



ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

**ANALYZING MIXED URBAN TRAFFIC BY
LINKING LARGE SCALE TRAJECTORY
DATASET TO UNDERLYING NETWORK**

EPFL MASTERS PROJECT FOR EXCHANGE STUDENTS

LANDTMETERS JOACHIM

SUPERVISOR: PROF. GEROLIMINIS NIKOLAS
Co-SUPERVISOR: BARMPOUNAKIS EMMANOUIL

ACKNOWLEDGEMENT

This project was extremely interesting to work on. I would like to thank my professor at KU Leuven Chris Tampère for the opportunity to go abroad and professor Nikolas Geroliminis for the academical support throughout my project. A special thank you goes to Emmanouil Barmpounakis who guided me through the process with very helpful advice and ideas. Lastly, without the support of my friends and parents this project would not have been possible.

Contents

List of Figures	iii
List of Tables	v
1 Introduction	2
1.1 Context	2
1.2 Motivation	3
1.3 Overview	3
2 Literature review	4
2.1 Fundamental Diagram	4
2.2 Macroscopic Fundamental Diagram	7
2.3 Summary	10
3 Extraction of Macroscopic Traffic Characteristics	11
3.1 pNEUMA Experiment	11
3.2 Framework	12
3.3 Application and Evaluation	20
3.4 Summary	26
4 Urban Traffic Analysis	30
4.1 Network	30
4.2 Fundamental Diagram	31
4.3 Analysis of arterials	42
4.4 Summary	46
5 Conclusion	49
Appendices	51
A Arterials	52
Bibliography	59

List of Figures

2.1	Tree of traffic models (Kessels, 2019)	5
2.2	General FD (as described in (Kessels, 2019))	6
2.3	Observing traffic in different ways (Kessels, 2019)	7
2.4	Flow-Occupancy representation for signalized intersection (Wu et al., 2011)	8
2.5	Empirical MFD for Yokohama of fused data (Geroliminis and Daganzo, 2008)	8
3.1	pNEUMA experiment: A. time periods of UAV data collection for weekday B. area of Athens, Greece with zones for every UAV (Barmpounakis and Geroliminis, 2020)	12
3.2	Program structure	14
3.3	Street network for different network types	15
3.4	Example of HMM taken from (Devos, 2018)	16
3.5	Result of map-matching for one trajectory	18
3.6	Detectors on used edges of network with specified DFI, ΔX and number of detectors per edge	19
3.7	Matching issues: a. maximum initial distance, the upper figure (1) has a too low initial distance b. maximum distance, the right figure (2) maps wrongly because of a too high distance c. missing edges, the trajectory is wrongly matched at the beginning and the end. It makes a U-turn on the pavement in the beginning (1) and at the end goes into a pedestrian street (2)	21
3.8	All the bus trajectories in the researched area	22
3.9	Network with pedestrian streets included	23
3.10	DBL in network: A. Newly added DBL B. All DBL in network	23
3.11	Histograms of variables used for evaluation: A, B and C show the median, mean and maximum straight-line distance to the matched edge, respectively, D. difference between trajectory and matched length, E and F. percentage of trajectory observations with bearing difference exceeding 45 and 90 degrees, respectively	28
3.12	Maximum distance histograms: A. Trajectories not travelling on Omonoia edges B. Only trajectories travelling on Omonoia edges C. Edges of Omonoia square D. Width in black and maximum distance in red of all trajectories matched to an edge	29
3.13	Space-time area: A. Flow-density scatter plots for different ΔT B. Flow-density scatter plots for different ΔX	29
4.1	Used network: A. Topology highlighted B. Edges with top 25 % matches on total and per lane basis C. Selected arterials	32
4.2	Link FDs for OM 1 (A), PAN 14 (B) and ALE 2 (C)	34
4.3	35
4.4	Adjusting for stalling vehicles in space-time area	36
4.5	Effect of PTWs on density and flow	37
4.6	Density and flow over time for second detector PAN 14	38
4.7	FDs with and without PTWs: A. ALE 2 detector 1 B. OM 1 detector 2	39

4.8	MFD for different adjustments and aggregation steps: A. filtering stalling vehicles, B. aggregating over longer time periods, C. combination, D. difference between detectors	40
4.9	Comparison between road types: A. Speed-density MFD B. Average speed over time	41
4.10	Speed-accumulation diagram for every arterial (aggregation step of 180 sec)	42
4.11	Average speed over time for different modes	43
4.12	Evolution of characteristics over whole OCT arterial: A. total density, B. flow, C. flow per lane	44
4.13	Boxplots for average lane flows for every arterial	45
4.14	Histograms of unit travel time for every arterial	47
4.15	Evolution of average unit travel time for consecutive sections along an arterial	48
A.1	Number of lanes for every arterial with DBL sections marked in red	53
A.2	54
A.3	55
A.4	56
A.5	Car share	57
A.6	PTW share	57
A.7	Taxi share	58
A.8	Bus share	58

List of Tables

3.1	Columns of collected dataset	12
3.2	Columns of combined dataframe of nodes and edges	13
3.3	Comparison of network types	15
3.4	Proportion and dimensions in meters of vehicle types in the dataset	20
3.5	Comparison of maximum distance after map-matching for the original and adjusted network	25
3.6	Correlation matrices for two selection procedures, left: maximum distance, right: average speed (MD: maximum distance, LD: length difference, LDr: relative length difference, BD: bearing difference (>90 deg)	25
4.1	Length of used network and arterials (*DBL are not part of total length, corresponding edge is counted)	31
4.2	Effect of stalling vehicle with flow, density and speeds in veh/h, veh/km and km/h, respectively (*IC = set of all individual vehicle speeds crossing during the time step)	33
4.3	Speed of PTWs at two selected locations (only time steps with non-zero density are kept), in km/h	37
4.4	Number of trajectories traveling whole arterial	45
A.1	Arterials with labels	52

ABSTRACT

In recent years new technology makes it possible to acquire trajectory-data more easily and on a wider scale. The pNEUMA experiment collected full trajectory data on a large scale by using a swarm of drones in the city of Athens, Greece. These trajectories do not exhibit a link with the underlying road network so far. In this research a tool is developed to easily get traffic characteristics at any location in the network and used later on to analyze urban traffic. The framework presented extracts a road network with all needed attributes, matches the trajectories to the underlying network using a map-matching algorithm and places virtual loops anywhere in the network to measure the traffic characteristics. From the fundamental diagrams (FD) it follows that a lot of heterogeneity is present in urban traffic. Special attention goes to the behaviour of powered two-wheelers (PTW), since this mode moves differently through urban traffic and travels at higher speeds than conventional traffic. Afterwards, the macroscopic fundamental diagram (MFD) is estimated and different aspects that influence the aggregated values are discussed.

SUMMARY

The pNEUMA experiment presents a new way of analyzing urban traffic. Using a swarm of drones traffic in dense urban areas gives very detailed trajectory data on a larger scale than ever before. These trajectory data, however, do not have a link with the underlying road network. It is thus not known on which road a vehicle is driving at a specific time. In this thesis a tool is developed to link this trajectory data and get the macroscopic traffic characteristics at any location in the road network. A first step is constructing a detailed graph with all needed attributes, the open-access OpenStreetMap (OSM) database is used to extract all the roads of the researched area in the pNEUMA experiment. Map-matching the trajectories to this road network forms the second step and is not a trivial process. With a probabilistic map-matching algorithm using hidden Markov Models (HMM) (Meert and Verbeke, 2018) the trajectories are matched in an efficient way. The last two steps of the presented tool are placing virtual loops on all selected roads of the network and measuring the traffic characteristics.

The traffic characteristics are determined for a large portion of the network to analyze individual fundamental diagrams (FD) and estimate an empirical macroscopic fundamental diagram (MFD). It is immediately clear that a lot of scatter is present due to different sources of heterogeneity. Adjusting for some heterogeneity reduces the scatter and shows a clearer shape in accordance with proposed traffic flow theory for urban roads. Powered two-wheelers (PTW) show a very different behaviour in the urban environment, they travel at higher speeds and use their smaller size to pass through slow-moving traffic.

First MFD results show scatter for short time steps, reducing the scatter through aggregation gives very few data points to analyze the shape of the MFD. Using the detectors for a brief travel time analysis confirm the heterogeneity between vehicle types and the higher speeds of PTWs in urban traffic.

CHAPTER 1

INTRODUCTION

1.1 CONTEXT

In recent years, new technologies and software developments made it possible to acquire and process large amounts of high quality data. In order to manage traffic systems in an ideal way such high quality data is of crucial importance. Especially full vehicle trajectories become more easily available due to these technologies (Kong et al., 2018). Vehicle trajectory data is the most comprehensive data available and allows direct measurement of density and detection of more detailed phenomena (Treiber and Kesting, 2013).

Several technologies can be used to acquire trajectory data. In numerous cases it is collected for large areas by tracking vehicles with GPS, called probe vehicles. Although very useful, it gives rise to a missing data problem, since only a small part of vehicles is tracked and thus no information is collected for a large share of the network at every time instance (Yang et al., 2018). Alternatively, video images provide the most complete form of trajectory data, since all movements in a certain area are recorded. For this technology having a wide enough area coverage to research macroscopic relations forms the main issue. The NGSIM program which consists of a large number of high-resolution cameras installed across different arterial roads and freeways is used for microscopic modelling, it can be used as a benchmark for accuracy of vehicle trajectory data collected through video images, but falls short for usage at the macroscopic network level (Punzo et al., 2011). Recently new research used unmanned aerial vehicles (UAV) to acquire trajectory data for different purposes as described in (Barmounakis et al., 2016). Salvo et al. (2017) used a UAV for analyzing intersections, they concluded that microscopic traffic characteristics can be accurately measured but are susceptible to different environmental factors, like weather and buildings for example. In Krajewski et al. (2018) and Barmounakis et al. (2019) a UAV is used on longer road lengths and shows promising results, they also mention that fast and accurate computer vision algorithms are needed for automatic vehicle detection. The article of Stofan (2018) explains the benefits these new intelligent algorithms have over traditional methods and stresses the accuracy gain that can be reached by these new methods.

The availability of a high quality trajectory dataset opens up a lot of research opportunities. The presented research uses data collected during the pNEUMA (New Era of Urban traffic Monitoring with Aerial footage) experiment, a first-of-its-kind large-scale vehicle trajectory dataset collected by a swarm of drones in Athens, Greece (Barmounakis and Geroliminis, 2020). The richness of this collected data can difficultly be overrated. The paper describes a lot of possible research directions, going from microscopic driver behaviour's analysis to macroscopic fundamental diagrams (MFD) for the whole multi-modal urban network, showing the uniqueness of this experiment.

1.2 MOTIVATION

In order to be able to manage traffic better, especially the multi-modal urban environment, this high-quality vehicle trajectories play a vital role. The last couple of years MFDs have been estimated for different cities around the world, Yokohama, Toulouse, Brisbane (Geroliminis and Daganzo, 2008; Buisson and Ladier, 2009; Tsubota et al., 2014). The existence of a fundamental relation between flow and density on the network level dates back to the 1950's and 1960's, Edie (1963) proposed generalized definitions to calculate network wide flow and density. Since the work of Daganzo (2007) and Daganzo and Geroliminis (2008) research of MFD and network wide traffic control have gained a lot of interest. The effect of different modes on the macroscopic scale becomes more important to gain more insights in the movement of vehicles in dense urban networks. Buisson and Ladier (2009) show that heterogeneity for different aspects have a big influence on the eventual macroscopic results, therefore, gaining more insight in how heterogeneity in this dataset and the underlying network affect the results is an interesting research direction.

The goal of this thesis is to get the needed traffic flow parameters for every link in the urban network to estimate fundamental diagrams (FD) and an MFD in a fast and easy way by linking the collected data to a road network and research different heterogeneity aspects, with a main focus on modal differences and their effect on the measured characteristics.

1.3 OVERVIEW

This thesis has the following structure. The next chapter, chapter 2, consists of a literature review about FDs and MFDs, and the incorporation of multi-modality. Chapter 3 explains the developed tool to extract the needed characteristics. A first section, briefly discusses the used dataset of the pNEUMA experiment. Second, the overall framework of the developed tool is presented and subdivided in four main steps, namely the extraction of the road network using the OpenStreetMap¹ (OSM) database, map-matching all trajectories to the underlying network in a fast and accurate way, placing virtual detectors at specified locations in the network and determining the traffic characteristic for every detector. The last section discusses the application of the tool for the used trajectory dataset, showing encountered issues and their solutions, and presents accuracy measures to define the overall performance. Chapter 4 analyzes urban traffic using the developed tool for selected arterials in the network. An analysis of the individual FD of one detector gives more insight in the causes of scatter due to different heterogeneities, afterwards the attained MFD is discussed and macroscopic differences between the main modes are presented. The last part shows for individual arterials the evolution of the traffic characteristics and travel time.

¹<https://www.openstreetmap.org>

CHAPTER 2

LITERATURE REVIEW

The FD forms the basis of current traffic flow theories. Greenshields et al. (1935) marks the start of traffic flow theory, he saw a relation between capacities on roads and speeds by the use of photographs of moving vehicles on traffic lanes. Later, Lighthill and Whitham (1955); Richards (1956) presented independently from each other a macroscopic model describing kinematic waves between aggregated variables of traffic flow, known as the Lighthill-Whitham-Richards (LWR) model. This model is still extensively used in present traffic flow research for its simplicity. Kessels (2019) gives an overview of the evolution in traffic flow modelling, showing the wide variety of models currently available in academic research. Figure 2.1 shows the tree with the different modelling families. It is stated that of the four presented families, all can be linked back to the FD (which is also one of the four families) and that the FD family does not describe traffic state changes but relates the traffic variables in a static way. Micro-, meso- and macroscopic traffic models then use this relation to model traffic state changes over time, e.g. creation and dissolution of congestion.

2.1 FUNDAMENTAL DIAGRAM

By plotting macroscopic traffic variables such as average speed, traffic volume and density it became clear that a relation existed, the FD. Finding a function that describes this relationship has been subject of numerous studies, this function is only valid for equilibrium states and homogeneous conditions. In Dhingra and Gull (2011) a history of proposed theoretical functions is presented. They compare Greenshields' linear speed-density function with other proposed relations, exponential and bounded curves. This shows some of the different theoretical approaches that exist to describe traffic flow. The LWR-model based on hydrodynamic theory is not perfect and has its shortcomings, Daganzo (1994) acknowledges this and proposes a new model, the cell transmission model, which still follows the hydrodynamic theory of traffic flow. The generalized flow-density relation for this model is triangular with four degrees of freedom: free flow speed, maximum flow, jam density and wave speed. The triangular representation is still heavily used in present research for its wide application and simple form. Many other diagram representations are proposed, an overview of different shapes is given in chapter four in Knoop (2017). Kessels (2019) poses three main requirements for the shape of the FD:

- finite maximum speed exists and is reached when density approaches zero
- finite maximum density exists and the speed is zero at this density
- speed does not increase when density increases and it does not decrease when density decreases

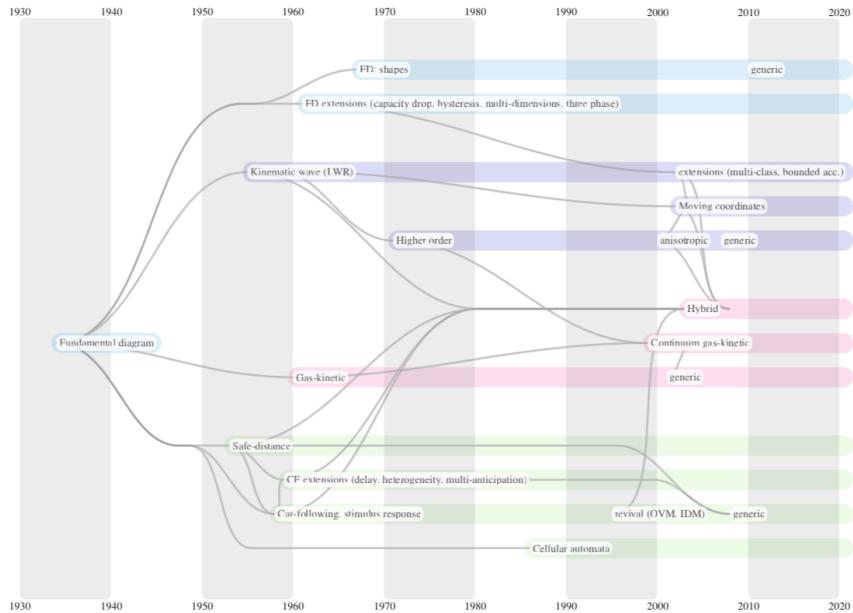


FIGURE 2.1
Tree of traffic models (Kessels, 2019)

These are basic requirements for the FD, more bounding requirements can also be stated to assure a unique solution to simple problems. Having a continuous diagram is one of those extra requirements, this makes it possible to describe a diagram without having two branches, so it would include bi-linear functions as well (Kessels, 2019). Figure 2.2 shows a general shaped FD for the density-flow relation. The idealized form of this diagram is valid for steady state equilibrium and homogeneous traffic.

2.1.1 EMPIRICAL FUNDAMENTAL DIAGRAM

The developed models and theories for the fundamental traffic relation $q = ku$ are based on empirical results. In order to have a valid theory a validation step using real-life data is necessary. Plots of these traffic variables, however, show a lot of scatter. Treiber and Kesting (2013) list reasons for the scatter plots not coinciding with an FD.

- systematic errors in the measurement process
- non-equilibrium traffic
- inhomogeneities in space and driver-behaviour

Schnetzler and Louis (2013) describe three main causes of wide scattering. The first one concerns the variability of traffic conditions due to weather conditions or congestion phases. The congestion phases exist due to accelerating and decelerating phases leading to a capacity drop and hysteresis, Daganzo (2002) addresses this and proposes a new behavioural model for multi-lane freeway traffic. A second one is the mixture of vehicles in traffic and variability in driver-behaviour. The presence of different types of vehicles has an influence on the occupancy rate because of the different vehicle lengths. Microscopic phenomena, e.g. lane changing, overtaking, form the third cause.

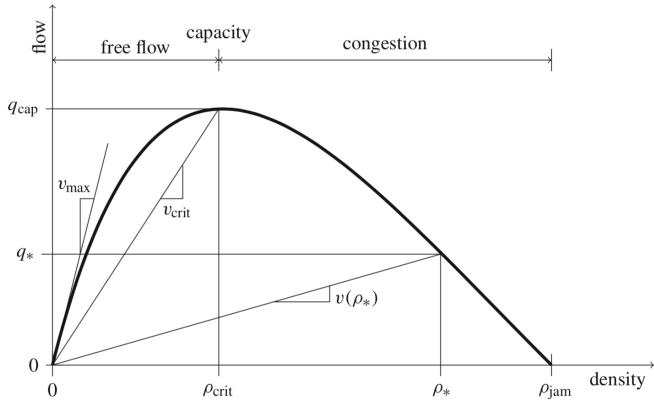


FIGURE 2.2
General FD (as described in (Kessels, 2019))

As stated in the introduction data collection methods are constantly improving and new methods arise to process all the gathered traffic data sources. The way data is collected determines for a large part the attainable accuracy of the study, see reasons for wide scattering. Aggregated data, for example, cannot be used accurately to model microscopical phenomena, although disaggregation is possible it falls back on necessary assumptions. Figure 2.3 shows the different ways of observing traffic.

- local (fixed position)
- instantaneous (fixed in time)
- trajectory (moving with vehicle)

Depending on the data acquiring method used, different traffic variables can be measured. Treiber and Kesting (2013) discuss the different measured variables for cross-sectional data, i.a. loop detectors. These kind of detectors are widely used for both real-time and offline analysis. Double inductive loops perform point measurements for flow (eq. 2.1), occupancy (eq. 2.2) and speed (eq. 2.3) with high accuracy.

$$Q(x, t) = \frac{\Delta N}{\Delta t} \quad (2.1)$$

$$O(x, t) = \frac{1}{\Delta t} \sum_{\alpha=\alpha_0}^{\alpha_0+\Delta N-1} (t_{\alpha}^1 - t_{\alpha}^0) \quad (2.2)$$

$$V_H(x, t) = \frac{\Delta N}{\sum_{\alpha=\alpha_0}^{\alpha_0+\Delta N-1} \frac{1}{v_{\alpha}}} \quad (2.3)$$

The main disadvantage of these detectors is the impossibility to measure traffic density directly. Instead occupancy or the fundamental relation with flow and speed can be used as a proxy. To make use of the fundamental relation space mean speed is needed, which is not directly available from detector loops but can be approximated by the harmonic time mean speed (eq. 3.5), neglecting accelerations. In Knoop and Daamen (2017) this approximation is also mentioned as a possible downside to attain high accuracy for FD estimation. To go from occupancy to density the relation can be assumed linear but for low speeds a non-linear relation is detected (Kim and Hall, 2004). Qiu et al. (2010) show that density estimation of loop detector data performs less than a combination of this data with probe vehicles. Calculating density with point measurements should thus be done with care to rule out errors as much as possible. Video footage over a larger area combines the local and instantaneous measurement methods, e.g. NGSIM experiment

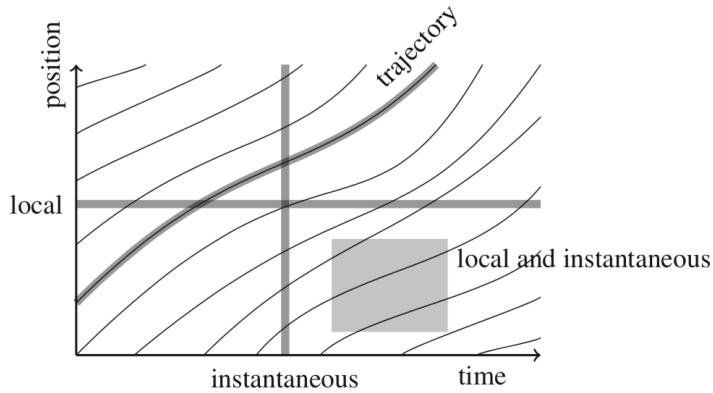


FIGURE 2.3
Observing traffic in different ways (Kessels, 2019)

and pNEUMA experiment, and is depicted generally as the grey rectangle in figure 2.3. These kinds of data acquisition make a direct measurement of density possible. Usage of different data collection methods can also help to better estimate the shape of the macroscopic traffic relations.

Li and Zhang (2011) estimate an FD with single loop detector data and find an accurate fit by separating traffic into free-flow and congested regimes. Seo et al. (2019) use multiple probe vehicles to estimate an FD of a link. They conclude that with some exogenous assumptions it is possible to estimate an FD almost everywhere in the network, depending on the trajectories of the probe vehicles. Automatic fitting of FDs is done for Wu's FD shape and should be possible for other shapes as well (Knoop and Daamen, 2017). These researches show the different ways the existence of FDs is proved with empirical data. Nevertheless, this empirical evidence is mainly for freeway traffic.

Wu et al. (2011) address the empirical FD for arterial roads, called the arterial fundamental diagram (AFD). Because of many signalized intersections aggregated plots show wider scatter with higher occupancy. Figure 2.4 shows a representation of flow-occupancy at a signalized intersection. The figure is based on findings from empirical data, two main phenomena were observed: two-capacity phenomenon and queued-over-detector (QOD). The former concerns the signal coordination and the turning movements effect on the capacity, the latter describes vehicles stopped over detectors during red-cycles without spill-backs, resulting in occupancy expansion. After accounting for those two phenomena a stable AFD is found and is proposed to follow a trapezoidal shape, although not enough data is available in the heavily congested region. Gan and Skabardonis (2019) try to estimate trapezoidal FDs for different intersections and conclude that this shape does exist for arterial links. However, some intersections show bad matches for the congested branch which is mainly due to bad coordination causing QOD due to the red signal phase as mentioned in Wu et al. (2011). In Barmpounakis and Geroliminis (2020) different FDs are depicted for different locations in the Athens' region and approximately follow the trapezoidal shape, nevertheless a clear congested branch due to spill-over is not always visible.

2.2 MACROSCOPIC FUNDAMENTAL DIAGRAM

The concept of MFD shifts the focus to the network scale. Daganzo (2007) renewed interest in the traffic control of large scale areas by proposing accumulation based control of reservoirs or neighbourhoods to prevent them from going into gridlock. In the conclusion it is mentioned that city traffic control can be

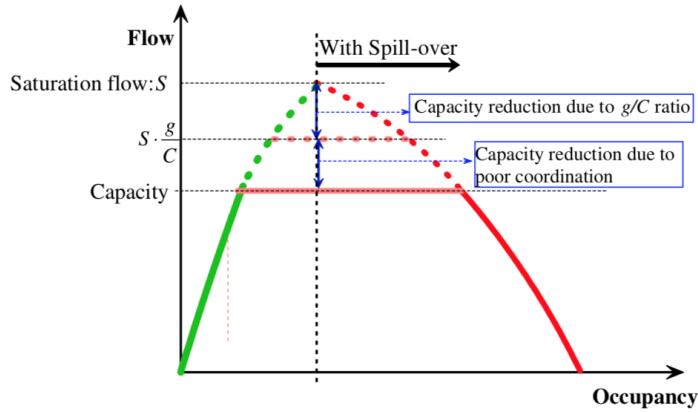


FIGURE 2.4
Flow-Occupancy representation for signalized intersection (Wu et al., 2011)

achieved with trial and error if enough data is available to represent the real-time situation. Follow-up research builds further on the idea of network-wide control by proving the existence of an MFD. First, simulation shows a relation between production and accumulation and the same three regimes as for individual links are observed. Equations 2.4 and 2.5 give the definitions of production and accumulation following Edie (1963).

$$P = \sum_i q_i l_i \quad (2.4)$$

$$N = \sum_i k_i l_i \quad (2.5)$$

Also, comparison between individual link plots and the aggregated plots for the whole simulated network show clear scatter differences in favor of the bigger scale (Geroliminis et al., 2007). Second, by using detector and probe vehicle data the existence of the MFD is empirically proved for the city of Yokohama (Geroliminis and Daganzo, 2008). In figure 2.5 the empirical MFD, speed-accumulation, is shown for the fused data, detectors and probe vehicles. An important finding is the absence of big differences between the scatter plots for different demand, stating that the MFD is a property of the network.

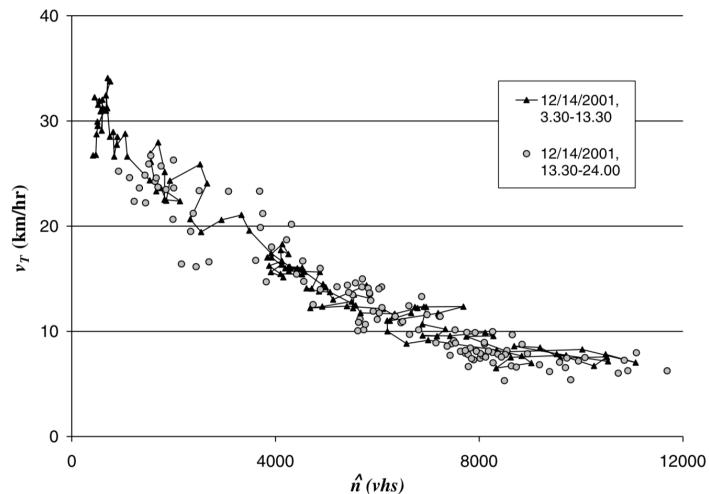


FIGURE 2.5
Empirical MFD for Yokohama of fused data (Geroliminis and Daganzo, 2008)

2.2.1 PROPERTIES

Estimating an accurate MFD is subject to some properties which have to be fulfilled. Buisson and Ladier (2009) present an empirical MFD for Toulouse and show that it contains a lot of scatter which is not explained by experimental errors. By researching the sensitivity of the MFD to different aspects of the data acquisition and network topology, the importance of homogeneity for these different aspects, namely detection position for signalized links, network topology and congestion distribution, is shown. For a well-defined MFD regularity conditions are: slow varying and distributed demand, homogeneous spatial distribution of congestion, many route choices and homogeneous topology (Daganzo and Geroliminis, 2008). The congestion homogeneity is relaxed to the property that a well-defined MFD exists when for two different time intervals with the same average network flow and density the spatial distribution of link densities is similar, equation 2.6 (Geroliminis and Sun, 2011):

$$Q(t_1) = Q(t_2) \& O(t_1) = O(t_2) \iff d_r(t_1) \sim d_r(t_2) \quad (2.6)$$

When this sufficient property is not met or the variance of the density distributions is constant but high, clustering the network into homogeneously loaded regions is needed (Ji and Geroliminis, 2012). This partitioning results in MFDs for every homogeneous region and shows that congestion is reached at different time steps during the total time period. More complex algorithms take also the spatiotemporal dynamics of density into account to form homogeneous clusters for different time steps (Saeedmanesh and Geroliminis, 2017).

2.2.2 DATA

The used data have an influence on the accuracy of MFD estimation, complete trajectory data shows better results than loop detector data and the location of the detectors played a big role (Courbon and Leclercq, 2011). A good spreading of detectors over links renders better estimates, since more traffic states are measured with the different detectors. However, Leclercq et al. (2014) conclude that using only loop detectors gives no better estimation than using probe vehicles combined with loop detectors. A more relaxed combination of these two data sources is presented in Du et al. (2016) and shows promising results for estimating MFDs. Optimal placement of detectors in the network forms one of the concluding questions. Through the form of a resource allocation problem Zockaei et al. (2018) try to find the optimal locations for fixed point measurements and the amount of probe vehicles. It follows that increasing the amount of probe vehicles when having a constant number of fixed locations does not automatically improve the estimation, conversely adding detector locations with a constant amount of probe trajectories does improve the estimation.

2.2.3 3D-MFD

Previous sections discussed traffic flow relations without incorporating modal differences. In recent years, multi-modal traffic gains more importance in MFD research. Geroliminis et al. (2014) presented a bi-modal MFD called 3D-MFD to examine the traffic flow relations for buses and cars and their interaction on the network level. Through simulation of bi-modal traffic a three-dimensional point cloud reveals the existence of a 3D-MFD with a shape dependent on the accumulation of both modes. A functional form for this MFD is proposed and the interaction between both modes is dynamically modelled. Loder et al. (2017) give the first empirical estimate of a 3D-MFD for the city of Zürich.

2.3 SUMMARY

In order to improve macroscopic models accurate data is of great importance. This literature review incorporates a discussion of the FD with its different proposed forms and how it is empirically estimated. Large scatter is present in most real FD plots, caused by different phenomena and data acquisition inaccuracies and approximations. The triangular shape of the FD is observed for uninterrupted traffic flow, for signalized arterials this shape is trapezoidal.

Controlling network-wide traffic presents a big challenge for policy makers, by estimating an MFD controlling large-scale networks becomes possible. The existence of such a diagram has been empirically proved and has since gained a lot of attention in traffic flow theory research. Homogeneity forms an important aspect to derive MFDs. New methods are being designed to account for possible heterogeneity in topology, congestion distribution and demand. Acquiring a complete and robust dataset stays a challenge, different aggregation methods and combinations of data sources try to fill the gaps.

CHAPTER 3

EXTRACTION OF MACROSCOPIC TRAFFIC CHARACTERISTICS

This chapter explains the framework of processing the trajectory dataset to get the macroscopic traffic characteristics and perform an elaborate traffic analysis afterwards. First the acquired dataset of the pNEUMA experiment is discussed, afterwards the framework of the developed tool is presented and the last part evaluates the tool for the considered dataset.

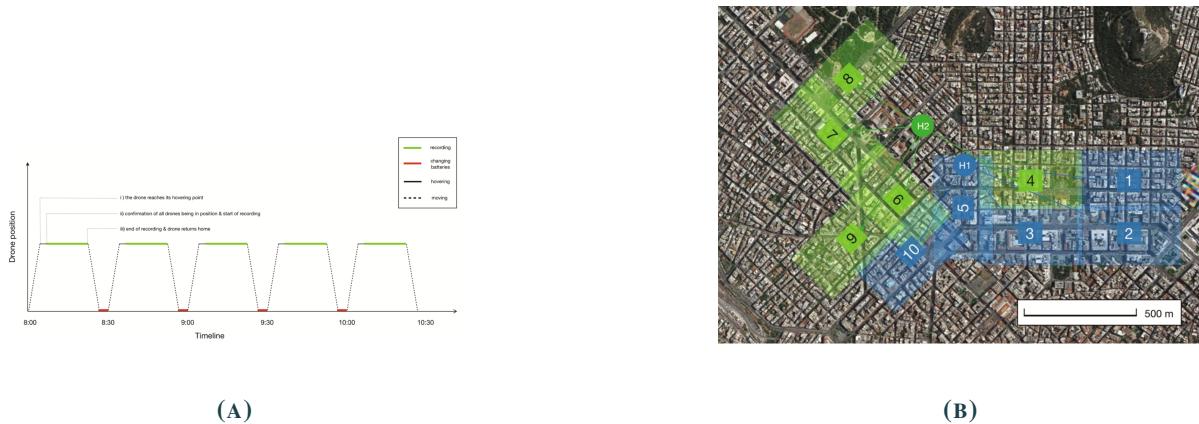
The unique dataset used in this research presents enormous opportunities to describe a large-scale network in more detail than ever before. As was explained in the previous chapter, current data sources contain inaccuracies with respect to the traffic flow characteristics and can cause a lot of scatter. The availability of traffic data of every vehicle in the network makes it possible to get the macroscopic parameters for every road with high accuracy. This is the real-life equivalent of the ground truth data in network simulations, since all OD's and routes through the network are known.

3.1 PNEUMA EXPERIMENT

As mentioned in the introduction the pNEUMA experiment (Bampounakis and Geroliminis, 2020) acquired a large-scale video-image dataset by using ten UAVs during 2.5 hours in the morning peak for all weekdays for the city of Athens, Greece. The dataset is a first-of-its-kind because of the wide scale. A complete trajectory dataset contains an enormous amount of information for the microscopic and macroscopic level because vehicles' movements can accurately be analyzed and less assumptions are needed to measure different characteristics.

Figure 3.1 shows the flight periods of every UAV and the examined area in Athens. The area spans 1.3 km^2 with over 100 km of roads, 100 busy intersections and more than 30 bus stops. With computer-vision software all vehicles are tracked and a distinct datapoint is created for every second, containing its position and multiple other variables. Urban traffic is heavily heterogeneous with different vehicle types present in the same spatial environment. In this dataset six vehicle types are distinguished: cars, taxis, buses, powered two-wheelers (PTWs or motorcycles), medium and heavy vehicles.

The collected data consists for every individual trajectory of eight fixed columns and six variable columns with values for every time stamp, table 3.1 lists all these columns. The fixed columns are only labeled or measured once for the whole length of the trajectory. Afterwards a column giving the bearing of every vehicle is added using two consecutive data points, giving the direction of travel for that particular point.


FIGURE 3.1

pNEUMA experiment: A. time periods of UAV data collection for weekday B. area of Athens, Greece with zones for every UAV (Bampounakis and Geroliminis, 2020)

These trajectories are mapped in WGS-84 coordinates and there is no link with the underlying network so far.

Fixed	Variable
Tracked vehicle	Latitude
Type	Longitude
Entry gate	Speed
Entry time	Tan. acceleration
Exit gate	Lat. acceleration
Exit time	Time
Travelled distance	Bearing
Average speed	

TABLE 3.1
Columns of collected dataset

3.2 FRAMEWORK

This section presents the main work flow to get the traffic characteristics out of the trajectory dataset. In chapter 2 an overview of research in FDs and MFDs explains the different aspects needed to come to accurate and reliable traffic control. Extracting the traffic characteristics out of the trajectory dataset to estimate link FDs and afterwards the MFD for the studied region forms the scope of this research. The data points (latitude-longitude) are not linked yet with the road network, therefore, a matching of the trajectories to the network to know where every vehicle is located at time t is needed. Figure 3.2 shows the flow chart of the process to get from rough data to traffic flow characteristics for every used edge in the network.

The main steps depicted in the flow chart are:

- A Create network graph
- B Map match trajectories

- C Install virtual detectors
- D Calculate traffic characteristics

Each of these processes is explained in more detail hereafter.

3.2.1 CREATING GRAPH

Since the coordinates of the collected data points are not linked to a network topology, it is not known on which specific road a vehicle is driving. A network graph is needed to match every trajectory data point to a physical road. In recent years GIS applications made it easier to extract network structures with their corresponding attributes. Since programming is done in python, the OSMnx package (Boeing, 2017) is an interesting tool to extract road networks all around the world. This recent package extracts a map of the network from OpenStreetMap¹ (OSM) data and simplifies the network topology to a valid graph structure of nodes and edges. Roads in OSM are depicted as edges of a graph structure connected to nodes, which represent the physical intersections. Two crossing edges without a node present on the crossing are not connected and going from one to the other is not directly possible. The OSM database has different topology keys, this makes the extraction of only some types of streets possible. OSMnx has some predefined lists of keys to extract specific network types, e.g. all walkable paths, only driveable streets, etc. Adding or removing keys is easily done to extract a customized selection of streets. In order to extract only streets of a specific area, a bounding-box comprising the most extreme coordinates for every direction (north, east, south, west) is defined.

Position	Attributes
osmid edge	length
N1	lanes
latitude N1	oneway
longitude N1	bearing
N2	highway
latitude N2	dbl left
longitude N2	dbl right
geometry	

TABLE 3.2
Columns of combined dataframe of nodes and edges

The constructed graph is a multidigraph, meaning that the graph consists only of directed edges allowing one direction of travel and multiple directed edges in the same direction between two nodes can exist, i.e. different attributes and weights. When travel is allowed in both directions the edge is split in reciprocal directed edges. In order to work in a more structured way with the graph, the nodes and edges are converted to geodataframes². These dataframes consist of rows for every unique node or edge in the network and columns for every attribute associated with it. Joining these two dataframes and selecting the needed attributes results in a more compact dataframe that is used in later steps. Table 3.2 shows the chosen columns of the resulting dataframe, the first column gives the coordinates and IDs of the nodes making up a directed edge, the second one consists of all the chosen attributes for the specific edge. The bearing specifies the direction of the street with respect to the north, measured counter-clockwise in

¹<https://www.openstreetmap.org>

²Syntax from pandas package: A dataframe with a geometry column

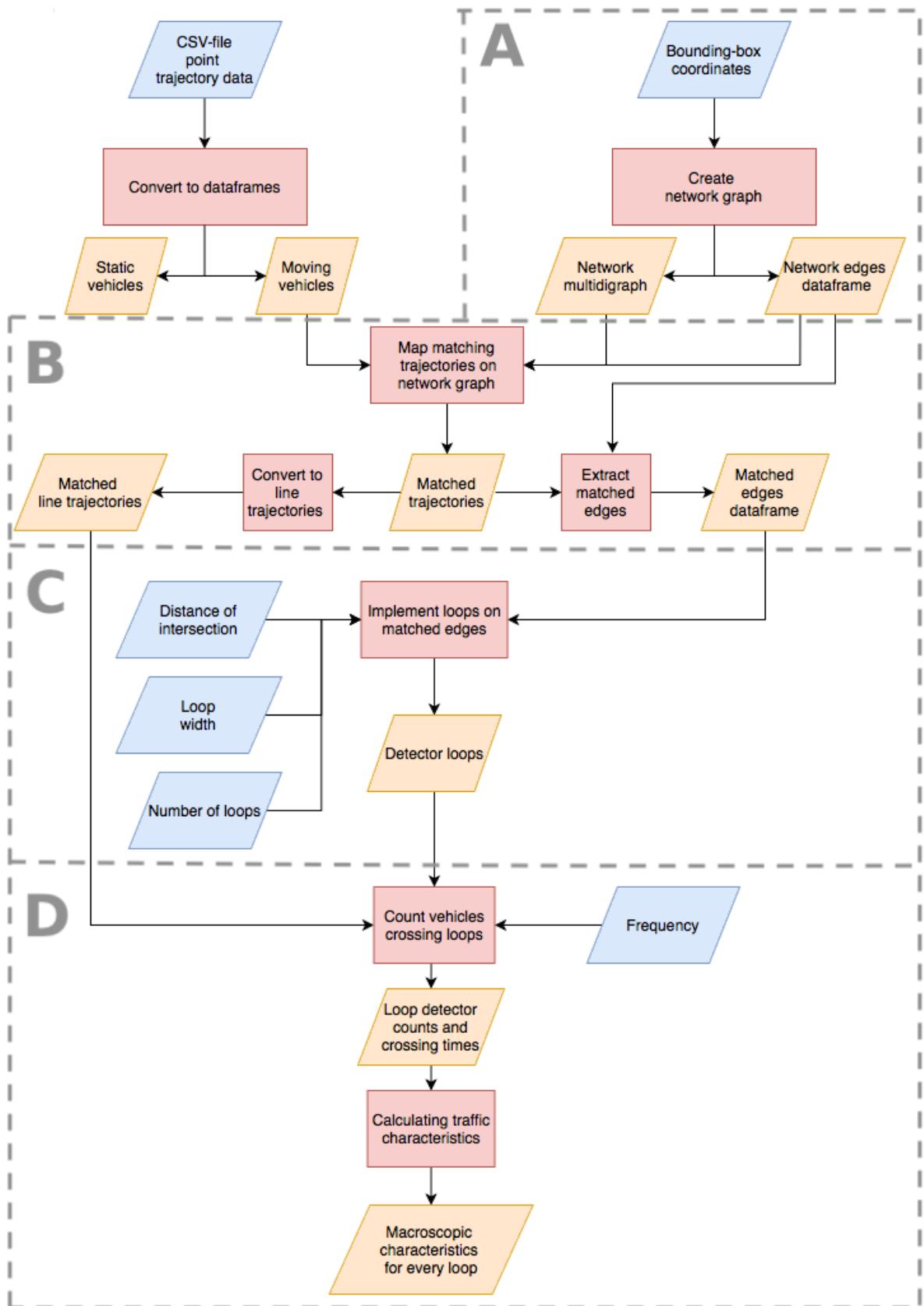


FIGURE 3.2
Program structure

degrees, and is determined with the coordinates of the start and end node. The presence of dedicated bus lanes (DBL) is split up in with flow (right) and contra flow (left) lanes.

Since OSM is a collaborative mapping initiative, updating or adding attributes and streets is very easy, this makes the map very dynamic but also prone to errors. Therefore it is important to keep in mind that the network structure may have changed over time, e.g one-way streets and pedestrian zones. Customizing the keys for the network type and the attribute tags, which can be very interesting for future iterations to extract a more accurate network or research extra attributes, is easily done using OSMnx.

Network type	Nodes	Edges
Drivable	1028	1872
All	2426	5613

TABLE 3.3
Comparison of network types

Figure 3.3 shows the street network and the difference between a chosen network type for Athens in a specified bounding-box. It is clear that extracting all streets results in a very dense network with a lot of edges around intersections, table 3.3 quantifies this difference. Note that the drivable network is a subset from the network containing all streets.



FIGURE 3.3
Street network for different network types

3.2.2 MAP-MATCHING

Knowing on which road a vehicle drives at a specific time gives valuable microscopic traffic information that can be translated to the macroscopic level. In order to achieve this, the acquired trajectory data with a latitude and longitude for every point has to be mapped to its corresponding edge of the underlying road network. In recent years numerous map-matching algorithms have been developed. Ranging from purely geometrical algorithms to probabilistic ones. For this research no new map-matching algorithm is developed, instead a publicly available open-source algorithm for python is chosen that is easily

compatible with the OSMnx package. The *leuvenmapmatching* package (Meert and Verbeke, 2018) fits these requirements. This algorithm is based on the hidden Markov model (HMM) approach presented in (Newson and Krumm, 2009).

3.2.2.1 ALGORITHM BACKGROUND

An HMM is a probabilistic model that is an extension of a Markov model but now the states are hidden and only observations associated with these states are observed. Following Rabiner (1989) an HMM comprises five elements:

1. N hidden states S_1, \dots, S_N
2. A transition model $a_{ij} = P(q_{t+1} = S_j | q_t = S_i)$
3. An initial probability distribution $\pi_i = P(q_1 = S_i)$
4. Domain of possible observations
5. An observation model $b_i = P(o_t | q_t = S_i)$

At every time instance the model is in one of the possible states and the transition model defines the probability of the model being in another state at the next time instance given the current state, known as the Markov property. The initial probabilities are necessary to know the state of the model at the first time step. The difference with the standard Markov model is the observation model, since it is not known directly in which state the model is at a given time but only observations are available, the probability of an observation corresponding to a state has to be defined. When all the parameters of the model are defined some inference problems need to be solved, depending on the application. An overview of these problems is given in (Devos, 2018).

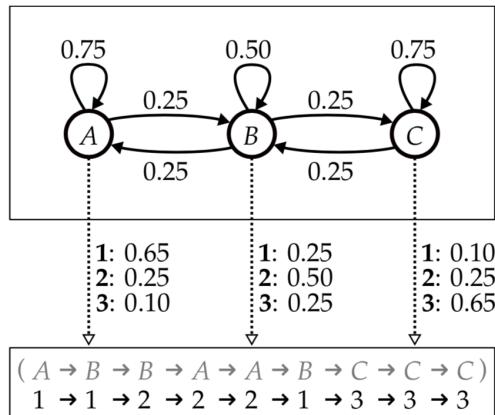


FIGURE 3.4
 Example of HMM taken from (Devos, 2018)

In figure 3.4 an example of an HMM is shown. Instead of observing states A, B and C directly, observations 1, 2 and 3 are observed with a certain probability associated with the states. It is said that the states emit the observations with this probability. For this example, the probabilities to observe 1 are 0.65, 0.25, 0.10 for A, B and C respectively.

In Devos (2018) the theoretical principles of the used algorithm are elaborately explained³. For map-

³For more explanation about their *leuvenmapmatching* algorithm the authors referred to (Devos, 2018)

matching, the desired output is a sequence of states that corresponds to a sequence of observations, this is one of the stated inference problems, referred to as the *best-state-sequence* problem.

$$\underset{q_1, \dots, q_T}{\operatorname{argmax}} P(q_1, \dots, q_T | o_1, \dots, o_T, \theta) \quad (3.1)$$

θ in equation 3.1 depicts the parameters of the HMM. Using dynamic programming this problem is solved by the Viterbi algorithm (Forney, 1973). This algorithm uses recursion to determine the *best-state-sequence* to arrive in state S_j , see equation 3.2 with ξ the Bayes normalization factor as stated in (Devos, 2018).

$$\delta_t(i) = \begin{cases} \frac{1}{\xi_1} \pi_i b_i & t = 1 \\ \frac{1}{\xi_t} \max_i [\delta_{t-1}(i) a_{ij}] b_j & 1 < t < T \end{cases} \quad (3.2)$$

This theoretical formulation is translated to the map-matching case. For equation 3.2 this means that for every edge the best route of edges is calculated to maximize the probability given a sequence of observations.

- edges ($N1, N2$) form the set of states
- all points of the map are in the domain of observations, representation depends on chosen projection (lat-lon, euclidean). Trajectories and the map thus need to be represented in the same projection.
- the transition model to move from one edge to the next is modeled with an exponential distribution of the difference between the distance of consecutive observations and the corresponding distance along the network edges (Newson and Krumm, 2009)
- the observation model to link an observation to a state with certain probability, modeled with an exponential distribution of the straight-line distance of the observation to the edge (Newson and Krumm, 2009)
- initial probabilities are defined by the observation probability of the first observation of a trajectory

3.2.2.2 IMPLEMENTATION

Firstly, the needed network graph for the algorithm is constructed from the extracted OSM graph by taking the node id, node location (lat-lon) and all the directed edges start and end nodes. Secondly, to map every individual trajectory some parameters have to be specified, namely an initial searching distance and a maximum searching distance. The former sets a searching radius to find all possible starting edges for the first observation, the latter specifies the maximum allowable straight-line distance of the observation to one of the possible edges. Translated to the theoretical model, the set of possible states for the first observation is constructed with the initial maximal distance and subsequently the initial probabilities for these states are calculated for those edges within the maximal allowed distance. It is clear that these two parameters limit the number of possible states for every observation and thus have a great impact on the speed of the algorithm. Since urban networks are quite dense, the number of possible edges can increase rapidly with slightly higher parameter values, it is thus not advisable to set general values for the whole trajectory dataset. By looping over all trajectories, starting with the same initial values for the parameters these parameters can be increased for every trajectory individually if the algorithm cannot find a matching result. The result of matching a trajectory to the network is depicted in figure 3.5.

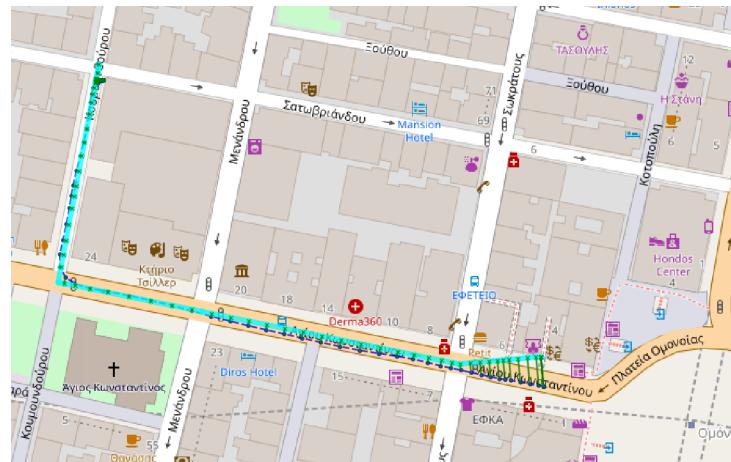


FIGURE 3.5
Result of map-matching for one trajectory

3.2.3 VIRTUAL LOOPS

At this stage every trajectory has for every data point a matched edge. Extracting macroscopic traffic characteristics for a specific location is achieved through placing virtual loops, referred to as detectors hereafter, over the corresponding edge. These detectors are pairs of parallel straight lines perpendicular to the corresponding edge and should at least cover the whole road width. The flow chart, figure 3.2, states three input variables that are freely chosen, namely a specified distance from the upstream and downstream intersection, the width of a virtual detector pair making up the measurement region (ΔX) and the number of detectors on an individual edge. It should be noted that the intersections of the network graph are point approximations of the real intersection, substantiating the necessity of the 'distance from intersection' (DFI) parameter. When the chosen number of detectors is higher detectors are evenly distributed over the length between the first and last detector.

3.2.3.1 IMPLEMENTATION

From the map-matching result a subset of the street network gives all the edges used by at least one vehicle. This prevents placing virtual loops on edges with no matched vehicles. Since the two distance input parameters are in meters the network is projected to euclidean space to assure right placement of the loops. The first step is finding the right locations of every detector pair on the edge given the three input parameters.

Once the right locations are found, the points are transformed to lines with a length long enough to detect every vehicle matched to the edge and a bearing that is perpendicular to the edge. Afterwards, the dataframe is again converted to the right coordinate reference system (crs). When one of the input parameters makes a right placement impossible, some adjustments are made:

1. **Distance from intersection (DFI)** When the edge is not long enough to assure the specified distance from the intersection, all detectors, depending on the specified number, are placed in the middle of the edge.
2. **Detector width (ΔX)** If an edge is shorter than ΔX , the detector is placed in the middle with a newly specified ΔX relative to the length
3. **Number of detectors too high** Given the number of detectors, an inter-detector length is determined

and if this causes a violation for one of the other two parameters, all detectors are again placed in the middle

The adjustments assure that the ΔX is preserved as long as the edge length allows it, making comparisons possible between different detector locations because the space-time window is the same. Courbon and Leclercq (2011) state that specifying this window is difficult which implies that comparing windows of a different size is inaccurate. As an example figure 3.6 depicts the placement of three detectors per edge.



FIGURE 3.6

Detectors on used edges of network with specified DFI, ΔX and number of detectors per edge

3.2.4 TRAFFIC CHARACTERISTICS MEASUREMENT

These detectors differ from the traditional double loop detectors on roads because the availability of all trajectories makes it possible to use Edie (1963) for every chosen space-time area on every edge by using the crossing times of every individual vehicle trajectory and calculating their time spent (VHT) and distance travelled (VKT) when inside the space-time area.

Until now the trajectories are made up of chronological points and a conversion to line segments for consecutive points is needed. The crossings of these lines with the created detectors give the exact location and time of a vehicle entering and leaving a detector, assuming constant movement between consecutive trajectory points allows the usage of linear interpolation. The equations 3.3, 3.4 and 3.5 give the average flow, density and speed for every detector, respectively.

$$q_i = \frac{\sum_k x_k}{\Delta X \Delta T} \quad (3.3)$$

$$k_i = \frac{\sum_k t_k}{\Delta X \Delta T} \quad (3.4)$$

$$v_i = \frac{q_i}{k_i} \quad (3.5)$$

with x_k , t_k the VKT and VHT of every vehicle k in space-time area $\Delta X \Delta T$, respectively. The average speed is equal to the space-mean speed. In contrast to traditional loop detectors, ΔX can be as large

as preferred, no assumption of constant speed is needed since the VKT of data points of a trajectory in between the detector edges is determined for every time interval ΔT .

3.2.4.1 IMPLEMENTATION

In order to determine the macroscopic traffic characteristics enumerating all the selected edges with their detectors and the trajectories is done for a chosen space-time area. VKT and VHT are calculated for every individual trajectory and the values are assigned to a specific time step.

To prevent negative values, tags are used to assure that VHT and VKT are only measured when the detector is crossed correctly. This tag is also necessary to include VKT if a vehicle's line segment is completely inside the space-time area. After the vehicle crossed the detector, these tags are reset to allow the same vehicle to cross the detector again, i.e. vehicle performing loops in the network.

All the values for one edge are saved in a dataframe with one time step per row and lists in every column comprising every individual trajectory's VHT and VKT for that particular time step. Keeping the individual trajectory values gives more possibilities for further analysis, for example splitting by mode or a travel time analysis based on entry and exit times of specific detectors. For different edges this results in a list of dataframes and additionally a new list of dataframes with the macroscopic value for every time step by using the equations 3.3, 3.4 and 3.5.

3.3 APPLICATION AND EVALUATION

This section describes the process to attain accurate values of the macroscopic traffic characteristics used for later analysis. Applying what was discussed earlier to the dataset and doing some extra iterations results in an eventual street network with an accurate map-matching of all the trajectories.

3.3.1 DATA

The used dataset in this research consists of 10659 different trajectories with the longest trajectories having a total time of fifteen minutes. Leading to a total set of over 1.5 million datapoints. Of these trajectories, 69 did not move during the whole recording period and are therefore not considered anymore. For the remaining dataset the share of every vehicle together with its corresponding dimension, length and width in meters, is shown in table 3.4. Four trajectories are not any of these vehicle types and are not used for further analysis. Cars, PTWs and taxis make up more than ninety percent of all the trajectories.

	Car	Taxi	Bus	PTW	Medium Vehicle	Heavy Vehicle	Total
Share	3989 (38 %)	1739 (16 %)	198 (2 %)	4037 (38 %)	526 (5 %)	97 (1 %)	10586 (100 %)
Dimension (l,w)	5,2	5,2	12.5,4	2.5,1	5.83,2.67	12.5,3.3	

TABLE 3.4
Proportion and dimensions in meters of vehicle types in the dataset

3.3.2 MAP-MATCHING ISSUES

Selecting the right network type is of crucial importance to get accurate results. As mentioned earlier there can exist a high difference in the number of nodes and edges due to the chosen network type. The highly accurate trajectory dataset asks for an accurate enough street network, however choosing a too detailed map will result in long computational time for little added value. For the given dataset the network type should comprise all the roads whereon the recorded vehicles can drive. In the first try, the chosen type contains all drivable roads which results in extracting all the streets for motorized traffic for the selected region. Performing the map-matching algorithm brings forward some issues, as listed below and depicted in figure 3.7.

1. **Too low initial distance** When the maximum initial distance is not high enough it can happen that the desired edge is not in the set of starting states for the first observation. Since the graph is build solely on the location of the nodes, the found edges are those of which the starting node is in the set of possible starting nodes. Therefore, when the starting node is not inside the initial searching area the corresponding directed edge will not be considered when the initial probabilities are determined for this first observation and the point is assigned to the wrong edge.
2. **Too high maximum distance** The straight-line distance from the observation to all the edges within the set of states cannot exceed the maximum distance parameter. When this happens the algorithm will fail to find a corresponding edge for that observation. Therefore the maximum distance should be high enough for all the points along the trajectory but this, on the other hand, can result in incorrect matching.
3. **Missing edges** It can happen that for some trajectories the correct edge is missing in the graph, two sorts of cases were encountered. The first one is due to vehicles driving on the pavement in the opposite direction of the corresponding road, ignoring one-way street signs or using pedestrian streets or some other streets which were not extracted from OSM because they are not part of the chosen network type, see section 3.2.1. The second case consists of roads missing that should be part of the network type. This is mainly observed for locations where a DBL is present.

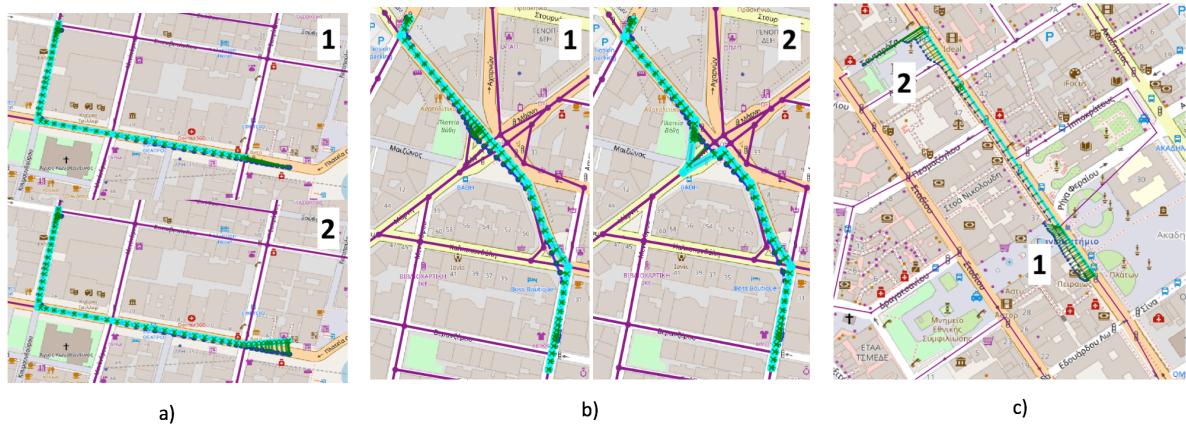


FIGURE 3.7

Matching issues: a. maximum initial distance, the upper figure (1) has a too low initial distance b. maximum distance, the right figure (2) maps wrongly because of a too high distance c. missing edges, the trajectory is wrongly matched at the beginning and the end. It makes a U-turn on the pavement in the beginning (1) and at the end goes into a pedestrian street (2)

The initial maximum distance is set to the maximum length of an edge in the network to solve the first issue. Assuming that an observation will not be further away from its correct match than this maximum

distance ensures that the right edge is always part of the initial set of possible states. Situation 2 in figure 3.7a shows that the first observations of the trajectory match to the right initial edge.

For the second issue the reason of the incorrect matching shows a possible shortcoming of using the HMM algorithm for this dataset with a one second sample rate. Since the observation probabilities are determined for all the states in the vicinity of the observation, it can happen that the trajectory is closer to the incorrect edge, given the dense street network. When the vehicle stops afterwards in between the two edges, the Viterbi algorithm will define the incorrect matching as the *best-state-sequence* up until the vehicle starts moving again. Because of the Markov property this sequence is not shifted to the correct one since no other route maximizes the inference problem, due to the stopping, and a detour is made. By increasing this distance when the algorithm fails this issue is resolved in an iterative way, although increasing it too slowly will impact the speed of the algorithm, increasing it too fast can again lead to the incorrect matching, especially at dense areas in the network, e.g. intersections. Situation 1 in figure 3.7b now does not make a detour at the intersection. In the comments the authors acknowledge that setting the maximum distance is a hard cut and quite arbitrary and propose using a minimum probability but choosing this value is also arbitrary and less interpretable. Using the maximum distance is thus preferred.

Solving the last issue is split up, given the two encountered cases. When connections are missing that should be part of the network, the open access of OSM allows adding these connections. For this dataset mainly DBLs are missing and are added by changing the respective tag and using this to create an extra edge if it is a contra-flow lane in a one-way street. The bundled bus trajectories show where buses drive and should overlap with the bus lanes, see figure 3.8. The cases where roads are used that are not part of the network, customizing the list of keys and tags can help to extract the best network, e.g. including pedestrian streets in the drivable network. Nevertheless, this only solves the most clear cases, when vehicles are maneuvering and using the pavement or ignoring one-way street directionality changing the network structure is not advisable. Instead a new column is created in the trajectory dataframe containing the difference between the bearing of the trajectory point and the matched edge, quantifying the mismatch.



FIGURE 3.8
All the bus trajectories in the researched area

Taking into account the solutions for the issues mentioned, the parameter values for the map-matching algorithm are set. For the initial maximum distance this is the maximum length of all the edges in the

network and the maximum of the straight-line distance to a possible matching edge is set at five meters and increased by five every time the algorithm fails. Five meters seems a reasonable distance knowing the high sample rate of the trajectory dataset and the dense network structure.

3.3.2.1 GRAPH ADJUSTMENTS

After applying the proposed adjustments, the street network is extracted a second time with a customized set of keys, including pedestrian streets, and the DBL tag is used to create missing edges where needed. Figure 3.9 shows the network with the pedestrian streets added to the network type.



FIGURE 3.9
Network with pedestrian streets included

In figure 3.10, the newly added bus lanes are depicted on the left and the total of all DBLs after the addition on the right. These bus lanes can be used by taxis and PTWs, making these missing links not negligible knowing the high shares of these two modes in the dataset.



FIGURE 3.10
DBL in network: A. Newly added DBL B. All DBL in network

3.3.2.2 EVALUATION

After running the algorithm with the right parameter values, taking into account the encountered issues, the whole trajectory dataset is matched to the network. However, it is not assured that even after addressing these issues every trajectory is correctly matched. Evaluating the matching result is therefore essential in order to be sure that the matched result corresponds with the observations. Since ground truth data is not available, the video footage is not analyzed in this research, some other defined variables can shed more light on the accuracy. In Brakatsoulas et al. (2005) a map-matching algorithm uses the Fréchet distance to match trajectories to a map. This distance gives a quantitative measure of similarity between two curves and is better known as the walking-dog problem: what is the longest length of a leash if a dog walks on one curve and a person on the other without the ability of backtracking. Calculating this distance continuously is computationally expensive and Eiter and Mannila (1994) present a discretized version of this distance, but interpreting this distance on its absolute value is difficult because the discretization seems to give high distances even for well-matched trajectories. Keeping the walking-dog problem in mind, a combination of other variables is used to evaluate the overall map-matching and find the badly matched trajectories. The variables are:

- median straight-line distance
- mean straight-line distance
- maximum straight-line distance
- difference between trajectory and matched length
- bearing difference between trajectory point and matched edge

Figure 3.11 shows the histograms for the different variables after map-matching all the trajectories. These clearly show that the overall accuracy of the map-matching is good, as was expected given the high sample rate of the trajectories and the well-considered setting of the parameters discussed earlier. The median and mean follow the same shape and have almost equal values indicating that for most trajectories possible mismatches are not due to a couple of large outliers. The maximum distance has its peak around ten meters which is reasonable given that road width is not given in the network graph, but it does show a second peak around twenty meters. Overall the bearing difference is very low for a large majority of the trajectories. A high value for any of these variables can point to a mismatch between the trajectory and the network somewhere along the path.

The second peak observed in the maximum distance histogram shows that there still are some possible mismatches between trajectories and the network, although the adjustments of the graph have improved the overall map-matching a lot, see table 3.5 for the comparison of the maximum distance for both network types. The most prominent improvement is the value for the 75th-percentile, with the adjusted network three quarters of the trajectories have a maximum distance lower than ten meters. Figure 3.12d shows that ten meters is a very accurate value because the width of the roads is not taken into account and the location of the edges extracted from OSM may not always be in the middle of the ground truth road, resulting in a systematic deflection.

A deeper look into the trajectories having a higher maximum distance than ten meters gives valuable information about the reasons of mismatches. The reason for the second peak around twenty meters follows from the way the graph of the map-matching algorithm is built. Only the node locations are used to define an edge, omitting any curvature of the edges. Figure 3.12b shows the situation when the edges around Omonoia square are left out, it is clear that the peak disappeared for the trajectories not travelling on one of the Omonoia edges and that for the trajectories passing Omonoia, figure 3.12a, the maximum distance lies around twenty meters.

	Original [m]	Adjusted [m]
mean	10.73	9.18
std	9.51	7.85
min	0.14	0.14
25 %	4.54	4.36
50 %	7.35	6.80
75 %	16.51	9.92
95 %	24.68	24.10
max	122.63	106.96

TABLE 3.5

Comparison of maximum distance after map-matching for the original and adjusted network

Further analyzing the dataset without the trajectories passing Omonoia, the correlation between the evaluation variables provides a clear picture of the variables that can flag a mismatch. In tables 3.6 the correlation matrices are shown for two different selection procedures. The first one, left figure, contains all the trajectories with a maximum distance higher than ten meters. The other one, right figure, has all the trajectories with an average speed higher than one kilometer per hour when their bearing difference exceeds the ninety degrees threshold. The reason for this last selection comes from the fact that vehicles stopping for a longer time have a different bearing because of maneuvers and parking in a direction leading to a high bearing difference with the matched edge. Therefore these vehicles are not mismatches and are left out in the second procedure.

	MD	LD	LDr	BD
MD	1	0.5	0.28	0.26
LD	0.5	1	0.68	0.37
LDr	0.28	0.68	1	0.59
BD	0.26	0.37	0.59	1

	MD	LD	LDr	BD
MD	1	0.53	0.21	0.26
LD	0.53	1	0.63	0.5
LDr	0.21	0.63	1	0.83
BD	0.26	0.5	0.83	1

TABLE 3.6

Correlation matrices for two selection procedures, left: maximum distance, right: average speed (MD: maximum distance, LD: length difference, LDr: relative length difference, BD: bearing difference (>90 deg)

All correlations are positive, as is expected, since high values for any of the variables points out possible mismatches. From these tables the most interesting correlation is the one between the relative length difference and the bearing difference, vehicles traveling in a direction that does not correspond with the direction of the matched edge have a very high difference in their actual and matched path. The reason for these mismatches is due to the third issue explained earlier, see 3.3.2. where vehicles ignore one-way streets or use the pavement to drive in another direction than the orientation of the corresponding road. A majority of those cases are PTWs, they use more frequently the pedestrian streets or other paths connecting two roads, leading to very high length and bearing differences.

In conclusion the algorithm gives accurate results, when mismatches occur it is because some edges are missing in the network, e.g. pavements and other walking paths, or maneuvers and driving behaviour that lead to high bearing differences. A maximum distance of ten meters is the reasonable threshold, excluding those edges with a lot of curvature, and additionally high length differences and bearing differences allow to assume some directionality issue with the matched edges.

3.3.3 TRAFFIC CHARACTERISTICS

After improving the map-matching result, the traffic characteristics are determined for every detector on the edges of the network. Taking into account possible mismatches, the tags prevent from counting crossings in the other direction and trajectories with a bearing difference higher than ninety degrees are skipped, choosing a lower difference is possible but this can wrongfully leave out trajectories due to lane changes, for example.

3.3.3.1 SPACE-TIME AREA

Selecting a window to aggregate all VKT and VHT is an important but also arbitrary task. ΔX can be very short or as long as the edge length, similarly ΔT can take one second steps or be fifteen minutes long. Figure 3.13a shows scatter plots for different ΔT . With only fifteen minutes of available data, taking too long ΔT results in only a few data points, and too short in very scattered data with peaks that are an overestimation of the ground truth. Aggregating the data to a multiple of ΔT is still possible without making any assumptions or incorrect averaging, i.e. time-mean speed instead of space-mean speed, because the individual VKT and VHT are available.

The plots in figure 3.13b show FDs for different ΔX . It is clear that wider loops show more scatter because different traffic states are mixed more frequently. The width for the detector also determines how many detectors per edge can physically be placed. As shortly mentioned in the previous chapter, Courbon and Leclercq (2011) state that placing many detectors in order to measure as many traffic states as possible gives good MFD estimates. Additionally, vehicles that do not cross both edges of the detector are left out for the traffic characteristics calculation, this would happen more frequently with wider detectors.

3.3.3.2 MIXED TRAFFIC

The aggregation of VHT and VKT can also easily be split up by mode, equations 3.6 and 3.7

$$\sum_j^N t_j = \sum_m \sum_j^{N_m} t_j \quad (3.6)$$

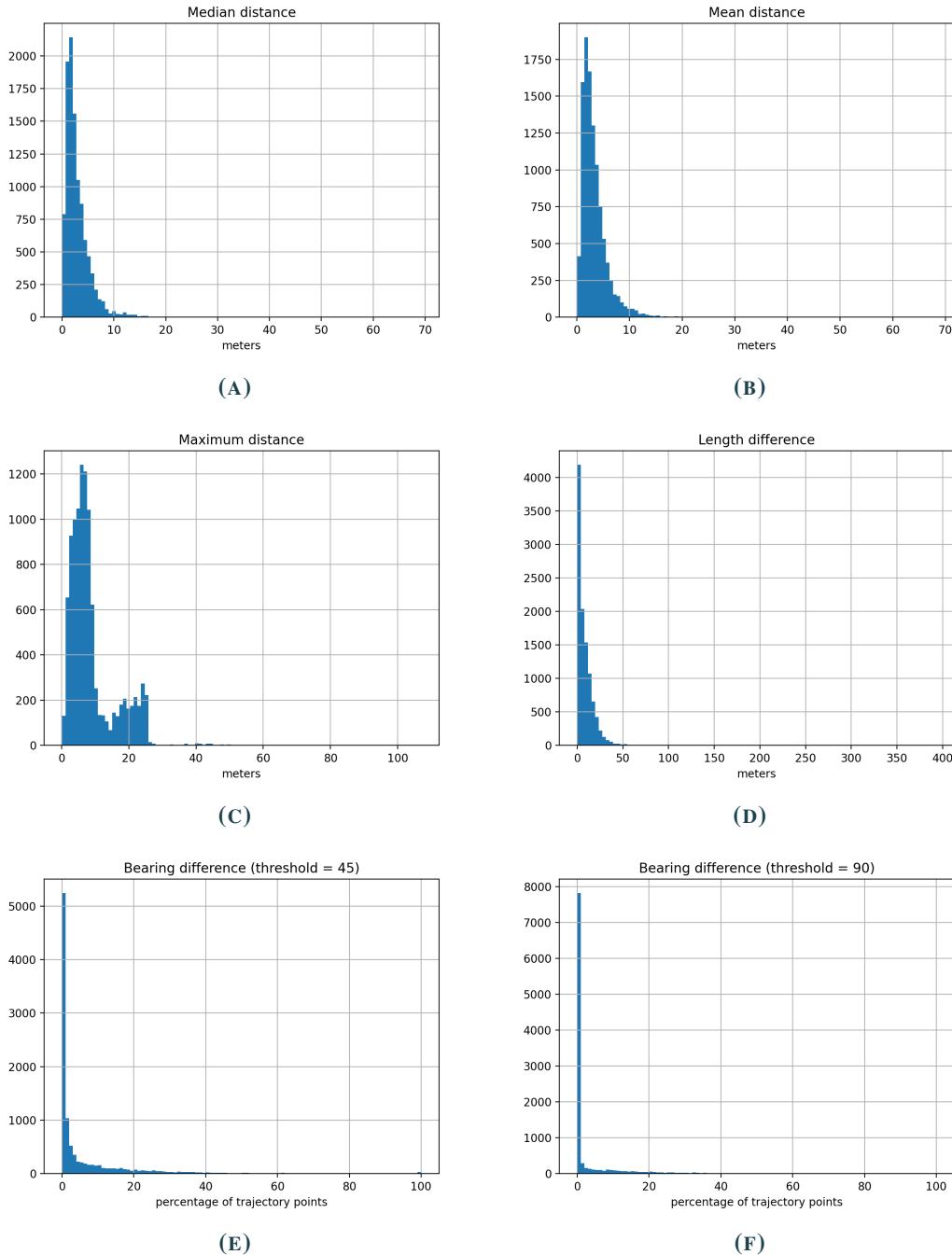
$$\sum_j^N x_j = \sum_m \sum_j^{N_m} x_j \quad (3.7)$$

With m for the different modes and N_m the subset of trajectories for a specific mode. Because temporal changes in the number of vehicles passing influences the magnitude of the flow and density, comparing the space-mean speed between modes is preferred.

3.4 SUMMARY

The presented framework leads to the extraction of traffic characteristics, i.e. flow, density and space-mean speed, for the highly-detailed trajectory dataset of the pNEUMA experiment. In order to know where vehicles are at a particular moment in time, a map-matching algorithm is used to get the most probable edge for the vehicle at every time instance. With detectors placed on every used edge, calculating VKT and VHT to come to the macroscopic traffic characteristics becomes very easy for selected locations in the network. Afterwards some issues are elaborately discussed and resolved, improving the result and

reliability of the map-matching output a lot. For this case the initial searching distance is set equal to the longest edge in the network and the maximum allowed straight-line distance starts at five meters increasing with five when the algorithm fails. Additionally, some evaluation variables for the map-matching result are presented to quantify persisting inaccuracies due to the chosen network type and the behaviour of some vehicles.


FIGURE 3.11

Histograms of variables used for evaluation: A, B and C show the median, mean and maximum straight-line distance to the matched edge, respectively, D. difference between trajectory and matched length, E and F. percentage of trajectory observations with bearing difference exceeding 45 and 90 degrees, respectively

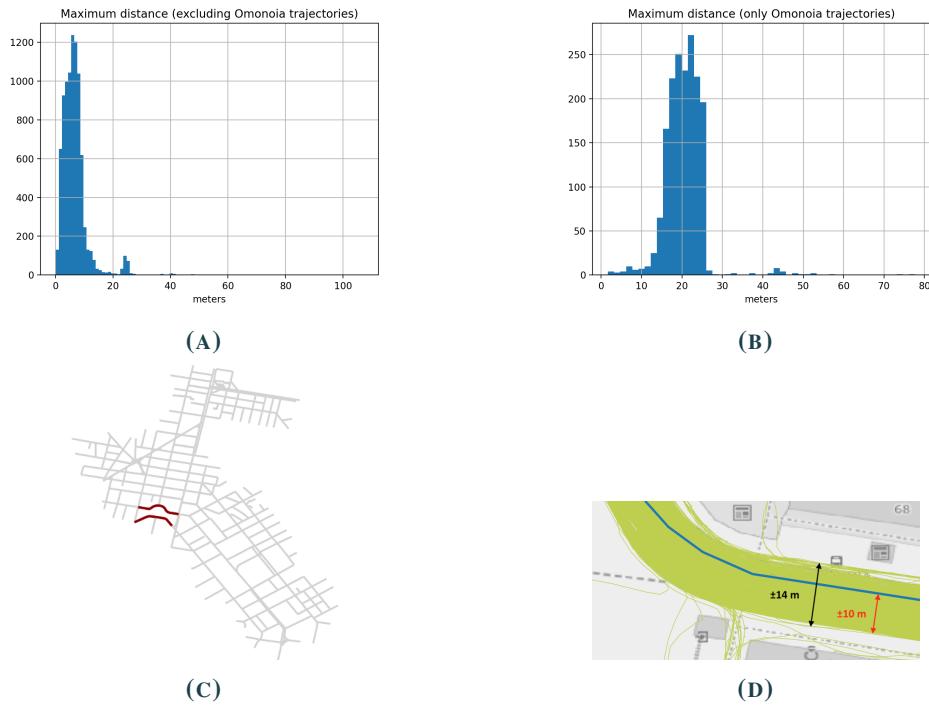


FIGURE 3.12

Maximum distance histograms: A. Trajectories not travelling on Omonoia edges B. Only trajectories travelling on Omonoia edges C. Edges of Omonoia square D. Width in black and maximum distance in red of all trajectories matched to an edge

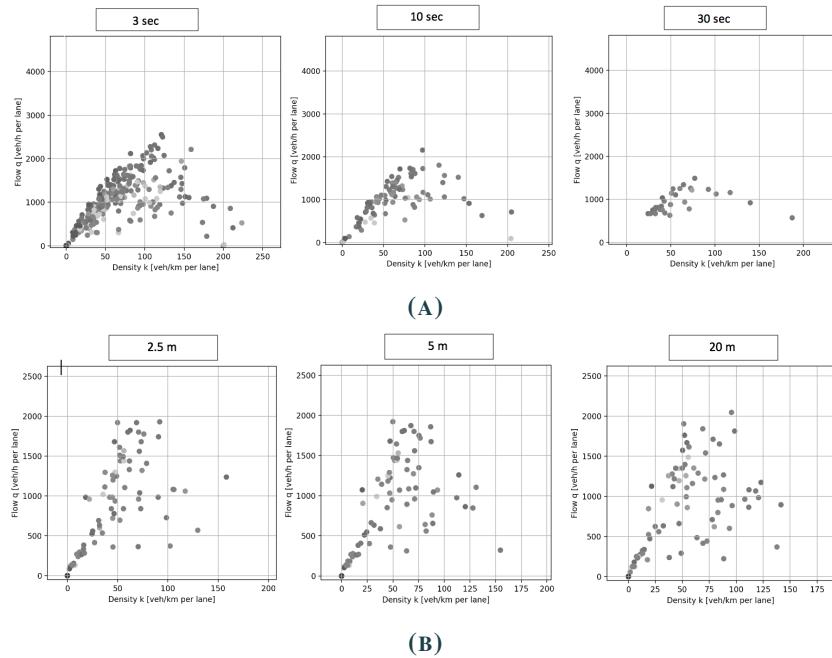


FIGURE 3.13

Space-time area: A. Flow-density scatter plots for different ΔT B. Flow-density scatter plots for different ΔX

CHAPTER 4

URBAN TRAFFIC ANALYSIS

In this chapter the traffic characteristics are determined for a selected portion of the network edges. The first part discusses the link fundamental diagram, subsequently the attained MFD is discussed. The last part describes the evolution of the characteristics and travel time over different arterials. Since the dataset contains different vehicle types, the modal effects are also analyzed. The high heterogeneity in driving behaviour, vehicle types and congestion has a large effect on the empirical results and is discussed.

Based on section 3.3.3.1 values of $\Delta X = 5m$ and $\Delta T = 10s$ are chosen for the space-time area to aggregate all individual VKT and VHT. This time step results in ninety-two data points of density, flow and speed for every detector. Three detectors are placed with a DFI of fifteen meters giving measurements at the begin, middle and end of every edge, except for short edges where the detector is only placed in the middle.

The computational time to determine traffic characteristics for all the detectors on the edges evolved during the development of the tool. Therefore, a large part of the analysis uses only a selection of the main arterials, but a network comprising all drivable roads excluding residential ones is also used for empirical MFD estimation¹.

4.1 NETWORK

Keeping only the edges with at least one vehicle matched to it results in figure 4.1a. This shows the main arterials, primary roads, and the secondary and tertiary roads through the researched area, comparing this with figure 4.1b confirms that the highest number of matches corresponds more or less to these highlighted roads. Figure 4.1c shows the location of the selected arterials in the network used for further analysis and table 4.1 gives the length of the used network and the selected arterials.

The total length of the selected arterials spans 45 % of the combined length of primary and secondary edges in the used network and DBL are present on the Alexandras, Akadimias and 28 October arterials. The total lane-length is around 57 km for the whole used network and 19 km, one third of the total, for all the arterials combined. In appendix A detailed figures of all the arterials are shown with corresponding labels for every individual edge being part of the arterial. The larger network used for the MFD has a total lane-length of 42 km and consists of all the roads with a primary, secondary or tertiary OSM tag, this corresponds to the network of all highlighted edges in figure 4.1a.

¹Eventually, the developed code allowed a faster analysis of the larger network

Used Network	Length (km)	Lane-length (km)	Arterials	Length (km)	Lane-length (km)
Primary	8.7	23.4	Akademias	1.22	3.67
Secondary	4.7	10.9	Alexandras EW	0.37	1.02
Tertiary	4.1	7.3	Alexandras WE	0.33	0.69
Residential	10.4	10.9	October 28	0.97	2.62
Pedestrian	3	-	Omonoia	0.36	1.21
DBL*	3.2	-	Panepistimiou	1.03	4.54
			Stadiou	1.04	3.12
			Titris	0.80	2.07
Total	32.5	57.2		6.12	18.9

TABLE 4.1

Length of used network and arterials (*DBL are not part of total length, corresponding edge is counted)

4.2 FUNDAMENTAL DIAGRAM

As mentioned in chapter 2, the FD is different for signalized arterials, leading to the trapezoidal FD. The link FD is analyzed and the effects of the different types of heterogeneity are discussed. Subsequently, all measurements are aggregated to estimate the empirical MFD and different aspects are discussed.

4.2.1 LINK FUNDAMENTAL DIAGRAM

This section researches if this trapezoidal shape is observed and what the effect of the heterogeneity is on the individual diagrams. For this analysis individual diagrams of different sections of the different arterials are used.

For OM 1, PAN 14 and ALE 2 the flow-density and speed-density diagrams are shown in figure 4.2. The high scatter is immediately visible for all three sections, mostly caused by the short time steps and heterogeneity in vehicle type and traffic states in different lanes, as well as the presence of traffic signals. Averaging over the number of lanes makes comparisons between different detectors on other edges possible but is susceptible to incorrect values in OSM and is thus used carefully before drawing conclusions. Differences between lanes are also not distinguishable because the trajectories are matched to a road and not an individual lane.

From the individual diagrams the free flow branch is clearly visible and congested states seem present for some time steps at all three locations, having very low speeds. In the flow-density diagrams for the second and third detector of OM 1 a parallel branch shifted to the right is visible, flows still increase linearly but the density has an extra offset. In the second and third flow-density diagram of PAN 14, on the other hand, very high densities are observed at very low flows, averaged over the number of lanes this gives more than 400 veh/km which is physically not possible when assuming a passenger car as the standard vehicle dimension.

PARALLEL BRANCH Since the detector aggregates all VKT and VHT over the total road width, differences between lanes are not directly distinguishable. If high heterogeneity exists between lanes, the overall traffic characteristics are heavily affected and represent a traffic state deflecting strongly from actual traffic states in the different lanes. In figure 4.3a the number of vehicles present inside the

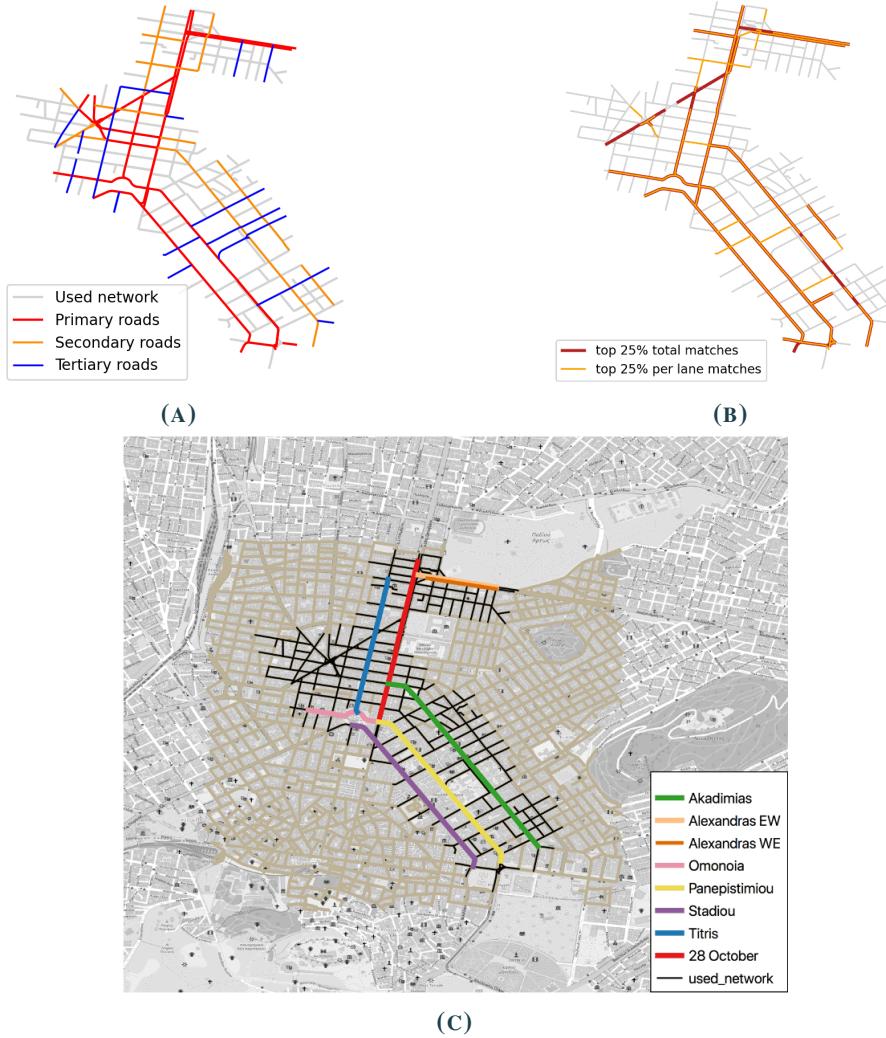


FIGURE 4.1

Used network: A. Topology highlighted B. Edges with top 25 % matches on total and per lane basis C. Selected arterials

space-time area for a whole time step and the total number of vehicles crossing that area are depicted for the second detector of OM 1. From time steps 27 to 44 and 57 to 74, one vehicle is constantly inside the space-time area but the number of vehicles crossing the area still follows variation due to traffic signals and is not significantly lower than time steps without a vehicle stalling. This implies that the traffic stream is not really affected and the stalling vehicle itself is not creating a bottleneck. Comparing two specific time steps, without and with a stalling vehicle (steps 25 and 27, respectively), results in table 4.2 and histograms 4.3b with the distribution of the individual vehicle speeds. The speed distributions are for both time steps more or less the same with most speeds between 15 and 25 km/h but the average space-mean speeds are very different. Leading to the conclusion that the stalling vehicle is not affecting the traffic stream. Adjusting for this stalling vehicle to get the FD for the remaining lanes is shown in figure 4.4. The FD shifts to the left and shows a clearer trapezoidal shape, the average speed is higher than 20 km/h for densities up to 200 veh/km and decreases to speeds below 10 km/h for higher densities aggregated over all remaining lanes. Note that there are numerous reasons of a vehicle stalling which are not immediately clear at the macroscopic level and it is also not known where exactly the vehicle is standing. A vehicle blocking an active lane, for example, should result in a disturbance of the flow

and a shockwave travelling upstream but when a vehicle parks somewhere next to the road, traffic is not affected. The simple adjustment performed here shows how the FD is affected but in order to accurately filter out these instances individual lane behaviour should be incorporated.

Parameter	No stalling	One vehicle stalling
Time step	25	27
Vehicles crossing	17	17
Flow	5602	5774
Density	289	461
VKT/VHT	19.4	12.5
Mean speed IC*	21.7	21.4
Std speed IC*	5.5	6.6

TABLE 4.2

Effect of stalling vehicle with flow, density and speeds in veh/h, veh/km and km/h, respectively (*IC = set of all individual vehicle speeds crossing during the time step)

HIGH DENSITIES The very high densities, some with a non-zero flow, present in the FDs of PAN 14 are not caused by lane heterogeneity but rather by the different vehicle types waiting during a red cycle. Splitting up the calculation of the traffic characteristics by mode provides more insight into these high values. Figure 4.6 depicts the evolution of density and flow over time for the second detector, upstream of the traffic signal, of edge PAN 14. Knowing that this edge has four lanes and the value of a stopping vehicle for one time step adds 200 veh/km to the aggregated density for the chosen space-time area, a jam density of 800 veh/km is expected if only cars, length of a car is equal to chosen ΔX , are in the measurement area of the detector. Though for some time steps the peaks go up to 1000 and more. Analyzing these peak values shows that the presence of PTWs increases the density a lot, their higher maneuverability and smaller dimension allows them to ignore lane discipline, e.g. stand shoulder to shoulder in one lane, during a red cycle and increase the density.

Coifman (2015) shows with empirical data the density's dependency on vehicle length, longer vehicles have lower jam densities, for loop detector data and validates this with the NGSIM video dataset. Therefore, making abstraction of the vehicle dimensions leads to scattered FDs, in Edie's generalized definitions vehicle heterogeneity is not incorporated, and explains the strong variation that is observed for jam densities during the red cycle at detectors close to a signalized intersection, visible in figure 4.6a where different high peaks, at least 800 veh/km, have a different vehicle type mix. Interestingly, some of these peaks have non-zero flows, the dashed vertical lines in figure 4.6 highlight two time periods during an assumed red cycle, the exact cycles are not addressed in this research, in which this is the case. From the flow plot it is clear that only PTWs are moving during these time steps, this effect is known as filtering, PTWs move between stationary vehicles filling the small gaps. In figure 4.5 all the flow-density points of every time step with an average density per lane higher than 200 veh/km are shown. It is clear that the share of PTWs has an influence on the magnitude of the density and that non-zero flows at very high densities only happen when PTWs are present. There are, however, also some points with no PTWs present, the most probable reason is an inaccurate number of lanes assigned to some edges.

From the above it is clear that the individual FDs presented experience a lot of distortions due to lane and vehicle type differences. In conclusion, the scatter of the FDs is due to:

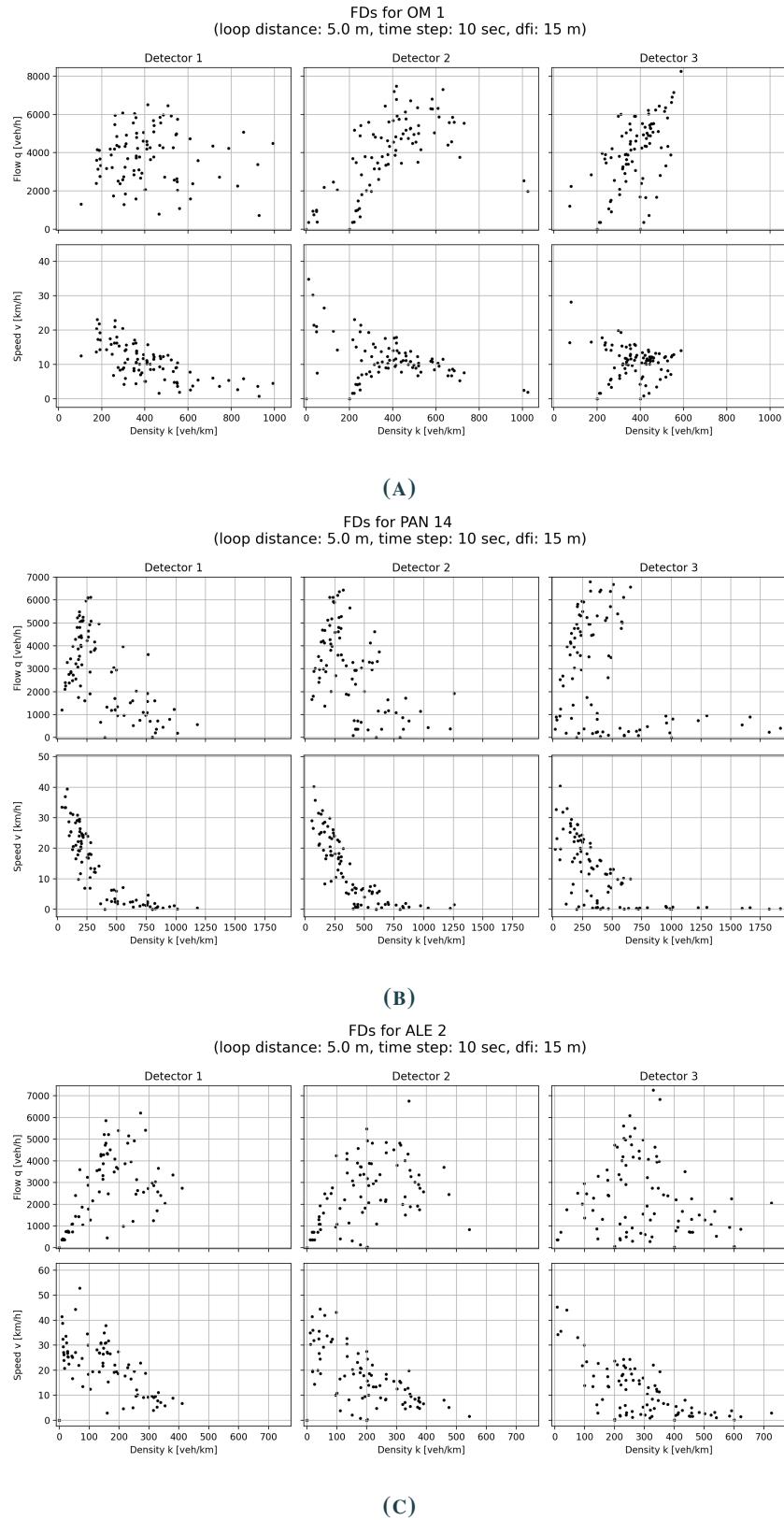
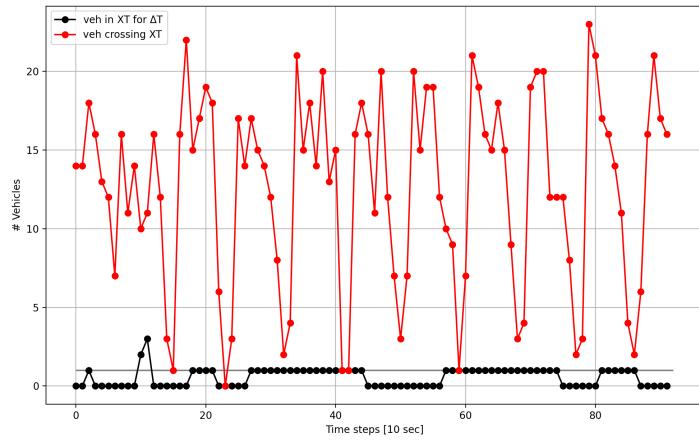
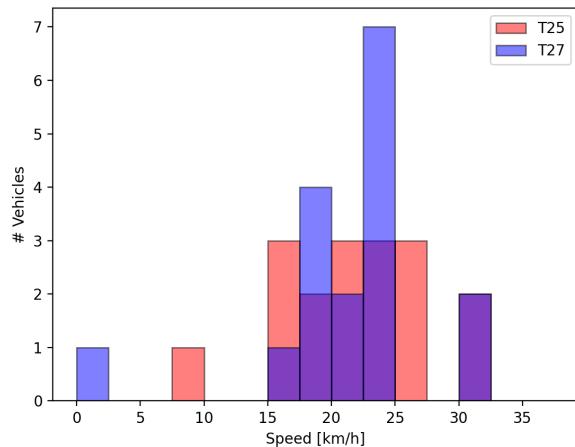


FIGURE 4.2
Link FDs for OM 1 (A), PAN 14 (B) and ALE 2 (C)



(A) Number of vehicles inside space-time area for whole time step and crossing area during time step



(B) Speed distribution of vehicles crossing space-time area

FIGURE 4.3

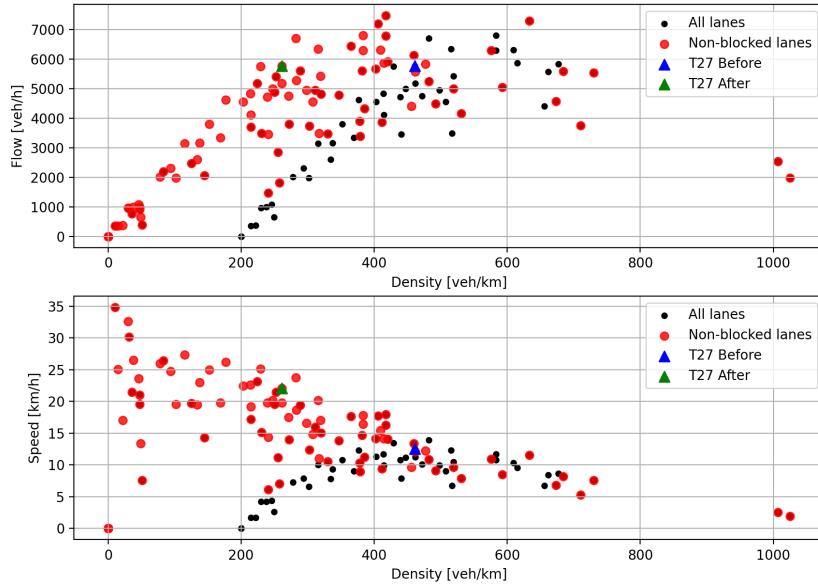


FIGURE 4.4
Adjusting for stalling vehicles in space-time area

- lane heterogeneity
- different vehicle types with different driving behaviour and dimensions
- non-stationary traffic states in same time step
- short time step, less scatter is observed for longer aggregation periods

The expected trapezoidal shape is not always clearly visible due to the scatter or the absence of the top and congested branch. For the second detector of OM 1 the trapezoidal shape is visible after the adjustment for the parallel branch but no heavily congested branch is present. The other arterials also have more detectors with only a free flow branch, especially for the detectors at the beginning and in the middle of an edge, and the highly congested states observed are due to the red cycles rather than over-saturated traffic.

4.2.1.1 SPLITTING MODES

Aggregating different modes has a large effect on the FDs, especially the jam densities. Selecting specific modes and constructing speed-density diagrams gives more insight in how the different modes move in the urban environment. In homogeneous traffic streams only counting a part of the traffic gives the traffic characteristics up to a factor for the full stream. This is not the case in real traffic because of the different types of heterogeneity present.

ONLY PTWs Splitting the trajectories in a group with and without PTWs results in different traffic characteristics. In figure 4.7 the FDs for ALE 2 detector 1 and OM 1 detector 2 are shown. The homogeneous FDs are constructed by multiplying the flow and density of every time step with the share of the corresponding mode, PTWs in this case. For ALE 2 detector 1, figure 4.7a, traffic without PTWs coincides more or less with the homogeneous case for the free flow branch but has higher densities in the

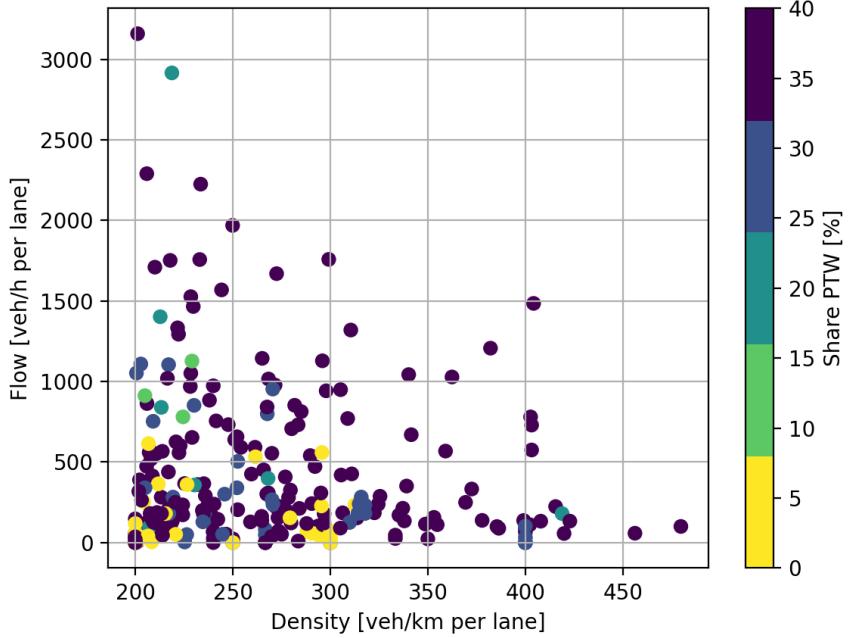


FIGURE 4.5
Effect of PTWs on density and flow

congested branch that are as high as the densities in the overall FD. The FDs with only PTWs do indeed show that they are always in free flow and have no congested states. Interestingly, their overall modal share is above thirty percent for every edge of the arterial, see appendix A, concluding that PTWs do not experience the same obstruction as other traffic. The same is found for the OM 1 case, although the speed of PTWs is lower. Table 4.3 contains basic statistics of the PTW's speed for the two locations and shows the large speed difference. PTWs tend to move between vehicles, known as filtering, filling the gaps but it seems unrealistic that this happens at free flow speeds as would be the case in ALE 2. Instead, it is possible that the DBL present at the ALE arterial gives extra space to PTWs and allows them to keep travelling at high speeds. Checking for the eleven buses travelling on the ALE arterial the minimum speed is 19 km/h with a mean of 25 km/h, thus confirming the DBL does not experience the same congestion as the other lanes (for those time steps that a bus crosses the detector) and making the effect of a DBL on the movement of PTWs more probable.

Section	ALE 2	OM 1
Time steps ($k > 0$)	57	85
Mean	30.0	20.4
Std	8.2	5.6
Median	29.5	19.9
Max	52.4	34.8

TABLE 4.3
Speed of PTWs at two selected locations (only time steps with non-zero density are kept), in km/h

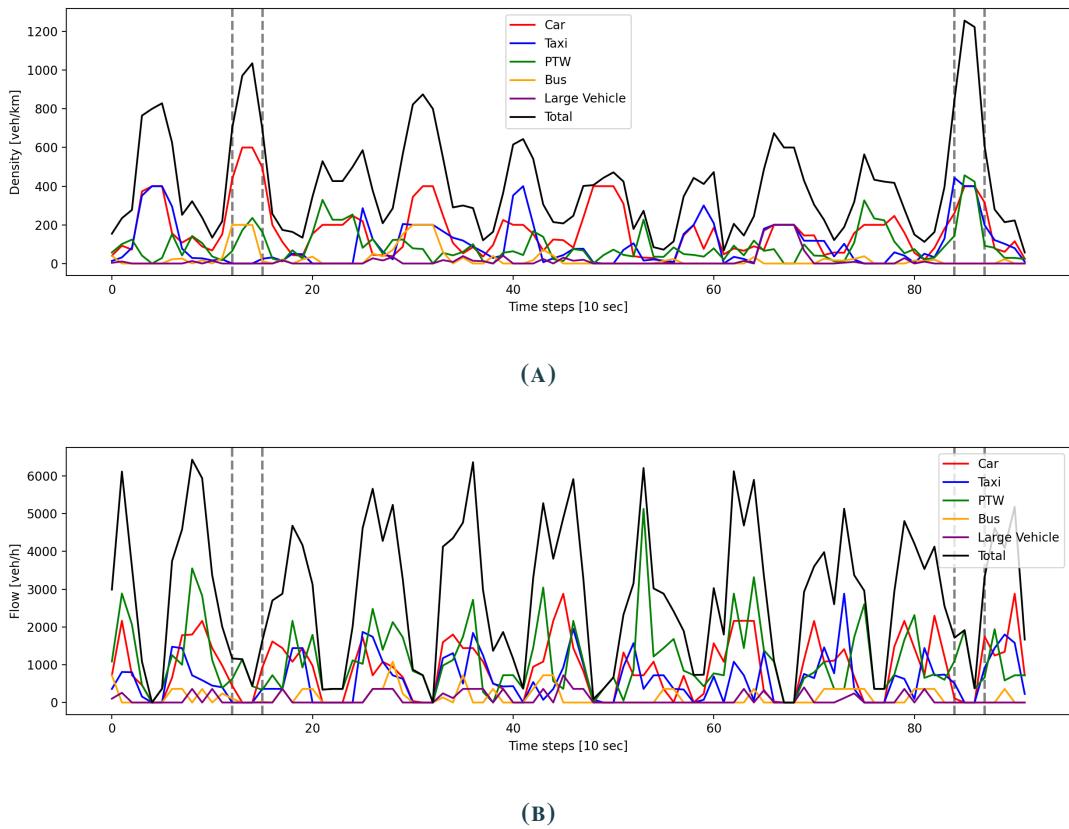


FIGURE 4.6
Density and flow over time for second detector PAN 14

4.2.2 MACROSCOPIC FUNDAMENTAL DIAGRAM

After discussing the individual FDs for an edge of an arterial, aggregating all measurements of every time step for all detectors leads to a well-defined MFD if the congestion is evenly distributed over the network, i.e. not heavily congested on some edges and uncongested on the others as is described in (Geroliminis and Sun, 2011). For the pNEUMA experiment only an uncongested branch with a lot of scatter was observed due to the presence of more uncongested residential traffic (Bampounakis and Geroliminis, 2020). In this section the network comprises all the primary, secondary and tertiary edges, afterwards some aggregated FDs are shown for the different arterials.

For the MFD, production and accumulation follow from equations 4.1 and 4.2 as defined in chapter 2. No distinction is made for individual lanes for the traffic characteristics calculation, flow and density are already aggregated over all lanes and multiplying with n_i is not needed. Dividing production and accumulation results in the average network speed and dividing them by the total lane-kilometers of the considered part of the network results in the average network flow and density, respectively (Mahmassani et al., 1984).

$$P = \sum l_i n_i q_i \quad (4.1)$$

$$N = \sum l_i n_i k_i \quad (4.2)$$

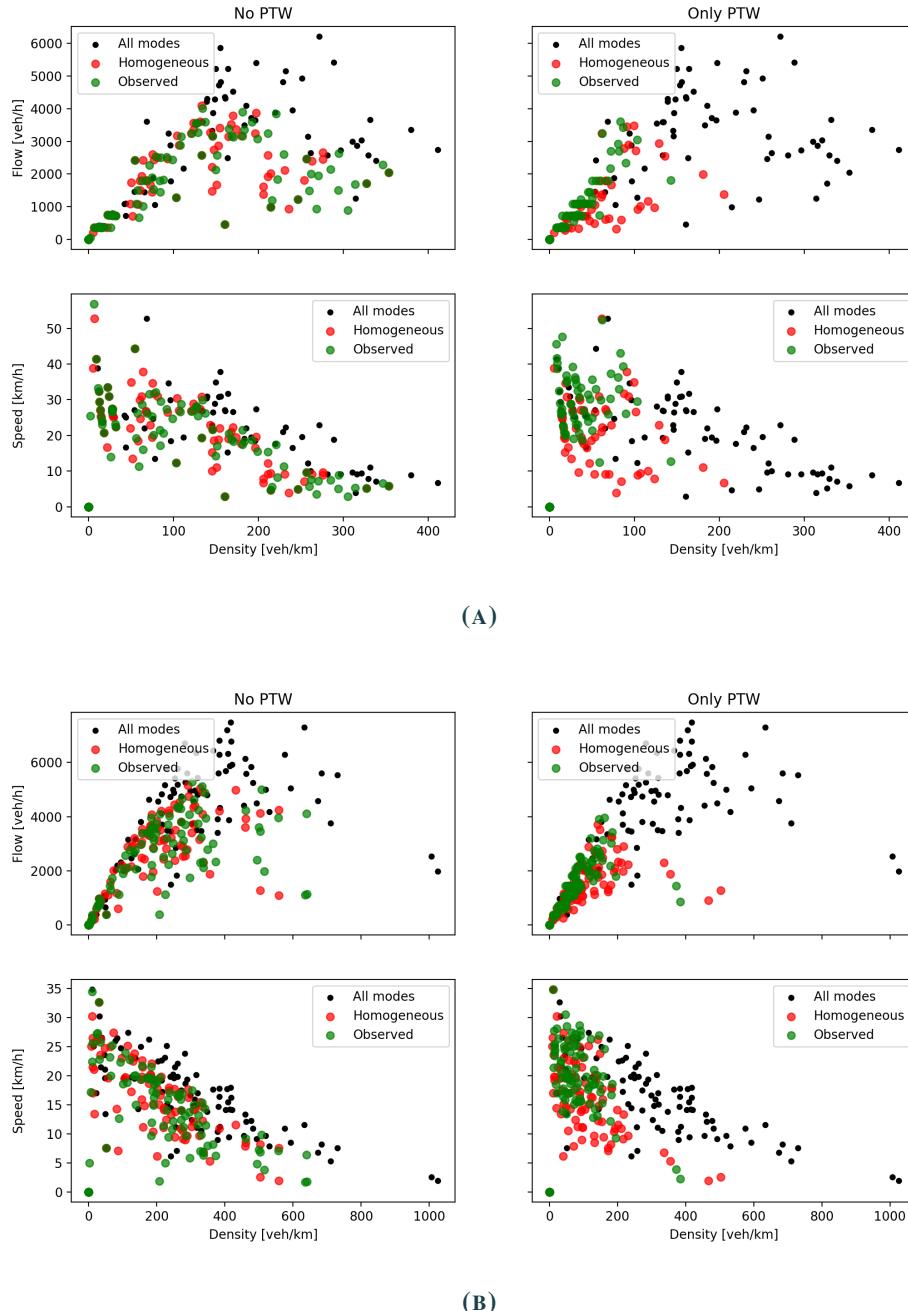


FIGURE 4.7
FDs with and without PTWs: A. ALE 2 detector 1 B. OM 1 detector 2

$$\bar{q} = \frac{\sum l_i n_i q_i}{\sum l_i n_i} \quad (4.3)$$

$$\bar{k} = \frac{\sum l_i n_i k_i}{\sum l_i n_i} \quad (4.4)$$

Plotting the MFD without further aggregating over a longer time period, results in a scattered cloud but nevertheless the bigger scale has a positive effect on the overall scatter when compared with the individual

FDs. In figure 4.8 the speed-accumulation MFD is shown with different adjustments that have an effect on the shape.

- Figure 4.8a, adjusting for stalling vehicles inside the detector area clearly shifts the MFD upwards. The magnitude of this effect shows that for all roads together there are numerous cases where only one vehicle is stalling.
- Figure 4.8b, aggregating over a longer time period reduces the scatter a lot. For a period of three minutes the points seem to follow a line, although there are not enough points to say something about the shape of the MFD.
- Figure 4.8c, as expected the combination of aggregation and adjusting for stalling vehicles shifts upwards. The effect of vehicles stalling persists, even after aggregation over longer time periods.
- Figure 4.8d, using only data from a specific detector on an edge has a very strong influence on the MFD. For the detectors placed at the beginning and in the middle of every edge the average network speed is significantly higher than for the end detector. This shows the effect of traffic signals on the MFD as well as the importance of placing multiple detectors on every edge, this confirms what was also found in Courbon and Leclercq (2011).

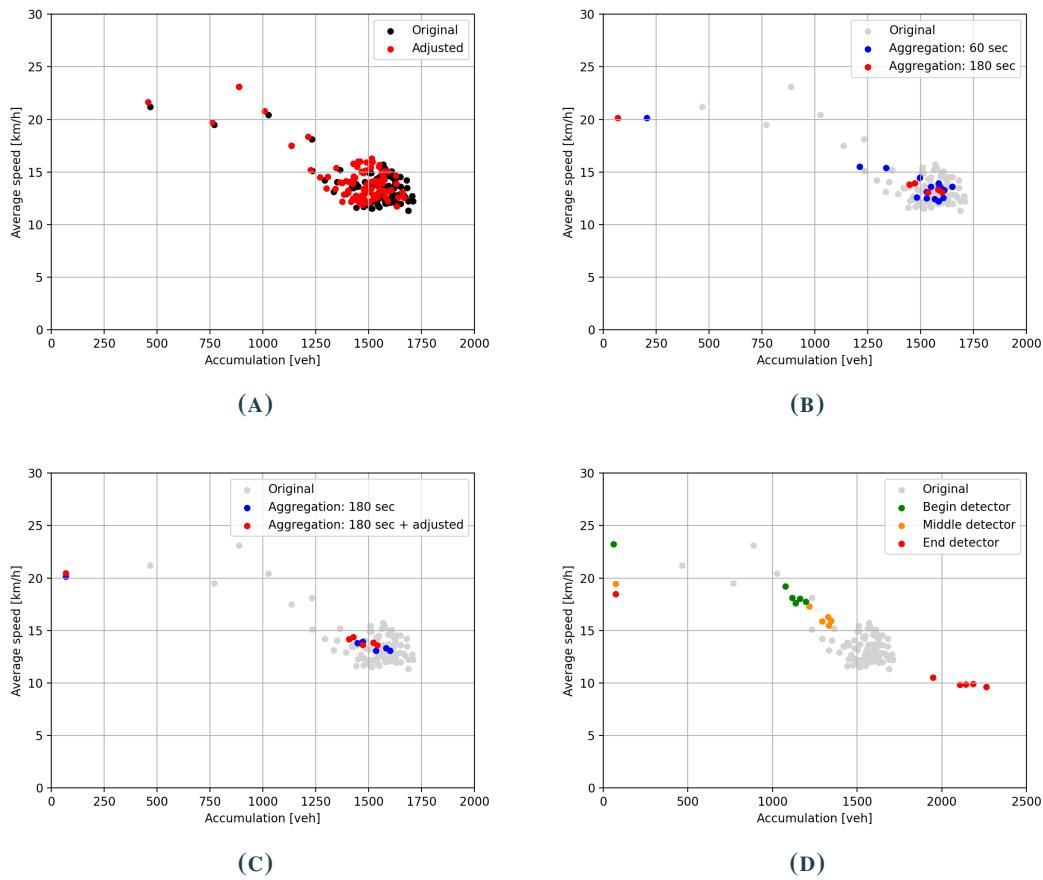


FIGURE 4.8

MFD for different adjustments and aggregation steps: A. filtering stalling vehicles, B. aggregating over longer time periods, C. combination, D. difference between detectors

The different road types give heterogeneity in topology, estimating the speed-density MFD, average density is used to adjust for the different network sizes, for every different type in the network shows

clear differences, see figure 4.9a. The primary roads have the highest average speed but it is only slightly higher than the secondary type, see figure 4.9b, for the tertiary roads the average speed lies significantly lower. The figure also shows that the scatter is highest for the primary type, high and low speeds exist for the same average density, the other two types follow a clearer trend in line with previous empirical MFD results (Geroliminis and Daganzo, 2008; Loder et al., 2017).

The differences are due to less heterogeneity for lower road types. The lane heterogeneity is more present for the primary arterials, e.g. slow and fast vehicles are present in different lanes during the same time step or vehicles preparing for a turning movement only block one lane, and affects the traffic characteristics. Also the vehicle mix can vary more on the wider primary roads, on the other hand the lower average speed of the tertiary roads can be due to short green cycles.

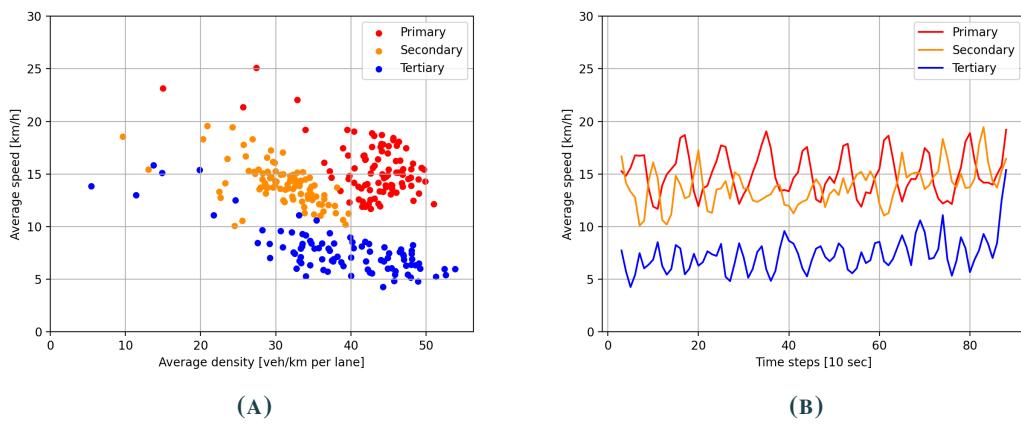


FIGURE 4.9
Comparison between road types: A. Speed-density MFD B. Average speed over time

In figure 4.10, the speed-accumulation diagram for every arterial is given with an aggregation step of 180 seconds. The average speeds are the lowest for the ALE and OM arterial, both have speeds approaching 10 km/h. For PAN and STA the points are clustered around 20 km/h and the rest of the arterials have values between 15 and 20 km/h. The expected relation of average speed in function of accumulation is somewhat visible for five of the eight arterials but shows only a small part of the under-saturated branch.

Splitting up by mode to find the 3D-MFD is not done, since a lot of scatter is present for the small time step and aggregating gives too few points to visualize any real trends. Also, from the uneven distribution of congestion it follows that clustering of those sections with higher congestion levels is needed to get less scatter and more parts of the MFD shape, which is not a trivial process (Saeedmanesh and Geroliminis, 2016). Instead figure 4.11 shows the evolution of average speed over time for different modes, namely cars, taxis, PTWs and buses. It seems that PTWs have a higher average speed for the selected arterials in the network. In Perco (2008) a comparison of the average speeds of cars and PTWs showed that the latter have a significantly higher speed in urban traffic, which is supported by the presented figure for this dataset. The smaller size and higher maneuverability offer PTWs the ability to travel at higher speeds, especially in the urban environment. Additionally, another factor having a positive influence on the speed of PTWs might be the DBL present on some arterials, see also section 4.2.1. In Athens, both buses and PTWs can use the DBL, therefore PTWs have more space to travel even when arterials are congested, although it should be noted that this is not visible in the figure and also from the measurements it is not directly visible since vehicles are not matched to a specific lane. The average speeds of buses crossing all detectors over time is lower on average than the other modes, also the high variation of the average speed of buses is very clear, this is due to the low number of buses present in the dataset, i.e. one bus more or less has a large effect on the speed of that time step, and possible service stops along the arterials.

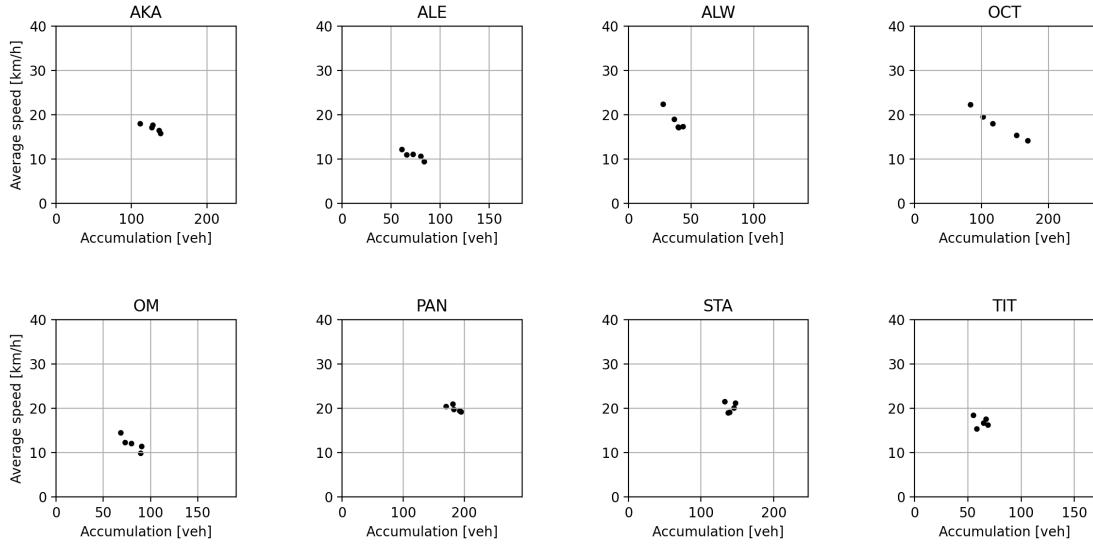


FIGURE 4.10
Speed-accumulation diagram for every arterial (aggregation step of 180 sec)

The effect of the traffic signals on the speed also follows from the figure, cars and PTWs follow the same pattern of low average speeds and immediately afterwards a peak of higher speeds. Taxis deflect somewhat from this pattern, since stalling vehicles for a longer time were filtered the reason of this deflection is not immediately clear.

Choosing a specific detector to aggregate all characteristics has a big influence on the MFD, the same effect is present for all modes. They all show the same behaviour, having high speed for the begin and middle detector and a low speed for the last one.

4.3 ANALYSIS OF ARTERIALS

The MFD only shows a very small part, making it not very useful to analyze the dataset in further detail. Comparing arterials also gives a lot of information about urban traffic. This section uses the FDs from the previous discussion to analyze consecutive edges, afterwards the travel time over arterials is discussed with special attention for the different vehicle types.

The arterials differ from each other in the number of lanes, presence of bus lanes, number of intersections and the vehicle mix. Looking at the shares of the different modes in appendix A, the arterials more to the north of the center, ALE, ALW, OCT begin and TIT end, tend to have a higher and lower share of cars and taxis, respectively. The share of PTWs is for all arterials above twenty-five percent but is very high for STA and the part of OM towards OM, for buses the share is for most arterials the same, between two and three percent.

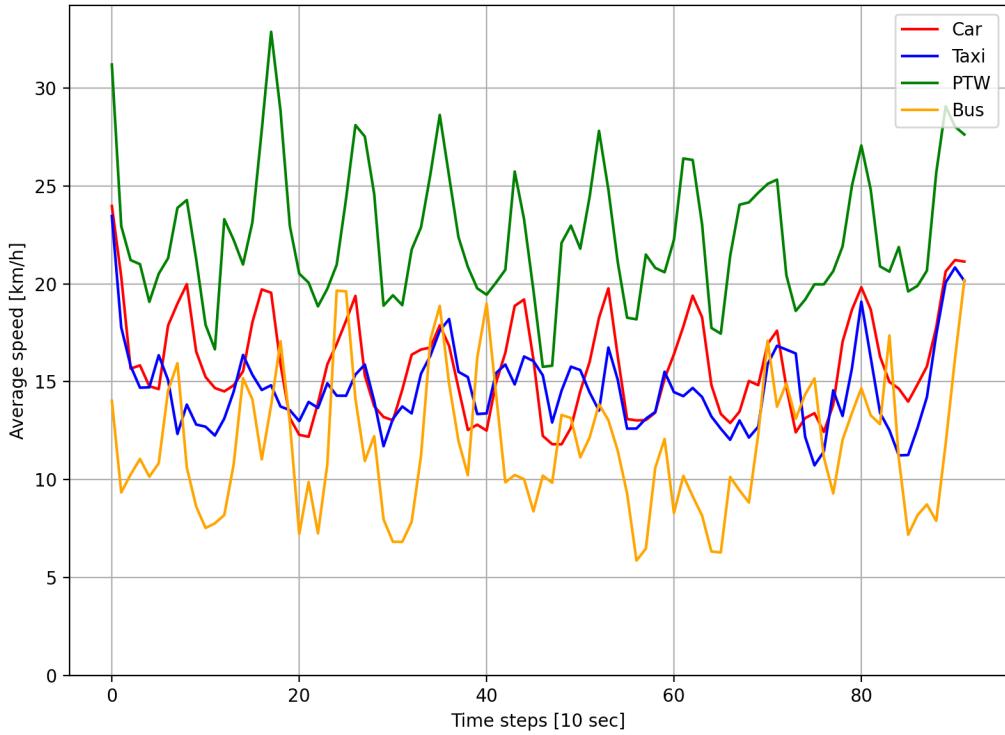


FIGURE 4.11
Average speed over time for different modes

4.3.1 ARTERIAL DIAGRAMS

It is impractical to go over all the FDs of a whole arterial, the high scatter makes it difficult to detect differences between detectors apart from the discussed cases in 4.2.1. Therefore the mean and maximum of every FD are determined, plotting these values over the length of the arterial makes analysis on the macroscopic scale and comparisons between arterials easier. The mean of the characteristics corresponds to measuring the traffic characteristics over the whole time period, the maximum value gives more insight about peaks at specific points.

The spatial evolution over the OCT arterial is visible in figures 4.12. The traffic density figure clearly shows the high peaks at every intersection due to traffic signals, traffic states vary a lot over one section of the arterial, resulting in fluctuating values for density. For the flow these fluctuations over a section disappear, therefore the average speed is lower more downstream on every edge, as is expected given the stop-and-go due to the signal cycles. In the flow figure turning movements are visible by the increase or decrease from one section to the other, a big drop is observed for traffic at sections five and six due to a street connecting OCT and TIT and afterwards some smaller fluctuations are present.

Throughout this chapter all FDs depicted showed measurements over the whole road width making no distinction between lanes. In the last figure, 4.12c, a large part of the arterial has very high flows per lane. Although it may be that such variations occur, in this case it is due to a wrong number of lanes, comparing the sections where the very high flows occur with the number of lanes over the arterials depicted in appendix A, shows that a DBL is present over most of this part. Checking for other arterials the same

behaviour is observed for those with a DBL, namely AKA, ALE and ALW. From an analysis point of view this observation is not very helpful but it does help to get the right number of lanes to make rightful comparisons between different edges. A comparison of the average flows for the whole time period over all arterials after adjusting for the DBL is shown in figure 4.13. The mean lane flow for a section varies over the arterial, low variation indicates that on average the traffic stream is not heavily disturbed by turning movements or the signal cycles keeping a constant traffic flow for consecutive sections.

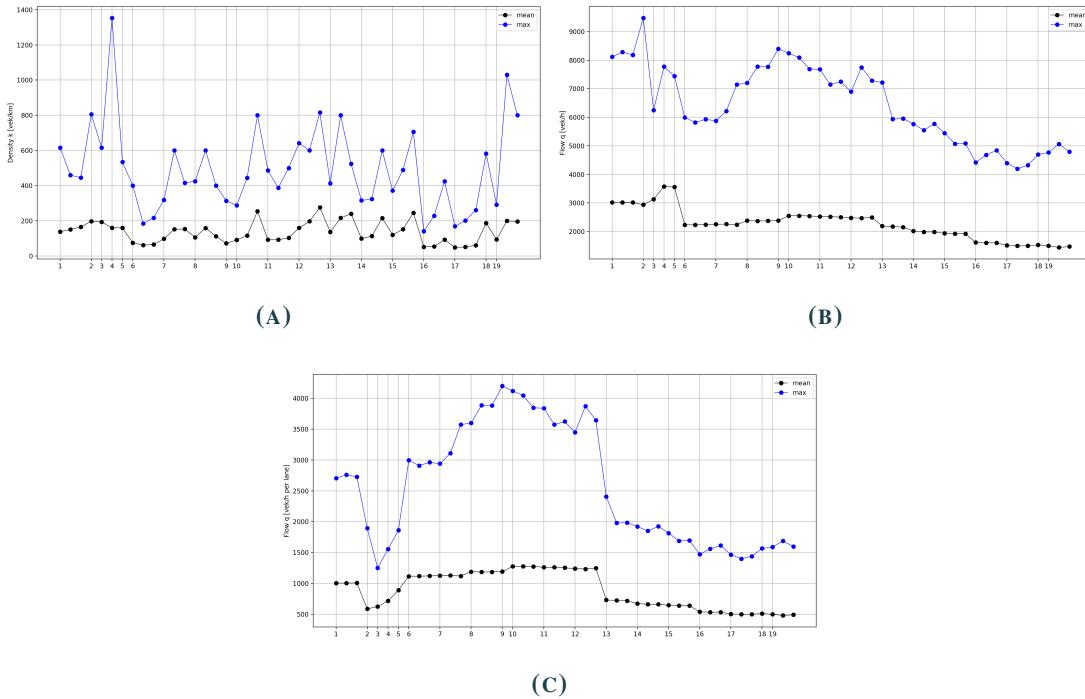


FIGURE 4.12

Evolution of characteristics over whole OCT arterial: A. total density, B. flow, C. flow per lane

The constructed diagrams can show in a simple way how the characteristics evolve over the arterial but the short time period of data available does not allow to detect temporal trends. Nevertheless these diagrams link the geographical location to the characteristics in a visual simple way.

4.3.2 TRAVEL TIME

In 4.2.1 it is clear that the mix of different vehicle types causes a lot of scatter in the FDs. Taxis and buses have service stops from time to time, inflating the density for that section when the other lanes are still in free flow, PTWs on the other hand give very high peaks in density due to filtering at traffic signals and can also stand shoulder to shoulder in the same lane. With the presented framework of the previous chapter, performing travel time analyses becomes easier because detectors are everywhere in the network. Selecting specific edges in the network provides the travel time of the trajectories crossing the corresponding detectors in a very fast way.

A very detailed travel time analysis falls out of the scope of this research, therefore the remainder of this section discusses only some empirical results related to travel time with a focus on modal differences. There are a lot of possible ways to analyze travel times over the arterials by using histograms, boxplots, the method proposed in (Ramezani and Geroliminis, 2012), etc. Also the length over which to measure the travel times can be varied to the liking since detectors are available along the whole arterial length.

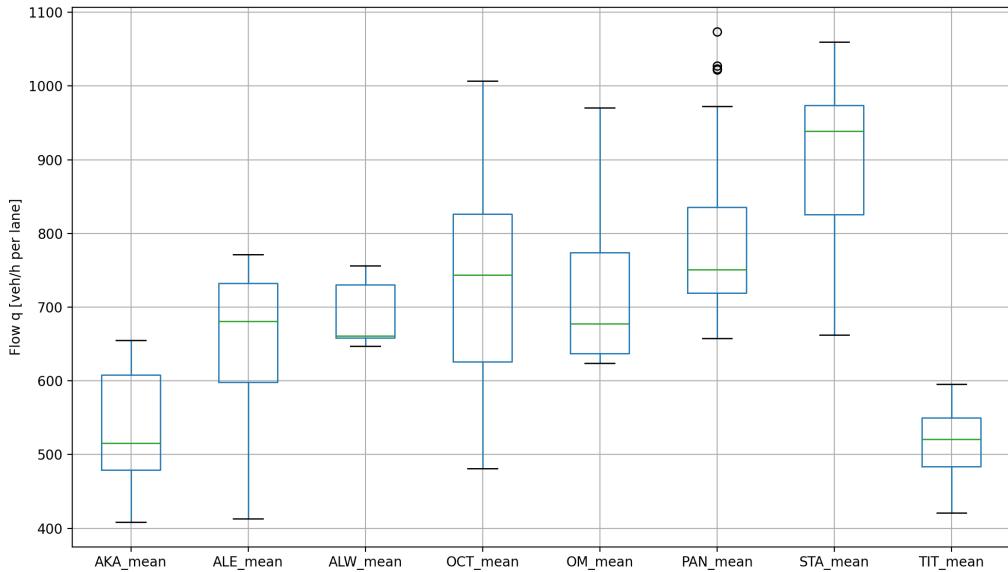


FIGURE 4.13
Boxplots for average lane flows for every arterial

For every arterial the total number of vehicles traveling the whole length, the distance from the intersection of the start and end detector is subtracted from the total length, are shown in table 4.4 and the histograms in figure 4.14. From the histograms it is clear that the distributions vary between the arterials. For those trajectories crossing a whole arterial PTWs always have the shortest travel times.

Arterial	Length (m)	# Trajectories
AKA	1187	65
ALE	344	352
ALW	297	409
OCT	938	156
OM	333	381
PAN	995	342
STA	1010	138
TIT	769	104

TABLE 4.4
Number of trajectories traveling whole arterial

Comparable to the diagrams from the previous section, figure 4.15 depicts the mean unit travel time on every section of an arterial split up by vehicle type. PTWs have the shortest travel time for almost all sections of every arterial, some exceptions for short sections, confirming the finding of the macroscopic average speed in section 4.2.2. Another interesting observation are the highest peaks of buses, ignoring those peaks that come from traffic signals and short edges when the other modes also show a peak or sudden increase. These show bus stops in the network, examples are AKA 2, 7 and 13, PAN 6, STA 5 and TIT 3.

4.4 SUMMARY

After matching all vehicles to their corresponding edge the analysis of the fifteen minutes of data is performed by selecting some arterials and a network with the most important roads. The detectors placed on these arterials produce individual FDs, a lot of scatter is immediately visible due to different reasons. The influence of stalling vehicles and the vehicle mix provided some answers to this high scatter, but a clear trapezoidal shape is not always observed. For the macroscopic analysis on the network level all detector measurements are aggregated to produce the MFD, however the dataset only contains fifteen minutes of data resulting in only a small part of the curve. Aggregation over three minute time steps gives not enough data points to draw any conclusions, nevertheless the average network speed seems to lie between 10 and 15 km/h for the network comprising the most important road types. Splitting up by mode shows that PTWs travel faster in the network and influence the density measurement because of their smaller size and higher maneuverability. The higher speed of PTWs is confirmed when briefly researching travel time for the different arterials.

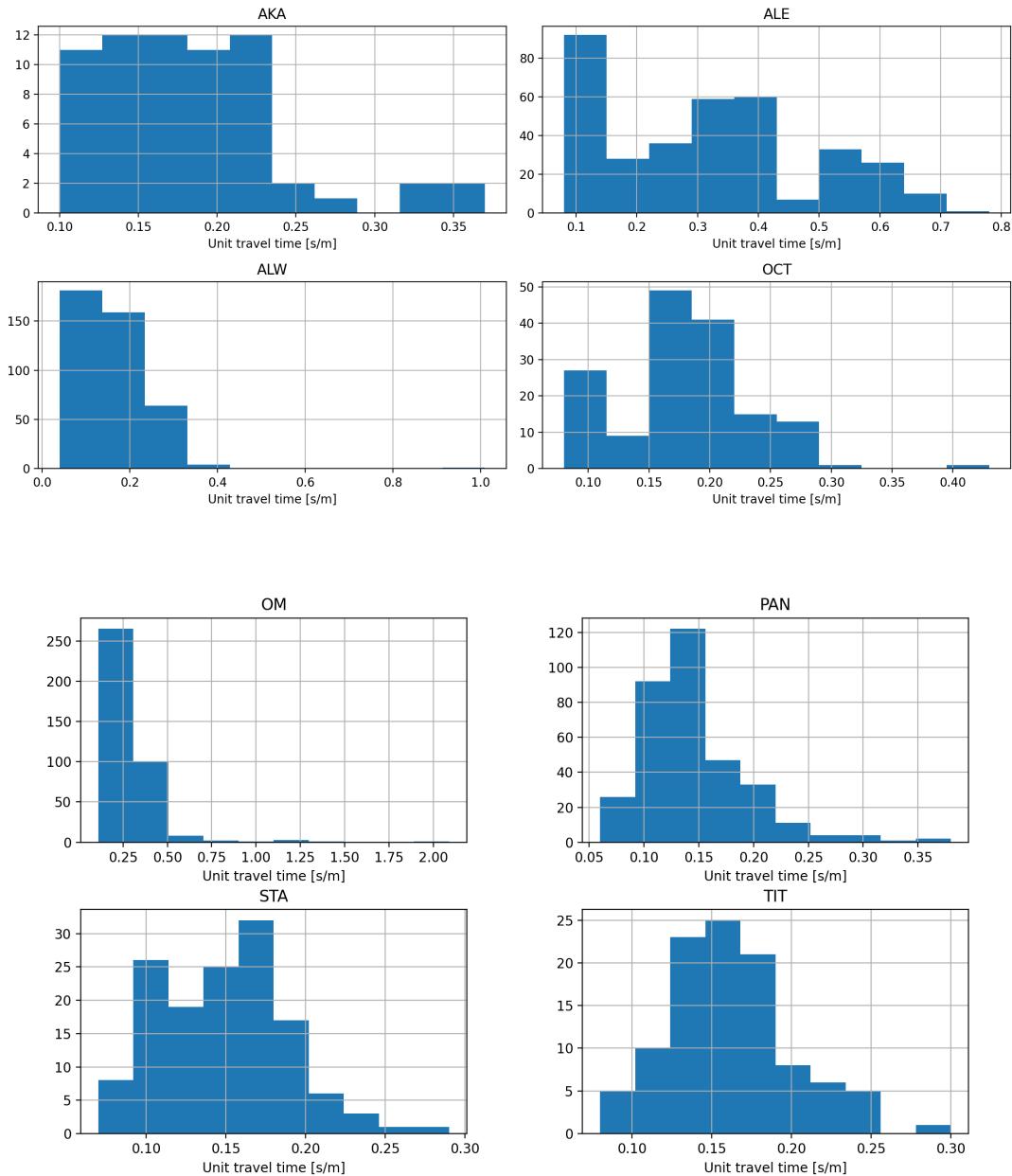


FIGURE 4.14
Histograms of unit travel time for every arterial

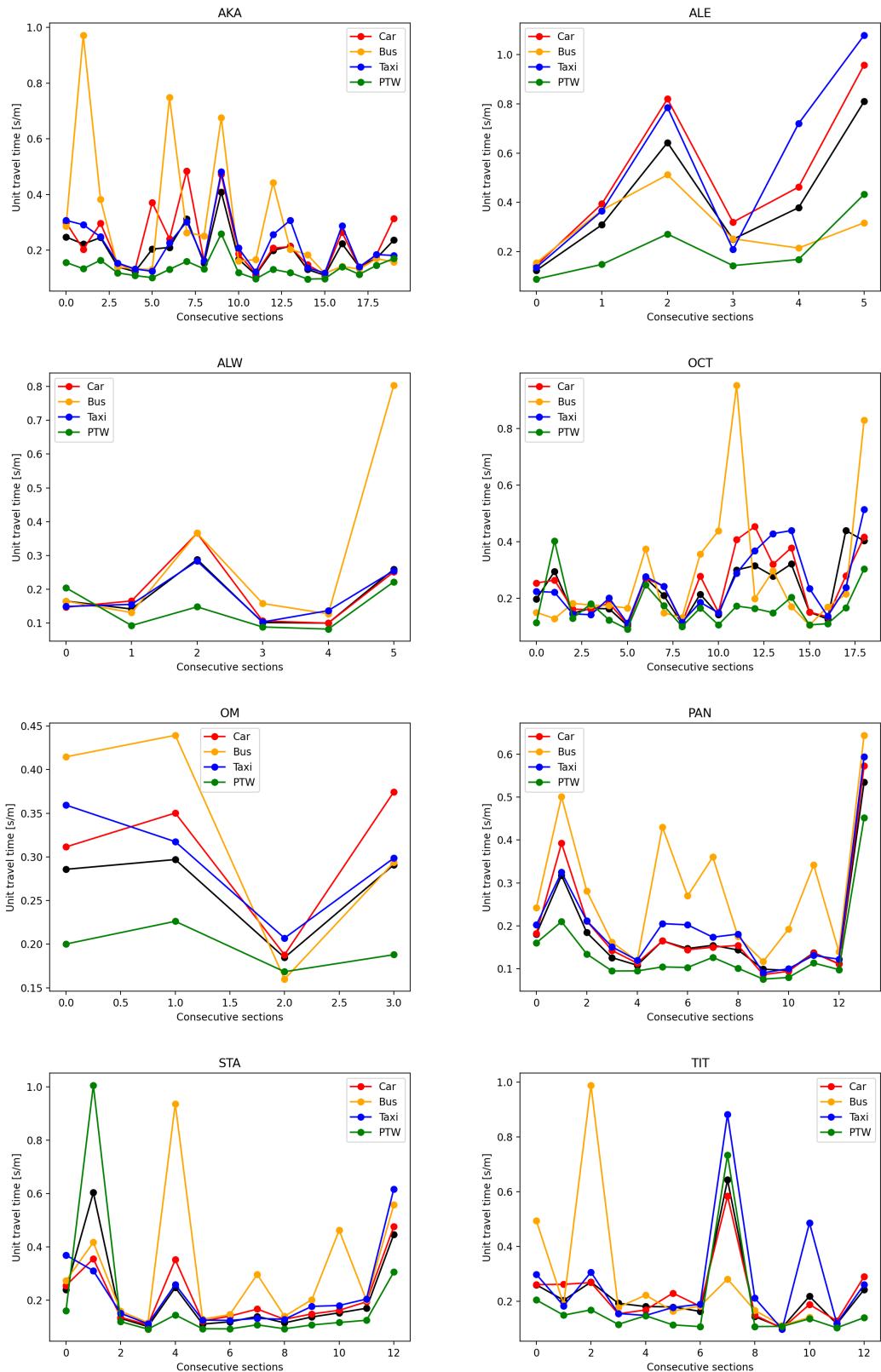


FIGURE 4.15
Evolution of average unit travel time for consecutive sections along an arterial

CHAPTER 5

CONCLUSION

This thesis uses the dataset collected by the pNEUMA experiment to estimate empirical FDs at any location in the network to get to an MFD for the whole network. In order to attain this empirical MFD, the trajectories are matched to the underlying network by using the open mapping database OSM. A python package, called OSMnx, helps to extract all the needed edges and nodes to build a sufficiently detailed network with the needed attributes. Having this network, the trajectories are matched by using a map-matching algorithm based on HMM. This results in a most-probable path along the extracted network for every trajectory. Although the trajectories have a very high sample rate, quantifying the accuracy of the map-matching algorithm was not trivial and different variables can point to possible mismatches between the trajectory and the network. Adjusting the network to solve these issues increased the overall accuracy of the map-matching further. Having a good-enough matching result, implementing virtual loops on every used edge gives measurements of the macroscopic traffic characteristics with Edie's generalized definitions.

Selecting main arterials for the network to analyze the urban traffic shows FDs with a lot of scatter. Different types of heterogeneity give rise to high variation, namely signalized intersections, individual lane behaviour, modal differences and topology differences. Adjusting for some heterogeneities improves the individual FDs and makes the trapezoidal shape for signalized roads clearer, although no congested branch is present for a large majority of the observed FDs. Selecting all the important road types for the MFD estimation results in only a small and scattered part of the curve. The speed-accumulation diagram gives average speeds between 10 and 15 km/h and the effect of the placement of virtual loops is apparent. Higher average speeds are observed for the primary road type compared to the other types but it seems that more heterogeneity is present due to the higher number of lanes and vehicle mix. The scatter still present prevents the observation of a useful 3D-MFD to check bi-modal network effects. Instead of this, looking to different modes separately shows the large differences between PTWs and normal cars, for all arterials PTWs have a higher speed than any other mode. Their higher maneuverability and smaller size prove very helpful in dense urban traffic with a lot of intersections and stop-and-go traffic due to traffic signals.

In essence this research makes it easier to analyze large trajectory datasets and helps to better analyze multi-modal urban traffic. Apart from having macroscopic traffic measurements for any location in the network, the virtual loops are also very helpful to analyze travel times over specific stretches of roads for the different vehicle types. However, the computation time of the tool to extract all needed information evolved a lot during the development process, influencing the scope of the research and the achieved results.

5.0.1 FUTURE WORK

This research only takes a first step to analyze large trajectory datasets in a structured way. For now the map-matching only matches vehicles to a road but not an individual lane, therefore incorporating a lane assignment algorithm to every vehicle makes more accurate measurements possible, especially for the microscopic scale. In a next step, adding traffic signals to the network structure will improve the placement of virtual loops to get more homogeneous measurements, additionally an analysis of the effect of the cycle times on the macroscopic characteristics is possible.

One of the main conclusions of this thesis is the very different behaviour of PTWs in urban traffic, a more detailed analysis of the effect of different modes on the measured traffic characteristics can help to better analyze modal differences on the network level. Incorporating mode specific attributes to the network structure, e.g. bus stops, delivery points, etc., also seems very helpful to better understand the modal differences, especially with respect to their travel times.

Appendices

APPENDIX A

ARTERIALS

The figures in this appendix show all the arterials in more detail. Every edge of an arterial gets a unique label used in the traffic analysis to refer to the right edge. Labels with a red background are short edges where the three detectors are all placed at the same location as described in 3.2.3. In figure A.1 the number of lanes extracted from OSM of every arterial is shown with DBL sections marked in red.

Arterial	Label	# Edges	Short edges	DBL
Akadimias	AKA	20	5,6,12,19,20	3-10
Alexandras EW	ALE	6	4,5,6	1-2
Alexandras WE	ALW	6	1,5,6	2-6
October 28	OCT	19	2,3,4,5,9,18	10-15
Omonoia	OM	4	None	None
Panepistimiou	PAN	14	1,2,10	None
Stadiou	STA	13	2,3	None
Titris	TIT	13	8,9,13	None

TABLE A.1
Arterials with labels

The vehicle mix for all the selected arterials and evolution over the consecutive edges is shown below, share is determined from the total number of matches for the whole period for a specific edge of an arterial. The number on the x-axis corresponds to the number of the arterial labels.

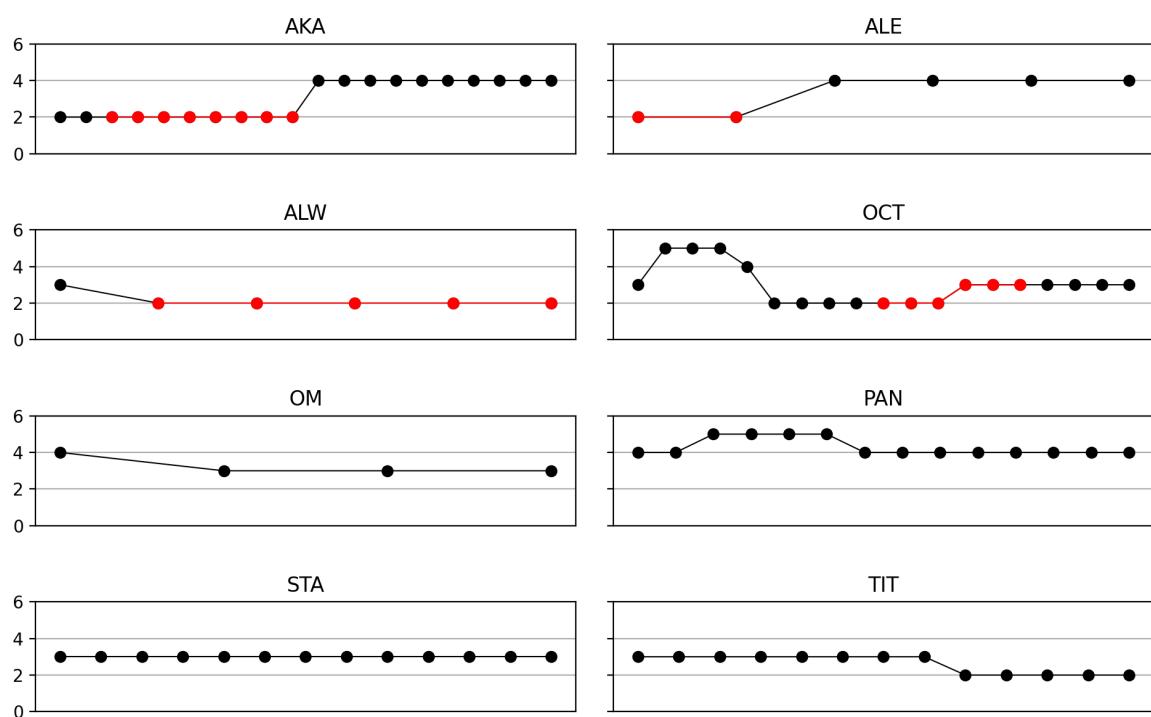


FIGURE A.1
Number of lanes for every arterial with DBL sections marked in red

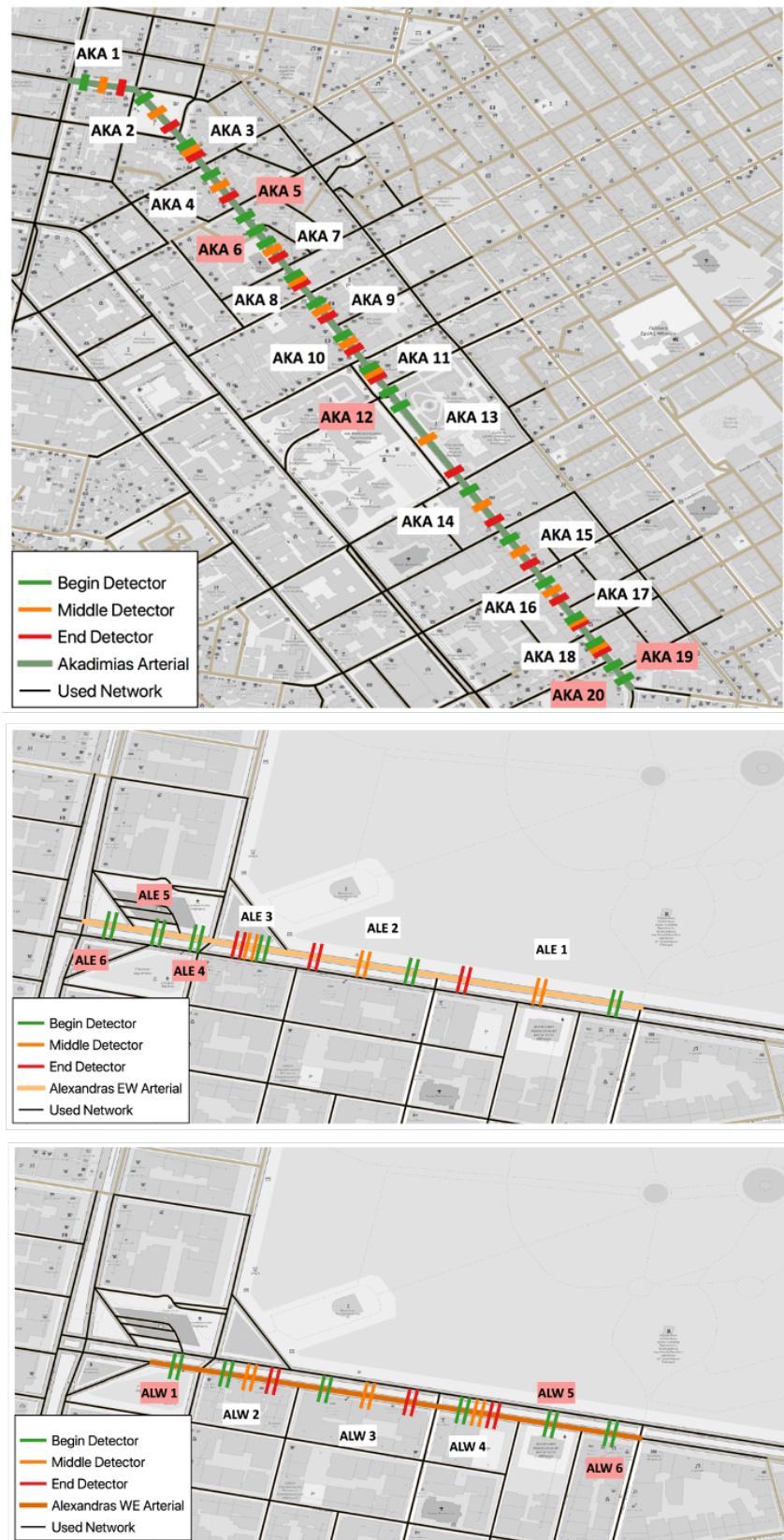


FIGURE A.2

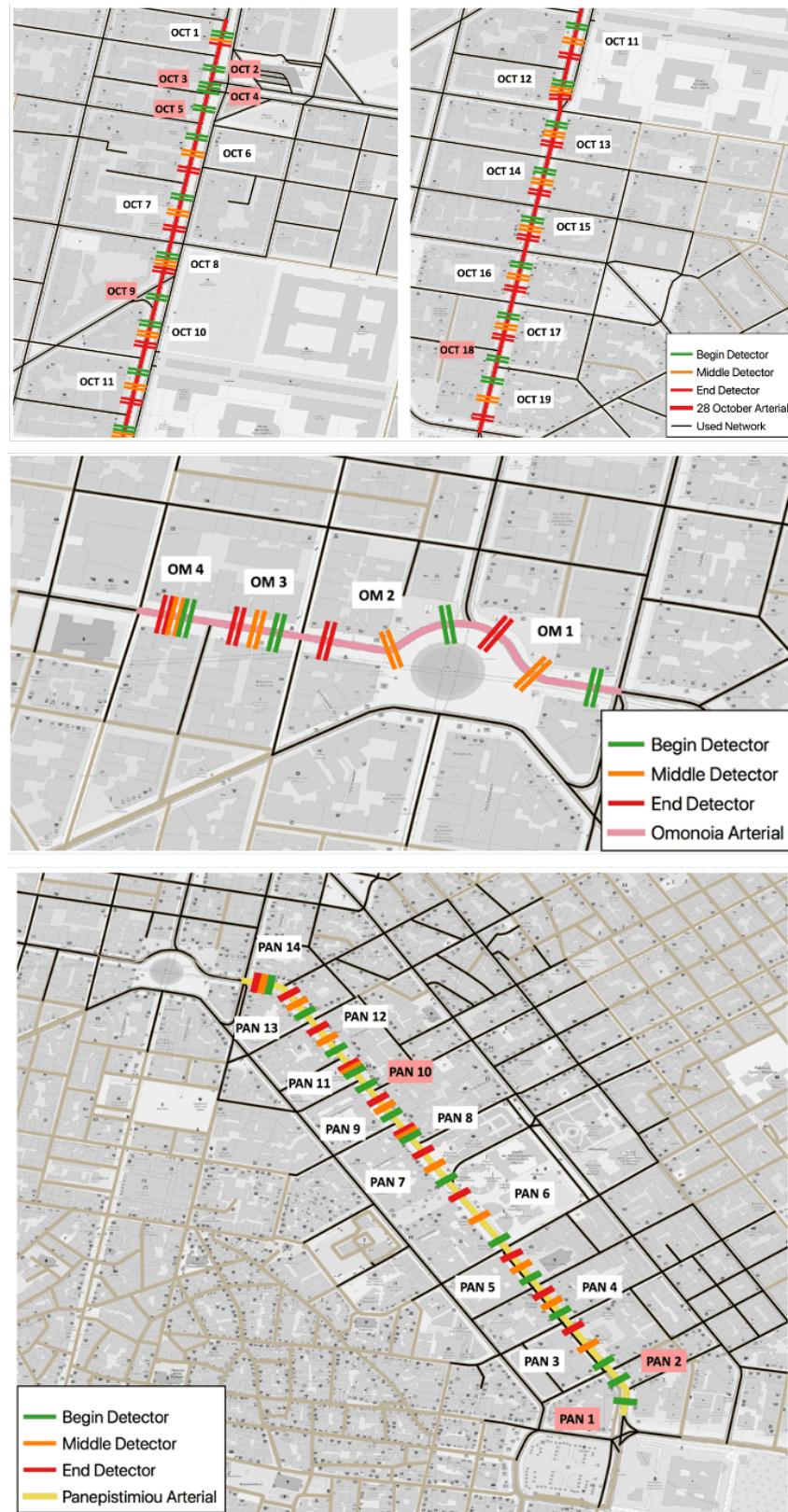


FIGURE A.3

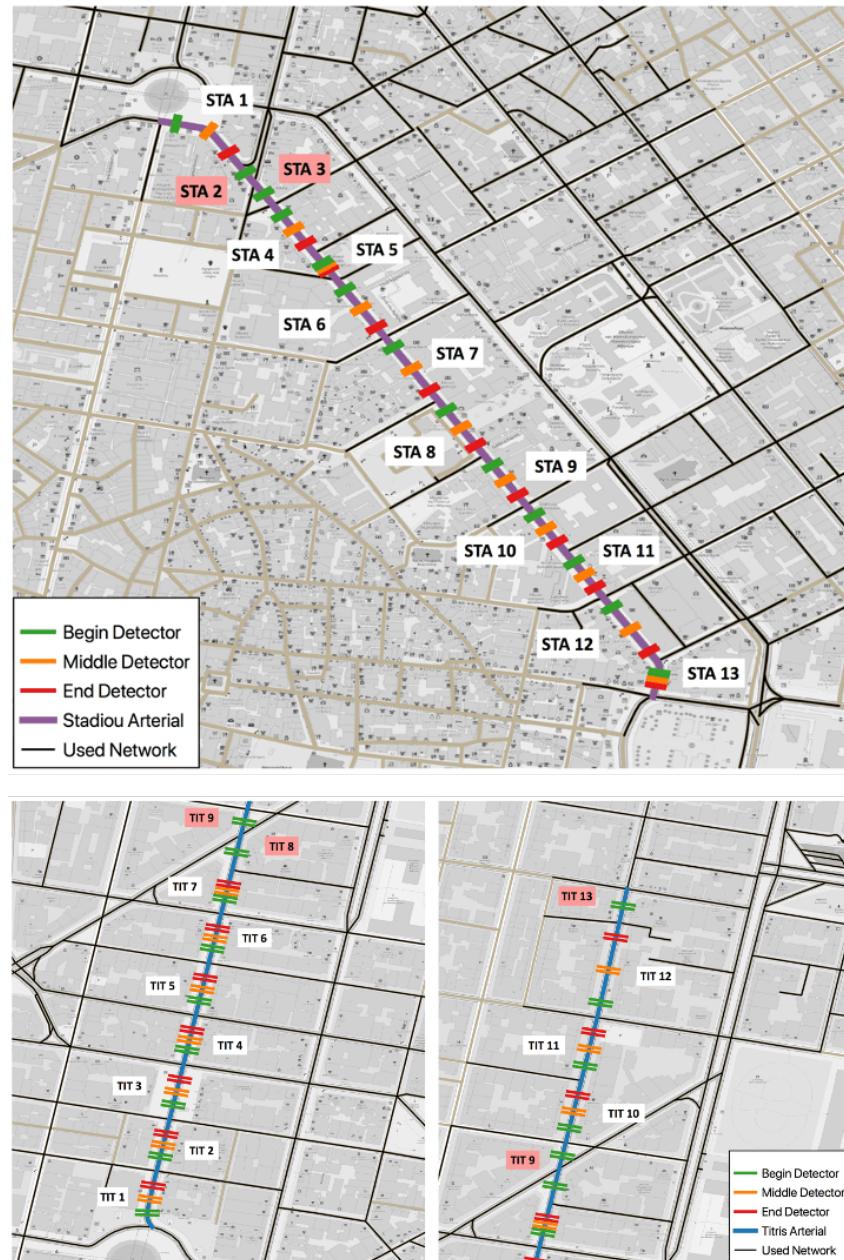


FIGURE A.4

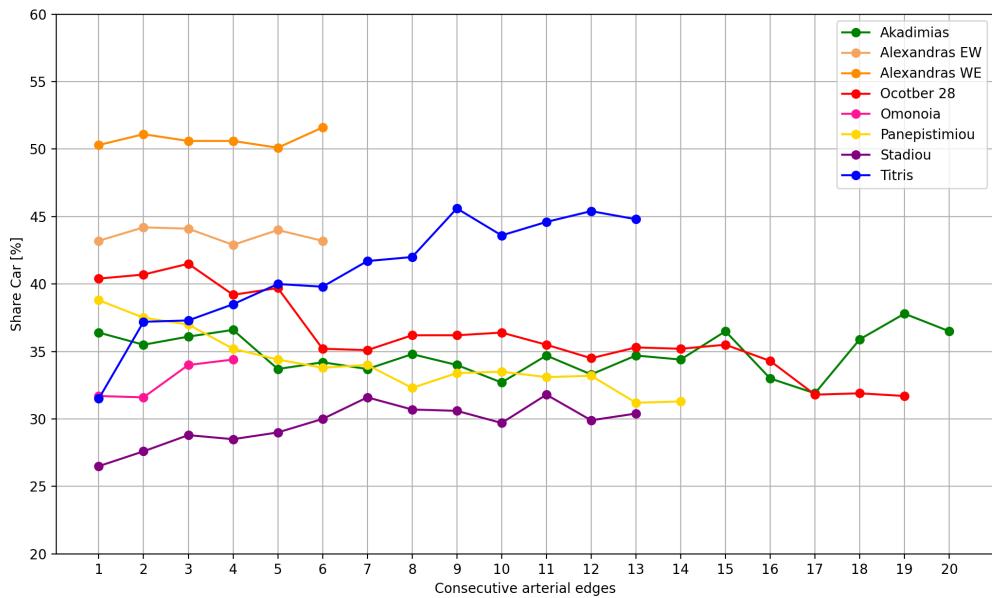


FIGURE A.5
Car share

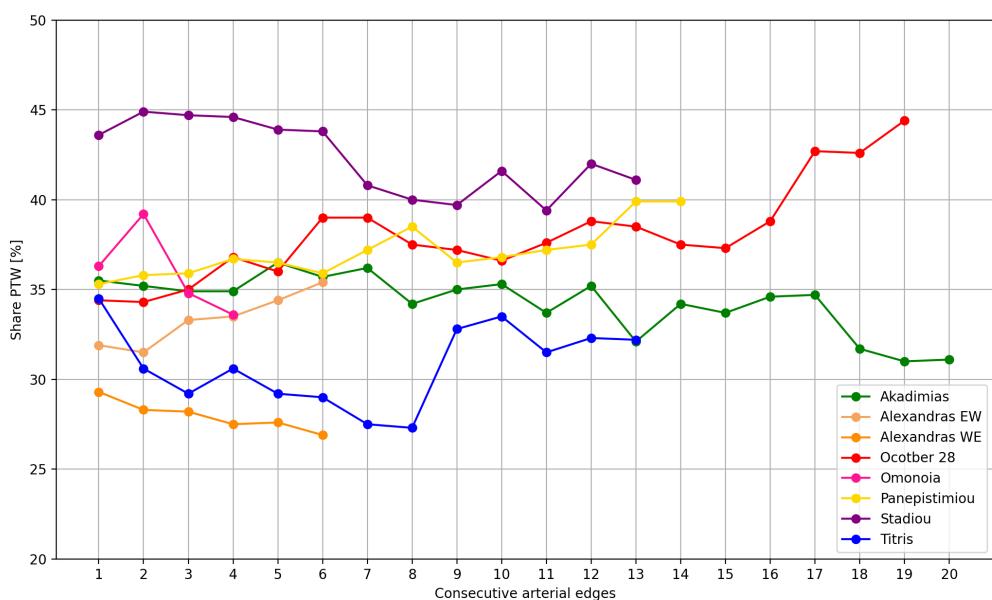


FIGURE A.6
PTW share

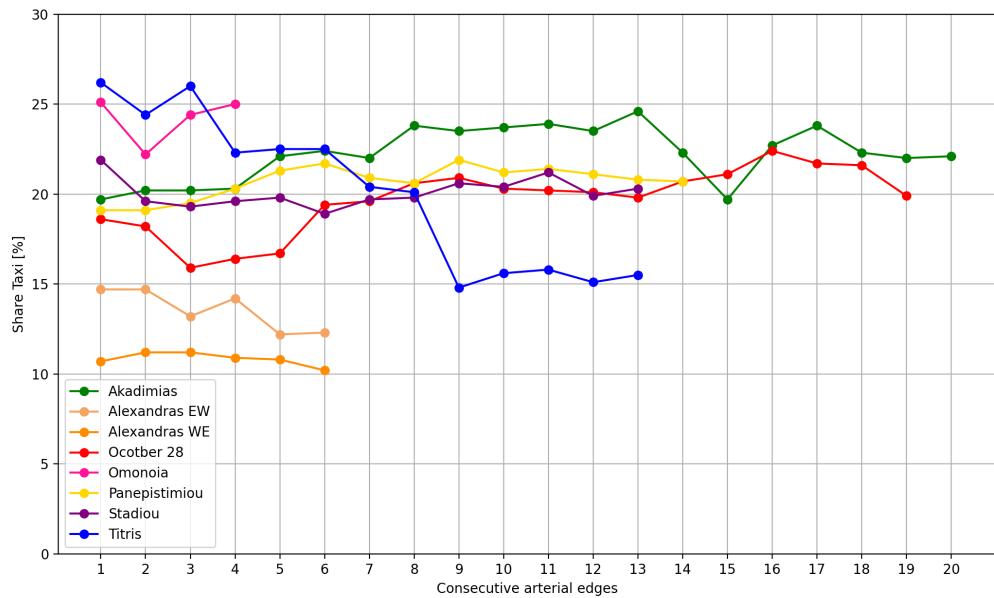


FIGURE A.7
Taxi share

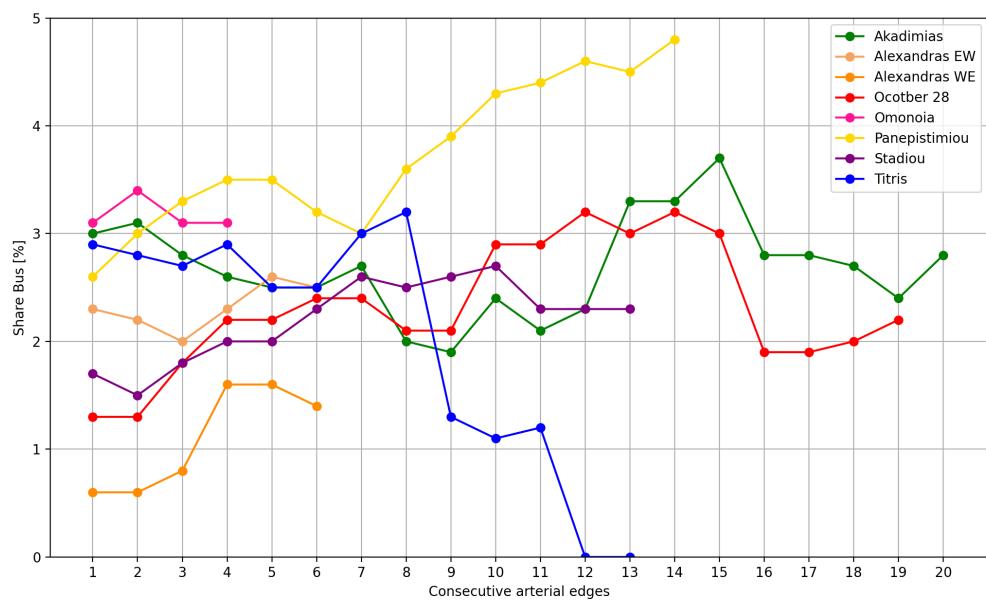


FIGURE A.8
Bus share

BIBLIOGRAPHY

- Barmpounakis, E. and Geroliminis, N. (2020). On the new era of urban traffic monitoring with massive drone data: The pneuma large-scale field experiment. *Transportation Research Part c*, 111:50–71.
- Barmpounakis, E. N., Vlahogianni, E. I., and Golias, J. C. (2016). Unmanned aerial aircraft systems for transportation engineering: Current practice and future challenges. *International Journal of Transportation Science and Technology*, 5(3):111–122.
- Barmpounakis, E. N., Vlahogianni, E. I., Golias, J. C., and Babinec, A. (2019). How accurate are small drones for measuring microscopic traffic parameters? *Transportation letters*, 11(6):332–340.
- Boeing, G. (2017). OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems*, 65:126–139.
- Brakatsoulas, S., Pfoser, D., Salas, R., and Wenk, C. (2005). On map-matching vehicle tracking data. In *Proceedings of the 31st international conference on Very large data bases*, pages 853–864.
- Buisson, C. and Ladier, C. (2009). Exploring the impact of homogeneity of traffic measurements on the existence of macroscopic fundamental diagrams. *Transportation Research Record*, 2124(1):127–136.
- Coifman, B. (2015). Empirical flow-density and speed-spacing relationships: Evidence of vehicle length dependency. *Transportation Research Part B: Methodological*, 78:54–65.
- Courbon, T. and Leclercq, L. (2011). Cross-comparison of macroscopic fundamental diagram estimation methods. *Procedia-Social and Behavioral Sciences*, 20:417–426.
- Daganzo, C. (1994). The cell transmission model: a dynamic representation of highway traffic consistent with the hydrodynamic theory. *transpn. res.-b*. 28b (4), 269–287.
- Daganzo, C. F. (2002). A behavioral theory of multi-lane traffic flow. part i: Long homogeneous freeway sections. *Transportation Research Part B: Methodological*, 36(2):131–158.
- Daganzo, C. F. (2007). Urban gridlock: Macroscopic modeling and mitigation approaches. *Transportation Research Part B: Methodological*, 41(1):49–62.
- Daganzo, C. F. and Geroliminis, N. (2008). An analytical approximation for the macroscopic fundamental diagram of urban traffic. *Transportation Research Part B: Methodological*, 42(9):771–781.
- Devos, L. (2018). Traffic patterns revealed through matrix functions and map matching. Master's thesis, Leuven: KU Leuven. Faculteit Ingenieurswetenschappen.
- Dhingra, S. and Gull, I. (2011). Traffic flow theory historical research perspectives. *The fundamental diagram for traffic flow theory*, page 45.
- Du, J., Rakha, H., and Gayah, V. V. (2016). Deriving macroscopic fundamental diagrams from probe data: Issues and proposed solutions. *Transportation Research Part C: Emerging Technologies*, 66:136–149.

- Edie, L. C. (1963). *Discussion of traffic stream measurements and definitions*. Port of New York Authority.
- Eiter, T. and Mannila, H. (1994). Computing discrete fréchet distance. Technical report, Citeseer.
- Forney, G. D. (1973). The viterbi algorithm. *Proceedings of the IEEE*, 61(3):268–278.
- Gan, Q. and Skabardonis, A. (2019). Arterial traffic estimation using field detector and signal phasing data.
- Geroliminis, N. and Daganzo, C. F. (2008). Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transportation Research Part B: Methodological*, 42(9):759–770.
- Geroliminis, N., Daganzo, C. F., et al. (2007). Macroscopic modeling of traffic in cities. In *Transportation Research Board 86th Annual Meeting*, number 07-0413. No. 07-0413.
- Geroliminis, N. and Sun, J. (2011). Properties of a well-defined macroscopic fundamental diagram for urban traffic. *Transportation Research Part B: Methodological*, 45(3):605–617.
- Geroliminis, N., Zheng, N., and Ampountolas, K. (2014). A three-dimensional macroscopic fundamental diagram for mixed bi-modal urban networks. *Transportation Research Part C: Emerging Technologies*, 42:168–181.
- Greenshields, B., Channing, W., Miller, H., et al. (1935). A study of traffic capacity. In *Highway research board proceedings*, volume 1935. National Research Council (USA), Highway Research Board.
- Ji, Y. and Geroliminis, N. (2012). On the spatial partitioning of urban transportation networks. *Transportation Research Part B: Methodological*, 46(10):1639–1656.
- Kessels, F. (2019). *Traffic Flow Modelling: Introduction to Traffic Flow Theory Through a Genealogy of Models*. Springer.
- Kim, Y. and Hall, F. L. (2004). Relationships between occupancy and density reflecting average vehicle lengths. *Transportation research record*, 1883(1):85–93.
- Knoop, V. (2017). Introduction to traffic flow theory: An introduction with exercises.
- Knoop, V. L. and Daamen, W. (2017). Automatic fitting procedure for the fundamental diagram. *Transportmetrica B: Transport Dynamics*, 5(2):129–144.
- Kong, X., Li, M., Ma, K., Tian, K., Wang, M., Ning, Z., and Xia, F. (2018). Big trajectory data: A survey of applications and services. *IEEE Access*, 6:58295–58306.
- Krajewski, R., Bock, J., Kloeker, L., and Eckstein, L. (2018). The highd dataset: A drone dataset of naturalistic vehicle trajectories on german highways for validation of highly automated driving systems. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pages 2118–2125. IEEE.
- Leclercq, L., Chiabaut, N., and Trinquier, B. (2014). Macroscopic fundamental diagrams: A cross-comparison of estimation methods. *Transportation Research Part B: Methodological*, 62:1–12.
- Li, J. and Zhang, H. M. (2011). Fundamental diagram of traffic flow: new identification scheme and further evidence from empirical data. *Transportation research record*, 2260(1):50–59.
- Lighthill, M. J. and Whitham, G. B. (1955). On kinematic waves ii. a theory of traffic flow on long crowded roads. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 229(1178):317–345.

- Loder, A., Ambühl, L., Menendez, M., and Axhausen, K. W. (2017). Empirics of multi-modal traffic networks—using the 3d macroscopic fundamental diagram. *Transportation Research Part C: Emerging Technologies*, 82:88–101.
- Mahmassani, H. S., Williams, J. C., and Herman, R. (1984). Investigation of network-level traffic flow relationships: some simulation results. *Transportation Research Record*, 971:121–130.
- Meert, W. and Verbeke, M. (2018). Hmm with non-emitting states for map matching. In *European Conference on Data Analysis (ECDA)*, Date: 2018/07/04-2018/07/06, Location: Paderborn, Germany.
- Newson, P. and Krumm, J. (2009). Hidden markov map matching through noise and sparseness. In *Proceedings of the 17th ACM SIGSPATIAL international conference on advances in geographic information systems*, pages 336–343.
- Perco, P. (2008). Comparison between powered two-wheeler and passenger car free-flow speeds in urban areas. *Transportation Research Record*, 2074(1):77–84.
- Punzo, V., Borzacchiello, M. T., and Ciuffo, B. (2011). On the assessment of vehicle trajectory data accuracy and application to the next generation simulation (ngsim) program data. *Transportation Research Part C: Emerging Technologies*, 19(6):1243–1262.
- Qiu, T. Z., Lu, X.-Y., Chow, A. H., and Shladover, S. E. (2010). Estimation of freeway traffic density with loop detector and probe vehicle data. *Transportation Research Record*, 2178(1):21–29.
- Rabiner, L. R. (1989). A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286.
- Ramezani, M. and Geroliminis, N. (2012). On the estimation of arterial route travel time distribution with markov chains. *Transportation Research Part B: Methodological*, 46(10):1576–1590.
- Richards, P. I. (1956). Shock waves on the highway. *Operations research*, 4(1):42–51.
- Saeedmanesh, M. and Geroliminis, N. (2016). Clustering of heterogeneous networks with directional flows based on “snake” similarities. *Transportation Research Part B: Methodological*, 91:250–269.
- Saeedmanesh, M. and Geroliminis, N. (2017). Dynamic clustering and propagation of congestion in heterogeneously congested urban traffic networks. *Transportation research procedia*, 23:962–979.
- Salvo, G., Caruso, L., Scordo, A., Guido, G., and Vitale, A. (2017). Traffic data acquirement by unmanned aerial vehicle. *European journal of remote sensing*, 50(1):343–351.
- Schnetzler, B. and Louis, X. (2013). Anisotropic second-order models and associated fundamental diagrams. *Transportation Research Part C: Emerging Technologies*, 27:131–139.
- Seo, T., Kawasaki, Y., Kusakabe, T., and Asakura, Y. (2019). Fundamental diagram estimation by using trajectories of probe vehicles. *Transportation Research Part B: Methodological*, 122:40–56.
- Stofan, D. (2018). The development of traffic data collection methods. Available at <https://medium.com/goodvision/the-development-of-traffic-data-collection-cd87cc65aaab> (2020/01/18).
- Treiber, M. and Kesting, A. (2013). *Traffic Flow Dynamics: Data, Models and Simulation*. Berlin, Heidelberg: Springer Berlin Heidelberg : Imprint: Springer, 1st ed. 2013. edition.
- Tsubota, T., Bhaskar, A., and Chung, E. (2014). Macroscopic fundamental diagram for brisbane, australia: empirical findings on network partitioning and incident detection. *Transportation Research Record*, 2421(1):12–21.

- Wu, X., Liu, H. X., and Geroliminis, N. (2011). An empirical analysis on the arterial fundamental diagram. *Transportation Research Part B: Methodological*, 45(1):255–266.
- Yang, Q., Hu, J., and Peng, Q. (2018). Solving the missing data problem in urban traffic estimation with principal component analysis. In *Proceedings of the 2018 2nd International Conference on Big Data and Internet of Things*, pages 23–28. ACM.
- Zockaie, A., Saberi, M., and Saedi, R. (2018). A resource allocation problem to estimate network fundamental diagram in heterogeneous networks: Optimal locating of fixed measurement points and sampling of probe trajectories. *Transportation Research Part C: Emerging Technologies*, 86:245–262.