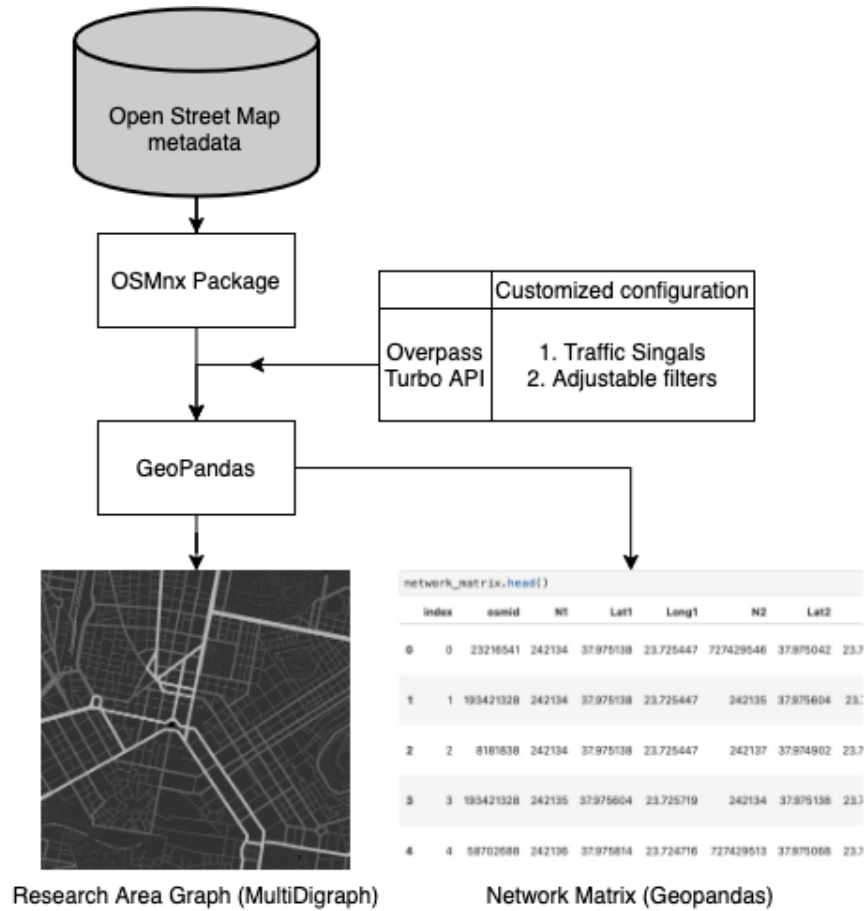


Figure 3.4 Working flow for network retrieving

The resulting network graph of research area in city of Athens is shown as Figure 3.7, the matrix is structured as following:

The positional columns include edges, nodes and geometries. The edge identity is recorded in osmid, indicating each directed segments of roads. Each edge consists of two nodes, N1 and N2, associated with their own latitude and longitude respectively. The attribute columns contain the length of edges, number of lanes, the allowed travel direction (oneway or not), the bearings between two consecutive edges, classification of streets and customized tags to present designated bus lanes. The bearings are calculated counter-clockwise with two pairs of coordinates of start and end nodes. Table 4 describes

Table 4 Statistics overview on resulting road network matrix

| | |
|-------------------------|---------|
| Number of Nodes | 3986 |
| Number of Edges | 1993 |
| Maximal edge length [m] | 292.137 |
| Minimal edge length [m] | 1.155 |
| Average edge length [m] | 55.428 |

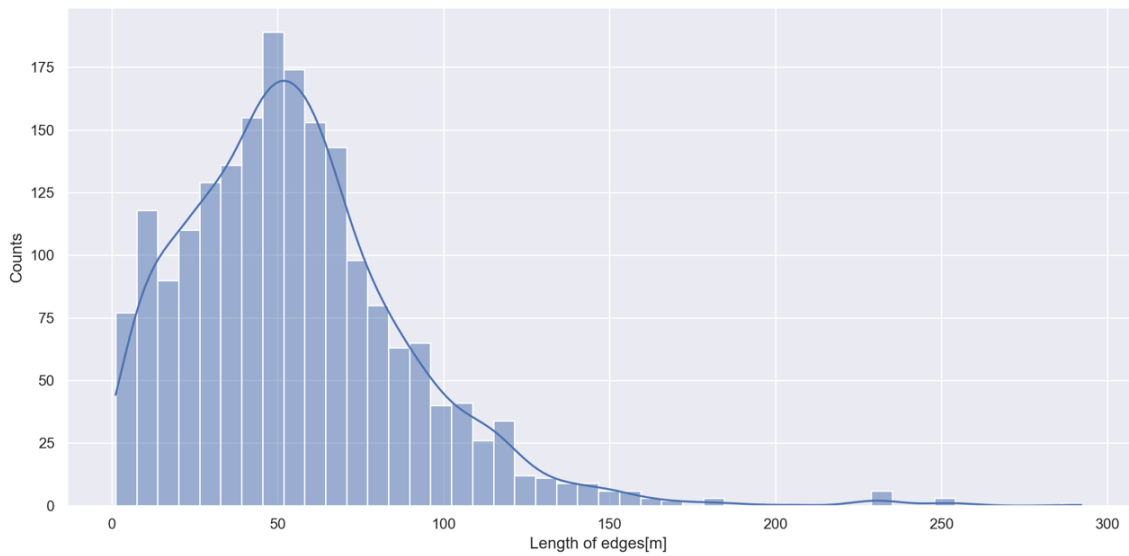


Figure 3.5 Histogram for edge length with KDE

The histogram (Figure 3.5) about length of edges shows, that most of the links were between 50 and 60 meters in the extracted road network. The maximal length of a single edge reached about 300 meters, which indicating that the network is widely dispersed. It is necessary to observe this attribute since the following section of map-matching will utilize the upper limit value as the searching radius in the HMM algorithms. Besides, with the knowledge of the length distribution will help to implement the map-matching process. Another important attribute in the network matrix is the hierarchy of the road system in this urban context.

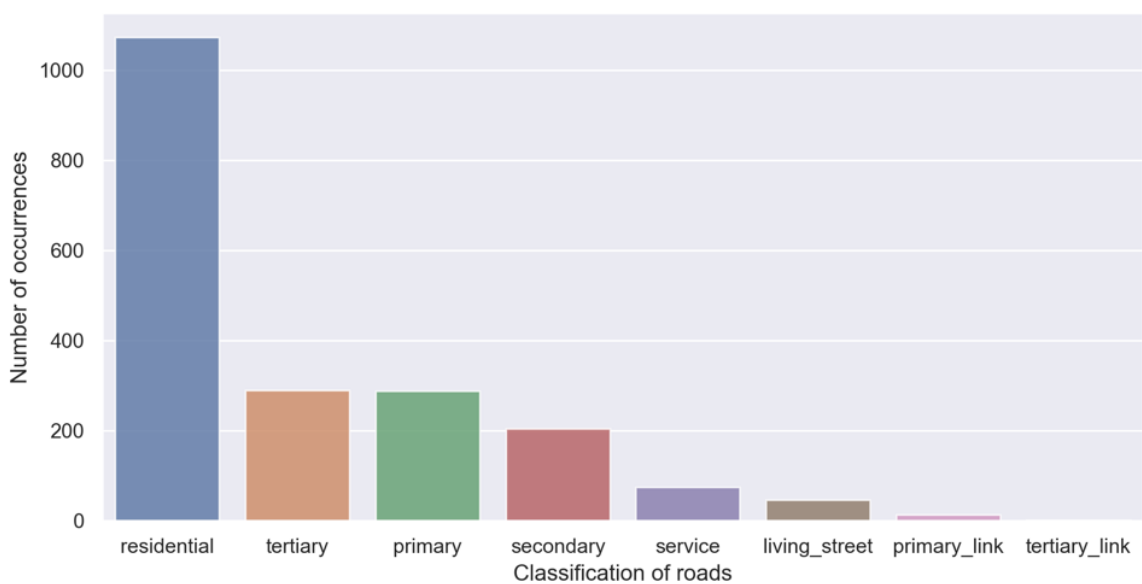


Figure 3.6 Counts of different types of roads

From Figure 3.6, it can be concluded that the residential streets take up the most counts among all kinds of roads, which confirms the research area is a dense and populated zone. The primary type represents the highest level of service in the resulting road network, usually found on the arterials in the city, as highlighted Figure 3.7. These trunk lines generally consist of more than three lanes, labeled with color codes. According to traffic flow theory discussed in Chapter 2, the lane-specific traffic characteristics are particularly important when investigating traffic flow state. To handle the traffic data on such a wide of variety of road classification, the metrics for congestion and traffic state should be cautiously chosen for such system with multi-modal traffic and hierarchical roads. More detailed parameters about traffic state estimation and congestion detection will be explored in the next chapter.

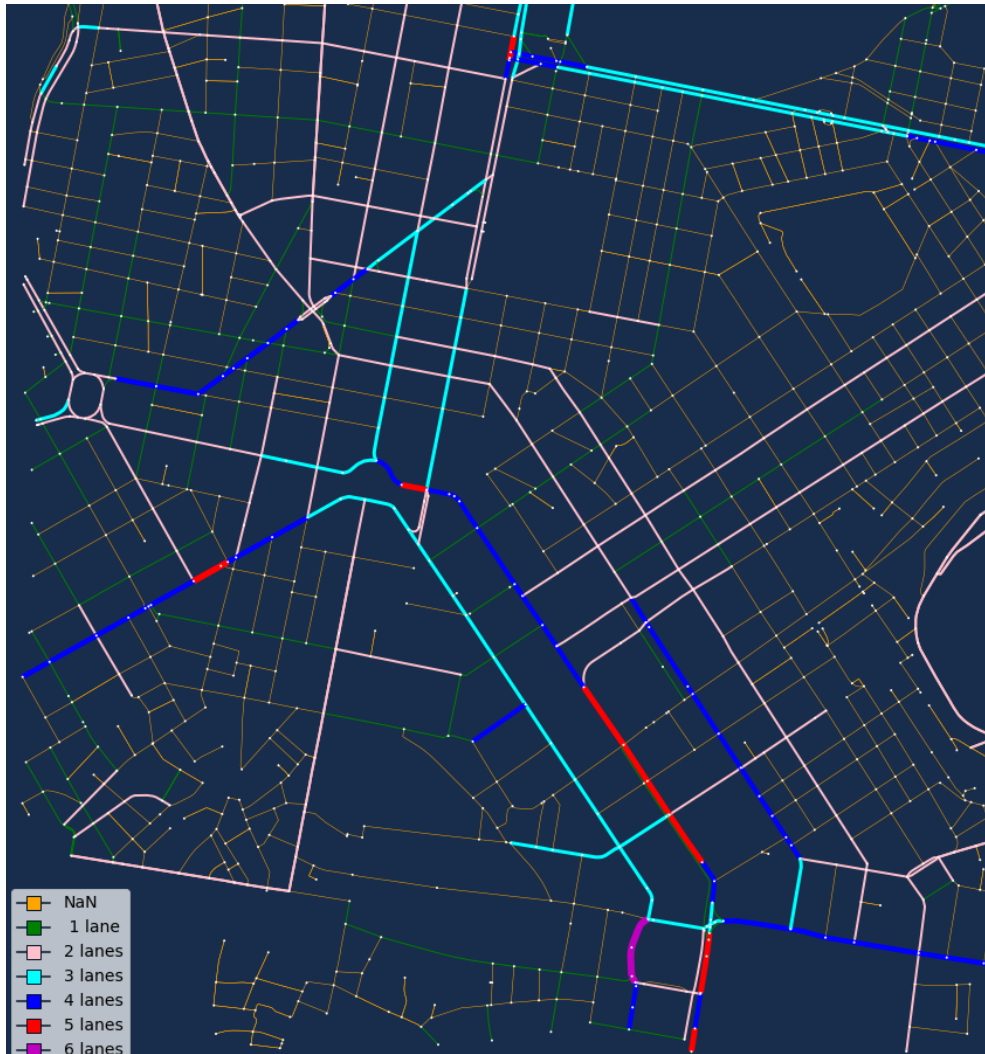


Figure 3.7 Road network labeled with number of lanes

3.4.2 Extract specified road for further research

The complete road network matrix is available for network-wide map matching. However, in order to investigate the traffic characteristics on any designated roads, it is required to create indices for interested roads by their name. The Overpass turbo, a web-based data mining tool for OSM, is employed to acquire road identifications i.e. osmid, associated with name attributes. The osmid is the key, which shared by indices as well as the network matrix. In the road

network matrix, one specific road may consist of multiple edges with various osmid. By constructing such index, the selection of targeted streets will be conducted precisely and correctly by nominate the street name. In the road classification system of OSM, primary roads are those of secondary importance roads in a country's system, often linking larger towns. In the context of research area in Athens, the indices include both primary and secondary roads for further exploration. Another crucial information is the location of traffic signals. It is desired in the later study for queueing at intersections. The traffic signal data matrix consists of nodes with coordinates and geometry columns. Figure 3.8 illustrated the results of query for primary and secondary roads and traffic signals, respectively. These data contain the key attribute which will be used to join with resulting network matrix.

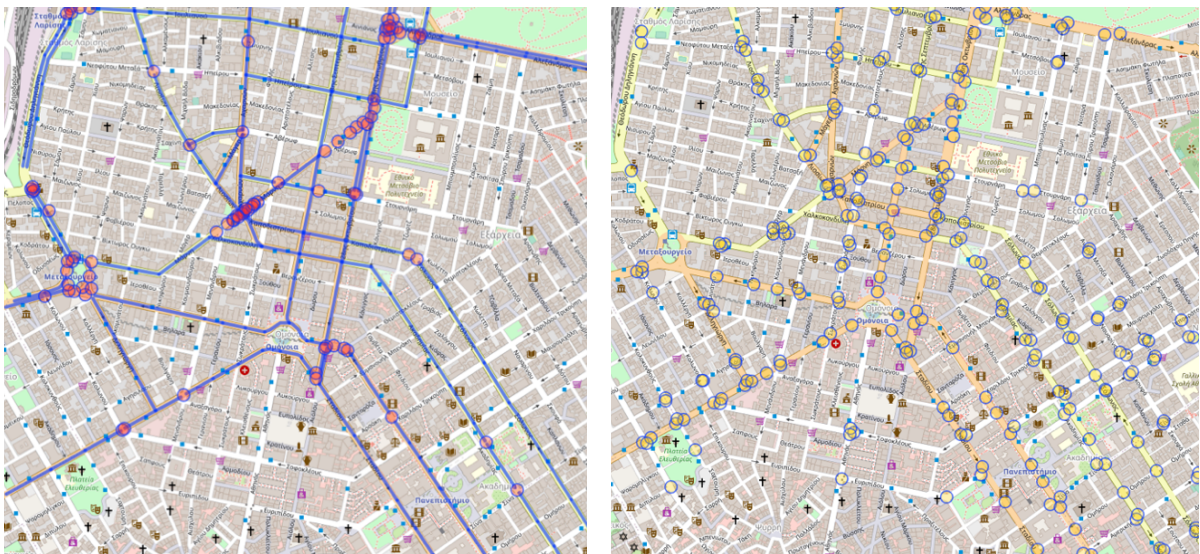


Figure 3.8 Primary and secondary roads (left) and Traffic signals (right)

3.5 Map-matching of trajectory metadata

The trajectory dataset is not linked with the road network. Although the quality and precise of the metadata is very high, it is also observed that there exists misalignment when plotting the trajectory straight on the road map. For instance, there is a noticeable lateral offset on road 'Sokratous' as shown in Figure 3.9, where the line in dark black is the trajectory of one vehicle. In order to know more exactly which specific road each vehicle is driving on, we need match the trajectory to the resulting map. To conclude, map matching is required to deal with identifying the most likely position of vehicle on the expected road links according to collected GPS data with errors. Without map matching, GPS data would be mismatched, thus influencing traffic states estimation.



Figure 3.9 Plot one trajectory on the road network without map-matching

3.5.1 Map-matching theory and methodology

It is possible to classify map-matching algorithms as global or incremental ones. While producing the solution, global algorithms batch process the whole input trajectory. Incremental algorithms use localization techniques that partition and sequentially filter the input trajectory into smaller segments, often resulting in a suboptimal solution. Theories that have been applied in map-matching range from geometrical analysis [CHEN, Daniel ET AL., p. 75-83], belief function theory [NASSREDDINE ET AL., 2009, p. 494-499], Extended Kalman Filter [OBRADOVIC ET AL., 2006, p. 111-122] to Hidden Markov Model (HMM) [PINK ET AL., 2008, p. 862-867]. In particular, HMM-based algorithms is robust to noise and sparseness.

Researchers [MEERT ET AL., 2018] proposed a new map matching approach using HMMs with non-emitting states, a smart pruning strategy such that only non-emitting states with the highest relevance are visited. Such properties are promising to apply on map-matching for GPS traces. In this part, an open-source algorithm based on python, Leuven Map-Matching [MEERT ET AL., 2018] algorithm will be adopted for linking the trajectories with existing road network.

Basic flow of the HMM algorithm:

The Hidden Markov Model (HMM) is a Markov mathematical model in which it is presumed that the system being modelled is a Markov process with intangible ("hidden") states. HMM implies that there is another element whose activity "depends" on a process of Markov. In this case, the candidate paths are generated and evaluated on the basis of their probabilities sequentially. Once a new point of trajectory occurs, older hypotheses are updated to cover the newly observed results. The surviving route with the greatest joint likelihood is then chosen as the final solution of all candidates in the last point. The optimal output for map-matching is a sequence of states that corresponds to a sequence of observations.

Basic components in a map-matching problem:

- a. Trajectory is a sequence of trace points which is already prepared as the resulting trajectories from section 3.3. Each data point is associated with attributes of longitude, latitude in degree at every time step. The coordinate pairs of longitude and latitude are stored in the new column of 'geometry' as mentioned before.
- b. Edge is a polyline indicating a road segment, which defined by two nodes in the network matrix. These edges composed a sequence of states. Besides the coordinates of nodes, edges contain the other information of the representing road segment, such as bearings, permissibility of bi-directional travel, etc.
- c. Map is the umbrella collection of edges, representing the whole road network of the research area.

3.5.2 Implementation and results

Above introduced elements form the input data of the algorithms. Before proceeding the map-matching program, two more parameters are remaining to define. For each trajectory point, the algorithm will find all candidate segments within a radius of a searching distance around it. The reasons for imposing this distance are two-fold: (1) to discard all candidates with very low emission probabilities, which indicates observing the GPS point conditional on the candidate segment is a false match. (2) to prevent penalty in execution speed as a consequence of excess candidates [GOH ET AL., 2012, p. 776-781]. Another parameter is the maximum distance of a single observation, which restricts the allowable length on an observing candidate of edges. These two parameters perform the function in the theoretical model.

At the beginning of an observation, the set of potential states is built based on the searching radius and the initial probabilities for these states are progressively determined for all edges within the maximum permitted distance. Noticing the high sampling rate of the trajectory dataset and the compact network architecture, five metres seems to be a considerable distance. As depicted in Figure 3.5, the longest edge in the network is about three hundred meters. This value is chosen as the first parameters to limit the searching horizon.

Figure 3.10 demonstrates the working flow of map-matching procedure. The input data are the resulting network graph and matrix as well as the trajectory files for all five time blocks. Thus, the implementation will produce five matched network and five matched trajectory files respectively.

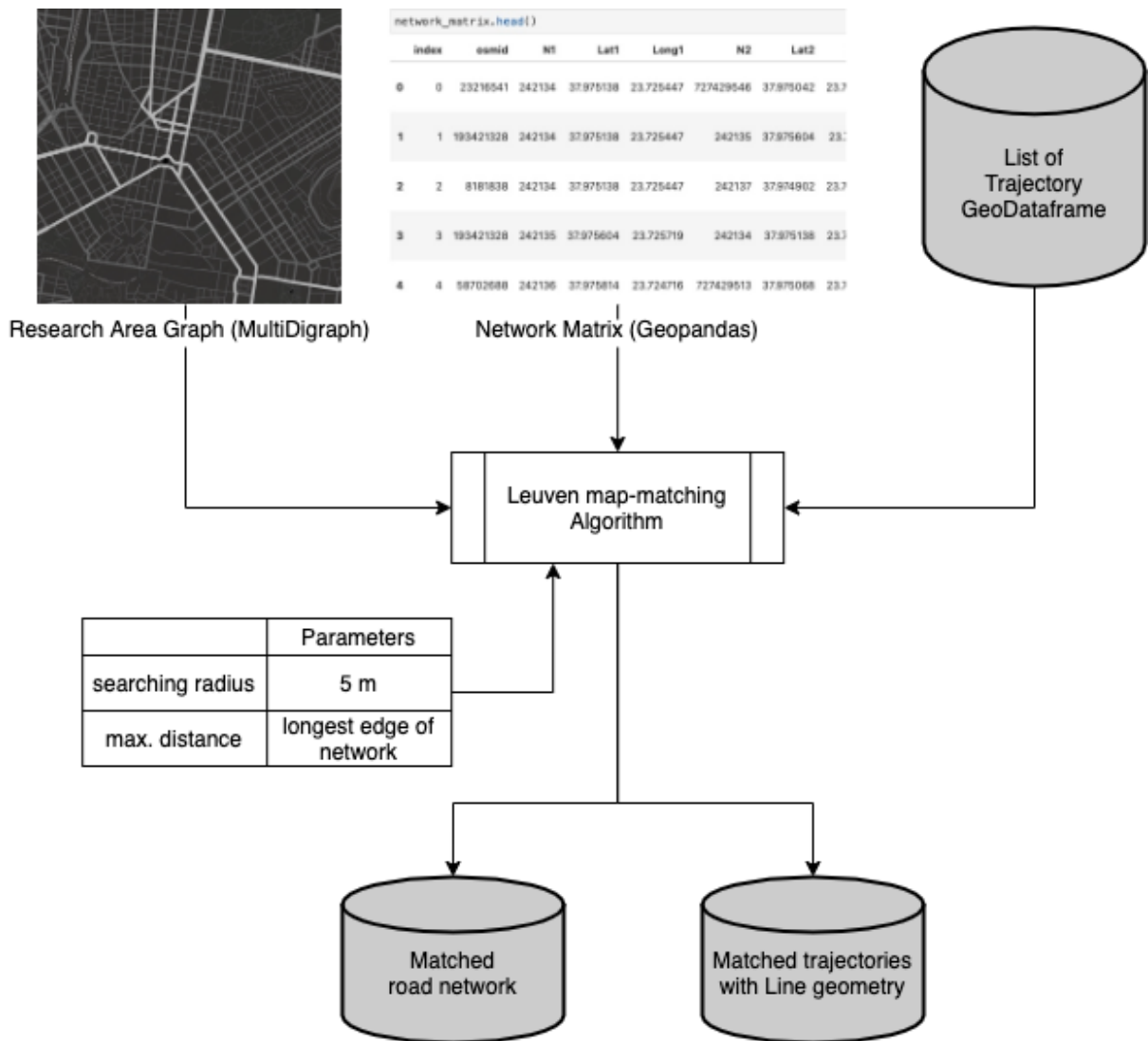


Figure 3.10 Implementation of map-matching algorithm

3.6 Extraction of traffic characteristics

3.6.1 Virtual loop detector

This part is inspired by the methods proposed by Mr. Joachim Landtmeters in his thesis “Analyzing mixed urban traffic by linking large scale trajectory dataset to underlying network” (source: https://pNEUMA_mastersproject.readthedocs.io/). By installing virtual loop detector on the road network, it is possible to extract traffic characteristics at any location in the network.

Loop detector is commonly used for traffic management as a data base for a range of uses in organisational and planning analysis of traffic knowledge generation, validation of transport demand models. The dual-loop detector consists of two single loop detectors that are separated at a fixed short distance, and this arrangement provides a possible real-time data source for speed and vehicle classifications for the dual-loop detector data.

Based on the concept and function of loop detector, virtual loops are special edges installed on the digital road network map. They are pair of two lines orthogonally placed on a network edge. The configuration includes the following parameters: number of detectors per road edge, width and spacing of loop detectors. According to Joachim Landtmeters’s design, the distance from intersection is also required to prevent any misplacing due to the simplified geometry representation. Figure 3.11 shows the configuration of virtual loop detector on one segment of 3rd Septemvriou Street. The three parallel lines in orange are the virtual detectors, with a spacing of 4 meters for each pair. The underlying blue line represents the edge of road network and white dots are a trajectory of vehicle which travelled through these three detectors. Table 5 gives all geometric settings for all virtual loop detectors in the network. The resulting of implementation is shown in Figure 3.12.

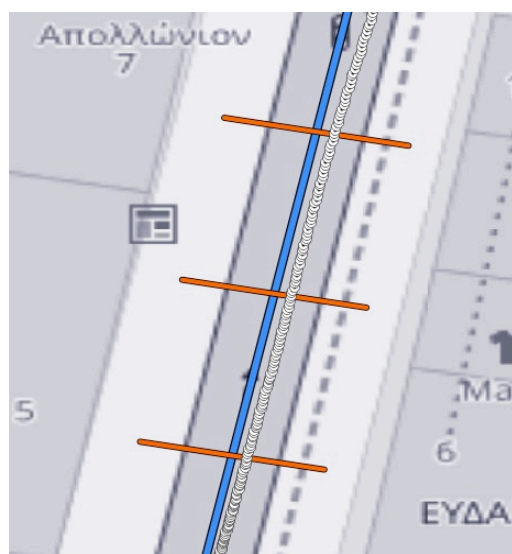


Figure 3.11 Configuration of virtual loop detector

Table 5 Parameters of the virtual loop

| Parameters | Value |
|----------------------------|------------|
| Number of detectors | 3 per edge |
| Double loops | True |
| Distance from intersection | 15 m |
| Width of loops | 15 m |
| Spacing of loop pairs | 4 m |

**Figure 3.12 Result of virtual detector installation**

3.6.2 Traffic parameters extraction

Traffic states can be expressed with three basic characteristics, namely, flow (quantity per unit time), density (quantity per unit space) and speed (space per unit time) by assuming the traffic stream as a fluid continuum [EDIE ET AL., 1965, p. 139-154]. The average flow and density are defined respectively as:

$$\text{Flow} = \frac{\text{Total distance travelled by counted vehicles}}{\text{Space - time window}} \quad (3.1)$$

$$q = \frac{\sum_{i=1}^n x_i}{|A|} \quad (3.2)$$

$$\text{Density} = \frac{\text{Total time spent by counted vehicles}}{\text{Space - time window}} \quad (3.3)$$

$$k = \frac{\sum_{i=1}^n t_i}{|A|} \quad (3.4)$$

And the average speed of traffic flow can therefore be deduced:

$$\text{Average speed} = \frac{\text{Flow}}{\text{Density}} \quad (3.5)$$

$$v = \frac{q}{k} = \frac{\sum_{i=1}^n x_i}{\sum_{i=1}^n t_i} \quad (3.6)$$

The space-time window is determined by the sampling interval of loop detectors and the spacing of double loops. As shown in Figure 3.13, the vehicle trace is represented in blue dotted line. When it travels across the space-time window, the x_i, t_i will be counted by the virtual detector. The width of the detector was set to 15 m to cover the whole width of the lane.

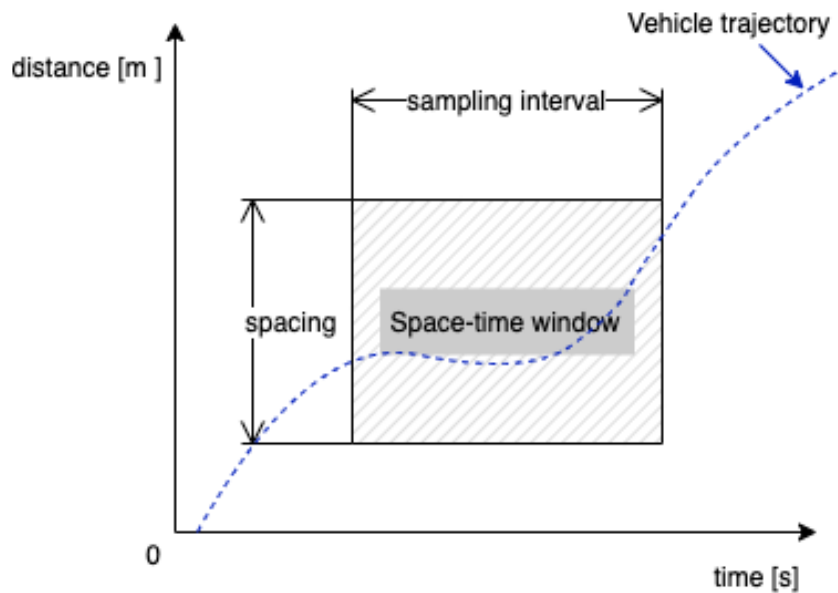


Figure 3.13 Trajectory of the i^{th} vehicle in the time-space window

3.6.3 Results analysis

The input is the trajectory data which has been transformed into connected line segments in section 3.5.2 and the detector information generated in section 3.6.1. Considering of the duration of each single trajectory file is only fifteen minutes, the sampling interval is set as thirty seconds. Too long interval will consume considerable computing time and high missing rate. The output matrix includes all traffic parameters from installed detectors with timestamp. For instance, the traffic characteristics extracted by detector 188 (red line in) on road 'Stadiou' at intersection of 'Karagiorgi Servias'.

Table 6 Statistics of traffic parameters of detector 188

| | Flow [veh/h per lane] | Density [veh/km per lane] | Speed [km/h] |
|---------|-----------------------|---------------------------|--------------|
| Mean | 183.993962 | 9.471069 | 13.139245 |
| Std. | 232.143146 | 13.546938 | 20.393256 |
| Minimum | 244.020000 | 0.000000 | 0.000000 |
| Maximum | 735.840000 | 48.283333 | 100.000000 |



Figure 3.14 Location of detector 188 (red line)

On the left of Figure 3.15, the graphs demonstrate the traffic density, flow rate and average speed of the vehicles that passed through detector 188 of the period from 8:00 to 8:30 am on 24/10/2018, respectively. On the right, the chart reflects the next time session from 8:30 to 9:00 am at the same location. It is need to pointed out that the time axis is only an approximation of

the metadata from pNEUMA experiment, since the metadata only provides a time interval of sampling, without a starting time or end time. Comparing two flow charts, the mean flow rate increased by 223.9% as the rush hours reached.

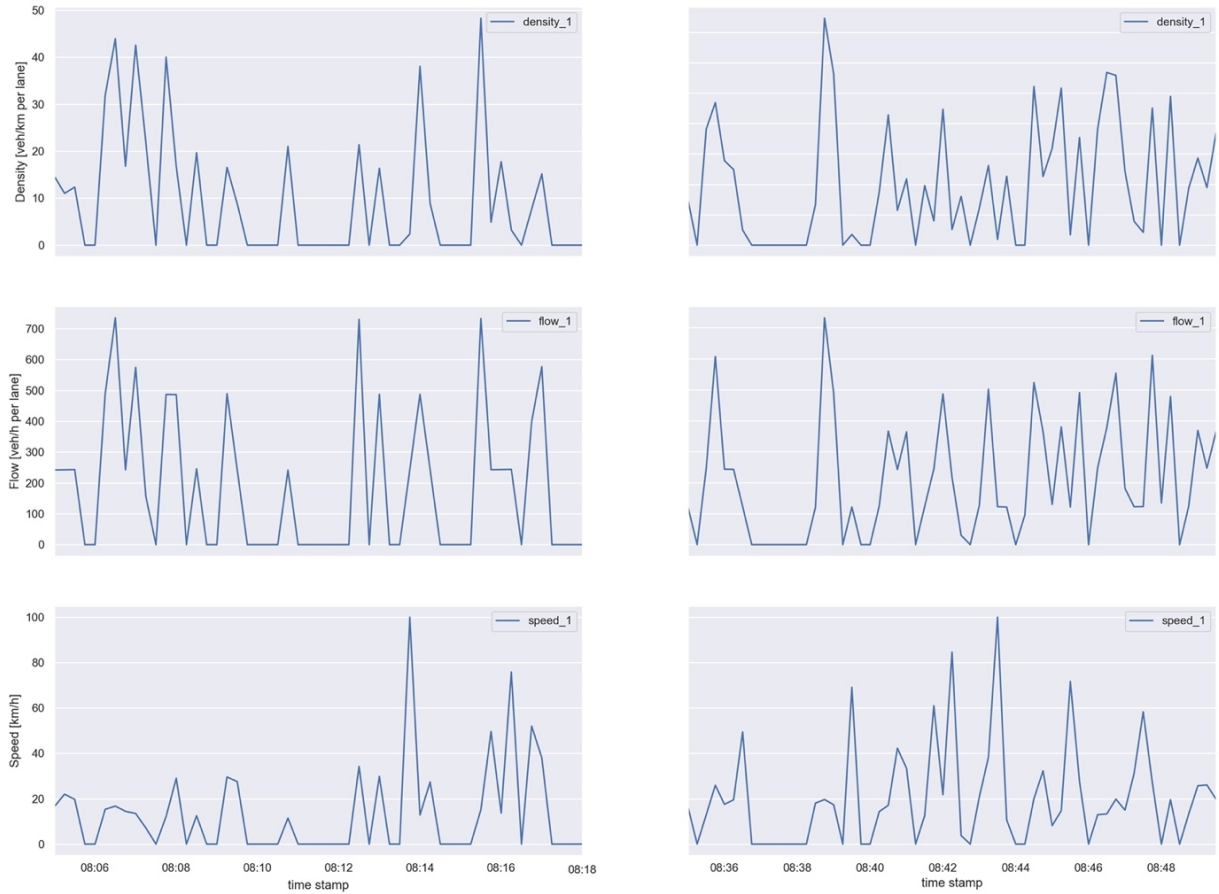


Figure 3.15 Flow, density and speed time series at detector 188

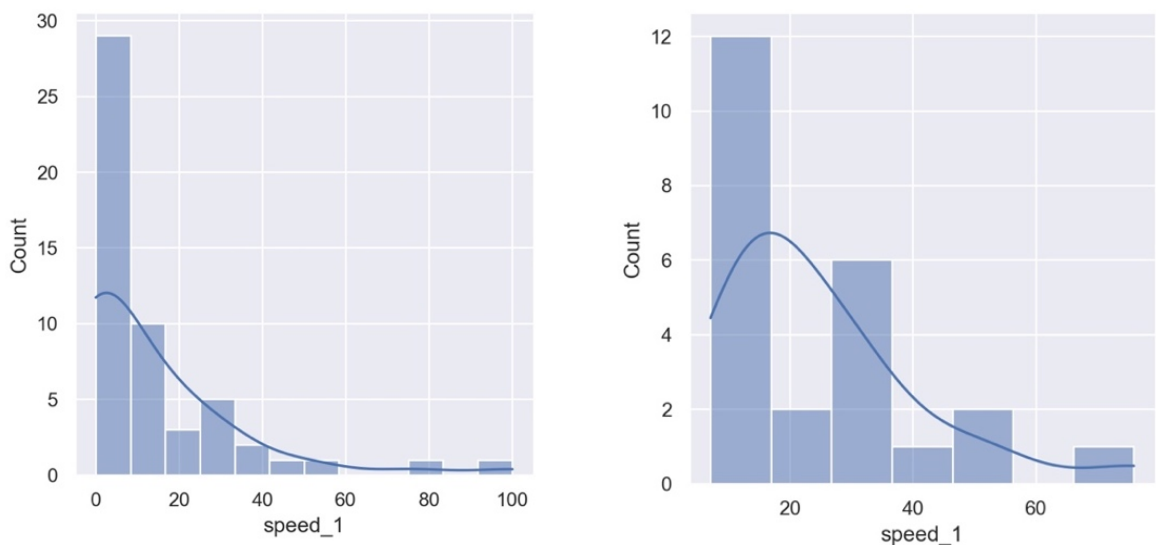


Figure 3.16 Histogram of speed
Left: with outliers, Right: after filtering

It is noticeable that the speed has outliers with a value of 100 km/h (see in Table 6), which needs to be filtered out. Those whose quantile falls between 0.01 and 0.99 will be kept and others will be discarded as outliers.(shown in Figure 3.16)

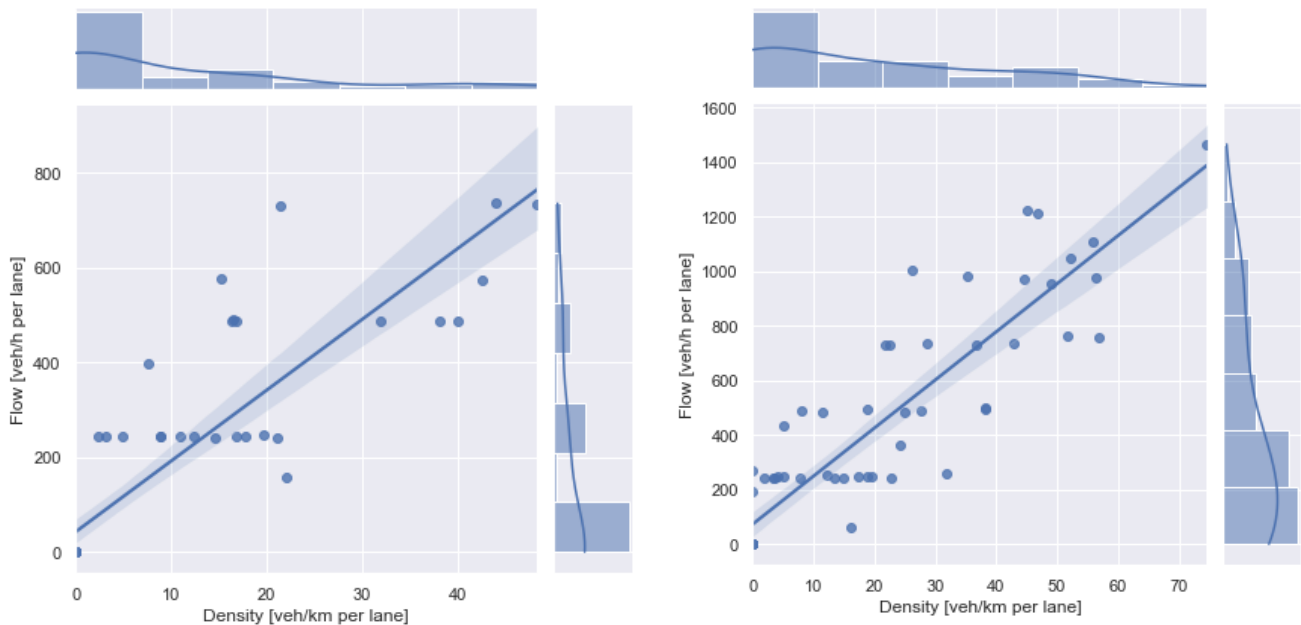


Figure 3.17 Relation between density and flow

Left: 8:30 – 9:00 AM, Right: 9:00-9:30 AM

Figure 3.17 demonstrated the density-flow relations at different time. On the left, it is clear that the flow and density of traffic on the road segment is under saturation and remains a medium speed during observing time. On the right chart, the highest flow once doubled the value from previous recording time phase, nearly reaching 1500 veh/hour/lane. This indicated that at this location, the traffic demand is increased significantly due to the morning rush. In addition, the speed remains at a relative high level which means the capacity is not fully exploited. It can be assumed that in the next time session, there might form congestion at this spot, as the demand will still be increasing for a while.

3.7 Summary

The main objective of this chapter is to extract the traffic parameters of the trajectory data in the road network. The traffic flow parameters are extracted for all matched road segments by map matching and placement of virtual loop detectors. The results include traffic flow, speed and density data of the placed detectors for each road section. In addition, the difference of the parameters of one of the loop detectors in two time periods is analysed and outliers are excluded.

The traffic indexes corresponding to all the road sectors where vehicles have travelled have been acquired, and this data is the base stone for the subsequent spatial and temporal analysis of congestion in the network.

4 Congestion Evolution in spatial temporal spectrum

4.1 Metrics for congestion identification

4.1.1 Theory background

Proper identification of various traffic flow states is one key task of the virtual loops. As stated in Chapter 2, it is difficult to describe distinct characteristics of various traffic flow states in only one parameter. In designing the algorithm to define traffic states, a heuristic approach with a mixture of two or three primary variables directly arising from detector data is implemented. And in such a dense urban area with hierarchical road system, thresholds should reflect the complexity of the network.

The results from extracting traffic parameters contain the major variables representing the characteristics of traffic streams, which includes density, speed, flow rate and vehicle counts. The state distinguishing puts the congestion detection as priority and other states defined as free flow and intermediate states.

4.1.2 Traffic state distinguished

According to Adolf D.[MAY ET AL., 1990] there are three types of traffic states:

- a. Free flow regime: high velocity with low traffic volume and density
- b. Intermediate regime: maximum traffic volume at critical density and velocity
- c. Congested-flow regime: high traffic density with low traffic volume and velocity

For congestion measures, both vehicle density and average speed can be used. Congestion can occur when a certain critical level is exceeded by the density of vehicles on the lane. If the observed density is greater than this threshold, this virtual detector and the its road segment will then be labelled as congestion-related. Additionally, when comparing the measured average speed and lower speed limit based on the road classification, if the measured average speed is smaller than this minimum free flow speed, this loop detector will then similarly be defined as congestion. This study adopts the “Urban Road Traffic Management Evaluation Index Systems Manual” as reference, combing the proposed values by other relevant researches [D'ANDREA ET AL., 2017, p. 43-56, BELLOCCHI ET AL., 2020, p. 4876]

4.1.3 Establishing of Thresholds

According above empirical studies, the following values are chosen to define the state of congestion in this research:

- a. Free flow criteria

$$\text{if free flow speed} \geq \text{speed} > 60\text{km/h} \quad (4.1)$$

Then the traffic stream at this specific location in time interval is considered as free flow. The time interval has been introduced before, representing a time window of observing and is set to 15 seconds.

- b. Intermediate criteria

$$\text{if } 60 \text{ [km/h]} > \text{speed} > 40\text{[km/h]} \text{ and flow} > 900\text{[veh/h/lane]} \quad (4.2)$$

the traffic stream at this specific location in time interval is considered intermediate state. The traffic speed of the such state of flow is relatively low compared with that of the free flow traffic.

- c. Congestion criteria

To handle with different road types, the detector index can be used to retrieve the location and return the classification of the specific road segment.

$$\text{if [detector index]} \in \text{[primary road index]}$$

$$\text{and if speed} < 30\text{[km/h]} \text{ and vehicle counts} > 0 \text{ and density} > 65\text{[veh/km/lane]} \quad (4.3)$$

Then the segment is recognized as congested state. And for secondary roads and others are calculated by the formular (4.4) and (4.5), respectively:

$$\text{if [detector index]} \in \text{[secondary road index]}$$

$$\text{and if speed} < 25\text{[km/h]} \text{ and vehicle counts} > 0 \text{ and density} > 60\text{[veh/km/lane]} \quad (4.4)$$

$$\text{if [detector index]} \in \text{[other index]}$$

$$\text{and if speed} < 20\text{[km/h]} \text{ and vehicle counts} > 0 \text{ and density} > 50\text{[veh/km/lane]} \quad (4.5)$$

The entire algorithm for the recognition of the above three distinct traffic states with dual-loop data represented by a graph based on the above rules.

4.2 Congestion identification

4.2.1 A closer look at time session 8:00 – 8:30 am

The Table 7 Table 7 Eligible congestion spots shows that there are five locations collected as congested spots. The mean value is calculated during the whole observation with a series sampling with a time step of fifteen seconds. The criterial proposed in the previous section focus on the individual space-time window. Due to the impact of traffic signals, the fluctuation can vary in a wide range. This is one of limitations of the identification algorithm.

Table 7 Eligible congestion spots

| Index of congested detectors | Mean speed [km/h] | Mean flow [veh/h/lane] | Mean density [veh/km/lane] |
|------------------------------|-------------------|------------------------|----------------------------|
| 135 | 4.88 | 84.01 | 79.66 |
| 186 | 21.79 | 415.32 | 41.34 |
| 365 | 23.77 | 311.96 | 30.53 |
| 82 | 16.28 | 379.22 | 50.42 |
| 70 | 28.62 | 505.34 | 39.01 |

Figure 4.1 shows the location of these detectors in the road network. Detector 135, 82 are on secondary roads and the rest of them are located at primary roads.



Figure 4.1 Location of congested spots in road network (red line)

To investigate the traffic situation, time series graph will demonstrate the trends and fluctuation of speed over time. Figure 4.2, Figure 4.3, Figure 4.4, Figure 4.5, Figure 4.6 and Figure 4.7 illustrated the speed-time changes during 8:00am to 8:30am on October 24th 2018 and histogram with KDE of speed respectively. The distribution of speed shows that majority falls between 10km/h and 20 km/h. As for detector 135, the traffic nearly stopped moving from 8:12 am and lasted for four minutes, which is an obvious jam on the link. Given the fact that detector

135 is located at a secondary link to a five-arm signalized intersection, the congestion might cause by the downstream from the connected primary roads.

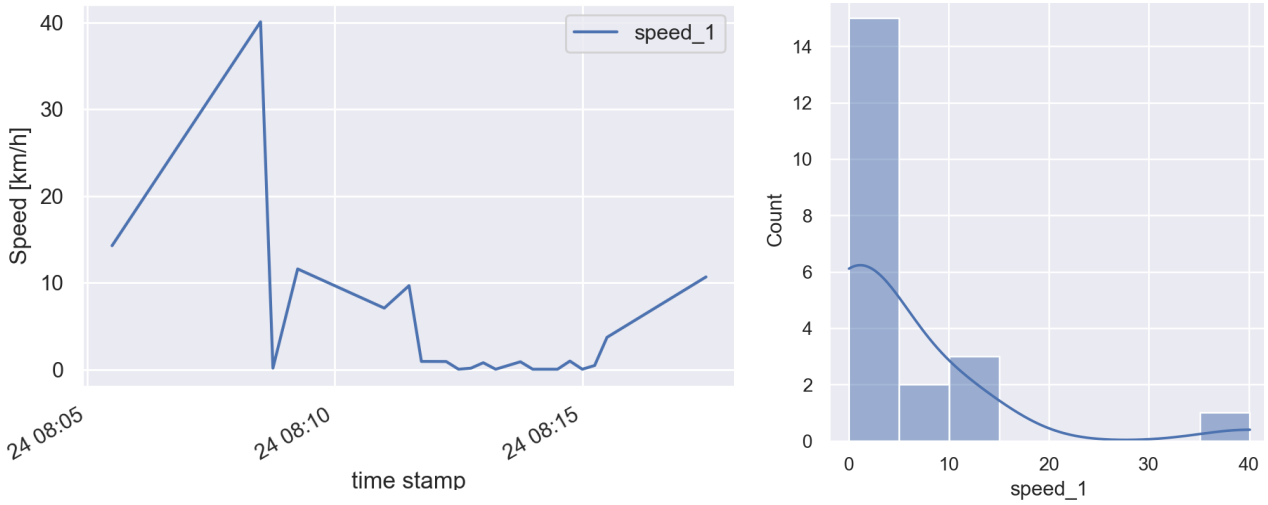


Figure 4.2 Detector 135: speed time plot (left) and histogram of speed (right)

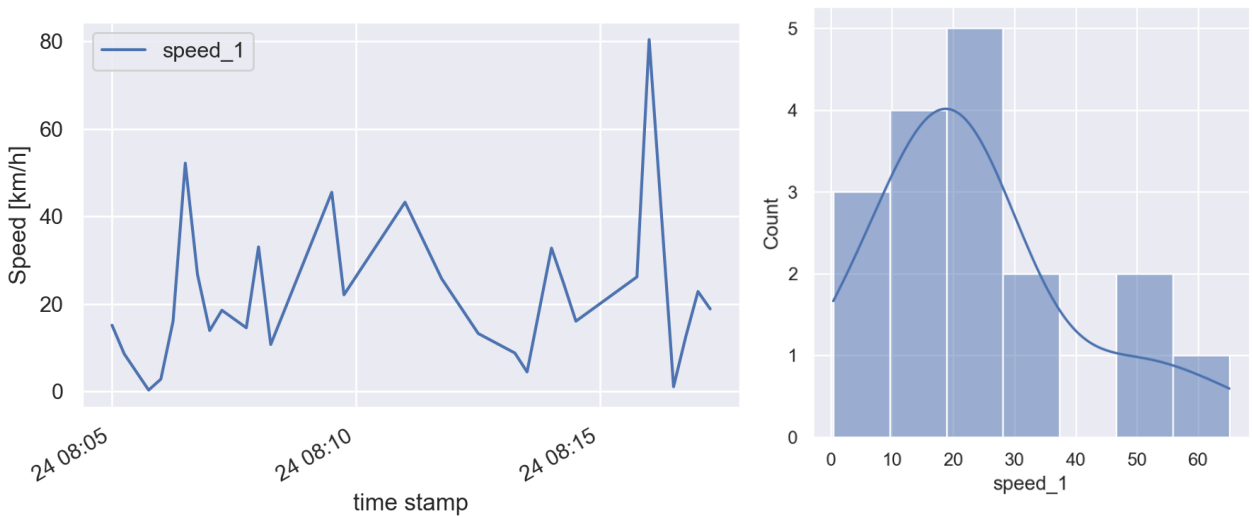


Figure 4.3 Detector 186: speed time plot (left) and histogram of speed (right)

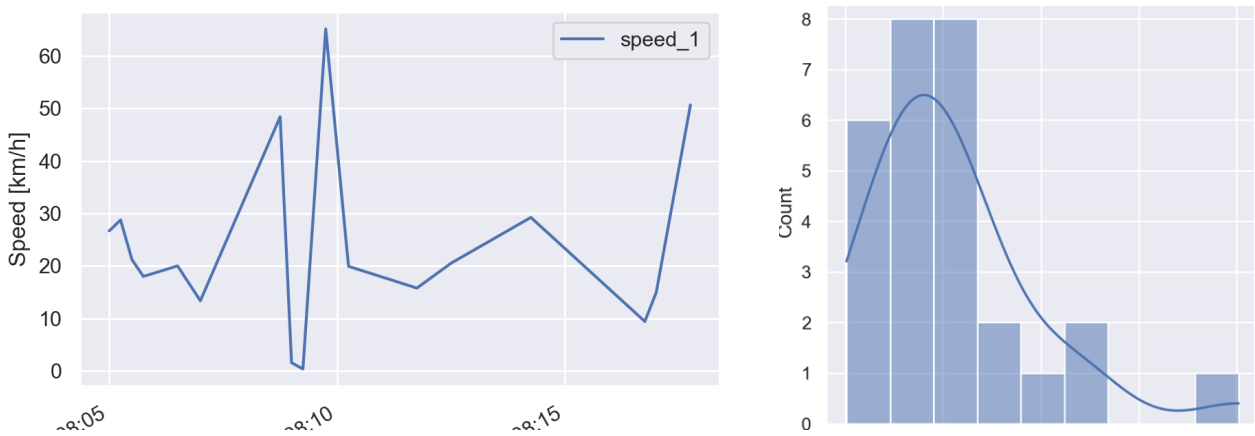


Figure 4.4 Detector 365: speed time plot (left) and histogram of speed (right)

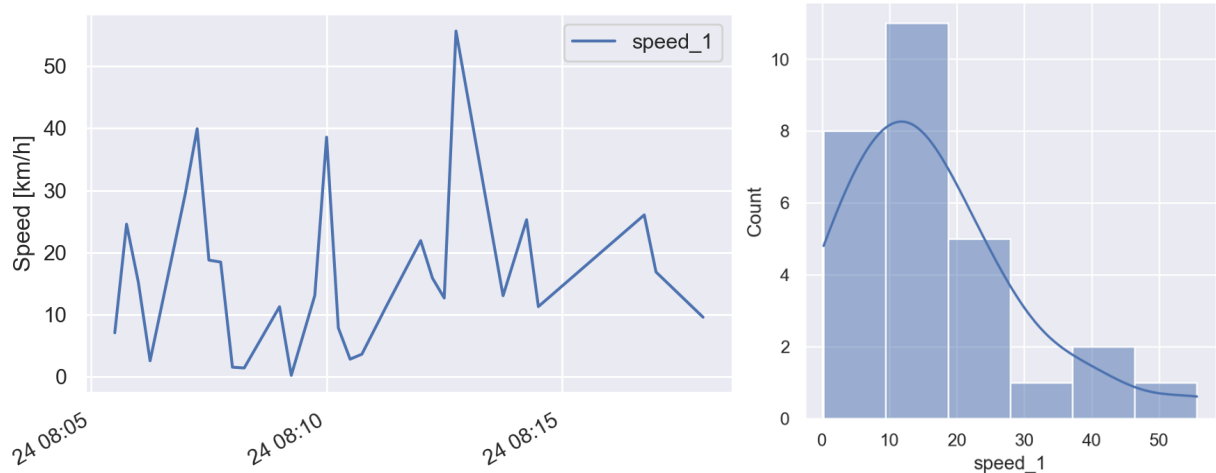


Figure 4.5 Detector 82: speed time plot (left) and histogram of speed (right)

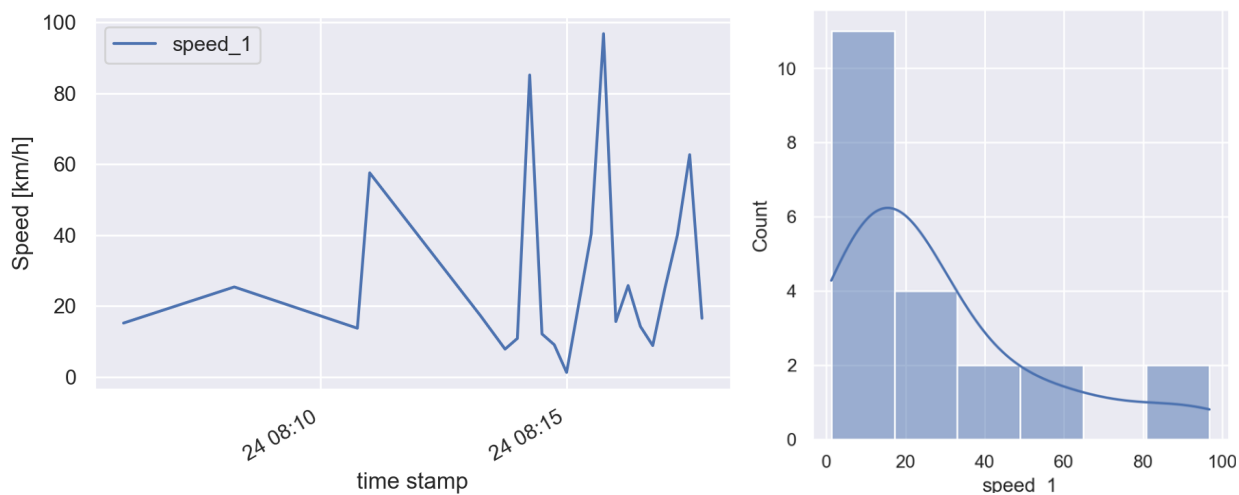


Figure 4.6 Detector 70: speed time plot (left) and histogram of speed (right)

4.3 Congestion evolution

4.3.1 Spatial distribution

This part will investigate a road as example to explore the spatial pattern of traffic flows. In the meanwhile, potential spot of traffic congestion will be identified.

The 3rd Septemvriou Street is chosen as the study object, a one-way road connecting to the Omonoia Square, which is an important shopping centre and transport hub in the study area of Athens. The Omonoia Square is located at the centre of the city at the intersection of six main streets; one of them is 3rd Septemvriou Street. Since the aim is to explore the spatial distribution of traffic jams, the time window is limited within a fifteen-minutes long observation. There are eleven virtual loop detectors installed along the street. Specifically, the classification of road changes from primary to secondary at from detector 32, which means a reduction of lanes from three to two. Consequentially, the threshold of congested state is varied along the entire street, as determined in section 4.1.3. As shown in Figure 4.4, only the travel from south to north is premissible. The city centre is located to eastsouth corner relatively (not shown in the figure). Table 8 lists a spectrum of data, which provides an index of detectors involved and the mean values of traffic characteristics for three time sections, 8:00-8:30 am, 8:30-9:00 am and 9:00-9:30 am. The mean values are calculated within an effective observation phase, i.e. 15 minutes. The orange-colored cells show all values higher than 400 veh/h/lane; the blue ones indicate that the speed is under 25km/h on secondary links and under 30km/h on primary links; and the dashed line means there were no records during observation.

Table 8 Traffic parameters from all eleven detectors on 3rd Septemvriou Street

| Index of De- tector (North to South) | Mean density [veh/km/lane] | | | Mean flow [veh/h/lane] | | | Mean speed [km/h] | | | |
|---|-------------------------------|---------------|---------------|---------------------------|---------------|---------------|----------------------|---------------|---------------|-------|
| | 8:00- 8:30 | 8:30- 9:00 | 9:00- 9:30 | 8:00- 8:30 | 8:30- 9:00 | 9:00- 9:30 | 8:00- 8:30 | 8:30- 9:00 | 9:00- 9:30 | |
| Secondary (2) | 1290 | 14.89 | 2.87 | -- | 204.24 | 76.26 | -- | 16.45 | 37.53 | -- |
| | 1033 | 40.11 | 27.84 | 22.52 | 384.92 | 643.92 | 534.39 | 20.06 | 25.55 | 26.28 |
| | 125 | 16.28 | 33.46 | 22.08 | 390.08 | 1061.48 | 506.04 | 27.42 | 33.05 | 27.08 |
| | 102 | 19.31 | 19.12 | 17.58 | 538.68 | 495.95 | 553.14 | 36.42 | 25.74 | 29.25 |
| | 40 | 19.31 | 17.80 | 16.95 | 444.17 | 542.77 | 542.72 | 26.02 | 29.42 | 31.18 |
| | 32 | 20.75 | 22.45 | 23.10 | 514.23 | 397.23 | 570.06 | 26.45 | 20.71 | 29.23 |
| Primary (3) | 58 | 17.05 | 13.61 | 13.16 | 393.26 | 348.93 | 355.86 | 24.23 | 35.46 | 27.49 |
| | 1540 | 9.77 | 15.57 | 14.03 | 122.12 | 353.41 | 300.72 | 18.96 | 28.58 | 23.77 |
| | 49 | 14.76 | 20.21 | 14.38 | 339.93 | 467.95 | 441.96 | 30.99 | 23.28 | 32.64 |
| | 51 | 13.47 | 11.51 | 17.78 | 279.85 | 188.89 | 275.23 | 37.39 | 19.17 | 18.79 |
| | 52 | 26.32 | 28.86 | 20.13 | 357.98 | 275.36 | 334.99 | 15.86 | 17.34 | 18.59 |

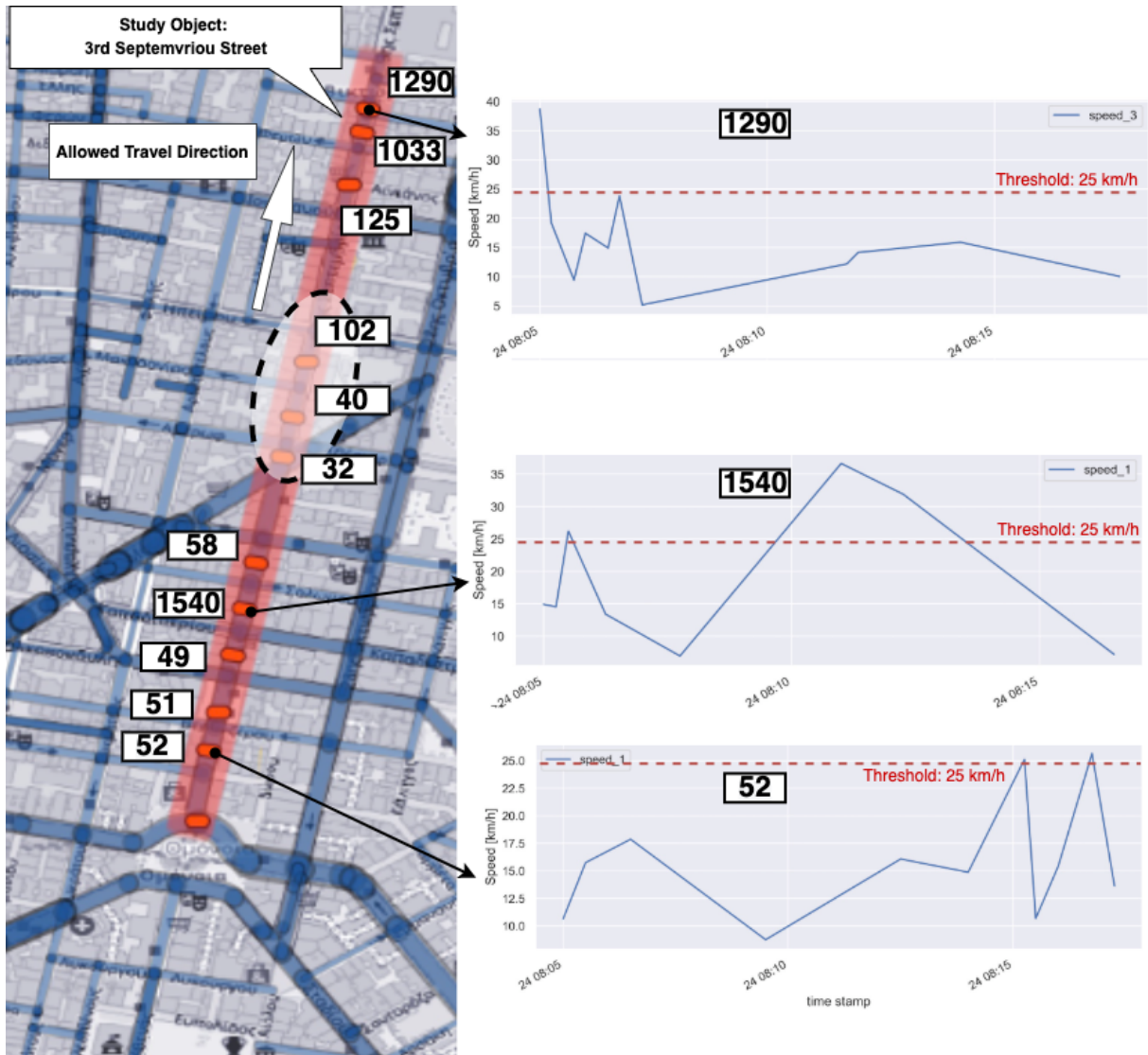


Figure 4.4 Layout of 3rd Septemvriou Street and the speed-time graph for detector 20, 8 and 105 (8:00 – 8:30)

Interestingly, the mean density at detector 1033 (purple-coded cell in Table 8) is exceptionally high. After investigating the metadata, there existed an almost stationary car that has been on the detector for thirty seconds long. According to equation (3.4), it is clear that the longer the time t_i of a vehicle travelled, the greater the density of traffic, since the ‘space-time’ window is constant. As a result, the local density reached 255.57 veh/km/lane, leveling up mean value in the whole duration.

From 8:00 to 8:30, the speed time series graphs in Figure 4.4 described the average traffic speed with sampling rate of fifteen seconds in duration of fifteen minutes. Specifically, the speed data of detector 1290 has a rather low-speed state of traffic stream for almost ten minutes. Their mean speeds during this period are all lower than 20km/h. To investigate these three spots which meet the critical threshold of congestion, the more information is plotted on a flow-density graph (Figure 4.5). On detector 1540, there are only eight records with ten vehicles passed through this spot in the whole observation phase. One of the reasons can be the

low level of traffic demand on this segment with an outbound travel direction. As for detector 52, the majority of speed records are below 20km/h. Particularly at around 8:15am, the speed nearly dropped to zero. By examining the other two detectors installed on the same edge, the speed is logged as 8.75km/h, 6.09km/h and eventually 0km/h. This is most likely a stop at signalized intersection. Focusing on detector 52, the relevant road segment is only one block away from the Omonia transport hub. The travel speed values are all slower than 25km/h due to radiative and aggregate effect of the gathering urban node.

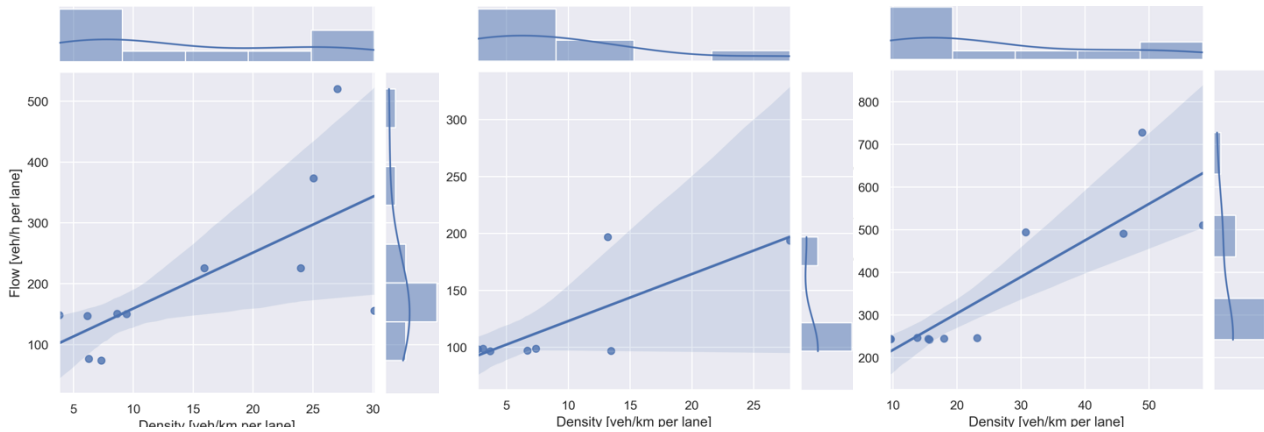


Figure 4.5 Flow-density joint distribution (8:00– 8:30)

Left:detector 1290

Middle:detector 1540

Right: detector 52

Given the fact that relatively high flow rate occurred at detector 32, 40 and 102, it can be implied that the spatial distribution of traffic streams is affected by the neighboring conjunctions, namely the diagonally crossed road intersecting with 3rd Septemvriou at detector 32. As the downstream of both 3rd Septemvriou and another arterial, the volume on detector 32, 40 and 102 (highlighted in dotted circle in Figure 4.4) will be increased consequently, reflected by the mean traffic flow rate in Table 8. Another factor also contributes to such phenomena, which is the classification of road downgrades to secondary with two lanes per direction from three-lane class.

By far, there is no clarified evidence of traffic congestion on this link road at this time duration except for detector 52. However, the spatial distribution of traffic flow reveals the potential bottleneck which prone to form an over-saturated traffic state, which is the road section of detector 32, 40 and 102. Here is only an assumption. And as time goes to 9:00, increasing traffic demand will push the capacity of road to its limit. As many studies and models proved previous, the bottleneck will be formed due to the road geometry or layout of network. As a result, the density will be rising synchronously until it reaches the maximum capacity of the road.

Therefore, the analysis on evolution will proceed in the next section, taking time variable into account.

4.3.2 Temporal distribution

As studied in empirical researches [SUN ET AL., 2011, p. 86-93, AN S ET AL., 2018], the evolution of congestion can be described as four or five phases by the status of traffic flow over time.

(1) Congestion start phase: travelling speed starts to decrease lower than free flow; travel time becomes longer; and queue tends to form. The traffic parameters collected from detectors start showing values which meet the criteria of a congestion state, yet the quantity and duration is at a low level with a growing trend.

(2) Congestion propagation phase: the congested state in one detector move to the surrounding ones within a certain time interval. Such propagation can be observed along a street, in a local area or in the whole network. The queue length increases steadily, but the length is still not very long. Vehicles barely have to stop twice within same intersection.

(3) Congestion Peak phase: traffic flow spreads easily to the upstream period, and almost all vehicles have to stop at least two times within the same intersection. The number of congested states reported from all detectors reaches the maximum within the same given time interval.

(4) Congestion Dissolution phase: it is accompanied by the decreasing traffic demand; queue length decreases and traffic flow is back to stabilize. The recurrent congestion commonly occurs in the context of urban area, due to the rush hours on every weekday. And the patterns can be therefore summarized as above.

In the case study, from the upcoming phases 9:00-9:30am and 9:30-10:00am, the congested detectors and their linkages can be identified as the rush hour comes. According to the threshold defined in section 4.1.3, detector 125 meet condition with all three parameters with a higher value than the critical ones. Detector 102,1033 and 1290, on the other hand, partly fulfil the condition for jam determination.



Figure 4.6 Spatial distribution of congested linkages at phase 8:30 to 9:00 (circled area)

Reviewing the traffic state of all detectors at phase one and phase two, it reveals a preliminary conclusion, that the formation of congestion along this specific road is from the intersection of 3rd Septemvriou Street and the diagonally crossed primary road, i.e. Marni Street to the downstream on the rest segments of the road. By now, the congestion is starting to form and propagate to the connected collector roads and the grided linkages. Figure 4.6 circled the congested detectors. From the spatial perspectives, the congestion formed itself as the traffic demand exceeds the service capacity of the infrastructure.

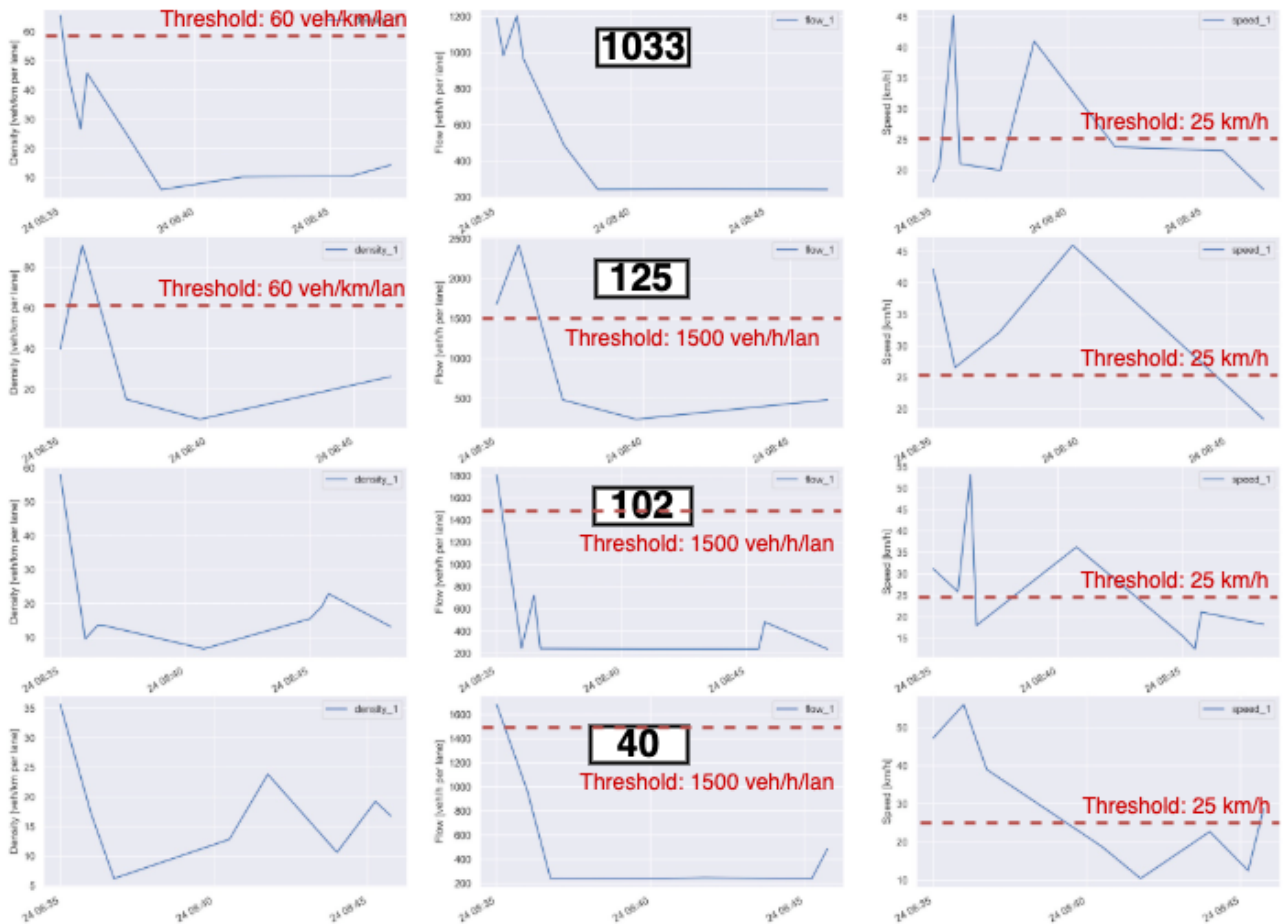


Figure 4.7 Density, flow and speed on congested location

From up to bottom: detector 1033, 125, 102 and 40

From left to right: density, flow, speed

The speed-time graph for detector 40 from 8:00 to 9:30 is shown in Figure 4.8. The fluctuation indicates that the traffic flow is becoming unstable. The part below threshold represent the duration of congested state, which has a decreased trend. The same situation also occurred to its upstream link, represented by speed-time of detector 32 (Figure 4.9), detector 58 (Figure 4.10). The locations of these detector are situated at both downstream and upstream of the merging point of two streets. The congestion pattern of these three consecutive spots reveals that the rush hour is passing and saturation of the segment is gradually falling back to its capacity.

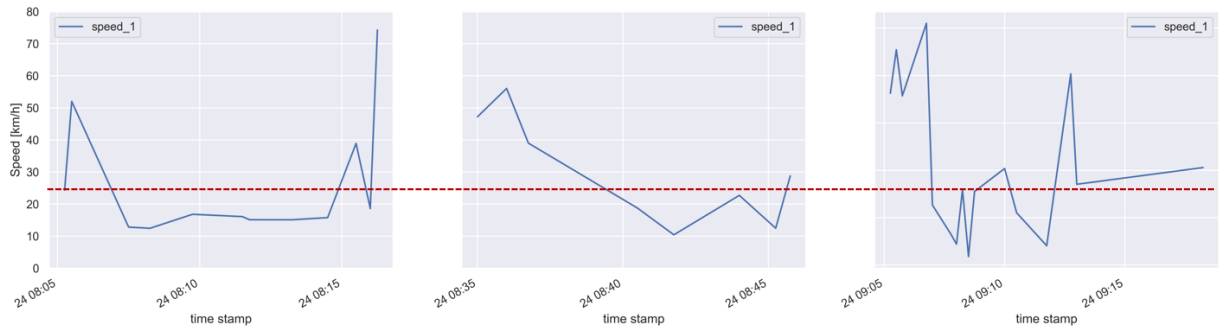


Figure 4.8 Temporal distribution of speed on detector 40 (threshold: red dotted line)

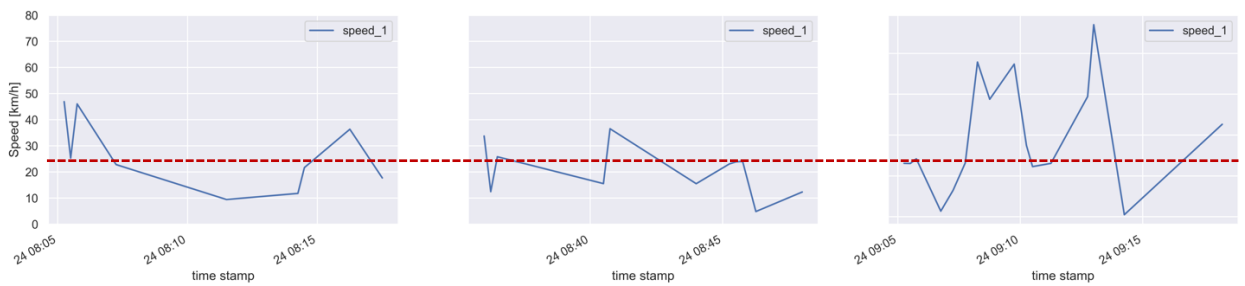


Figure 4.9 Temporal distribution of speed on detector 32 (threshold: red dotted line)

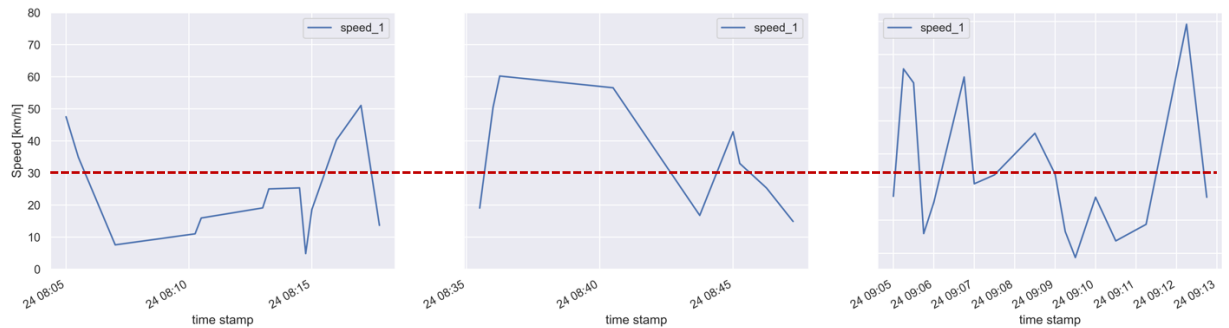


Figure 4.10 Temporal distribution of speed on detector 58 (threshold: red dotted line)

4.4 Summary

This chapter explored the evolution of congestion from the perspective of space and time. The case study focused on a south-north direction corridor, 3rd Septemvriou Street. The results reveal that, as the rush hour comes, the congestion starts to form initially at the segments of detector 32, 40 and 102 from 8:00 to 8:30am. The reason includes not only the increasing traffic demand but also the narrowing of road geometry, from a three-lane primary road to a two-lane secondary road. As for the temporal distribution of congestion, the duration of congested state of detector 40, 32 and 58 showed a decreasing trend through the observation period. The longest congested time was up to almost ten minutes, which found on detector 1290 at the first phase of surveillance.

5 Conclusion and future work

5.1 Conclusion

The thesis conducted a trajectory-based traffic state estimation and congestion identification.

Results can be concluded as four aspects:

1. Create a road network with geometry attributes and customized layers, e.g. tags for traffic signals, primary and secondary corridors.
2. Adopt Hidden Markov Model-based algorithm to match the trajectory data onto the road network; align the GPS point to the identify the true traversed paths.
3. Extract key traffic parameters, i.e. flow rate, density and average speed on each installed virtual detector
4. Analyzed the spatial and temporal congestion distribution and evolution by a newly proposed traffic state distinguish methodology.

These achievements will be benefit for the related academic research as well as the traffic management authorities. Negative effect of traffic congestion problem urges all of us to make breakthrough in every potential field of relevance.

5.2 Limitations and future work

In terms of data processing, the feed-back evaluation for results of the map-matching was not conducted yet. The mismatching of traces on the links may lower the precision when calculating the traffic flow characteristics, which is to be improved in the future. As for the conversion from point GPS data to lined-up traces, the microscopic attributes will be eliminated, including vehicular headways and maneuvering activities. After this transformation, the discrete state is invisible and unavailable from the processed dataset, which will limit the further research on queue profiling estimation and car following behaviors modeling and so on.

The extraction of traffic flow parameters only considered three key components, omitting travel time and occupancy. A congestion metrics should be extended to a multidimensional index table, which will be count for as many factors as possible. This is the trend of big data, which means the transition from deterministic modeling to a data-driven based solution. As well, a clustering method is preferable for mining and distinguishing different traffic states and potential congestion

In the part of congestion identification, the temporal analysis did not fully exploit the trajectory data, due to the gaps (about fifteen minutes) between every collection of datasets. The duration of non-surveillance is almost as long as the surveillance time. Consequently, the discontinuity in time series graph cannot be neglected. On the other hand, the smoothen algorithm and resampling may fail the task when re-aggregating the time-related parameters.

List of References

- Next Generation Simulation (Ngsim). [2006]: Federal Highway Administration. Available at: <https://ops.fhwa.dot.gov/trafficanalysisitools/ngsim.htm>.
- ALTSHULER, A. A. [1979]: *The Urban Transportation System : Politics and Policy Innovation* / Alan Altshuler, with James P. Womack, John R. Pucher. Publication of the Joint Center for Urban Studies., Cambridge, Mass, MIT Press, ISBN 0262010550.
- AMELIA, A.; SAPTAWATI, G. A. P. Detection of Potential Traffic Jam Based on Traffic Characteristic Data Analysis. 2014 International Conference on Data and Software Engineering (ICODSE), 1-5.
- AN S, Y. H., Wang J. [2018]: Revealing Recurrent Urban Congestion Evolution Patterns with Taxi Trajectories. In: ISPRS International Journal of Geo-Information, Vol. 7 (2018), 10.3390/ijgi7040128.
- ANTONIOU, C.; KOUTSOPOULOS, H. [2006]: Estimation of Traffic Dynamics Models with Machine-Learning Methods. In: *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1965 (2006), pages 103-111, 10.1177/0361198106196500111.
- BARMPOUNAKIS, E.; GEROLIMINIS, N. [2020]: On the New Era of Urban Traffic Monitoring with Massive Drone Data: The Pneuma Large-Scale Field Experiment. In: *Transportation Research Part C: Emerging Technologies*, Vol. 111 (2020), pages 50-71, 10.1016/j.trc.2019.11.023.
- BELLOCCHI, L.; GEROLIMINIS, N. [2020]: Unraveling Reaction-Diffusion-Like Dynamics in Urban Congestion Propagation: Insights from a Large-Scale Road Network. In: *Scientific Reports*, Vol. 10 (2020), pages 4876, 10.1038/s41598-020-61486-1.
- BERTHELIN, F.; DEGOND, P.; DELITALA, M.; RASCLE, M. [2008]: A Model for the Formation and Evolution of Traffic Jams. In: *Archive for Rational Mechanics and Analysis*, Vol. 187 (2008), pages 185-220, 10.1007/s00205-007-0061-9.
- BHASKAR, A.; CHUNG, E.; DUMONT, A.-G. [2011]: Fusing Loop Detector and Probe Vehicle Data to Estimate Travel Time Statistics on Signalized Urban Networks. In: *Computer-Aided Civil and Infrastructure Engineering*, Vol. 26 (2011), pages 433-450, <https://doi.org/10.1111/j.1467-8667.2010.00697.x>.
- BHASKAR, A.; QU, M.; CHUNG, E. [2015]: Bluetooth Vehicle Trajectory by Fusing Bluetooth and Loops: Motorway Travel Time Statistics. In: *IEEE Transactions on Intelligent Transportation Systems*, Vol. 16 (2015), pages 113-122, 10.1109/TITS.2014.2328373.
- BOEING, G. [2017]: Osmnx: New Methods for Acquiring, Constructing, Analyzing, and Visualizing Complex Street Networks. In: *Computers, Environment and Urban Systems*, Vol. 65 (2017), pages 126-139, 10.1016/j.compenvurbsys.2017.05.004.
- CHEN, D.; DRIEMEL, A.; GUIBAS, L. J.; NGUYEN, A.; WENK, C. Approximate Map Matching with Respect to the Fréchet Distance. *2011 Proceedings of the Workshop on Algorithm Engineering and Experiments (Alenex)*.

- CHEN, Z.; LIU, X. C.; ZHANG, G. [2016]: Non-Recurrent Congestion Analysis Using Data-Driven Spatiotemporal Approach for Information Construction. In: *Transportation Research Part C: Emerging Technologies*, Vol. 71 (2016), pages 19-31, 10.1016/j.trc.2016.07.002.
- CHENG, T.; TANAKSARANOND, G.; BRUNSDON, C.; HAWORTH, J. [2013]: Exploratory Visualisation of Congestion Evolutions on Urban Transport Networks. In: *Transportation Research Part C: Emerging Technologies*, Vol. 36 (2013), pages 296-306, 10.1016/j.trc.2013.09.001.
- D'ANDREA, E.; MARCELLONI, F. [2017]: Detection of Traffic Congestion and Incidents from Gps Trace Analysis. In: *Expert Systems with Applications*, Vol. 73 (2017), pages 43-56, 10.1016/j.eswa.2016.12.018.
- DERBEL, A.; BOUJELBENE, Y. 2020. A Systematic Literature Review of Studies on Road Congestion Modelling. *Distributed Computing for Emerging Smart Networks*.
- EDIE, L. C. [1965]: Discussion of Traffic Stream Measurements and Definitions. In: *Proceedings of the Second International Symposium on the Theory of Traffic Flow, London 1963.*, Vol. (1965), pages 139-154.
- FALCOCCHIO, J. C.; LEVINSON, H. S. 2015. Measuring Traffic Congestion. In: FALCOCCHIO, J. C. & LEVINSON, H. S. (eds.) *Road Traffic Congestion: A Concise Guide*. Cham: Springer International Publishing.
- FALCOCCHIO, J. C.; LEVINSON, H. S. 2015. Overview of the Causes of Congestion. In: FALCOCCHIO, J. C. & LEVINSON, H. S. (eds.) *Road Traffic Congestion: A Concise Guide*. Cham: Springer International Publishing.
- FALCOCCHIO, J. C.; LEVINSON, H. S. [2015]: *Road Traffic Congestion - a Concise Guide*. Springer, Cham, ISBN 978-3-319-15165-6.
- GOH, C. Y.; DAUWELS, J.; MITROVIC, N.; ASIF, M. T.; ORAN, A.; JAILLET, P. Online Map-Matching Based on Hidden Markov Model for Real-Time Traffic Sensing Applications. 2012 15th International IEEE Conference on Intelligent Transportation Systems, 776-781.
- KAFFASH, S.; NGUYEN, A. T.; ZHU, J. [2021]: Big Data Algorithms and Applications in Intelligent Transportation System: A Review and Bibliometric Analysis. In: *International Journal of Production Economics*, Vol. 231 (2021), 10.1016/j.ijpe.2020.107868.
- KAN, Z.; TANG, L.; KWAN, M.-P.; REN, C.; LIU, D.; LI, Q. [2019]: Traffic Congestion Analysis at the Turn Level Using Taxis' Gps Trajectory Data. In: *Computers, Environment and Urban Systems*, Vol. 74 (2019), pages 229-243, 10.1016/j.compenvurbsys.2018.11.007.
- LI, L.; JIANG, R.; HE, Z.; CHEN, X.; ZHOU, X. [2020]: Trajectory Data-Based Traffic Flow Studies: A Revisit. In: *Transportation Research Part C: Emerging Technologies*, Vol. 114 (2020), pages 225-240, 10.1016/j.trc.2020.02.016.
- LI, W.; NIE, D.; WILKIE, D.; LIN, M. C. [2017]: Citywide Estimation of Traffic Dynamics Via Sparse Gps Traces. In: *IEEE Intelligent Transportation Systems Magazine*, Vol. 9 (2017), pages 100-113, 10.1109/MITS.2017.2709804.

- LIU, Y.; YAN, X.; WANG, Y.; YANG, Z.; WU, J. [2017]: Grid Mapping for Spatial Pattern Analyses of Recurrent Urban Traffic Congestion Based on Taxi Gps Sensing Data. In: *Sustainability*, Vol. 9 (2017), 10.3390/su9040533.
- LOMAX, T. J.; TEXAS TRANSPORTATION, I.; NATIONAL RESEARCH, C.; TRANSPORTATION RESEARCH, B. [1997]: Quantifying Congestion. In: Vol. (1997).
- MAY, A. D. [1990]: *Traffic Flow Fundamentals*. Englewood Cliffs, N.J., Prentice Hall, ISBN 0139260722 9780139260728.
- MCCLINTOCK, M. [1925]: *Street Traffic Control*. New York, McGraw-Hill book company, inc.
- MEERT, W.; VERBEKE, M. [2018]: Hmm with Non-Emitting States for Map Matching. In: Vol. (2018).
- MEYER, M. A Toolbox for Alleviating Traffic Congestion and Enhancing Mobility.
- MOHAN RAO, A.; RAMACHANDRA RAO, K. [2012]: Measuring Urban Traffic Congestion – a Review. In: *International Journal for Traffic and Transport Engineering*, Vol. 2 (2012), pages 286-305, 10.7708/ijtte.2012.2(4).01.
- NASSREDDINE, G.; ABDALLAH, F.; DENOEU, T. Map Matching Algorithm Using Interval Analysis and Dempster-Shafer Theory. 2009 IEEE Intelligent Vehicles Symposium, 494-499.
- OBRADOVIC, D.; LENZ, H.; SCHUPFNER, M. [2006]: Fusion of Map and Sensor Data in a Modern Car Navigation System. In: *VLSI Signal Processing*, Vol. 45 (2006), pages 111-122, 10.1007/s11265-006-9775-4.
- PINK, O.; HUMMEL, B. A Statistical Approach to Map Matching Using Road Network Geometry, Topology and Vehicular Motion Constraints. 2008 11th International IEEE Conference on Intelligent Transportation Systems, 862-867.
- QUEK, W. L.; CHEW, L. Y. [2014]: Mechanism of Traffic Jams at Speed Bottlenecks. In: *Procedia Computer Science*, Vol. 29 (2014), pages 289-298, 10.1016/j.procs.2014.05.026.
- RAMEZANI, M.; GEROLIMINIS, N. [2015]: Queue Profile Estimation in Congested Urban Networks with Probe Data. In: *Computer-Aided Civil and Infrastructure Engineering*, Vol. 30 (2015), pages 414-432, <https://doi.org/10.1111/mice.12095>.
- SIGUA, R. G. [2008]: *Fundamentals of Traffic Engineering*. University of the Philippines Press, ISBN 9789715425520.
- STATHOPOULOS, A.; KARLAFTIS, M. G. [2003]: A Multivariate State Space Approach for Urban Traffic Flow Modeling and Prediction. In: *Transportation Research Part C: Emerging Technologies*, Vol. 11 (2003), pages 121-135, [https://doi.org/10.1016/S0968-090X\(03\)00004-4](https://doi.org/10.1016/S0968-090X(03)00004-4).
- SUN, J.; LIU, Q.; PENG, Z. [2011]: Research and Analysis on Causality and Spatial-Temporal Evolution of Urban Traffic Congestions—a Case Study on Shenzhen of China. In: *Journal of Transportation Systems Engineering and Information Technology*, Vol. 11 (2011), pages 86-93, 10.1016/s1570-6672(10)60143-2.
- TANG, J. 2019. Chapter 6 - Urban Travel Mobility Exploring with Large-Scale Trajectory Data. In: WANG, Y. & ZENG, Z. (eds.) *Data-Driven Solutions to Transportation Problems*. Elsevier.

- WANG, J.; GU, Q.; WU, J.; LIU, G.; XIONG, Z. Traffic Speed Prediction and Congestion Source Exploration: A Deep Learning Method. 2016 IEEE 16th International Conference on Data Mining (ICDM), 499-508.
- XU, T.; OU, D.; YANG, Y. 2013. An Arcgis-Based Hybrid Topological Map Matching Algorithm for Gps Probe Data. *Icte 2013*.
- YU, J.; STETTLER, M. E. J.; ANGELOUDIS, P.; HU, S.; CHEN, X. [2020]: Urban Network-Wide Traffic Speed Estimation with Massive Ride-Sourcing Gps Traces. In: *Transportation Research Part C: Emerging Technologies*, Vol. 112 (2020), pages 136-152, <https://doi.org/10.1016/j.trc.2020.01.023>.
- ZHANG, S.; LI, S.; LI, X.; YAO, Y. [2020]: Representation of Traffic Congestion Data for Urban Road Traffic Networks Based on Pooling Operations. In: *Algorithms*, Vol. 13 (2020), 10.3390/a13040084.
- ZHENG, Y.; LI, Q.; CHEN, Y.; XIE, X.; MA, W.-Y. 2008. Understanding Mobility Based on Gps Data. *Proceedings of the 10th international conference on Ubiquitous computing - UbiComp '08*.

List of Abbreviations

| | |
|--------|---|
| ALENEX | Algorithm Engineering and Experiments |
| CSV | Comma-separated Values |
| GIS | Geographic Information System |
| GPS | Global Positioning System |
| HMM | Hidden Markov Model |
| IID | Incident_Induced Delay |
| KDE | Kernel Density Estimation |
| pNEUMA | New Era of Urban Traffic Monitoring with Aerial Footage |
| NGSIM | Next Generation Simulation |
| NRC | Non-recurring Congestion |
| OSM | Open Street Map |
| RC | Recurrent Congestion |
| UAS | Unmanned Aerial System |
| UAV | Unmanned Aerial Vehicle |
| VKT | Vehicle Kilometres of Travel |
| WGS | World Geodetic System |

List of Symbols

| | | |
|---|---------------|-------------------------------|
| q | [veh/h/lane] | Volume of traffic flow |
| k | [veh/km/lane] | Density of traffic flow |
| v | [km/h] | Average speed of traffic flow |

List of Figures

| | |
|---|----|
| Figure 3.1 Researched area in Athens divided into 10 zones | 10 |
| Figure 3.2 Time phases of UAV data collection..... | 10 |
| Figure 3.3 Procedure of preprocessing of metadata | 13 |
| Figure 3.4 Working flow for network retrieving | 15 |
| Figure 3.5 Histogram for edge length with KDE | 16 |
| Figure 3.6 Counts of different types of roads..... | 16 |
| Figure 3.7 Road network labeled with number of lanes..... | 17 |
| Figure 3.8 Primary and secondary roads (left) and Traffic signals (right)..... | 18 |
| Figure 3.9 Plot one trajectory on the road network without map-matching..... | 19 |
| Figure 3.10 Implementation of map-matching algorithm | 21 |
| Figure 3.11 Configuration of virtual loop detector..... | 22 |
| Figure 3.12 Result of virtual detector installation..... | 23 |
| Figure 3.13 Trajectory of the i^{th} vehicle in the time-space window | 24 |
| Figure 3.14 Location of detector 188 (red line)..... | 25 |
| Figure 3.15 Flow, density and speed time series at detector 188 | 26 |
| Figure 3.16 Histogram of speed | 26 |
| Figure 3.17 Relation between density and flow | 27 |
| Figure 4.1 Location of congested spots in road network (red line)..... | 31 |
| Figure 4.2 Detector 135: speed time plot (left) and histogram of speed (right) | 32 |
| Figure 4.3 Detector 186: speed time plot (left) and histogram of speed (right) | 32 |
| Figure 4.4 Layout of 3rd Septemvriou Street and the speed-time graph for..... | 35 |
| Figure 4.5 Flow-density joint distribution (8:00– 8:30)..... | 36 |
| Figure 4.6 Spatial distribution of congested linkages at phase 8:30 to 9:00 (circled area)..... | 37 |
| Figure 4.7 Density, flow and speed on congested location | 38 |
| Figure 4.8 Temporal distribution of speed on detector 40 (threshold: red dotted line) | 39 |
| Figure 4.9 Temporal distribution of speed on detector 32 (threshold: red dotted line) | 39 |
| Figure 4.10 Temporal distribution of speed on detector 58 (threshold: red dotted line) | 39 |

List of Tables

| | |
|--|----|
| Table 3.1 Main Python Modules | 11 |
| Table 3.2 Columns of pNEUMA trajectory dataset | 12 |
| Table 3.3 Example of one resulting trajectory matrix..... | 14 |
| Table 3.4 Statistics overview on resulting road network matrix | 15 |
| Table 3.5 Parameters of the virtual loop..... | 23 |
| Table 3.6 Statistics of traffic parameters of detector 188 | 25 |
| Table 4.1 Eligible congestion spots | 31 |
| Table 4.2 Traffic parameters from all eleven detectors on 3rd Septemvriou Street | 34 |

Appendix A: Retrieved road network matrix

| index | osmid | N1 | Lat1 | Long1 | N2 | Lat2 | Long2 | length | lanes | oneway | bearing | highway | dbl_left | dbl_right | geometry |
|-------|-----------|------------|-----------|-----------|-----------|-----------|-----------|--------|-------|--------|---------|---------------|----------|-----------|--|
| 0 | 23216541 | 242134 | 37.975138 | 23.725447 | 727429546 | 37.975042 | 23.725384 | 11.989 | NaN | FALSE | 207.2 | living_street | FALSE | FALSE | LINSTRING (23.7254537.97514, 23.7253837.97504) |
| 1 | 193421328 | 242134 | 37.975138 | 23.725447 | 242135 | 37.975604 | 23.725719 | 57.011 | NaN | FALSE | 24.7 | living_street | FALSE | FALSE | LINSTRING (23.7254537.97514, 23.7257237.97560) |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 1991 | 896087076 | 8330742371 | 37.978338 | 23.741607 | 250712281 | 37.978422 | 23.741560 | 10.178 | NaN | TRUE | 336.1 | residential | FALSE | FALSE | LINSTRING (23.7416137.97834, 23.7415637.97842) |
| 1992 | 23184195 | 8330742371 | 37.978338 | 23.741607 | 250712262 | 37.978652 | 23.741038 | 61.259 | NaN | TRUE | 305.0 | residential | FALSE | FALSE | LINSTRING (23.7416137.97834, 23.7415337.978... |

Appendix B: Detector matrix on 3rd Septemvriou Street

Table: Detector matrix on 3rd Septemvriou Street

| index | N1 | N2 | highway | lanes | length | loop_distance |
|-------|-----------|-----------|-----------|-------|--------|---------------|
| 800 | 262236447 | 95663423 | primary | 3 | 67.279 | 4.0 |
| 102 | 95663631 | 95663394 | secondary | 2 | 74.219 | 4.0 |
| 32 | 95663392 | 95663400 | secondary | 2 | 30.891 | 4.0 |
| 40 | 95663400 | 95663631 | secondary | 2 | 74.195 | 4.0 |
| 125 | 97788471 | 358465410 | secondary | 2 | 86.607 | 4.0 |
| 1033 | 358465410 | 635122963 | secondary | 2 | 45.282 | 4.0 |
| 1290 | 635122963 | 635122988 | secondary | 2 | 15.972 | 4.0 |
| 58 | 95663454 | 95663467 | primary | 3 | 59.110 | 4.0 |
| 1540 | 962356679 | 95663454 | primary | 3 | 62.364 | 4.0 |
| 49 | 95663420 | 962356679 | primary | 3 | 58.782 | 4.0 |
| 51 | 95663422 | 95663420 | primary | 3 | 65.230 | 4.0 |
| 52 | 95663423 | 95663422 | primary | 3 | 61.694 | 4.0 |

Table: Detector matrix on 3rd Septemvriou Street (continued)

| det_edge_1 | det_edge_1bis | det_bearing_1 | det_bearing_1bis |
|---|---|---------------|------------------|
| LINestring (23.72809581607818 37.9844766, 23.7... | LINestring (23.72809581607818 37.9844766, 23.7... | 270.000000 | 270.000000 |
| LINestring (23.72932095350776 37.9896441544338... | LINestring (23.72932984313156 37.9896794869473... | 281.215768 | 281.215768 |
| LINestring (23.72908422601225 37.9887034281759... | LINestring (23.7290931228358 37.98873875954592... | 281.225115 | 281.225115 |
| LINestring (23.72915625979005 37.9889896553393... | LINestring (23.72916515081267 37.9890249876265... | 281.217655 | 281.217655 |
| LINestring (23.72980141293278 37.9916533337178... | LINestring (23.72981026828561 37.9916886716393... | 281.171654 | 281.171654 |
| LINestring (23.72999304268926 37.9924174381495... | LINestring (23.73000190557651 37.9924527748946... | 281.181169 | 281.181169 |
| LINestring (23.73007329208044 37.9927373185528... | LINestring (23.73007601018145 37.9927481400229... | 281.196963 | 281.196963 |
| LINestring (23.72874853686437 37.9873682371191... | LINestring (23.72875739712687 37.9874035742317... | 281.178578 | 281.178578 |
| LINestring (23.72861055692018 37.9868180400698... | LINestring (23.72861942201067 37.9868533764190... | 281.184833 | 281.184833 |
| LINestring (23.72848045519132 37.9862994397968... | LINestring (23.72848931973282 37.9863347762277... | 281.184209 | 281.184209 |
| LINestring (23.72833619261658 37.9857239453133... | LINestring (23.72834181634903 37.9857463390799... | 281.195948 | 281.195948 |
| LINestring (23.728199644413 37.98517963816457,... | LINestring (23.72820850602772 37.9852149750454... | 281.180637 | 281.180637 |

Declaration concerning the Master's Thesis

I hereby confirm that the presented thesis work has been done independently and using only the sources and resources as are listed. This thesis has not previously been submitted elsewhere for purposes of assessment.

Munich, February 5th, 2021



Zhichun Liu