

Calibration of Multi-region MFD Models using Mobile Phone Data

Mahendra Paipuri^{a,*}, Yanyan Xu^b, Marta C. González^b, Ludovic Leclercq^a

^aUniv. Gustave Eiffel, Univ. Lyon, ENTPE, LICIT, F-69518, Lyon, France.

^bDepartment of City and Regional Planning, University of California, Berkeley, CA, 94720, USA.

Abstract

The present work proposes a framework to calibrate the MFD models using mobile phone data. A trip enrichment scheme based on map matching process is proposed for the trips that have sparser records. Time dependent penetration rates are estimated by fusing the OD matrix and the Loop Detector Data (LDD). Two different types of penetration rates are proposed based on the OD flow and the trips starting within an origin, respectively. The estimated MFDs based on two types of penetration rates are stable with very low scatter. Following, macro-paths and their corresponding trip lengths are estimated. Dynamic evolution of trip lengths is demonstrated using the present data, which is otherwise very difficult to capture with other types of data sources. Another important component that is calibrated is the time dependent path flow distribution between the different macro-paths for a given OD pair. The manuscript is concluded by presenting the time evolution of User Equilibrium (UE) gap for different macroscopic OD pairs.

Keywords:

2010 MSC: 00-01, 99-00

1. Introduction

Macroscopic Fundamental Diagram (MFD) relates average density to average flow in the urban network (Mahmassani *et al.*, 1984; Daganzo, 2007). It evolved as a promising tool for urban traffic management. The empirical existence of MFD was first reported by Geroliminis and Daganzo (2008) for the city of Yokohama, Japan under certain homogeneity assumptions. Since then, several applications like traffic state estimation (Knoop and Hoogendoorn, 2014; Yildirimoglu and Geroliminis, 2014; Kaviani-pour *et al.*, 2019), perimeter control (Keyvan-Ekbatani *et al.*, 2012; Haddad and Mirkin, 2017; Ampountolas *et al.*, 2017; Mohajerpoor *et al.*, 2019), congestion pricing (Gu *et al.*, 2018) and cruising for parking (Cao and Menendez, 2015; Leclercq *et al.*, 2017), *etc.* were proposed based on MFD approach. The three main elements for any MFD-based simulation are the underlying MFD, the trip lengths for each macro-path and the path flow distribution.

*Corresponding author

Email addresses: mahendra.paipuri@ifsttar.fr (Mahendra Paipuri), yanyanxu@berkeley.edu (Yanyan Xu), martag@berkeley.edu (Marta C. González), ludovic.leclercq@ifsttar.fr (Ludovic Leclercq)

Most of the applications founded on MFD assume a well-defined MFD relation for the urban network under consideration. However, in reality, the estimation of MFD for urban networks is far from trivial. There are two main types of data sources namely, Loop Detector Data (LDD) and Floating Car Data (FCD) for the estimation of empirical MFD. Most of the empirical MFDs proposed in the literature are based on either one of LDD, FCD or combination of both data sources (Ambühl *et al.*, 2017). Wang *et al.* (2015); Ampountolas and Kouvelas (2015) used the LDD to estimate the empirical MFDs for the urban networks of Sendai, China and Chania, Greece, respectively. The main limitation of the data derived from LDDs is the placement and distribution of loop-detectors. Buisson and Ladier (2009) demonstrated that the slope of MFD depends on the distance of loop-detectors to the downstream traffic signals. Even though a methodological framework was proposed by Leclercq *et al.* (2014) to compute the average density on the link based on the placement of loop-detectors, network coverage still remains a strong limitation to estimate an accurate MFD (Courbon and Leclercq, 2011; Ambühl and Menendez, 2016). More recently, Shim *et al.* (2019) studied the bifurcations in empirical MFDs by estimating them using roadside detectors and Alonso *et al.* (2019) analyzed the shape of empirical MFDs for the urban corridors using LDD. On the other hand, FCD is more attractive than its counterpart, as it provides the vehicular trajectory data. Typically, FCD is provided by either Global Positioning System (GPS) or mobile phones. FCD was used to estimate empirical MFD (Geroliminis and Daganzo, 2008; Bazzani *et al.*, 2011; Tsubota *et al.*, 2014), traffic monitoring (Herrera *et al.*, 2010) or travel time estimation (Jie *et al.*, 2011). Beibei *et al.* (2016); Ambühl and Menendez (2016) estimated an empirical MFD by data fusion of both LCD and FCD. Shoufeng *et al.* (2013) used the combination of GPS data and visually counted traffic to estimate MFD for the Central Business District (CBD) of Changsha city in China. Most of the works that employ FCD use only the GPS data from taxis to estimate MFD, as the GPS data of private cars is not readily available. Hence, the penetration rate of taxis is an important factor in accurate estimation of MFD from FCD (Du *et al.*, 2016). Knoop *et al.* (2018) used large scale FCD to estimate the empirical MFD of Amsterdam city by assuming a constant penetration rate. Recently, Huang *et al.* (2019) used the GPS data from taxis, private cars and public buses to estimate a 3D-MFD for the city of Shenzhen, China.

Another key ingredient of the MFD-based modeling framework is the set of macro-paths and their corresponding trip lengths. It is not trivial to estimate neither the macro-paths nor their trip lengths using LDD without any further equipment. On the other hand, FCD from taxis can be processed to obtain a distribution of trip lengths. However, FCD is generally sparse and it fails to capture the repetitive trips made by the frequent users/residents. Most of the works in the literature proposed in the context of MFD-based framework assumed a simple constant trip length inside the each reservoir. It was also concluded that using a single mean trip length inside each reservoir might result in significant error in the traffic dynamics (Yildirimoglu and Geroliminis, 2014; Kouvelas *et al.*, 2017). The importance of estimation of accurate and reliable trip

lengths in the context of MFD-based simulation was discussed in-detail in [Batista et al. \(2019\)](#) and it is as critical as estimation of the MFD *per se* for the accurate resolution of traffic dynamics. However, due to the lack of empirical data, the authors in the stated work built a virtual set of shortest path trips by randomly sampling the origins and destinations in the network. Similarly, it is difficult to observe the path flow distributions along each macro-path with the existing LDD or FCD data sources. [Yildirimoglu et al. \(2015\)](#) used Dynamic Traffic Assignment (DTA) can be used to estimate the path flow distribution in the context of route guidance. Most often path flow coefficients are estimated using DTA with User Equilibrium (UE), Stochastic User Equilibrium (SUE) or Bounded Rational User Equilibrium (BRUE) equilibration [Batista et al. \(2019\)](#). However, recent studies show that route choice discipline may not valid for macro-paths for large-scale networks ([Mariotte et al., 2020](#)). At the same time, there have not been any works proposed to estimate the path flow coefficient from the empirical data. These research gaps can be appropriately addressed by using the massive phone data to calibrate the above discussed models.

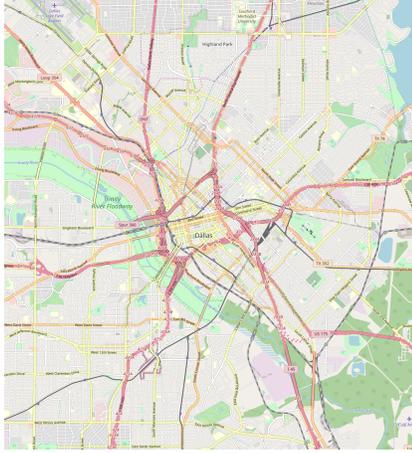
The present work proposes a method to estimate the empirical MFD of any urban network from the Location Based Service (LBS) data. This type of data is generated by the smart phone Apps, which share their location data actively to the App developer. The positioning of the user device is provided by either GPS or Wi-Fi. LBS data can be collected from the various applications on the smart phones, if the users allow the application to share the location data. Recently, LBS data has been used to propose frameworks on data driven metrics like Origin-Destination (OD) matrix estimation ([Jin et al., 2014](#)), travel route identification ([Hsieh et al., 2015](#)), *etc.* One obvious advantage of the LBS data over the FCD is that the LBS has wide coverage of the population in the urban network and therefore, high penetration rates are observed. Hence, LBS data has potential to accurately estimate the urban scale MFD. However, the major drawback from this data is the diversity of users providing the GPS locations. In other words, data contains not only the information about people traveling in private cars or taxis, but also pedestrians, bicyclists, people using public transport, *etc.* Hence, it is important to pre-process the data to extract the information of cars and taxis. Another important factor in the case of LBS data is the sampling rate, which can vary from few seconds to few minutes depending on the type of App being used. Even though travel time computation is not effected by the data with large sampling interval, inaccuracies in the traveled distance are inevitable in the urban network cases. This work proposes a method to enhance the data with poor sampling rates in order to obtain accurate travel distance within each interval. This work also introduces a framework to estimate the time dependent penetration rates. [Ji and Geroliminis \(2012\)](#) used GPS data of taxis with a constant penetration rate for the whole network. This resulted in a large uncertainty in the outcome of their study. The importance of computing accurate penetration rates for estimating MFD is discussed thoroughly in [Du and Rakha \(2019\)](#). This issue is addressed in this work by proposing a method to estimate penetration rates as a function of both time and OD pair. The secondary contribution of this work is the analysis of trip lengths inside the network. A static analysis is proposed, which yields

the major macro-paths (or regional paths) along with their trip length distributions for each OD
90 pair. Following, a dynamic analysis is also presented to demonstrate the dynamic effects in the
trip lengths during the peak hours. The final part of the manuscript discusses the estimation of
path flow distribution. The validity of UE hypothesis is discussed in-detail using the empirically
estimated path flow distribution. This work presents the techniques to segment the records into
individual trips, enhance the trips with sparse data, compute penetration rates, estimate MFD, trip
95 lengths and their corresponding path flow distributions. This is a generic framework and it can be
used to process any type of phone data that has records of GPS coordinates and timestamps.

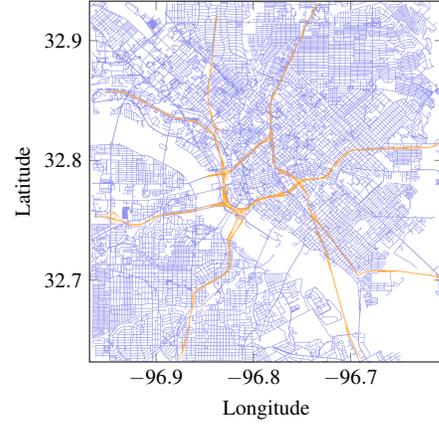
The paper is organized as follows: Section 2 presents the details about data and data process-
ing techniques, Section 3 discusses the trip enhancement method used to improve the trajectory
data for trips with sparse data points, Section 4 briefs about the computation of macroscopic
100 variables and penetration rates, Section 5 presents the estimated MFDs for the city of Dallas,
Section 6 presents the static and dynamic analysis of trip lengths and finally, Section 7 discusses
the estimation of path flow distributions from the data.

2. Data Description and Pre-processing

The data contains the positioning of the smart phone devices either by GPS or WiFi Position-
105 ing System (WPS) for the city of Dallas, Texas in United States for a period of 2 months from
March, 2017 to April, 2017. Dallas is one of the most populous cities in the US with an estimated
1.3 million inhabitants. In the present work, downtown Dallas and neighboring suburbs as shown
in Fig. 1a are considered to calibrate the MFD models. Fig. 1b shows the link level presentation
of the area under consideration, which extends across 160km² and contains 18386 nodes and
110 48287 links. The length of the road network is 4800km, which includes all types of roads . The
raw data contains anonymized user ID, time stamp, longitude, latitude and uncertainty of the lo-
cation. The data consists of 3.7 million users and there are around 4.5 billion records available for
processing. Fig. 2a shows the visitation map for the considered area for a period of 14 days. It
is clear from the heat map that the data is more concentrated in the downtown Dallas and major
115 arterials surrounding the downtown. As already stated, LBS data is only generated when the user
is interacting with the application on the smart phone. At the same time, different applications
use the location data at different frequencies. For instance, the applications that use map-related
services share their location more actively than the others. Therefore, each user in the LBS data
has different sampling intervals that range from few seconds to few minutes. A sampling interval
120 is defined as time difference between two consecutive records for the same user. Fig. 2b presents
the distribution of the sampling intervals of the raw data for the month of March, 2017. Even
though, major fraction of users have sampling interval around 100 sec, it can be noticed from
the distribution that there are peaks at 600 sec and around 1000 sec. A large sampling interval
is a limitation while computing macroscopic variables like total traveled distance, trip length,



(a) Map of the Dallas, TX ©OpenStreetMap 2019.



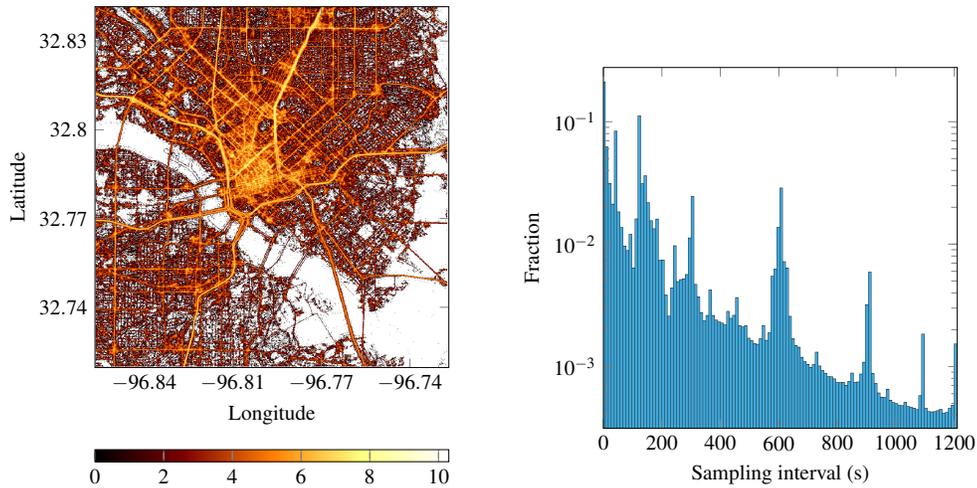
(b) Link level representation of the Dallas.

Figure 1: Dallas network: map of the area and its link level description.

125 *etc.* This issue is addressed in the following sections by introducing a map matching scheme by re-constructing the trajectory of the user for the records with large sampling intervals.

Data selection and segmentation of individual records into trips are discussed in-detail in [Xu et al. \(2019\)](#) and a similar approach is used in this work. Data processing is done in various stages to extract more useful and representative data for the estimation of different macroscopic traffic variables. Firstly, records that have an uncertainty of more than 100m are removed from the analysis. This LBS data contain different types of users like residents, tourists, people using motorways, *etc.* There are several users that leave either very few records or stay during very short period of time. In order to obtain a representative trips from regular users of the urban network, it is necessary to filter the users with few records from the frequent users. Hence, only users with more than 1 000 records that span across 30 days are considered in the present work. 135 The resulting fraction of active users and selected records are 25% and 93%, respectively.

The following step is to segment the records of each user into individual trips. This is done in two stages namely coarse segmentation and fine segmentation. As the names suggest, the first stage involves coarse segmentation of selected records into trips. The second stage deals with the refinement of already segmented trips either by splitting them into further trips or removing the records in the trips that do not comply with the user movement. In the coarse segmentation, the records are clustered into the trips based on the following assumptions. Firstly, the user starts a new trip if the time interval between two consecutive records is more than 30min apart. Since, the data can contain the information of pedestrians and bicyclists, only trips that have an average speed of more than 5 km h^{-1} over the course of whole trip are selected. Finally, only trips that have at least 5 records are selected in order to have more robust trip information. The second stage refines the already segmented trips in order to enhance the quality of the trip. In this stage, the mean speed between the consecutive records is monitored and the data points having a mean speed of less than 4 km h^{-1} are removed from a given trip. For example, if the data points in 145



(a) Heatmap of the data for randomly selected 14 days in March, 2017. The contour corresponds to log scale of number of data points observed. (b) Distribution of sampling rates for the raw data for the month of March, 2017.

Figure 2: Heat map and sampling rate distribution of the raw LBS data.

150 the middle of the trip are removed, the trip is further segmented into two individual trips. If the data points at the beginning or the end of the trip are removed, the method just truncates the trip removing the records, where user is idle. This process yields a total of 3.3 million number of trips. This method ensures the records are clustered into representative individual trips. However, few trips, especially from taxis or ride sharing vehicles, cannot be refined with the discussed strategy.

155 This is due to very short duration between two different trips. The occupancy of taxis tend to be high close to downtown and the idle time between the trips can be almost non-existent. This prove to be a bottleneck to segment the aggregation of several trips into individual trips. At the same time, having these type of aggregated trips have very little influence of computation of the macroscopic variables like distance and time traveled. On the other hand, they can introduce bias

160 in the estimation of the trip lengths. Hence, care is taken to remove these type of aggregated trips for the estimation of trip lengths. It is done by comparing the shortest path distance and actual trip distance for a given trip. If the actual trip length is more than twice the shortest path distance, that trip is removed for the trip length analysis. It is noticed in the present work that this technique removes a very small proportion of trips, improving the trip length analysis.

165 As already discussed, the sampling intervals of the raw LBS data vary across a wide range. Consequently, each segmented trip might have a different sampling interval, even within the trip. This can introduce considerable biases in the computation of travel distances. In the current work, the Haversine formula is used to compute the distance between given coordinates to take the curvature of the planet into account. When two GPS coordinates are far apart, even the

170 distance computed by the Haversine formula leads to inaccurate estimation of traveled distance, especially when the points lie on different links in the network. One trivial solution to minimize the biases in the traveled distances is to choose the trips that have records relatively close to each

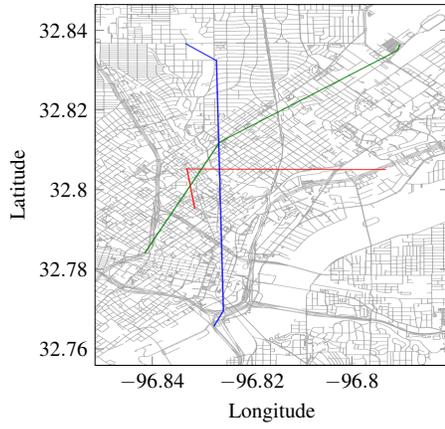
other. As a first step, trips where the consecutive records are within a radius of 500m to each other are selected. This filtration process reduced the total number of trips to 290 000, *i.e.*, 9% of the segmented raw trips. This averages to around 4 700 trips per day for a relatively large area under investigation. Hence, a trip enrichment method based on the map matching scheme is proposed in this work to process the trips with sparse records.

3. Trip Enrichment Method

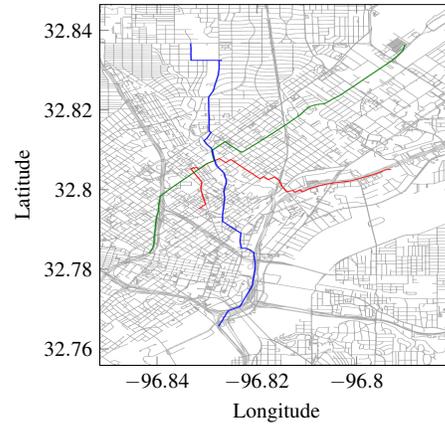
The main idea behind the enrichment method is to find the shortest path between the sparse records of a given trip. Two important tools are used in this context namely, OSMnx (Boeing, 2017), which is a Python package that is used to analyze the road networks and NetworkX (Hagberg *et al.*, 2008), another Python package used to study the dynamics of road networks. The enrichment method is explained with an example in the following.

Figure 3a shows three example traces of a sample trip with sparse records from the given data. All the three trips have either 3 or 4 records for relatively long trips lengths of around 10 km. The result is very poor resolution of the trajectory of the trip, which is shown in the Fig. 3a. It is also evident that the trips either start or end close to downtown region of Dallas and ends in the suburbs of the city. Hence, they are realistic trips across the city that spans considerable amount of time. A simple Haversine distance between successive data points of each trip introduce a huge approximation in the traveled distances. At the same time, the traveled time estimation is unaffected due to the sparse records. The combination of these two phenomenon can introduce a considerable scatter in the MFD and poor estimation of trip lengths. Hence, it is desirable to map the trajectory of the trip to the underlying road network as closely as possible.

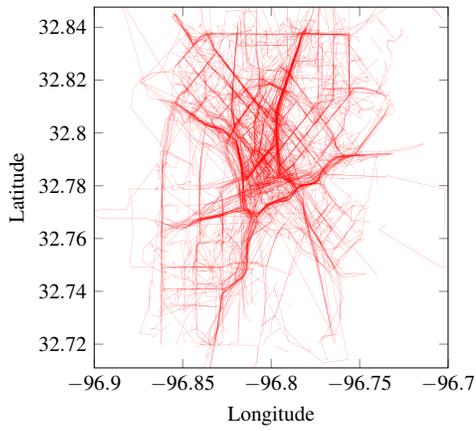
In the current work, the trips are enriched using the spatial geometry of the network. OSMnx contains the information of the whole network in the form of links and nodes. For the each trip, the distance between successive records is estimated. If the distance is bigger than a threshold, defined *a priori*, the location of nearest nodes close to the GPS positions of those records are obtained. The threshold distance to be defined depends on the size of the block in the road network. In the current case of Dallas city, the block size in the downtown is smaller than the suburban areas. It is possible to define different threshold distances based on the location of the trip in this framework. However, a constant threshold distance of 200m is chosen in the present work. Once the locations of the nodes are obtained between the sparse data records, a shortest path between those two nodes is computed using NetworkX shortest path subroutine. The shortest path is represented as the sequence of nodes at each intersection between the given two nodes. Consequently, the GPS locations of the nodes in the shortest path are added to the trip between those two sparse records. Fig. 3b presents the traces of considered trips after computing the shortest paths between the sparse records. It can be clearly noticed that the trajectory of each trip is matched to the network after the enrichment process.



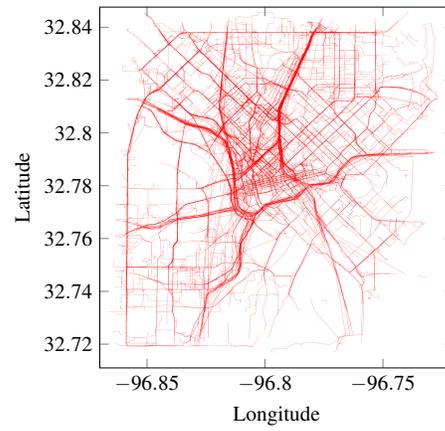
(a) Original traces of low resolution trips.



(b) Traces of the trips after enrichment method.



(c) Randomly sampled original traces.



(d) Randomly sampled traces after enrichment method.

Figure 3: Trip enhancement method: Original and enriched traces.

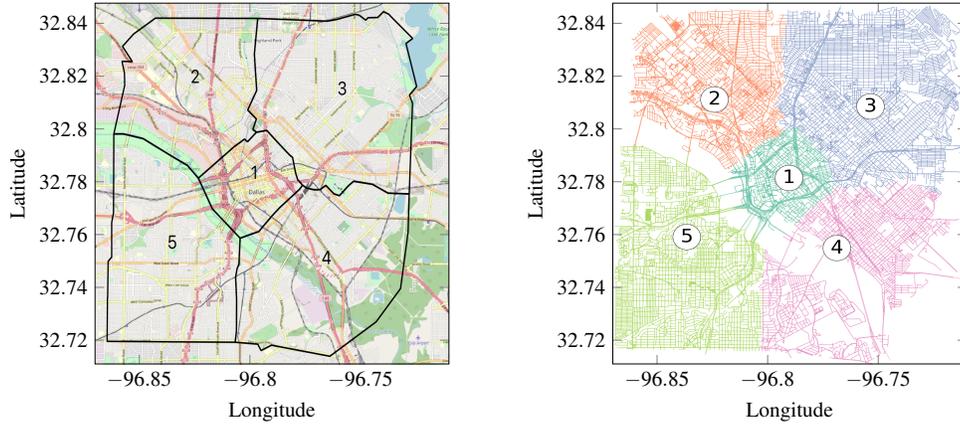
Another method to enrich the trips is to estimate the shortest path between the origin and the destination locations of each trip. However, this method fail to capture the longer paths that users tend to take during the peak hour congestion periods. However, the scheme proposed in the current work keeps the blueprint of the actual trip, while only adding the shortest path between the records that are sparsely placed. Hence, this can be considered as the closest approximation to the actual path that user had taken. The main advantage of this method is that it results in a trace that conforms the actual network, which can be observed from Fig. 3b. The main limitation is that the shortest path is computed between two *nodes* and not two *points*. Since there can be negligible distance between the actual GPS location and its closest node, this method introduce some errors in the traveled distances. Also, the curvature of links between two nodes, if exists, is neglected *i.e.*, the Haversine distance between two nodes is assumed to be the traveled distance. In spite of the limitations, it is shown in the subsequent sections that the errors of the present framework are within the acceptable tolerance.

Whenever the trace of the trip is enhanced between two sparse records, the time stamps are also interpolated to match the spatial data. The time stamps are interpolated based on the average speed between the two records in the original trace. Hence, this method transforms the sparse spatial and temporal data into dense data, thereby improving the overall accuracy and the representation of the MFD. Figs. 3c and 3d present a sample set of traces of trips before and after enrichment method, respectively. It is clear from the plots that original traces do have sparse records and the enhancement of trips result in the high resolution traces. It is to be noted that the trips in Figs. 3c and 3d are randomly sampled and do not correspond to the same set of trips.

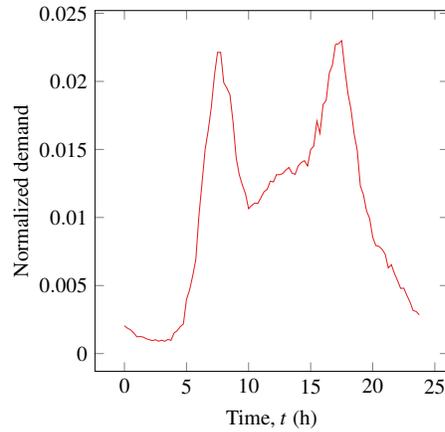
4. Computation of Macroscopic Variables

4.1. Partitioning of the Network

Partitioning of the considered network into homogeneous reservoirs is the first step to estimate the macroscopic traffic variables. A prerequisite to obtain a well-defined MFD is to partition the network into homogeneous subnetworks (Geroliminis and Sun, 2011). There have been partitioning algorithms proposed in the literature based on the properties of links (Ji and Geroliminis, 2012; Saeedmanesh and Geroliminis, 2016) and based on the traffic data (Ambühl *et al.*, 2019). As the primary objective of the current work is to propose a methodology to estimate the MFD and the trip lengths from mobile phone data, a simple partitioning scheme is assumed. However, it is to be noted that the present framework can be used with any of the partitioning schemes proposed in the literature. The considered network is divided into 5 reservoirs as shown in Fig. 4a. The rationale behind the partitioning scheme is to have one reservoir for the downtown region and divide the region around the downtown into similar sized reservoirs. It is worth noting that the boundaries of reservoirs are not placed along the road network, but in-between the road networks. Generally, partitioning of the network is made along the bi-directional roads, where each



(a) Partitioning of Dallas network: Representation of the reservoirs along with their identities. ©OpenStreetMap 2019. (b) Partitioning of Dallas network: Link level representation of the reservoirs.



(c) Normalized demand profile for 24h period.

Figure 4: Partitioning of Dallas network and normalized demand pattern for a typical day scenario.

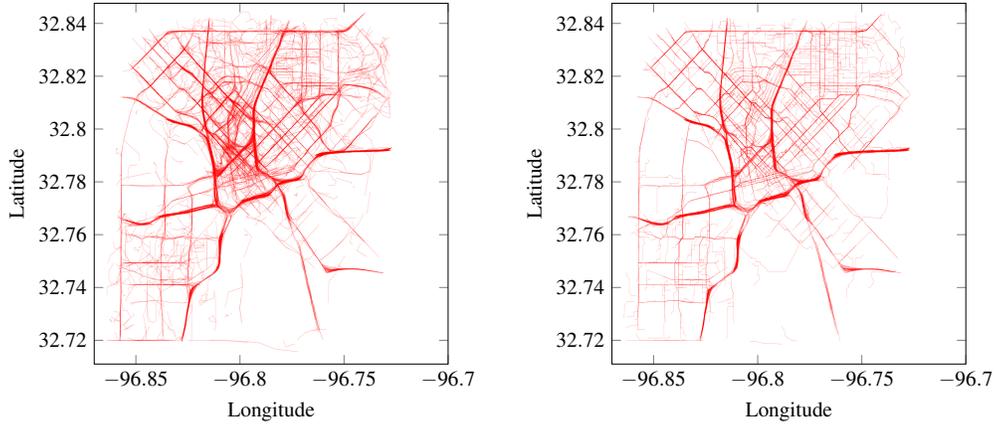
245 zone contains one direction of the road. Since GPS data has uncertainties, dividing the network along the road might result in trips that alternate between two adjacent reservoirs, even though they belong to one of those neighboring reservoirs in entirety. These type of alternating trips can be avoided by placing the boundaries within the blocks of the road network and it is the reason for having an irregular boundary in the present partitioning.

250 4.2. Error Estimation of Trip Enrichment Scheme

In this section, the error in traveled distances introduced by the proposed trip enrichment method is estimated. In order to do so, trips from the raw data with dense records, *i.e.*, each record that is within a radius of 200m its neighbor, are chosen for each OD pair. Since the records of these trips are relatively close to each other, no enrichment process is required to conform the trip
 255 to the network structure. This set of trips for each OD pair is considered as the reference trips to estimate the error in the traveled distances. The idea is to convert these high resolution reference trips to low resolution reference trips by randomly removing the intermediate records of each trip.

Table 1: Relative RMSE norm of trip lengths for the reference trips.

Origin	Destination	Error	No. of trips	Origin	Destination	Error	No. of trips
1	1	0.076	3629	3	1	0.062	1802
1	2	0.056	1052	3	3	0.049	10182
1	3	0.055	1766	4	4	0.055	1429
2	1	0.058	1321	5	5	0.050	4742
2	2	0.057	9403				



(a) Reference low resolution trip samples for all OD pairs. (b) Reference enriched trip samples for all OD pairs.

Figure 5: Randomly sampled reference low resolution and enriched trips for all OD pairs.

In this work, 70% of intermediate records are removed for each reference trip keeping the origin and the destination records unchanged. Now, these low resolution reference trips are enriched
260 using the enrichment method discussed in the previous section to obtain the map matched trip trajectories. The error is computed based on the difference in the traveled distances between the high resolution reference trips and the enriched reference trips.

Table 1 presents the relative Root Mean Square Error (RMSE) norm of the errors in the trip
265 lengths for the major OD pairs for the reference trips. It is clear from the error values that the enrichment scheme is very accurate in terms of traveled distances. Only the OD pairs, where there are more than 1 000 reference trips are presented in the table and the error values for the remaining OD pairs are also between 0.05 to 0.07. Fig. 5 shows the sample reference low resolution and enriched traces. Even though the major arterials are well represented in the low resolution traces, there is considerable scatter between the arterials. This is corrected in the enriched traces, where
270 the trip trajectories are well conformed to the underlying network. Hence, it can be concluded that the errors introduced from the proposed trip enrichment scheme are within the acceptable limits for the present purpose of estimation of the macroscopic variables.

Table 2: Total number of vehicles ($\times 10^4$) for 24h period between different macroscopic OD pairs.

Reservoir	1	2	3	4	5
1	7.9	3.4	2.6	0.5	0.7
2	4.8	7.2	1.8	0.2	0.4
3	7.0	4.4	6.5	0.4	0.2
4	1.7	0.9	0.7	0.7	0.1
5	4.0	3.0	0.7	0.2	2.2

4.3. Estimation of Penetration Rates

Since the available mobile phone data accounts only for the part of the total traffic, it is necessary to estimate the penetration rate (ρ) of the vehicles to compute a well-defined MFD and path flow distributions. An average penetration rate can be estimated by using the traffic counts from fixed loop detectors and aggregate number of vehicles passed through those points in the LBS data. However, this method gives a mean penetration rate across the network. Depending on the heterogeneity in the network under investigation, this mean penetration rate might not be accurate enough. For example, in the present case where network extends from downtown Dallas to suburbs, penetration rates can vary widely. It is already proposed that using an OD specific penetration rate is important to estimate an accurate MFD (Du *et al.*, 2016). In addition, penetration rates can vary with time, where higher penetration rates are normally observed at the peak hours compared to the off-peak hours.

In this work, penetration rates are estimated by fusing the OD matrix data, LDD and LBS data. The OD matrix data is obtained from North Central Texas Council of Governments (NCTCOG) for the morning peak period, *i.e.*, 06:30 AM to 08:59 AM. This data contains the total count of vehicles for different microscopic OD pairs (defined by NCTCOG) in the stated 2.5h period. Since this data does not have dynamic information of OD matrix, LDD is used to transform this static OD matrix into dynamic OD matrix. There are various loop-detectors placed all over the urban network of Dallas city and historical data is made available by NCTCOG. Several loop detectors are sampled all over the network and traffic counts for all vehicles for each 15 min interval in a day are considered. Each LDD is normalized with total number of vehicles throughout 24h period of that particular LDD. Finally, the mean of all normalized curves is approximated as the demand pattern of the city network. Fig. 4c presents the estimated normalized demand profile from the loop-detectors in the network. Now, using the OD matrix data for the given 2.5h period and the normalized demand profile, it is possible to expand the partial OD matrix to a typical full day OD matrix. Table 2 gives the total number of vehicles for the macroscopic OD pairs, *i.e.*, vehicle counts from one reservoir to another for 24h period.

According to the generalized definitions of Edie (1963), the average network density (k) and

flow (q) can be expressed as,

$$k = \frac{\sum_{i=1}^N TT_i}{L_n \Delta T} \text{ and } q = \frac{\sum_{i=1}^N TD_i}{L_n \Delta T}, \quad (1)$$

where TT_i and TD_i are time traveled and distance covered on a link l , respectively, L_n is the total network length, ΔT is the aggregation interval and N is the total number of links in the network under consideration. In the present work, an aggregation interval of 15 min is used. It is clear from eq. (1) that the density is computed using Total Traveled Time (TTT) and the flow is estimated by Total Traveled Distance (TTD) in the network by all vehicles within each aggregation interval. Density and flow of the network can be estimated using eq. (1) only if the trajectories of all the vehicles are known *a priori*. Often, that is not the case in reality and only trajectories of fraction of the vehicles are available. Hence, it is necessary to correct the expressions in eq. (1) with penetration rates. According to Nagle and Gayah (2014), density and flow of vehicles in the network can be re-defined as,

$$k = \frac{\sum_{i=1}^P TT_i/\rho}{L_n \Delta T} \text{ and } q = \frac{\sum_{i=1}^P TD_i/\rho}{L_n \Delta T}, \quad (2)$$

300 where ρ is the penetration rate and it is defined as number of probe vehicles to the total number of vehicles in the network. In eq. (2), the sum is made over the total number of probe vehicle trajectories P .

The final step to compute the macroscopic variables is to estimate the penetration rate. The current work proposes two different types of penetration rates namely, OD specific penetration rate (ρ_{od}) and origin specific penetration rate (ρ_o). It is possible to estimate the trip OD matrix based for each day from the mobile phone data based on the departure time of each trip. Since the number of trips between the same OD can vary from day-to-day, trip OD matrix based on the mobile phone data is estimated for each day separately. Let $N_{od,p}^I$ be number of trips from the data between origin o and destination d starting within the aggregation interval I . Similarly, $N_{od,n}^I$ is total number of trips estimated by data fusion of loop detectors and OD matrix from NCTCOG as elaborated earlier within a given interval I . Now, OD specific penetration rate at the aggregation interval I can be defined as,

$$\rho_{od}^I = \frac{N_{od,p}^I}{N_{od,n}^I}. \quad (3)$$

Similarly, let $N_{o,p}^I$ be the total number of trips from the data originating from origin o to all the destinations for a given interval I . In the same way, $N_{o,n}^I$ is the total number of trips from origin o to all the destinations computed from the NCTCOG data. Using these two quantities, origin specific penetration rate can be expressed as,

$$\rho_o^I = \frac{N_{o,p}^I}{N_{o,n}^I} \equiv \frac{\sum_{d=1}^r N_{od,p}^I}{\sum_{d=1}^r N_{od,n}^I}, \quad (4)$$

Table 3: Total network length in km for each reservoir.

Reservoir	Length of the network (km)
1	324.8
2	414.8
3	431.2
4	309.6
5	522.3

where r is the total number of macroscopic reservoirs in the network.

As the OD matrix provide the data of trip production to trip attraction, penetration rates should be used based on the departure time of the trip. Consider a trip i starting within an aggregation interval I between the OD pair od has traveled time and traveled distance of TT_i and TD_i , respectively. The expanded travel distance (TD_i^e) and the expanded travel time (TT_i^e) for the whole network considering the two types of penetration rates for that trajectory can be expressed as,

$$TT_i^{e,od} = \frac{TT_i}{\rho_{od}^{I_p}} \text{ and } TD_i^{e,od} = \frac{TD_i}{\rho_{od}^{I_p}}, \quad (5a)$$

$$TT_i^{e,o} = \frac{TT_i}{\rho_o^{I_p}} \text{ and } TD_i^{e,o} = \frac{TD_i}{\rho_o^{I_p}}, \quad (5b)$$

where suffixes/prefixes od and o represent that the expansion is done by OD and origin specific penetration factors, respectively and I_p is the aggregation interval corresponding to the departure time of trip i . Similarly, k and q can be expressed as,

$$k_{od} = \frac{\sum_{i=1}^P TT_i^{e,od}}{L_n \Delta T} \text{ and } q_{od} = \frac{\sum_{i=P}^N TD_i^{e,od}}{L_n \Delta T}, \quad (6a)$$

$$k_o = \frac{\sum_{i=1}^P TT_i^{e,o}}{L_n \Delta T} \text{ and } q_o = \frac{\sum_{i=P}^N TD_i^{e,o}}{L_n \Delta T}. \quad (6b)$$

Finally, the total length of the network L_n is estimated per reservoir to compute mean density and mean flow. Only major roads are considered and the roads connecting the residential areas are neglected to estimate the total network length. Table 3 shows the resulting lengths of the road networks per reservoir.

5. Estimated MFDs

The average values of density and flow for each day are estimated using eqs. (6). Only data from the weekdays, which are 43 in total, is considered in the computation of the macroscopic variables to estimate a stable and reproducible MFD. By computing the average of all the weekdays, a mean MFD for each reservoir can be estimated. However, there can be specific events like accidents, road works, *etc.* that can influence the shape of MFD on certain days. Since, the

information of these type of events that can influence the characteristics of the MFD is not available, the data is filtered based on the estimated MFD shape. Another important factor to consider in this context is the phenomenon of hysteresis in the MFD for the urban networks (Leclercq and Paipuri, 2019). It is normal to observe the clockwise hysteresis loops in the MFD due to network heterogeneity, demand pattern, driver's behavior, *etc.* However, the loading of the network from near empty state, which is observed during late night hours, is more stable and reproducible. Hence, a parabolic curve is fitted for the estimated MFD for each weekday for the data point ranging from the midnight till the morning peak hour, which is 08:00 AM in the present case. Only the days that show similar MFD fit characteristics are chosen to compute the mean MFD for each reservoir. Fig. 6 shows the estimated flow MFDs for the reservoirs, where the macroscopic variables are computed by the OD specific penetration rate. Firstly, it can be noticed that all the MFD curves are relatively stable with very less scatter. Reservoir 1 experiences the highest flow rate among all the reservoirs in the considered region. From Fig. 4a, it is clear that this reservoir corresponds to the downtown Dallas area and hence, higher mean flow in this reservoir is observed. Another important inference to be made from the MFD plots is the presence of clockwise hysteresis loops in the MFDs. Different colors of the data points correspond to the different times of the day to differentiate the loading and unloading phases of morning and evening peak hours. Reservoirs 1 and 3 in Figs. 6a and 6c, respectively exhibit more clearer hysteresis during the morning peak hour compared to others. In the case of reservoir 1, the difference in the flow during onset and offset of congestion is around $100 \text{ veh h}^{-1} \text{ lane}^{-1}$ for the same value of density. The hysteresis phenomenon is present in other reservoirs too, albeit the size of the loop is relatively smaller.

Figure 7 presents the mean speed MFDs using the OD specific penetration rate. The hysteresis phenomenon can be clearly observed in the speed MFDs, where in Figs. 7a and 7c, the mean speed during the loading is clearly higher than the unloading during the morning peak hour. The estimated speed MFD in the reservoir 4, which is shown in Fig. 7d has scatter in the shape of MFD. This is due to the sparser phone data available in this reservoir. The reservoirs surrounding the downtown Dallas area, *i.e.*, reservoirs 2 to 5 have freeways included as shown in Fig. 4a. This implies that the free-flow speed of reservoirs 2 to 5 should be higher compared to reservoir 1, which is downtown area. It can be observed from the plots that the free-flow speed of reservoirs 2 to 5 is higher than reservoir 1 and hence, the estimated MFDs are qualitatively verified.

Figure 8 presents the estimated flow MFDs using origin specific penetration rate for computing macroscopic variables. Firstly, it is clear that the MFDs are qualitatively and quantitatively very similar to the ones presented in Fig. 6, where OD specific penetration rate is used in computation of macroscopic variables. The hysteresis loops observed in the reservoirs 1 and 3 in the previous case are also noticed in the present case. However, the scatter in the reservoir 5 in Fig. 8e is comparatively larger than the scatter in Fig. 6e. This is true for not only reservoir 5, but also to all the reservoirs. This can be demonstrated using speed MFD plots shown in Fig. 9. Even though

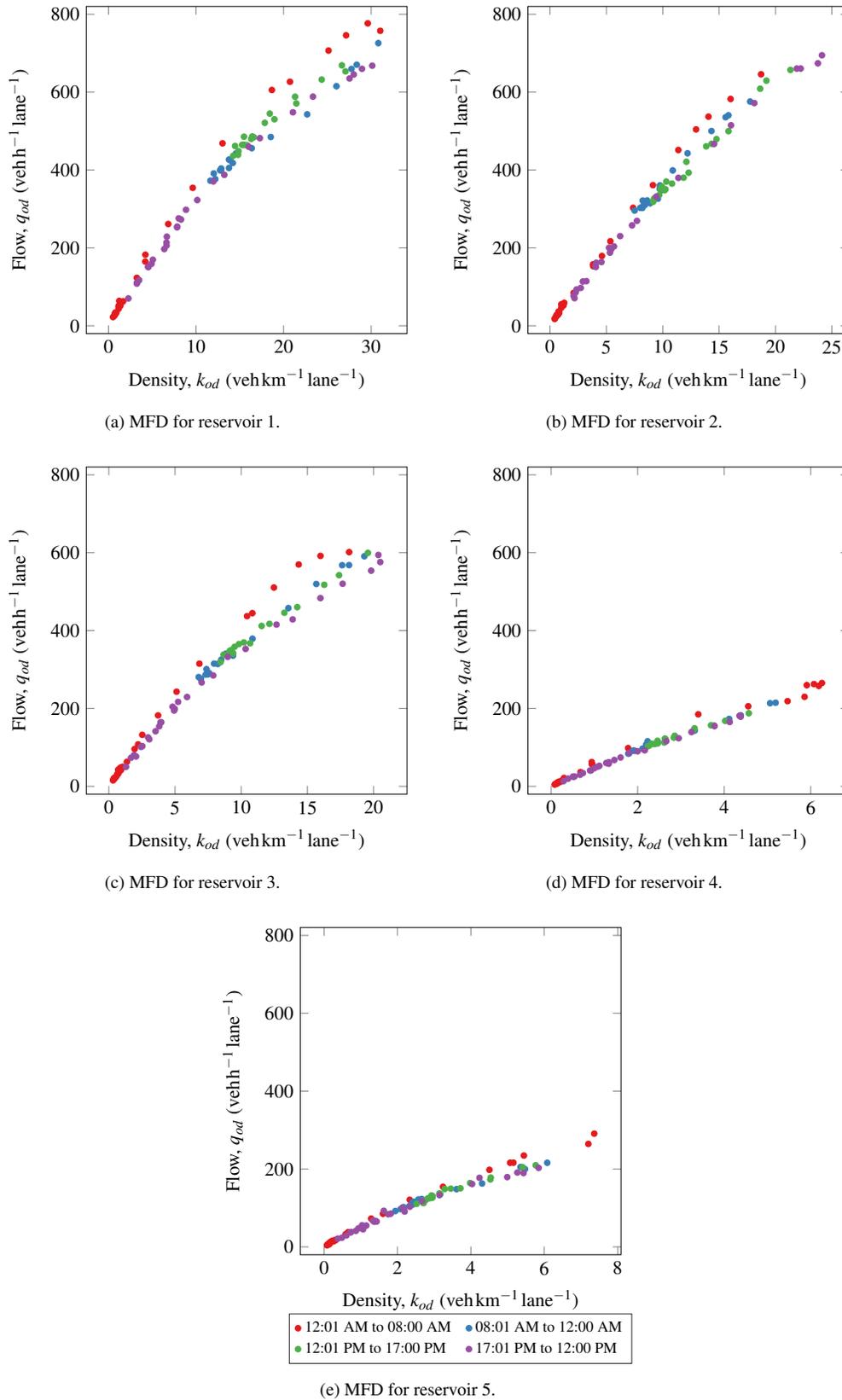
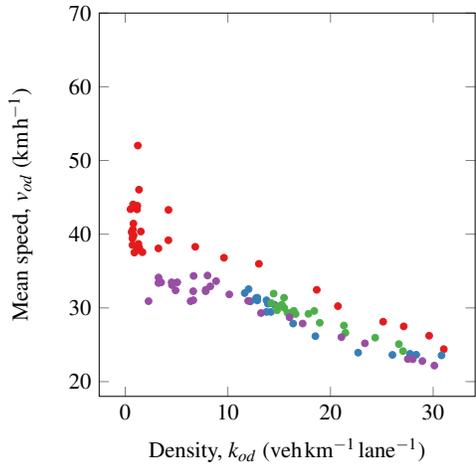
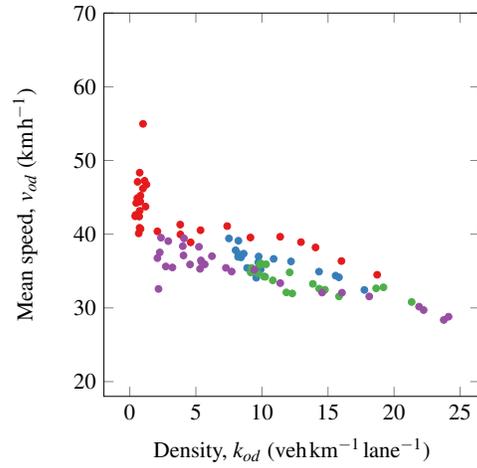


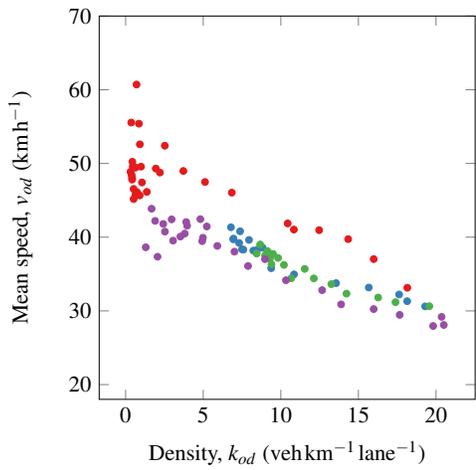
Figure 6: Flow MFD estimates using the OD specific penetration rate for computing the density and the flow.



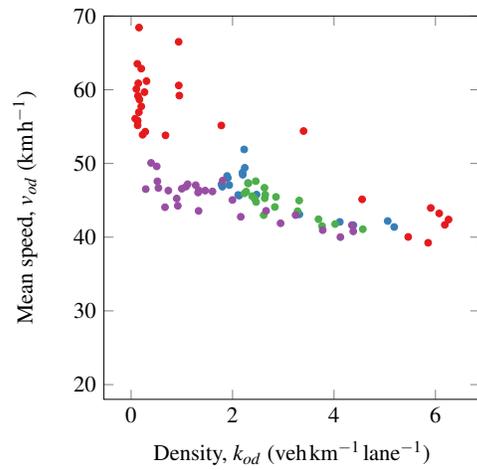
(a) Speed MFD for reservoir 1.



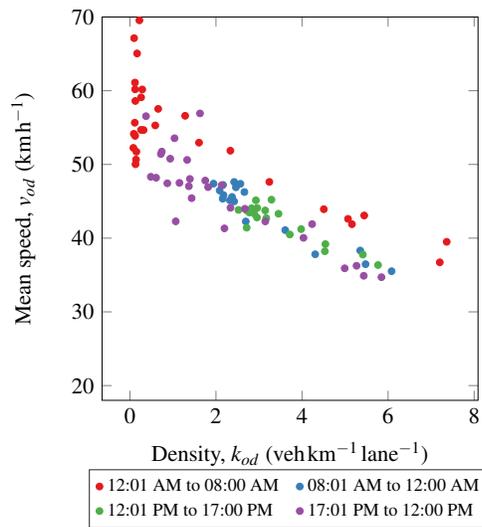
(b) Speed MFD for reservoir 2.



(c) Speed MFD for reservoir 3.



(d) MFD for reservoir 4.



(e) Speed MFD for reservoir 5.

Figure 7: Speed MFD estimates using the OD specific penetration rate for computing the density and the flow.

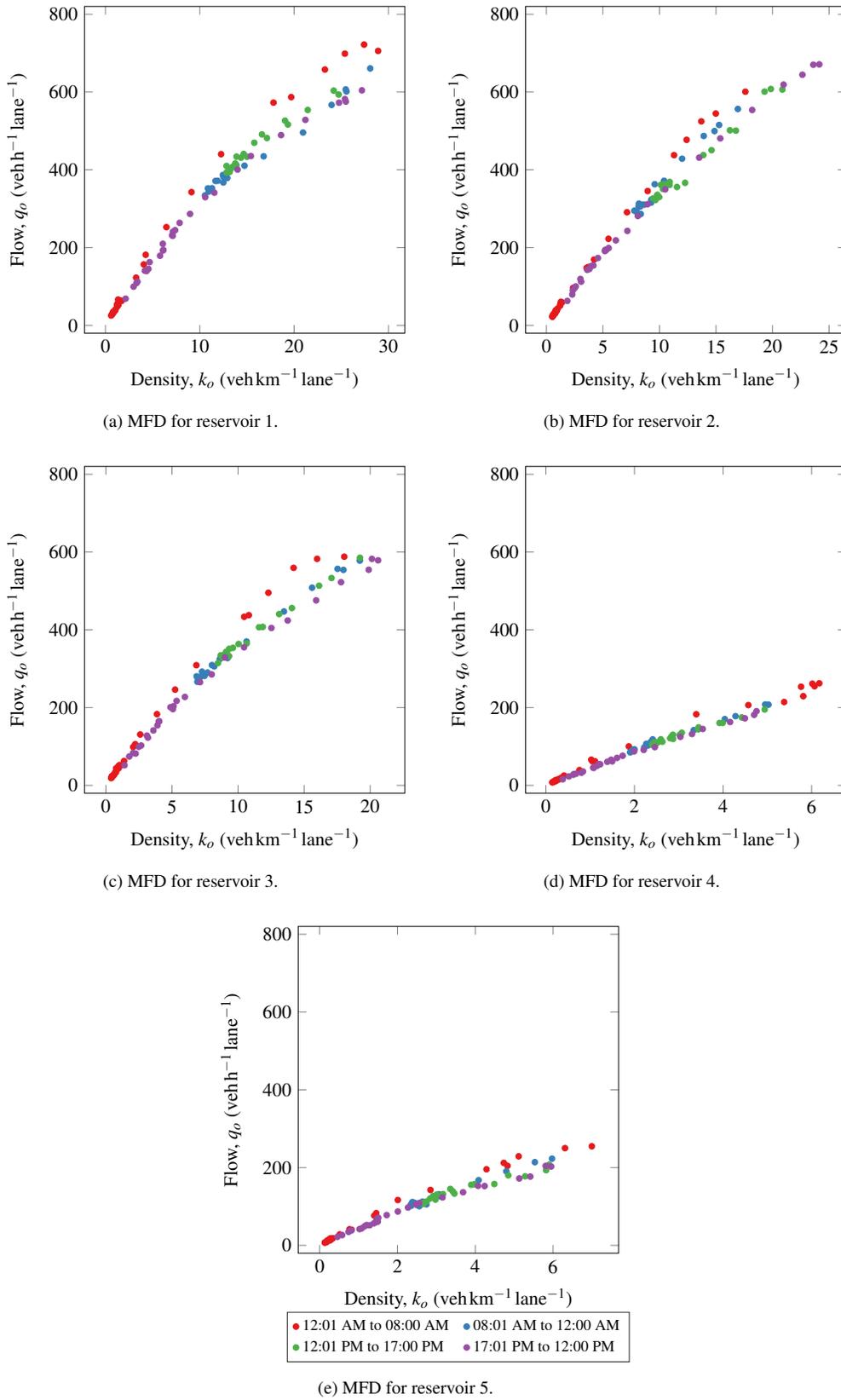


Figure 8: Flow MFD estimates using the origin specific penetration rate for computing the density and the flow.

all the plots show a good relation between mean speed and density, it is evident that the scatter in the present case is relatively larger than the previous case. This is expected as the OD specific penetration rate is more accurate than the origin specific counterpart. Expanding the macroscopic variables without considering the destination might result in overestimation in few OD pairs and underestimation in the rest. This can contribute to the scatter in the MFD, which is noticed in the Figs. 8 and 9. Despite the presence of scatter, both flow and speed MFDs show a reasonable correlation between the macroscopic variables.

Figure 10 shows the average penetration rates of probe vehicles from LBS data for each reservoir. Notice that according to the present framework, the penetration rates within same interval is not same and depends on the starting time of the trip as presented in eq. (5). Hence, multiple penetration rates are possible for same interval as multiple trips can transverse during that period. Average penetration rates are computed by comparing the expanded macroscopic variable like travel time and travel time values estimated from the LBS data. The estimated penetration rates for different days are again averaged to obtain an average trend of penetration rates during a typical day. In order to obtain a representative trend of the penetration rates, only days that show similar MFD shape is considered to estimate the average penetration rates. This is done for both OD specific and origin specific penetration rates. The first inference from the plots is that both proposed penetration rates show very similar trend as expected. It can be observed that peak penetration rates are observed during the morning and the evening peak hours. This phenomenon is clearly noticed in the reservoir 1, where two peaks, one at morning and another at evening, are noticed. It is also evident that the variation of penetration rate within day cannot be neglected when estimating the MFD and using a mean penetration for whole day can lead to erroneous results. In order to show the relative accuracy between using constant and time dependent penetration rates, MFD is calibrated using a constant penetration rate for whole 24 h period. Fig. 11 presents the estimated MFD for reservoir using time averaged constant and time dependent penetration rates for reservoir 1. It is clear from the plots that the MFD estimated using constant penetration in Fig. 11a has less scatter than its counterpart. In other words, the hysteresis phenomenon is absent in the case of constant penetration rate. This is expected outcome as using the constant penetration rate expands the macroscopic variables in the same proportion in all the aggregation intervals. However, this is not the case in the reality, where higher penetration rates are observed during the peak hours compared to the off-peak hours. Therefore, MFD estimated using constant penetration rate yields a mean MFD curve without hysteresis.

6. Trip Lengths Estimation

6.1. Static Analysis

The second part of the present work presents the details of trip lengths estimation from the mobile phone data. As already stated earlier, estimating trip lengths is impossible without massive

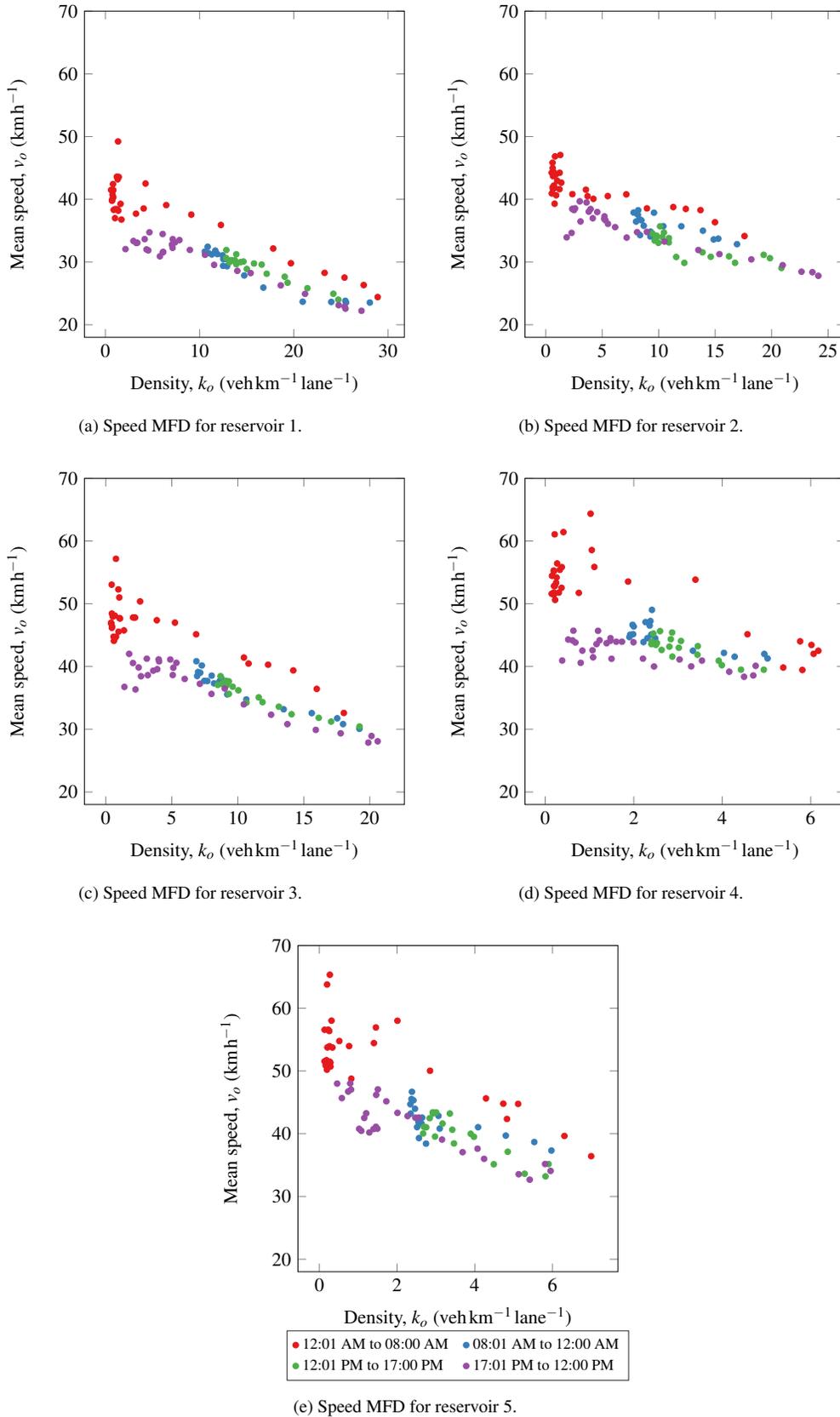


Figure 9: Speed MFD estimates using the origin specific penetration rate for computing the density and the flow.

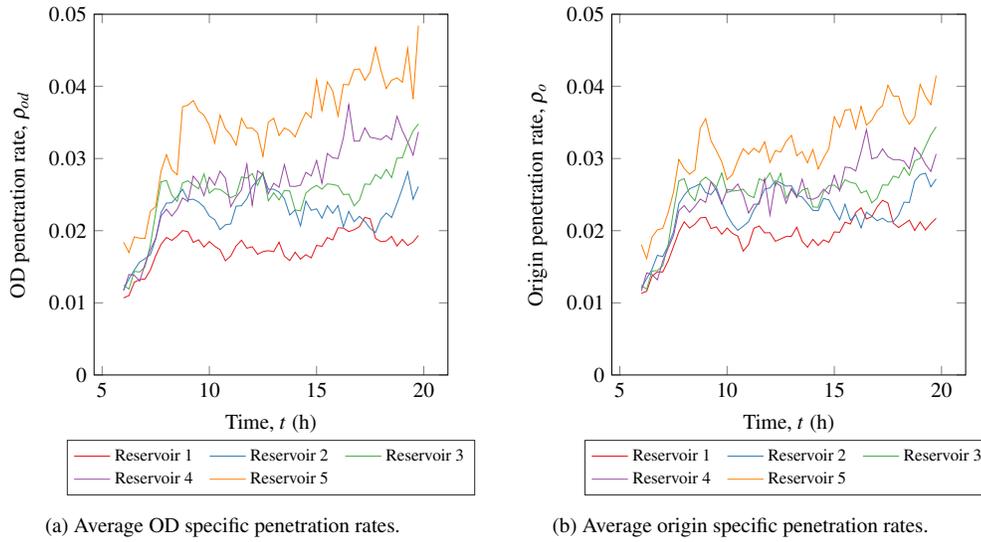


Figure 10: Average penetration rates for each reservoir.

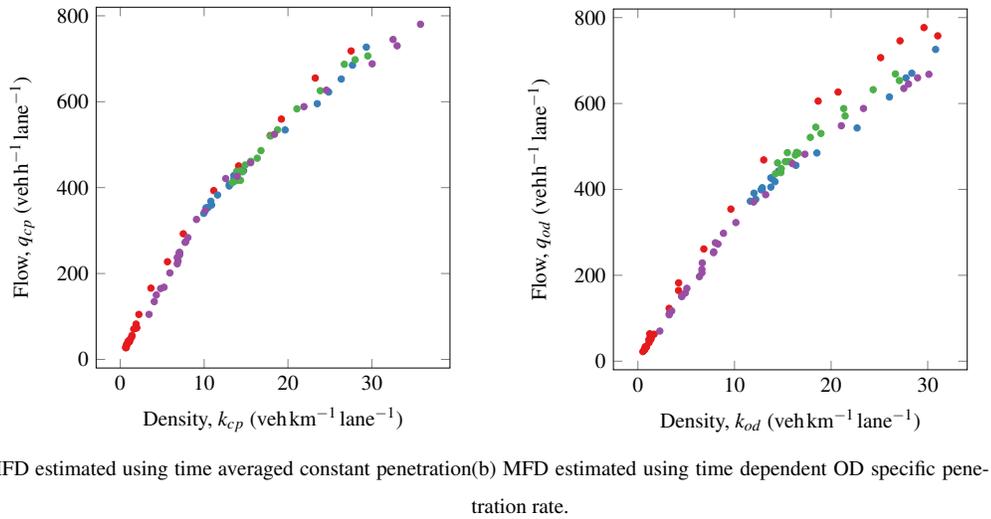
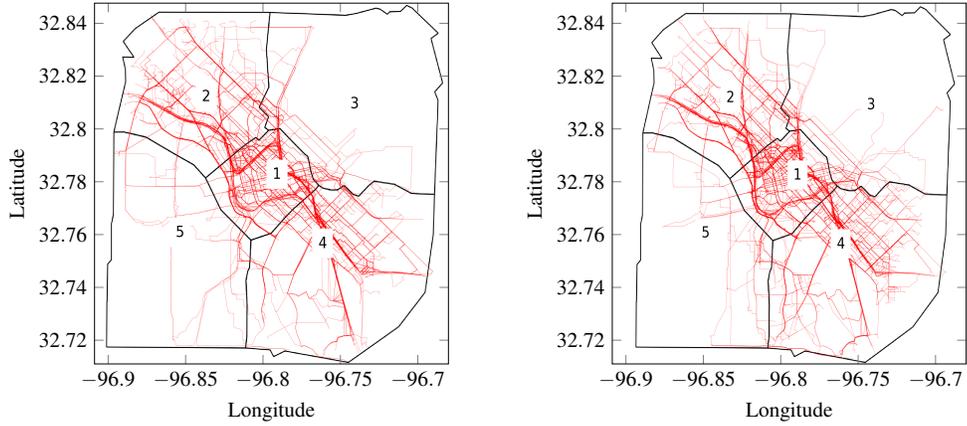


Figure 11: Comparison of MFDs using constant and time varying penetration rates.



(a) Randomly sampled trajectories between macroscopic OD pair 4 – 2. (b) Randomly sampled trajectories between macroscopic OD pair 2 – 4.

Figure 12: Sample set of trajectories between OD pairs 4 – 2 and 2 – 4.

individual data. Most of the works in the literature are based on the network exploration methods to build virtual trips. The present type of mobile phone data, *i.e.*, LBS data is a unique opportunity to fill the gap in the estimation of trip lengths. Since, the trajectory of each individual trip is readily available, it is possible to cluster the trips based on the reservoir sequence. Once the trips are clustered based on the reservoir sequence, major macro-paths between the macroscopic OD pairs can be identified. Note that in the present section, all the macro-paths are represented as the reservoir sequence. For instance, a trip that starts in reservoir 1 and ends in 3 by transversing through reservoir 2, it is named as $1 \rightarrow 2 \rightarrow 3$.

Consider the macroscopic OD pairs 4 – 2. From Fig. 4a, it is clear that there can be several possible macro-paths between the considered OD pair like $4 \rightarrow 1 \rightarrow 2$, $4 \rightarrow 3 \rightarrow 2$, *etc.* Notice that these are only few sample macro-paths and still more realistic combinations like $4 \rightarrow 3 \rightarrow 1 \rightarrow 2$, $4 \rightarrow 1 \rightarrow 3 \rightarrow 2$, *etc.* are possible from the considered clustering. Assuming a single macro-path per OD pair is too crude for accurate resolution of the traffic dynamics. On the other hand, considering all the possible and realistic macro-paths between OD pair adds complexity to the MFD-based simulation, thereby losing the spirit of the framework. This issue can be appropriately addressed by using the trajectory data of the mobile phones.

Figure 12a shows the randomly sampled trajectories estimated from the phone data between the OD pair 4 – 2. At first, it is clear that most of trips transverse the network in the sequence $4 \rightarrow 1 \rightarrow 2$. Hence, it can be considered as the major macro-path for the considered OD pair. However, it is clear from the plot that there are other macro-paths like $4 \rightarrow 3 \rightarrow 1 \rightarrow 2$, $4 \rightarrow 1 \rightarrow 3 \rightarrow 2$, $4 \rightarrow 5 \rightarrow 1 \rightarrow 2$ and $4 \rightarrow 1 \rightarrow 5 \rightarrow 2$, which contribute towards the OD flow. The trajectory data for all the days considered can be used here to rank the most used macro-path to the least used. Consequently, macro-paths with very few trips can be safely neglected without compromising the modeling framework. In the present example of OD pair 4 – 2, almost 70%

Table 4: Average trip length in m inside each reservoir using OD and origin specific penetration rates using method 1.

Reservoir	Average trip length using OD pen. rate (m)	Average trip length using origin pen. rate (m)
1	1957	1916
2	2178	2023
3	2125	2080
4	2580	2458
5	2742	2480

of the trips take the macro-path $4 \rightarrow 1 \rightarrow 2$ and the majority of the rest are distributed among $4 \rightarrow 3 \rightarrow 1 \rightarrow 2$ and $4 \rightarrow 1 \rightarrow 3 \rightarrow 2$. Fig. 12b presents the sampled trajectories for the OD pair $2 - 4$, which is symmetrically opposite to $4 - 2$. It can be observed that the trips have a similar
415 macroscopic behavior as the OD pair $4 - 2$. This can be verified by the trajectory data, where the major macro-path is $2 \rightarrow 1 \rightarrow 4$ and the rest of the flow is observed in $2 \rightarrow 1 \rightarrow 3 \rightarrow 4$ and $2 \rightarrow 3 \rightarrow 1 \rightarrow 4$. It is inferred from the data that the present network of Dallas city shows this symmetric property for most of the OD pairs. This might not be the case always depending on the structure of the network and its corresponding partitioning. It should be noted that whatever the
420 macro-path patterns are, the LBS data provide enough samples to define the major macro-paths. This may not be the case with other data sources, which leads to more uncertainties.

Once the macro-paths between the OD pairs are established, the following question is to estimate the lengths of each macro-path. A straightforward and simple method, say method 1, is to estimate the average trip length inside each reservoir taking into account all the trajectories.
425 This corresponds to the original approach proposed by Daganzo (2007). This can be considered as constant static trip lengths as dynamic changes in the trip lengths are neglected. This method does not account for origin and destination of the macro-path. For instance, if the average trip length in the reservoir 1 is 1 000 m, the same value is assigned for reservoir 1 for the macro-path $1 \rightarrow 2$ and $4 \rightarrow 3 \rightarrow 1 \rightarrow 2$. However, this can introduce big discrepancies in the MFD simulation
430 results (Batista *et al.*, 2019). Table 4 presents the average trip lengths inside each reservoir estimated using OD and origin specific penetration rates. In accordance with the previous results, the estimated average trip lengths are very similar using both types of penetration rates. The lengths are representative to the size of the reservoirs considered. A more intensive method, say method 2, is averaging the trip lengths inside each reservoir based on the macro-path, *i.e.*, taking the mean
435 of all trips that transverse a given reservoir for a given macro-path. Again, this approach is possible only because of the massive phone data available, which guarantees enough observations on all major macro-paths. The difference between two approaches is that the former considers the average of all trips inside each reservoir irrespective of the macro-path, while the later consider the mean trip length inside each reservoir for a given macro-path individually. In other words,

Table 5: Average trip length in m using OD specific penetration rate for selected macro-paths.

Macro-path	Avg. length per reservoir (m)	Avg. length per macro-path (m)	Rel. difference in total length
1 → 1	{1957}	{1366}	0.4331
1 → 3	{1957, 2125}	{1736, 2339}	0.001
3 → 1 → 5	{2125, 1957, 2742}	{2815, 4581, 3499}	0.3736
4 → 1 → 2	{2580, 1957, 2178}	{2961, 4546, 3138}	0.3692

Table 6: Average trip length in m using OD specific penetration rate for selected macro-paths along with their corresponding standard deviations.

Macro-path	Avg. length per reservoir (m)	Std. (m)	Coefficient of variability
1 → 1	{1366}	{1079}	{0.79}
2 → 2	{1738}	{1502}	{0.86}
2 → 1 → 5	{2898, 2751, 3183}	{1951, 907, 1922}	{0.67, 0.32, 0.59}
3 → 1 → 5	{2814, 4580, 3498}	{1941, 863, 1968}	{0.68, 0.18, 0.56}
4 → 1 → 5	{2910, 3861, 3310}	{2072, 1235, 1972}	{0.71, 0.32, 0.59}
5 → 1 → 4	{2961, 3734, 3333}	{1795, 1348, 2084}	{0.60, 0.36, 0.62}

440 the average trip length in reservoir, say 1, is same in all macro-paths in the first method. Whereas in the second one, mean trip length of reservoir 1 changes for each macro-path. It is already discussed in-detail in [Batista *et al.* \(2019\)](#) that the second method is more accurate than the first one based on the simulation results. In the current work, the discrepancies between two methods are demonstrated using the data.

445 Table 5 shows the average trip lengths of selected macro-paths using both approaches and the relative differences in the total trip lengths. Consider the internal macro-path 1 → 1. The mean trip length estimated by method 1 is 1957 m, while method 2 gives 1366 m, which is significantly lower than its counterpart. As reservoir 1 is the downtown area of the Dallas city, majority of the internal trips are between the freeways that encompass the area. One of the longest trips
450 possible, without considering freeway network, in this reservoir is around 2500 m. Taking this into the account an average trip length of 1957 m over more than 100000 trips is unrealistic. The reason for such a high mean trip length is due to averaging all the trips that transverse reservoir 1 irrespective of the OD pair. This can be elaborated clearly using macro-paths 3 → 1 → 5 and 4 → 1 → 2. From Fig. 4a, it is clear that both stated macro-paths need to cross the reservoir in its
455 entirety. As they are mostly long distance trips, users tend to use freeways, which are ring roads in the reservoir 1. Hence, longer average trip lengths are observed for these macro-paths in the reservoir 1, as vehicles need to circumnavigate the downtown area. In the method 1, these type of trips are aggregated along with internal trips of reservoir 1 and hence, a higher average trip length

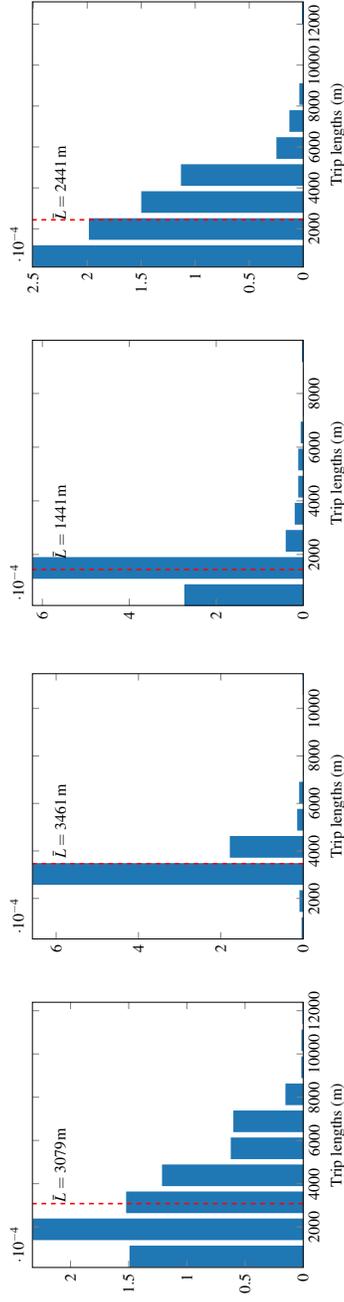
is estimated. On the other hand, the method 2 estimates a more representative shorter trip length
460 for the internal trip $1 \rightarrow 1$ and a longer trip length for macro-paths $3 \rightarrow 1 \rightarrow 5$ and $4 \rightarrow 1 \rightarrow 2$ as
expected. This conclusion is in-line with the results in the literature (Batista *et al.*, 2019) and the
present results demonstrate that phone data provides a practical and effective way to calibrate the
trip lengths.

The standard deviation of the trip lengths are estimated inside each reservoir in order to deter-
465 mine the accuracy of the estimated lengths. Table 6 presents the trip lengths for selected macro-
paths using method 2 and their corresponding standard deviations. In the case of internal trips,
i.e., $1 \rightarrow 1$ and $2 \rightarrow 2$, the standard deviation is quite large compared to the mean trip length. This
is expected as trip lengths vary widely across inside the reservoir and hence, higher coefficient
of variation. However, macro-paths that transverse across reservoir 1, for instance $2 \rightarrow 1 \rightarrow 5$,
470 $3 \rightarrow 1 \rightarrow 5$, *etc.*, the coefficient of variation inside the reservoir 1 is relatively low. This shows
that the estimated macro-paths captured the patterns of the trips effectively, where average trip
length across reservoir 1 is very similar for all the trips. The coefficient of variation in origin
and destination reservoirs for the stated macro-paths are relatively large. This is due to the exact
points of departure and arrival can vary across a wide range in the reservoirs and hence, wide
475 range of trip lengths. It is also to be noted that the considered reservoirs are relatively large in
area and using a partition with smaller reservoir will decrease the variability of the trip lengths.

In order to illustrate the importance of considering average trip length per macro-path, con-
sider two macro-paths $4 \rightarrow 3 \rightarrow 1 \rightarrow 2$ and $4 \rightarrow 1 \rightarrow 3 \rightarrow 2$. Even though same reservoirs are
transversed, albeit in different order, between the considered OD pair, the mean trip lengths inside
480 each reservoir for both trips are not same. Figs. 13 and 14 show the trip lengths distribution for
the stated two macro-paths, respectively. It is clear from the distributions that even for the same
OD pair, the mean trip lengths inside each reservoir depends on the macro-path. In the presented
trip length distributions, except for the mean trip length inside reservoir 2, other mean trip lengths
show significant differences. For instance, the mean trip length in reservoir 1 for the macro-path
485 $4 \rightarrow 1 \rightarrow 3 \rightarrow 2$ is 3461 m, whereas in the case of $4 \rightarrow 3 \rightarrow 1 \rightarrow 2$, it increases to 4080 m. Hence,
this inference reinforces the previous conclusion about the importance of considering mean trip
lengths per OD pair and per macro-path.

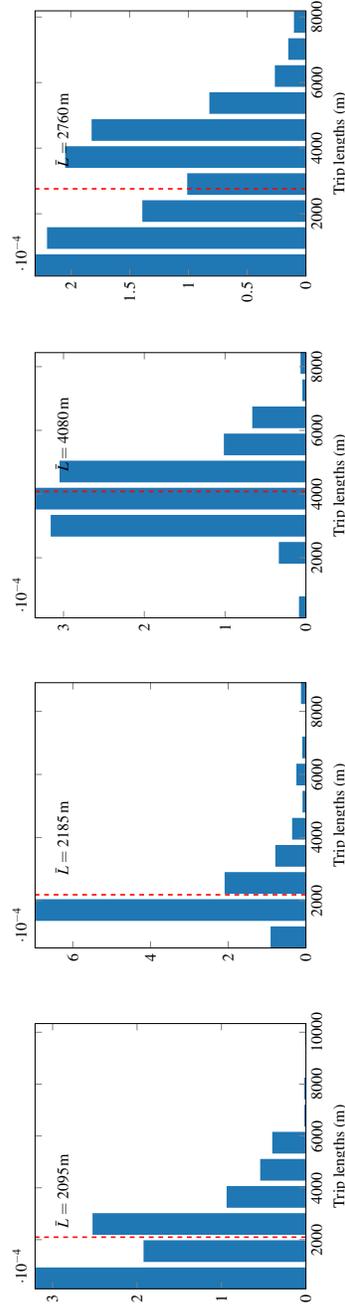
6.2. Dynamic Analysis

The analysis presented till now neglects the dynamic information of the trip lengths. There
490 had not been many works on dynamic variation of trip lengths using empirical data because of
lack of data. However, this is known as the important factor as local congestion changes the trip
length distributions, where users tend to avoid shorter congested routes to take longer ones. The
average trip lengths presented in Tables 4 and 5 and distribution in Figs. 13 and 14 takes into
the account all the trips observed during the whole 2 months period. However, it is intuitive that
495 trip lengths within and between the reservoirs tend to be dynamic, where the users tend to prefer



(a) Distribution in Reservoir 4. (b) Distribution in Reservoir 1. (c) Distribution in Reservoir 3. (d) Distribution in Reservoir 2.

Figure 13: Distribution of trip lengths inside each reservoir for the macro-path $4 \rightarrow 1 \rightarrow 3 \rightarrow 2$.

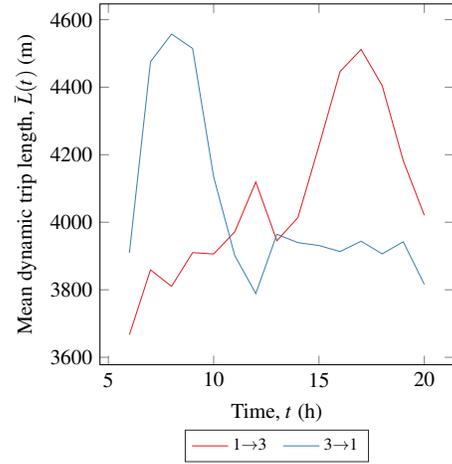
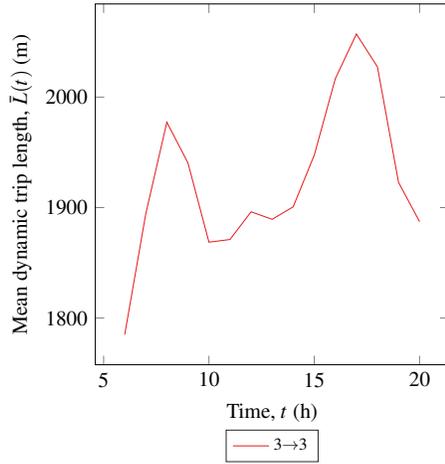


(a) Distribution in Reservoir 4. (b) Distribution in Reservoir 3. (c) Distribution in Reservoir 1. (d) Distribution in Reservoir 2.

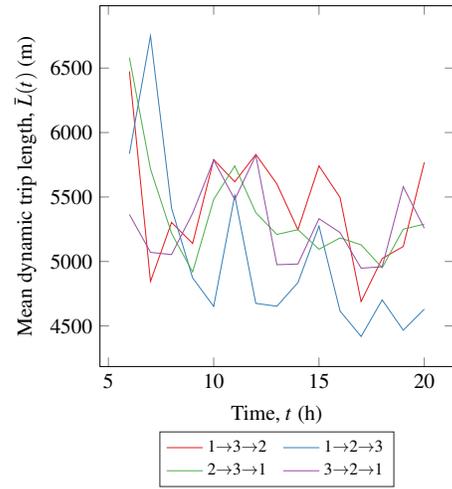
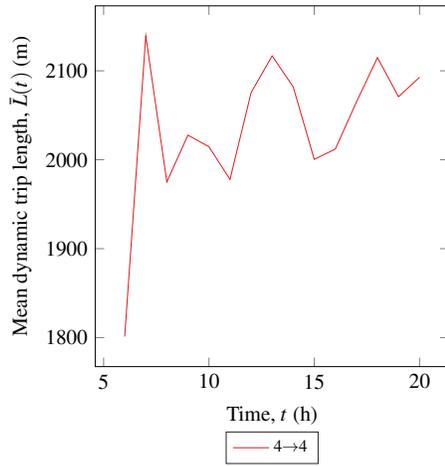
Figure 14: Distribution of trip lengths inside each reservoir for the macro-path $4 \rightarrow 3 \rightarrow 1 \rightarrow 2$.

the alternative paths during the congestion periods. [Batista et al. \(2020\)](#) proposed a framework to estimate the dynamic trip lengths explicitly and concluded that including the dynamic changes in the trip lengths improves the accuracy of the MFD-based simulations. However, as stated already, the authors built a virtual set of trips by network sampling to do this analysis in the stated work
500 and not real trip lengths.

In the present work, the dynamic trip lengths between macroscopic OD pairs is estimated using starting time of each trip. In order to do so, an aggregate time period of 60 min is considered and hence, for each day there are 24 aggregation periods. For a given OD pair, all the trips that are starting within a given aggregation period are selected and a mean trip length is estimated for
505 that given period. This is done for each weekday separately in the considered 2 month period. Finally, the mean trip length within each aggregation interval is estimated for all the weekdays. Fig. 15 presents the mean evolution of the trip length for different macroscopic OD pairs. The macro-paths that show different types of trends are selected for the discussion. Consider Fig. 15a, where the evolution of internal macro-path $3 \rightarrow 3$ is presented. Two peaks, one at the morning
510 peak hour and another at the evening peak hour, can be clearly noticed in the plot. This trend signify that the users perhaps take the longer paths during the peak hours to avoid the most used routes. The difference between the trip lengths at the peak and non-peak period is only less than 10% for the macro-path $3 \rightarrow 3$. However, this relatively insignificant difference can introduce a considerable bias in the traffic dynamics at the peak hour in the MFD-based simulations ([Batista et al., 2019](#)). Fig. 15b shows the dynamic trip lengths of the macro-paths $1 \rightarrow 3$ and $3 \rightarrow 1$. It is clear from the plots that the dynamic trip lengths evolution of the two macro-paths are nearly
515 symmetric. This is due to movement of the people from the suburban region to the downtown in the morning and *vice-versa* in the evening. The mean dynamic trip length evolution of the macro-path $4 \rightarrow 4$ is shown in Fig. 15c, where the changes in the trip length are relatively insignificant. This is due to the presence of large urban spaces for leisure activities, which decreases its overall
520 contribution to the mean flow of the network. This can be verified using the length of the network presented in Table 3, where reservoir 4 has the least network length amongst all the reservoirs. Hence, a more stable trip length evolution is noticed. Finally, Fig. 15d presents the dynamic trip lengths of the macro-paths that involve reservoirs 1, 2 and 3 in different sequences. Even though,
525 all the macro-paths show the variation within the same limits of trip lengths, no clear trend in the evolution is observed. It is clear that these types of behaviors are justified, given the topology of the network. However, it is difficult to predict these trends between different OD pairs *a priori* and appropriately calibrate the MFD models. This type of analysis has not been done before in the literature due to lack of sufficient and reliable data. Therefore, this framework estimates
530 the mean evolution for different macro-paths, which can be directly used in the MFD simulation framework.



(a) Evolution of trip length for internal macro-path $3 \rightarrow 3$. (b) Evolution of trip lengths for macro-paths $1 \rightarrow 3$ and $3 \rightarrow 1$.



(c) Evolution of trip length for internal macro-path $4 \rightarrow 4$. (d) Evolution of trip lengths for macro-paths crossing reservoirs 1, 2 and 3.

Figure 15: Mean dynamic trip lengths for different macro-paths.

6.3. Analysis on Path Flow Distribution

The principal input data for a MFD-based simulation are underlying MFD, macro-paths and their corresponding trip lengths, which are discussed until now. As already seen earlier, there can be more than one macro-path feasible between a given macroscopic OD pair. It is noticed that almost all the major macro-paths, *i.e.*, the ones that have higher flow compared to the others, in the present partitioning transverse across the reservoir 1, which is the downtown area. So, considering just one major macro-path between an OD pair and assigning the total flow to this path might lead to unrealistically high flows in the reservoir 1, which might result in the gridlock. Hence, depending on the relative flow between all the feasible macro-paths for a given OD pair, it is necessary to have more than one macro-path. For instance, for the internal trips that start and finish in the reservoir 1, it is noticed that 97% of the trips have the macro-path $1 \rightarrow 1$ and the rest follows $1 \rightarrow 2 \rightarrow 1$, $1 \rightarrow 3 \rightarrow 1$, *etc.* In this case, it is safe to neglect the other macro-paths and consider only the major macro-path. It is observed that all the internal trips in the present work shows this behavior and hence, only one major macro-path is considered for them. However, this is not true for the transfer trips that have origins and destinations in different reservoirs.

In the case of having more than one macro-path between an OD pair, it is essential to know the amount of flow to be assigned to each path. In this context, path flow coefficient for a macro-path p for a given OD pair, od can be defined as,

$$\alpha_{od}^p = \frac{N_{od}^p}{\sum_{i \in \mathcal{P}_{od}} N_{od}^i}, \quad (7)$$

where N_{od}^i is the number of trajectories on macro-path i and \mathcal{P}_{od} is the set of all major macro-paths between the OD pair, od . The path flow coefficient can be estimated using DTA determining the UE conditions. However, DTA can be computationally demanding depending on the size of the network under study and it is possible to extract the information on the path flow coefficients using the trajectory data. However, at the network level it is not possible to observe the path flow distribution between all local OD pairs. This becomes feasible only at the regional level and this is extracted using phone data. Since path flow distributions are computed assuming UE settings, it is possible to validate this hypothesis by empirically deriving the gap to the UE conditions from the phone data. The important research questions in this context are to determine (i) if the gap for macroscopic OD pairs are close to zero as assumed by UE hypothesis and (ii) if the gaps change in time.

Figure 16 shows the mean evolution of the path flow coefficients for few OD pairs. There are two major macro-paths for each between the OD pairs $1 - 2$ and $2 - 1$ and their corresponding path flow evolutions are presented in Fig. 16a. It can be observed that the macro-paths $1 \rightarrow 2$ and $2 \rightarrow 1$ has majority of the flow in both cases and they remain stable over the course of the day. On the other hand, path flow for the macro-paths between OD pair $3 - 4$ show a significant variation, which is shown in Fig. 16b. It is clear that during the peak hours, users tend to use the

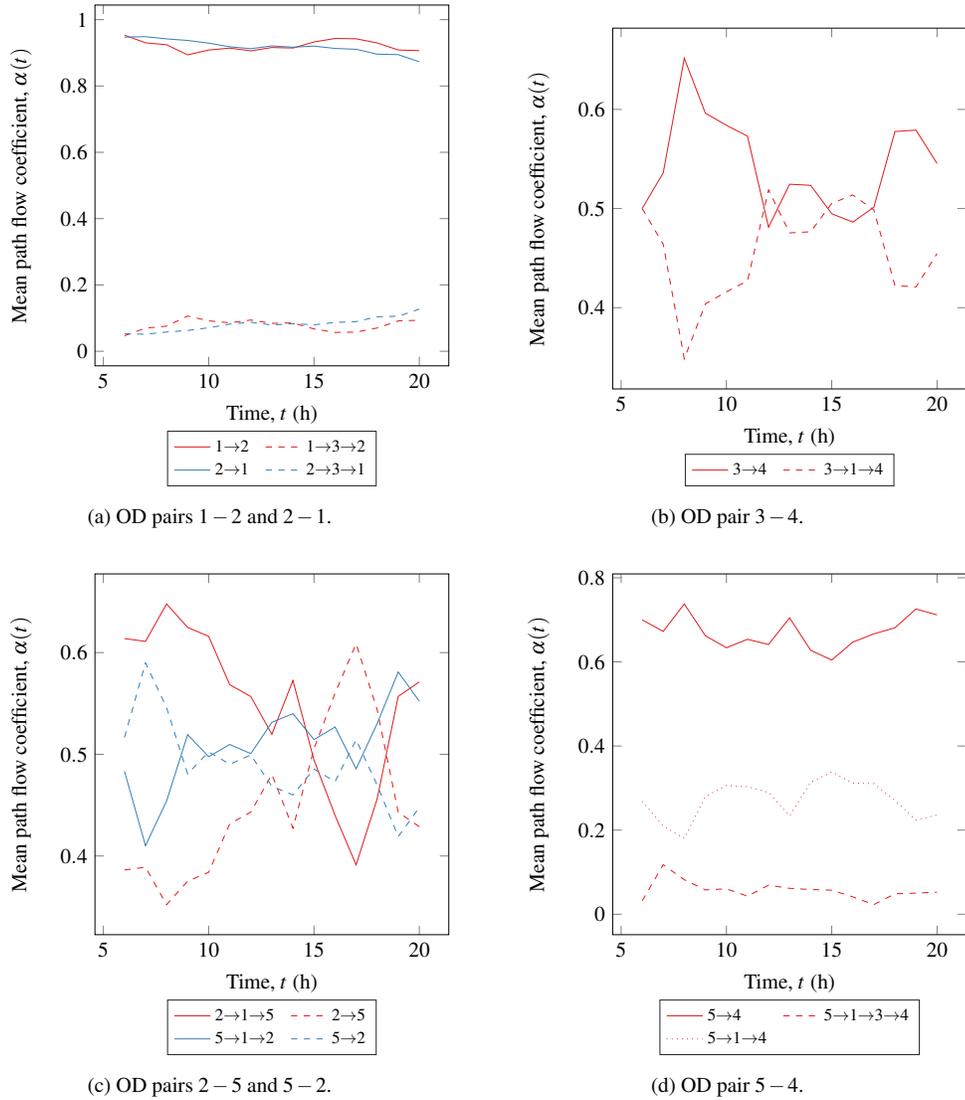


Figure 16: Mean evolution of path flow coefficients between different macroscopic OD pairs.

macro-path $3 \rightarrow 4$ over $3 \rightarrow 1 \rightarrow 4$, whereas both macro-paths experience nearly equal amount
 565 of flow during the off peak hours. Fig. 16c shows a similar trend, albeit, users tend to take one
 macro-path during the morning peak and another during the evening peak. Consider the OD
 pair 2 – 5, where the users prefer the macro-path $2 \rightarrow 1 \rightarrow 5$ over the macro-path $2 \rightarrow 5$ during
 the morning peak hour and *vice-versa* during the evening peak hour. A symmetrically opposite
 case is observed in the reverse trip, *i.e.*, for the OD pair 5 – 2, where users use the macro-path
 570 $5 \rightarrow 2$ in the morning over the macro-path $5 \rightarrow 1 \rightarrow 2$. Finally, Fig. 16d shows the path flow
 coefficients of the three major macro-paths for the OD pair 5 – 4, where a stable evolution in all
 the three macro-paths is observed. Therefore, it can be concluded that it is possible to estimate the
 valuable information like dynamic trip lengths and path flow coefficients from the mobile phone
 data, which is otherwise very difficult to obtain from other data sources.

Finally, the current work is concluded by presenting the estimates of User Equilibrium (UE)

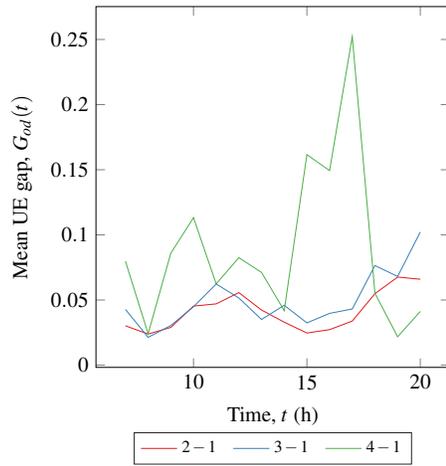
gaps for each macroscopic OD pair. The gap corresponds to the relative difference between the travel time on the macro-path and minimum travel time among all the macro-paths for a given OD pair (Sbayti *et al.*, 2007). The UE gap for a given OD pair, G_{od} can be expressed as,

$$G_{od} = \frac{1}{TT_{min}^{od}} \sum_{i \in \mathcal{P}_{od}} \alpha_{od}^i (TT_i^{od} - TT_{min}^{od}), \quad (8)$$

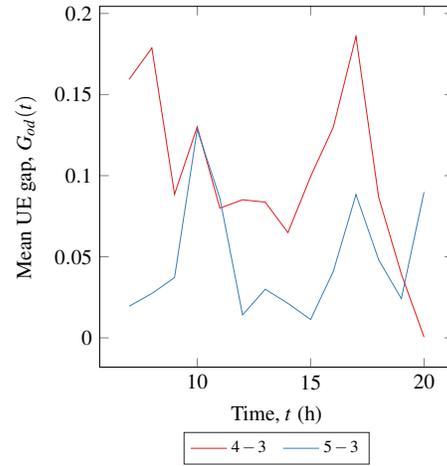
575 where TT_i^{od} is the travel time of the macro-path i , TT_{min}^{od} is the minimum travel time between the OD pair od and α_{od}^i is the path flow coefficient of the macro-path. If all the macro-paths between the OD pair are shortest paths in time, G_{od} is zero by the definition. On the other hand, if the macro-path that has longest travel time experiences higher flow, the gap is bigger and the network is far from the equilibrium. Hence, this parameter shows how far is the network from
 580 the UE conditions. The dynamic gap evolution can be computed by using the eq. (8) within each aggregation period of 60 min. The trips starting within a given aggregation period are collected for a given OD pair and the minimum travel time amongst all the macro-paths is estimated. Using this minimum travel time and path flow coefficients computed earlier, it is trivial to estimate the dynamic UE gap, $G_{od}(t)$.

585 Figure 17 presents the evolution of UE gaps for different macroscopic OD pairs. The selected OD pairs show different trends as discussed in the previous cases. Fig. 17a shows the OD pairs that involve the downtown Dallas city from the neighboring suburbans. Since the OD pair 5 – 1 has only one major macro-path, $G_{53}(t)$ is always zero. It can be observed that the OD pairs 2 – 1 and 3 – 1 are close to UE, where the observed gaps are less than 10%. However, OD pair
 590 4 – 1 shows the peaks in the morning and evening with a relatively high gap values. For the OD pairs 2 – 1 and 3 – 1, the major macro-paths are 2 → 1 and 3 → 1, respectively, where on an average more than 90% of the users choose this path. In the case of OD pair 4 – 1, the proportion of users choosing the macro-path 4 → 1 is comparatively low and more than 20% choose the longer macro-path of 4 → 3 → 1. Hence, bigger UE gaps are noticed for this particular OD pair.
 595 Fig. 17b presents the gaps for OD pairs 4 – 3 and 5 – 3, where peaks at the morning peak and the evening peak can be observed. Symmetric peak hours can be noticed for the OD pairs 2 – 3 and 3 – 2 in Fig. 17c, where OD pair 3 – 2 has the bigger gap during the morning peak hour and its counterpart has the bigger gap during the evening peak hour. This is clearly due to the difference in the direction of traffic flow during the morning and evening periods. Fig. 17d shows the mean
 600 UD gap evolution for OD pairs 3 – 5 and 5 – 2, where cyclic variations are obtained. At the same time, it is also clear that the magnitude of the gap is relatively low for the stated OD pairs, where the day averaged gap is close to 5%.

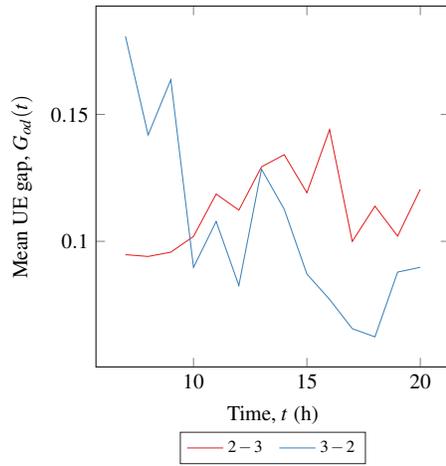
Finally, Fig. 17e presents the time evolution of relative number of ODs for a given gap condition. For instance, the red curve corresponds to the number of ODs with gap $G_{od} > 0.01$. It
 605 is clear from the plot that almost 70% of total OD pairs satisfy this condition. In other words, if the network is assumed to be in UE conditions when the gap less than or equal to 0.01, from the plot it is evident that less than 30% of ODs fulfill the condition. It is trivial that if the threshold



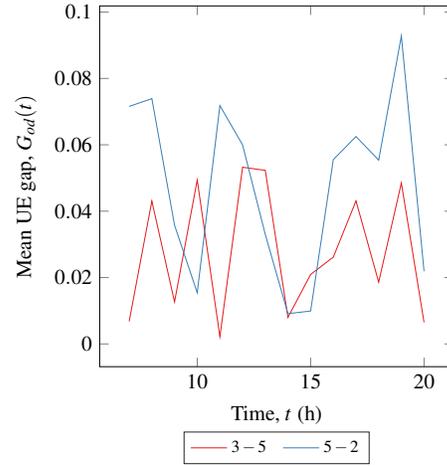
(a) OD pairs that involve downtown area, *i.e.*, reservoir 1.



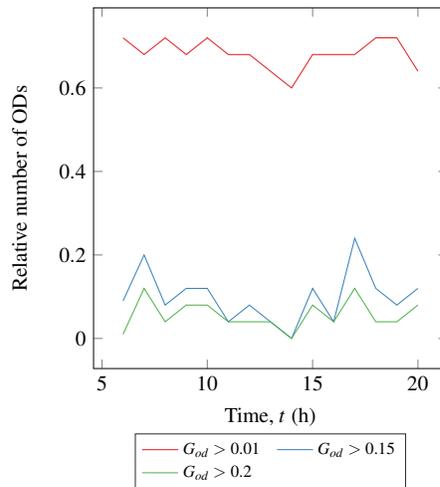
(b) OD pairs that have peak hour distributions.



(c) OD pairs that have symmetric peak hours.



(d) OD pairs with low mean gap.



(e) Relative number of ODs for a given gap condition.

Figure 17: Mean evolution of relative UE gap for different macroscopic OD pairs.

is increased, the number of ODs satisfying the UE conditions increase. It can be observed in the plot, where threshold gaps of 0.15 and 0.2 are also presented. In the cases of threshold gaps of 0.15 and 0.2, two peaks, one at morning rush hour and another at evening can be clearly noticed. This is an expected outcome as the network tends to far from the equilibrium during the peak hours. Overall, if the threshold gap of 0.15 or more is assumed, almost 90% of ODs pairs satisfy the UE conditions. However, this analysis can be refined by partitioning the network into smaller reservoirs and it is out of scope of the present work. Hence, the data can be used to extract the useful information like UE gap, which in other case is only possible to estimate by doing a DTA simulation. This type of data can also be used to validate DTA simulations and calibrate the input parameters in the simulation framework.

7. Conclusions

The present work proposes a framework to calibrate the multi-region MFD models using the mobile phone data. The methodology to select the data and segment the individual data records into representative trips is illustrated. Since LBS data is used in the current work, the frequency of data collection varies widely resulting in trips with very sparse records. A method to enrich these types of trips using map matching scheme is discussed in-detail.

Following, a simple partition of the Dallas city network is considered to estimate the macroscopic variables. It is important to state that the proposed framework is independent of partitioning scheme and can be used with any other network partition. Firstly, the error of the trip enhancement scheme is estimated using the set of high resolution trips from the *raw data*. It is concluded that from the relative errors that the map matching scheme introduces little or acceptable error in the traveled distances. The next step is to estimate the penetration rates of the data and to this extent, OD matrix data from the city council, LDD and present LBS data are fused together to obtain time dependent penetration rates. Two different types of penetration rates are proposed namely, OD specific penetration rate and origin specific penetration rate. As the names suggest, the OD specific rate takes into account the OD flow, while the origin specific rate is computed based on the flow that originates within a zone irrespective of its destination. Using the estimated time dependent penetration rates, mean density and mean flow of different reservoirs are computed to estimate the MFD for each day separately. Only days that show MFDs with similar characteristics are selected and a mean MFD is estimated. It is noticed that the MFDs for all the reservoirs are reasonably well-defined with low scatter. It is also observed that the MFDs computed from both OD specific and origin specific penetration rates are very similar. It is worth noting that the penetration rates in the present framework are not constant and they depend on the time. The following part of the work presents the analysis on the trip lengths between the macroscopic OD pairs. This type of analysis is only possible with the phone data because of the massive number of records available. The importance of considering the average trip length per macro-path instead of using mean trip length per reservoir is demonstrated. The evolution of

645 dynamic trip lengths that depends on the traffic conditions is discussed in-detail. The coefficient
of variation for selected macro-paths are presented to show the accuracy of the present methodol-
ogy. Finally, the third part discusses the evolution of the path flow distributions and UE gaps. The
evolution of mean path flow coefficient for different OD pairs are illustrated. It is noticed that the
path flow coefficient in few OD pairs exhibit a strong variation during the morning and evening
650 peak hours. Similarly, the evolution of UE gap is estimated based on the travel time information.
This gap parameter can be used to observe how far is the network from the UE and as well, it can
be used to validate the DTA simulations. It is also noticed that UE gap varies with time during a
typical day scenario and this variation is explained for few OD pairs. Finally, the proportion of
OD pairs that satisfy the given threshold UE gap condition is presented. When a higher threshold
655 gap is chosen (> 0.15), it is noticed that most of the OD pairs satisfy the UE condition. However,
it is also observed that during the peak hours the number of OD pairs satisfy the condition is less
than the off-peak hours suggesting that the network is relatively far from UE conditions during
peak hours.

The proposed framework is very generic and network independent. It can be applied to any
660 network that has sufficient phone data and all the parameters necessary to do a MFD simulation
can be calibrated. In the case of absence of OD matrix data, it can be replaced by the census
data of the network to estimate the penetration rates. Overall, the present framework can estimate
lot of interesting and useful parameters that can be used to perform MFD simulations as well to
validate them. Most of the analysis presented in the current framework cannot performed using
665 other types of data sources and it is to be re-iterated that the trip lengths and path flow distributions
analysis can only be achieved because of massive phone data.

Acknowledgment

M. Paipuri and L. Leclercq acknowledge the funding from the European Research Council
(ERC) under the European Union's Horizon 2020 research and innovation program (grant agree-
670 ment No 646592 – MAGnUM project). Y. Xu and M. González acknowledge the support by the
MIT Energy Initiative.

References

Alonso, B., Ibeas, A., Musolino, G., Rindone, C., Vitetta, A., 2019. Effects of traf-
fic control regulation on network macroscopic fundamental diagram: A statistical analy-
675 sis of real data. *Transportation Research Part A: Policy and Practice* 126, 136 – 151.
URL: <http://www.sciencedirect.com/science/article/pii/S0965856418303665>,
doi:<https://doi.org/10.1016/j.tra.2019.05.012>.

- Ambühl, L., Loder, A., Menendez, M., Axhausen, K.W., 2017. Empirical Macroscopic Fundamental Diagrams: New insights from loop detector and floating car data, in: Transportation Research Board 96th Annual Meeting Transportation Research Board, Washington.
- Ambühl, L., Loder, A., Zheng, N., Axhausen, K.W., Menendez, M., 2019. Approximative network partitioning for MFDs from stationary sensor data. Transportation Research Record URL: <https://doi.org/10.1177/0361198119843264>, doi:10.1177/0361198119843264, arXiv:<https://doi.org/10.1177/0361198119843264>. (In press).
- Ambühl, L., Menendez, M., 2016. Data fusion algorithm for Macroscopic Fundamental Diagram estimation. Transportation Research Part C: Emerging Technologies 71, 184 – 197. URL: <http://www.sciencedirect.com/science/article/pii/S0968090X16301267>, doi:<https://doi.org/10.1016/j.trc.2016.07.013>.
- Ampountolas, K., Kouvelas, A., 2015. Real-time estimation of critical values of the macroscopic fundamental diagram for maximum network throughput, in: Transportation Research Board 94th Annual Meeting Transportation Research Board, Washington.
- Ampountolas, K., Zheng, N., Geroliminis, N., 2017. Macroscopic modelling and robust control of bi-modal multi-region urban road networks. Transportation Research Part B: Methodological 104, 616 – 637. URL: <http://www.sciencedirect.com/science/article/pii/S0191261515300370>, doi:<https://doi.org/10.1016/j.trb.2017.05.007>.
- Batista, S., Leclercq, L., Geroliminis, N., 2019. Estimation of regional trip length distributions for the calibration of the aggregated network traffic models. Transportation Research Part B: Methodological 122, 192 – 217. URL: <http://www.sciencedirect.com/science/article/pii/S0191261518311603>, doi:<https://doi.org/10.1016/j.trb.2019.02.009>.
- Batista, S.F.A., Leclercq, L., Menendez, M., 2020. Regional dynamic traffic assignment framework with time-dependent trip lengths, in: Transportation Research Board 99th Annual Meeting Transportation Research Board, Washington.
- Bazzani, A., Giorgini, B., Gallotti, R., Giovannini, L., Marchioni, M., Rambaldi, S., 2011. Towards congestion detection in transportation networks using GPS data, in: 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing, pp. 1455–1459. doi:10.1109/PASSAT/SocialCom.2011.249.
- Beibei, J.Y., van Zuylen, H.J., Shoufeng, L., 2016. Determining the Macroscopic Fundamental Diagram on the basis of mixed and incomplete traffic data, in: Transportation Research Board 95th Annual Meeting Transportation Research Board, Washington.

- Boeing, G., 2017. OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems* 65, 126 – 139. URL: <http://www.sciencedirect.com/science/article/pii/S0198971516303970>, doi:<https://doi.org/10.1016/j.compenvurbsys.2017.05.004>.
715
- Buisson, C., Ladier, C., 2009. Exploring the impact of homogeneity of traffic measurements on the existence of Macroscopic Fundamental Diagrams. *Transportation Research Record: Journal of the Transportation Research Board* 2124, 127–136. URL: <https://doi.org/10.3141/2124-12>, doi:[10.3141/2124-12](https://doi.org/10.3141/2124-12), arXiv:<https://doi.org/10.3141/2124-12>.
- 720 Cao, J., Menendez, M., 2015. System dynamics of urban traffic based on its parking-related-states. *Transportation Research Part B: Methodological* 81, 718 – 736. URL: <http://www.sciencedirect.com/science/article/pii/S0191261515001654>, doi:<https://doi.org/10.1016/j.trb.2015.07.018>. iSTTT 21 for the year 2015.
- Courbon, T., Leclercq, L., 2011. Cross-comparison of Macroscopic Fundamental Diagram estimation methods. *Procedia - Social and Behavioral Sciences* 20, 417 – 426. URL: <http://www.sciencedirect.com/science/article/pii/S1877042811014285>, doi:<https://doi.org/10.1016/j.sbspro.2011.08.048>.
725
- Daganzo, C.F., 2007. Urban gridlock: Macroscopic modeling and mitigation approaches. *Transportation Research Part B: Methodological* 41, 49 – 62. URL: <http://www.sciencedirect.com/science/article/pii/S0191261506000282>, doi:<https://doi.org/10.1016/j.trb.2006.03.001>.
730
- Du, J., Rakha, H., Gayah, V.V., 2016. Deriving Macroscopic Fundamental Diagrams from probe data: Issues and proposed solutions. *Transportation Research Part C: Emerging Technologies* 66, 136 – 149. URL: <http://www.sciencedirect.com/science/article/pii/S0968090X15003162>, doi:<https://doi.org/10.1016/j.trc.2015.08.015>. advanced Network Traffic Management: From dynamic state estimation to traffic control.
735
- Du, J., Rakha, H.A., 2019. Constructing a network fundamental diagram: Synthetic origin–destination approach. *Transportation Research Record* URL: <https://doi.org/10.1177/0361198119851445>, doi:[10.1177/0361198119851445](https://doi.org/10.1177/0361198119851445), arXiv:<https://doi.org/10.1177/0361198119851445>. (In press).
740
- Edie, L.C., 1963. Discussion of traffic stream measurements and definitions, in: *The 2nd International Symposium on Theory of Traffic flow*, London.
- Geroliminis, N., Daganzo, C.F., 2008. Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transportation Research Part B: Methodolog-*
745

ical 42, 759 – 770. URL: <http://www.sciencedirect.com/science/article/pii/S0191261508000180>, doi:<https://doi.org/10.1016/j.trb.2008.02.002>.

Geroliminis, N., Sun, J., 2011. Properties of a well-defined macroscopic fundamental diagram for urban traffic. *Transportation Research Part B: Methodological* 45, 605 – 617.
750 URL: <http://www.sciencedirect.com/science/article/pii/S0191261510001372>, doi:<https://doi.org/10.1016/j.trb.2010.11.004>.

Gu, Z., Shafiei, S., Liu, Z., Saberi, M., 2018. Optimal distance- and time-dependent area-based pricing with the Network Fundamental Diagram. *Transportation Research Part C: Emerging Technologies* 95, 1 – 28. URL: <http://www.sciencedirect.com/science/article/pii/S0968090X18300573>, doi:<https://doi.org/10.1016/j.trc.2018.07.004>.
755

Haddad, J., Mirkin, B., 2017. Coordinated distributed adaptive perimeter control for large-scale urban road networks. *Transportation Research Part C: Emerging Technologies* 77, 495 – 515. URL: <http://www.sciencedirect.com/science/article/pii/S0968090X16302509>, doi:<https://doi.org/10.1016/j.trc.2016.12.002>.

760 Hagberg, A., Swart, P., S Chult, D., 2008. Exploring network structure, dynamics, and function using NetworkX.

Herrera, J.C., Work, D.B., Herring, R., Ban, X.J., Jacobson, Q., Bayen, A.M., 2010. Evaluation of traffic data obtained via gps-enabled mobile phones: The mobile century field experiment. *Transportation Research Part C: Emerging Technologies* 18, 568 – 583.
765 URL: <http://www.sciencedirect.com/science/article/pii/S0968090X09001430>, doi:<https://doi.org/10.1016/j.trc.2009.10.006>.

Hsieh, H.P., Li, C.T., L., S., 2015. Measuring and recommending time-sensitive routes from location-based data, in: Wooldridge, M., Yang, Q. (Eds.), *IJCAI 2015 - Proceedings of the 24th International Joint Conference on Artificial Intelligence, International Joint Conferences on Artificial Intelligence*. pp. 4193–4196.
770

Huang, C., Zheng, N., Zhang, J., 2019. Investigation on the bi-modal macroscopic fundamental diagrams in large-scale urban networks - an empirical study of Shenzhen city, in: *Transportation Research Board 98rd Annual Meeting Transportation Research Board, Washington*.

Ji, Y., Geroliminis, N., 2012. On the spatial partitioning of urban transportation networks. *Transportation Research Part B: Methodological* 46, 1639 – 1656. URL: <http://www.sciencedirect.com/science/article/pii/S0191261512001099>, doi:<https://doi.org/10.1016/j.trb.2012.08.005>.
775

Jie, L., van Zuylen, H., Chunhua, L., Shoufeng, L., 2011. Monitoring travel times in an urban network using video, GPS and bluetooth. *Procedia - Social and Behavioral Sciences* 20, 630 – 637.

- 780 URL: <http://www.sciencedirect.com/science/article/pii/S1877042811014509>,
doi:<https://doi.org/10.1016/j.sbspro.2011.08.070>. the State of the Art in the Euro-
pean Quantitative Oriented Transportation and Logistics Research – 14th Euro Working Group
on Transportation & 26th Mini Euro Conference & 1st European Scientific Conference on Air
Transport.
- 785 Jin, P.J., Cebelak, M., Yang, F., Zhang, J., Walton, C.M., Ran, B., 2014. Location-
based social networking data: Exploration into use of doubly constrained gravity
model for Origin–Destination estimation. *Transportation Research Record* 2430,
72–82. URL: <https://doi.org/10.3141/2430-08>, doi:10.3141/2430-08,
[arXiv:https://doi.org/10.3141/2430-08](https://doi.org/10.3141/2430-08).
- 790 Kavianipour, M., Saedi, R., Zockaie, A., Saberi, M., 2019. Traffic state estima-
tion in heterogeneous networks with stochastic demand and supply: Mixed la-
grangian–eulerian approach. *Transportation Research Record* 0, 0361198119850472. URL:
<https://doi.org/10.1177/0361198119850472>, doi:10.1177/0361198119850472,
[arXiv:https://doi.org/10.1177/0361198119850472](https://doi.org/10.1177/0361198119850472). (In press).
- 795 Keyvan-Ekbatani, M., Kouvelas, A., Papamichail, I., Papageorgiou, M., 2012. Exploiting the fun-
damental diagram of urban networks for feedback-based gating. *Transportation Research Part*
B: Methodological 46, 1393 – 1403. URL: [http://www.sciencedirect.com/science/
article/pii/S0191261512000926](http://www.sciencedirect.com/science/article/pii/S0191261512000926), doi:[https://doi.org/10.1016/j.trb.2012.06.
008](https://doi.org/10.1016/j.trb.2012.06.008).
- 800 Knoop, V.L., Hoogendoorn, S.P., 2014. Network transmission model: a dynamic traffic model
at network level, in: *Transportation Research Board 93rd Annual Meeting Transportation Re-
search Board*, Washington.
- Knoop, V.L., van Erp, P.B.C., Leclercq, L., Hoogendoorn, S.P., 2018. Empirical MFDs using
google traffic data, in: *2018 21st International Conference on Intelligent Transportation Sys-
805 tems (ITSC)*, pp. 3832–3839. doi:10.1109/ITSC.2018.8570005.
- Kouvelas, A., Saeedmanesh, M., Geroliminis, N., 2017. Enhancing model-based feedback
perimeter control with data-driven online adaptive optimization. *Transportation Research*
Part B: Methodological 96, 26 – 45. URL: [http://www.sciencedirect.com/science/
article/pii/S019126151630710X](http://www.sciencedirect.com/science/article/pii/S019126151630710X), doi:[https://doi.org/10.1016/j.trb.2016.10.
810 011](https://doi.org/10.1016/j.trb.2016.10.011).
- Leclercq, L., Chiabaut, N., Trinquier, B., 2014. Macroscopic Fundamental Diagrams: A cross-
comparison of estimation methods. *Transportation Research Part B: Methodological* 62, 1 – 12.
URL: <http://www.sciencedirect.com/science/article/pii/S0191261514000174>,
doi:<https://doi.org/10.1016/j.trb.2014.01.007>.

- 815 Leclercq, L., Paipuri, M., 2019. Macroscopic traffic dynamics under fast-varying demand. *Transportation Science* 53, 1526–1545. URL: <https://doi.org/10.1287/trsc.2019.0908>, doi:10.1287/trsc.2019.0908, arXiv:<https://doi.org/10.1287/trsc.2019.0908>.
- Leclercq, L., Sénécat, A., Mariotte, G., 2017. Dynamic macroscopic simulation of on-street parking search: A trip-based approach. *Transportation Research Part B: Methodological* 101, 268 – 282. URL: <http://www.sciencedirect.com/science/article/pii/S0191261516309717>, doi:<https://doi.org/10.1016/j.trb.2017.04.004>.
- 820 Mahmassani, H., Williams, J., Herman, R., 1984. Investigation of network-level traffic flow relationships: Some simulation results. *Transportation Research Record* , 121–130.
- Mariotte, G., Leclercq, L., Batista, S., Krug, J., Paipuri, M., 2020. Calibration and validation of multi-reservoir mfd models: A case study in Lyon. *Transportation Research Part B: Methodological* (Under Review).
- 825 Mohajerpoor, R., Saberi, M., Vu, H.L., Garoni, T.M., Ramezani, M., 2019. H_∞ robust perimeter flow control in urban networks with partial information feedback. *Transportation Research Part B: Methodological* URL: <http://www.sciencedirect.com/science/article/pii/S0191261518308609>, doi:<https://doi.org/10.1016/j.trb.2019.03.010>. (In press).
- 830 Nagle, A.S., Gayah, V.V., 2014. Accuracy of networkwide traffic states estimated from mobile probe data. *Transportation Research Record* 2421, 1–11. URL: <https://doi.org/10.3141/2421-01>, doi:10.3141/2421-01, arXiv:<https://doi.org/10.3141/2421-01>.
- NCTCOG, . The North Central Texas Council of Governments. <https://www.nctcog.org/>.
835 Accessed: 2018-09.
- Saeedmanesh, M., Geroliminis, N., 2016. Clustering of heterogeneous networks with directional flows based on “Snake” similarities. *Transportation Research Part B: Methodological* 91, 250 – 269. URL: <http://www.sciencedirect.com/science/article/pii/S0191261515302605>, doi:<https://doi.org/10.1016/j.trb.2016.05.008>.
- 840 Sbayti, H., Lu, C.C., Mahmassani, H.S., 2007. Efficient implementation of method of successive averages in simulation-based dynamic traffic assignment models for large-scale network applications. *Transportation Research Record* 2029, 22–30. URL: <https://doi.org/10.3141/2029-03>, doi:10.3141/2029-03, arXiv:<https://doi.org/10.3141/2029-03>.
- 845 Shim, J., Yeo, J., Lee, S., Hamdar, S.H., Jang, K., 2019. Empirical evaluation of influential factors on bifurcation in macroscopic fundamental diagrams. *Transportation Research Part C: Emerging Technologies* 102, 509 – 520. URL: <http://www.sciencedirect.com/science/article/pii/S0968090X18304042>, doi:<https://doi.org/10.1016/j.trc.2019.03.005>.

- Shoufeng, L., Jie, W., van Zuylen, H., Ximin, L., 2013. Deriving the macroscopic fundamental
850 diagram for an urban area using counted flows and taxi GPS, in: 16th International IEEE
Conference on Intelligent Transportation Systems (ITSC 2013), pp. 184–188. doi:[10.1109/
ITSC.2013.6728231](https://doi.org/10.1109/ITSC.2013.6728231).
- Tsubota, T., Bhaskar, A., Chung, E., 2014. Macroscopic Fundamental Diagram for Brisbane,
Australia: Empirical findings on network partitioning and incident detection. Transportation
855 Research Record 2421, 12–21. URL: <https://doi.org/10.3141/2421-02>, doi:[10.3141/
2421-02](https://doi.org/10.3141/2421-02), arXiv:<https://doi.org/10.3141/2421-02>.
- Wang, P., Wada, K., Akamatsu, T., Hara, Y., 2015. An empirical analysis of macroscopic funda-
mental diagrams for sendai road networks. Interdisciplinary Information Sciences 21, 49–61.
doi:[10.4036/iis.2015.49](https://doi.org/10.4036/iis.2015.49).
- 860 Xu, Y., Clemente, R.D., Gonzalez, M.C., 2019. Understanding route choice behavior with
location-based services data (In preparation).
- Yildirimoglu, M., Geroliminis, N., 2014. Approximating dynamic equilibrium conditions
with macroscopic fundamental diagrams. Transportation Research Part B: Methodologi-
cal 70, 186 – 200. URL: [http://www.sciencedirect.com/science/article/pii/
865 S0191261514001568](http://www.sciencedirect.com/science/article/pii/S0191261514001568), doi:<https://doi.org/10.1016/j.trb.2014.09.002>.
- Yildirimoglu, M., Ramezani, M., Geroliminis, N., 2015. Equilibrium analysis and route guidance
in large-scale networks with MFD dynamics. Transportation Research Part C: Emerging Tech-
nologies 59, 404 – 420. URL: [http://www.sciencedirect.com/science/article/pii/
S0968090X15001813](http://www.sciencedirect.com/science/article/pii/S0968090X15001813), doi:<https://doi.org/10.1016/j.trc.2015.05.009>. special Is-
870 sue on International Symposium on Transportation and Traffic Theory.