Empirical Macroscopic Fundamental Diagrams

New insights from loop detector and floating car data

Empirical Macroscopic Fundamental Diagrams: New Insights from Loop Detector and Floating Car Data

1. August 2016

Word Count: 5’703 words + 4 Figures + 3 Tables = 7’453 words

Corresponding Author:
Lukas Ambühl
IVT, ETH Zürich
CH-8093 Zürich
Phone: +41 44 633 32 51
lukas.ambuehl@ivt.baug.ethz.ch

Allister Loder
IVT, ETH Zürich
CH-8093 Zürich
Phone: +41 44 633 62 58
allister.loder@ivt.baug.ethz.ch

Monica Menendez
IVT, ETH Zürich
CH-8093 Zürich
Phone: +41 44 633 66 95
menendez@ivt.baug.ethz.ch

Kay W. Axhausen
IVT, ETH Zürich
CH-8093 Zürich
Phone: +41 44 633 39 43
axhausen@ivt.baug.ethz.ch

ABSTRACT

The macroscopic fundamental diagram, relating average flows and densities in an urban network, has been analyzed in some empirical studies and many simulations. It has been shown to be an efficient tool for traffic management and control or the estimation of travel times in a network. However, empirical studies remain scarce and are usually based on one single data source, such as loop detector data (LDD) or floating car data (FCD).

In this paper, we analyze an extensive data set based on both, LDD and FCD for the city of Zurich. We show that each source exhibits a well-defined and reproducible MFD. However, they differ from each other, due to limitations of the data sources. We identify a placement bias, and a link selection bias for LDD, which leads to an overestimation of occupancy or density values, respectively. In order to mitigate such biases we develop a methodology accounting for the relative position of a loop detector on links and their frequency at that position. Moreover, we investigate and validate common practices when transforming LDD occupancy and FCD flows, which are the space effective mean length of a vehicle and the probe penetration rate, respectively. We also apply a combination of LDD flows and FCD speeds to estimate the MFD, which partly eliminates key drawbacks of both data sources.

INTRODUCTION

The relationship between the accumulation of vehicles and their impact on speeds in urban networks raises the question of optimal congestion levels *(1, 2)*. In the end, urban congestion levels are key determinants of a city's productivity in terms of its transportation system *(3, 4)*. First advances in understanding urban congestion were made by Mahmassani et al. *(5)* based on simulations. They found that the macroscopic relations between traffic variables appear to behave in a similar manner as their link level counterparts. Empirical evidence for these macroscopic relations was absent for almost twenty years until Geroliminis and Daganzo estimated a macroscopic fundamental diagram (MFD) for Yokohama, Japan *(6)*. Subsequent findings were mostly based on simulations *(7)*, as empirical data remains scarce (see Table 1). However, for application purposes, such as traffic control, we need to better understand how cities can estimate their MFDs *(8–10)* from empirical data.

Table 1 Empirical studies on urban MFD estimation

City	Year	Data	Sample	Filter	Source
Yokohama, Japan	2008	LDD+FCD	500+140	Occupied taxis	(6)
Toulouse, France	2009	LDD	1000	Distance to signal	*(13)*
Rome, Italy	2011	FCD	N/A		*(23)*
Brisbane, Australia	2013	FCD	301		*(21)*
Shenzhen, China	2014	FCD	20000	Occupied taxis	*(31)*
Sendai, China	2015	LDD	1756		*(32)*
Chania, Greece	2015	LDD	70		*(33)*
Changsha, China	2016	LDD+FCD	N/A+6200	Taxis	*(27)*

LDD: Loop detector data; FCD: Floating car data

The existence of the MFD was originally based on the key assumptions that congestion spreads homogeneously across the network and that it is independent of demand patterns as long as average traveled distance remains unchanged. However, various findings challenge these assumptions. Urban networks might not be homogenously congested. Thus, efforts were made to partition networks according to the homogeneity of congestion, e.g. *(11)*. Moreover recent studies

show that the MFD is not invariant to changes in the origin-destination matrix *(12)*. In light of such limitations, the question arises how a well-defined and reproducible MFD can be estimated from available data.

There are typically two empirical data sources considered as viable for the estimation of the MFD: loop detector data (LDD) and floating car data (FCD).

Loop detectors are installed for traffic control and congestion monitoring. They typically report the traffic variables *flow* (i.e. number of vehicles passing a detector), and *occupancy* (i.e. share of time a detector is occupied). Loop detectors are mainly used for counting vehicles, detecting congestion and controlling traffic signals. They have been used to estimate the MFD empirically and through simulation e.g. *(6, 13–15)*. An important issue to consider is that their distance to the downstream traffic signal influences the shape of the MFD significantly *(13)*, but the only correction method proposed so far is more appropriate for corridors *(15)*. The network coverage and the spatial distribution of the LDD are critical for the estimation accuracy *(14, 16, 17)*. Moreover, the assumptions made to convert occupancy to density have not been validated and might underestimate the complexity of such conversion *(18–20)*.

FCD is collected from probe vehicles transmitting the data through a trajectory measurement device, today GPS *(6)*, or cellphones *(21)*. FCD requires a matching of the GPS trajectories to the road network. This comes with uncertainties and does not allow to match a measurements to a lane but only to road segment *(22)*. FCD has been used to estimate the MFD empirically and through simulation *(e.g., 18, 24, 25)*. Important issues to consider here are the probe penetration rate (i.e. the relative number of vehicles sending FCD), and its spatial distribution. The knowledge about both factors is crucial for the estimation accuracy *(25, 26)*.

In the literature, almost all MFDs are based on either one *or* the other source. A few studies cover both sources, some use loop detectors to estimate the probe penetration rate (ppr) of FCD *(27, 25, 26)*, and others aim at comparing, combining, or fusing *both* data sources in order to estimate a more accurate MFD *(16, 17)*. However, the latter efforts have been limited to simulations *(15)*.

In this paper, we investigate the differences between both data sources based on an extensive empirical dataset from the City of Zurich. We also apply an approach formulated by Leclercq et al. *(15)* that combines the two sources, and compare the results with those obtained from either data source used individually. More importantly, in order to construct the MFDs, we identify the limitations that arise in practice for each data source, and also validate the common practice in determining the probe penetration rate and the conversion of occupancy into density.

The remainder of this paper is organized as follows: We first present MFDs from both data sources. Subsequently, we compare both data sets by using appropriate parameters, and later combine them. The combined MFD is then used to validate the required parameters. Based on the results, we present the key findings of both datasets for the City of Zurich. Lastly, we discuss the appropriateness of each dataset to estimate a reproducible and well-defined MFD.

DATA

The city of Zurich, Switzerland, stretches across an area of 91.9km^2 with a population of around 400'000 inhabitants. The road network excluding motorways is 740km long. The traffic management system of Zurich operates 4852 traffic detectors at 384 intersections *(28)*. They detect either public transport vehicles, private motorized vehicles or a combination thereof. Their purpose is mainly to give priority to public transport, support traffic signal control algorithms, and identify congestion. For the analysis we concentrate on loop detectors that measure private motorized

Table 2 Overview of the data sets for Zurich

LDD	26/10/2015 to 01/11/2015 (Monday to Sunday), 3min intervals
	variables recorded: flow (number of vehicles passing a detector), occupancy in percentage (share of time vehicles occupy the detector)
	link attributes: GPS coordinates, lane type, distance to downstream traffic signal, link length, road class
FCD5	2014-2015, 5min intervals for an average week
	variables recorded: 2-year average speeds per segment, hits
	segment attributes: GPS coordinates, road class, segment length
FCD15	26/10/2015 to 01/11/2015 (Monday to Sunday), 15min intervals
	variables recorded: average speed per segment, hits
	segment attributes: GPS coordinates, road class, segment length

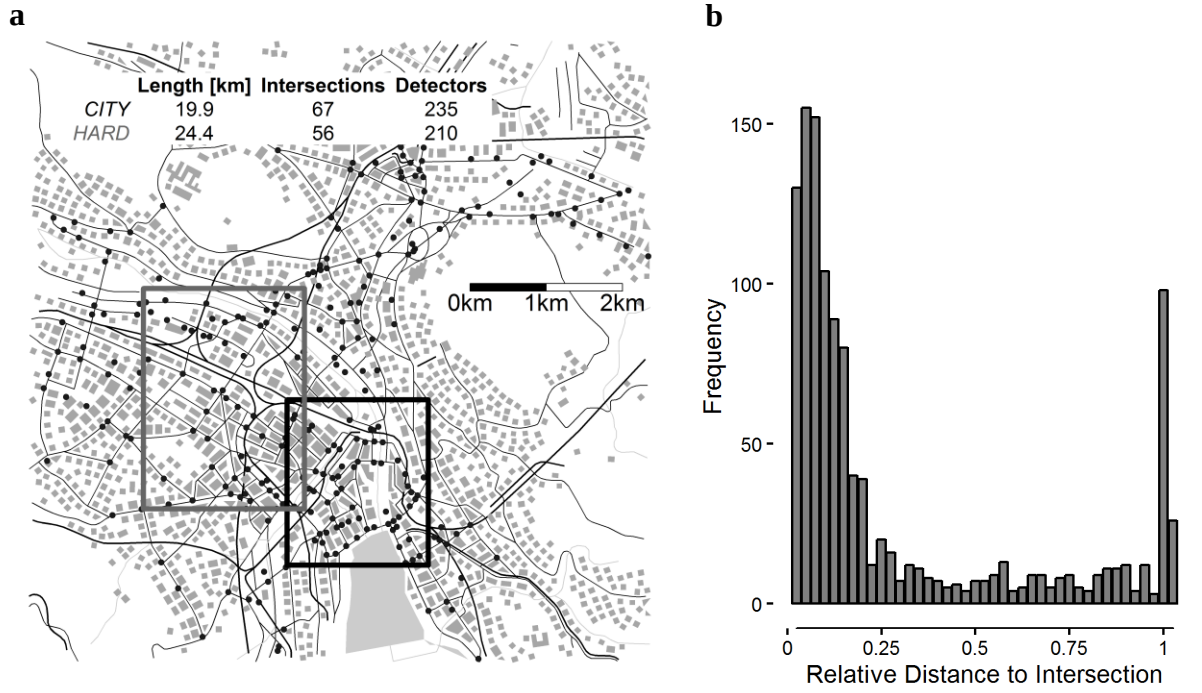

FIGURE 1 LDD (a) Case study network, intersections equipped with loop detectors, and the two study areas used in this paper; (b) Distribution of loop detectors position on links with respect to downstream traffic signal

transport. We have geo-coded all loop detectors and matched them to the corresponding links on the road network. We identified and removed 3.9% of all detectors due to defective measurements. Table 2 provides an overview of the variables and the time period recorded by LDD.

In Figure 1a, all intersections equipped with loop detectors are marked by black dots. Figure 1b shows the distribution of detectors across all city links by their relative distance to the downstream traffic signal (i.e. 0 means the detector is at the stop line of the downstream traffic signal). As most detectors are used for traffic signal control, their average location is rather close to the traffic signal.

As previously stated, FCD measurements can be recorded from navigation devices, smartphones, and fleet management systems, and can be matched to the road network map with an accuracy of 10m *(22)*. Table 2 provides an overview of the FCD used in the following sections *(22)*. Hits are the number of vehicles contributing to the average speed. In FCD15, the average speed for segment *i* on Monday between 8:00 to 8:15 is calculated from all probe vehicles passing

the segment in this exact time interval. On the other hand, in FCD5, the average speed for segment *i* between 8:05 and 8:10 is averaged over all Mondays in 2014 and 2015.

The mean length of an FCD segment is 59m, whereas the average length of a link with loop detectors is 220m. FCD segments are merged with the LDD links, based on a joint geo-reference system for both data sources.

For the MFD estimations in the next sections we focus on two specific regions within the overall network, "City" (4.3km^2) and "Hard" (6.4km^2). Both regions have a similar number of loop detectors and network length. Both are downtown-like with one important difference, region "City" has an adaptive traffic congestion management system (see *(14)* for details). Figure 1a summarizes the respective network length and number of detectors. We removed all data measured on motorways, their ramps, and on all local roads from the samples. Latter are excluded since they usually serve residential areas where traffic-calming (e.g. dead ends, etc.) measures were undertaken.

Note, for the region "C*ity"* construction work around *Bellevue* was finished shortly prior the LDD and FCD15 recording period. As a result, some of the most relevant arterial re-routings were only lifted during of the observation period *(28)*.

SINGLE SOURCE MFD

In the following, we estimate the MFD for the region "City". For clarity, we decided to plot only data for Mondays. We investigated the scatter plots for Tuesday to Friday as well and observed only marginal differences in the uncongested branch of the MFD and more scatter around the critical density.

Loop Detector Data

In this section we introduce different filters based on the placement of the loop detectors, and propose an approach on how to overcome the resulting biases.

As a base and in accordance with Geroliminis and Daganzo *(6)*, we calculate network flow, $\bar{q}_{LDD}$, and network occupancy, $\bar{o}_{LDD}$, including all loop detectors, as follows

$$\bar{q}_{LDD} = \frac{\sum q_i \, l_i}{\sum l_i} \qquad Eq\ 1$$

$$\bar{o}_{LDD} = \frac{\sum o_i \, l_i}{\sum l_i} \qquad Eq\ 2$$

where l_i stands for the length of link *i*. Hereafter this method is referred to as "base".

Since most loop detectors are used for traffic signal control, a subsample of detectors is located on turning pockets. Figure 2a shows the effect of excluding turns – the maximum flow is increased by 15%. This makes sense as turning lanes have in average less green time than straight-ahead lanes.

Buisson and Ladier show that the placement of loop detectors affects the shape of the MFD significantly *(13)*. We confirm these findings by restricting the position of loop detectors to more than 20m upstream of a traffic signal in Figure 2b. Loop detectors right in front of a signal register much higher occupancies than loops further upstream (x>20; x being the distance between loop detector and traffic signal). However, this high occupancy might only represent the periodic queue over the loop detector. Excluding such loop detectors will thus result in lower average occupancies across the network.

These results show a clear drawback of LDD. Loop detectors are representative of their exact location. However, due to traffic dynamics on a link, they cannot reproduce correctly the average occupancy for the entire link, which is actually necessary to accurately estimate the MFD *(17)*. The underlying assumption of representative occupancies (or densities) of the base method applied earlier is violated. Moreover, filtering data, as previously employed, might exclude valuable information. Therefore, we propose an approach that includes all loop detectors, but takes into account their relative position and their frequency.

By *projecting loop detectors on a virtual link*, we try to incorporate findings by Courbon et al. *(16)*. Their study shows that if distances to downstream traffic signals are uniformly distributed across the network, an LDD MFD is accurate. Thus, we propose to first, project the network onto a single virtual link of unity length including all loop detectors of the network at their relative positions. Then, we average the weighted values for evenly distributed link segments. In other words, all loop detectors are put on a virtual link based on their relative position, then we

split the virtual link in *J* segments, calculate the weighted flow and occupancy of all LDD in each segment, and take the average over all segments.

Evenly splitting the virtual link into *J* segments emulates a network where loop detectors are uniformly distributed. We tested for different values of J. We chose *J*=20, as this value ensures at least one loop detector in each segment. As seen, a majority of loop detectors are located in front of a traffic signal and overestimate the density for the whole link. It makes sense that average occupancy in Figure 2c is lower using our approach compared to the base method. Flow values measured on a road without any side entries are less susceptible to the location of the loop detector. However, in reality, roads in the network of Zurich are complex with frequent driveways and side entries. Thus the flow value is affected by the loop detector position as well and it makes sense to follow the aforementioned approach. In short:

$$\hat{q}_{LDD} = \frac{1}{J}\sum_{j=1}^{J}\sum_{i\in N_j}\frac{q_i\, l_i}{\sum l_i} \qquad \textit{Eq 3}$$

$$\hat{o}_{LDD} = \frac{1}{J}\sum_{j=1}^{J}\sum_{i\in N_j}\frac{o_i\, l_i}{\sum l_i} \qquad \textit{Eq 4}$$

$$\text{with } N = \bigcup_{j\in J} N_j \text{ and } N_j = \{i \in N \mid \frac{j-1}{J} < \frac{x_i}{l_i} < \frac{j}{J}\}$$

Floating Car Data

FCD5 provides 2-year daily averages of hits and speeds during 5min intervals (see section “Data”). In our case, only a fraction of vehicles is equipped with FCD generating devices. We show later in detail that the ppr can be estimated by a combination of both data sources and amounts to roughly 4% during peak hours. For now, we analyze the macroscopic relations with the following equations, not accounting for the ppr:

$$\hat{q}_{FCD} = \frac{\sum \max(H_i\, l_i\, , v_i\, T)}{T\sum l_i} \qquad \textit{Eq 5}$$

$$\hat{v}_{FCD} = \frac{\sum \bar{v}_i l_i}{\sum l_i} \qquad Eq\ 6$$

where $\hat{q}_{FCD}$ and $\hat{v}_{FCD}$ are network flow and network speed, respectively. H_i is the average number of probe vehicles during observed time *T* on link *i* with length l_i. $\bar{v}_i$ is the average speed of these vehicles. Note that the average speed was calculated by using $\bar{v}_i = \frac{\sum v_{p,i}}{H_i}$, where $v_{p,i}$ is the speed of a FCD probe on link *i*. Thus, it constitutes an upper bound to the FCD space-mean speed. Figure 2d shows the estimated MFD. Its density is calculated by dividing MFD flow by MFD speed. Due to the sample size, flow and density values are low. Nevertheless a low-scatter MFD is apparent. Interestingly, the MFD does not show strong indications of congestion. Still, during peak hour, we observe on certain links speeds of around 5km/h, confirming findings in *(29)*. This is the result of an inhomogeneous spread of congestion in the city of Zurich. While certain links are congested and show very low speeds, others remain uncongested and in free flow condition. A short analysis on the variance of the speeds confirms these findings. Although not shown here for brevity, results from such analysis reveal that the speed variance is 85 km^2/h^2 during peak hour 5-min intervals. This poses the question, whether it makes sense to further partition these areas as per *(11)*. Notice, however, that as of now the areas are relatively small, thus an additional partition could lead to very local results, defeating the idea of a macroscopic perspective. This dilemma is not unique to Zurich, as ideal homogeneous settings can hardly ever be expected in reality.

When highlighting the different peak periods, we observe a slight bifurcation, indicating a difference in congestion during morning and evening peak. However, since ppr is disregarded, we must ignore this phenomenon for the time being. Such a bifurcation is not necessarily present in real conditions. This shows that lack of knowledge of the probe penetration and its temporal or spatial distribution are key drawbacks of the FCD.

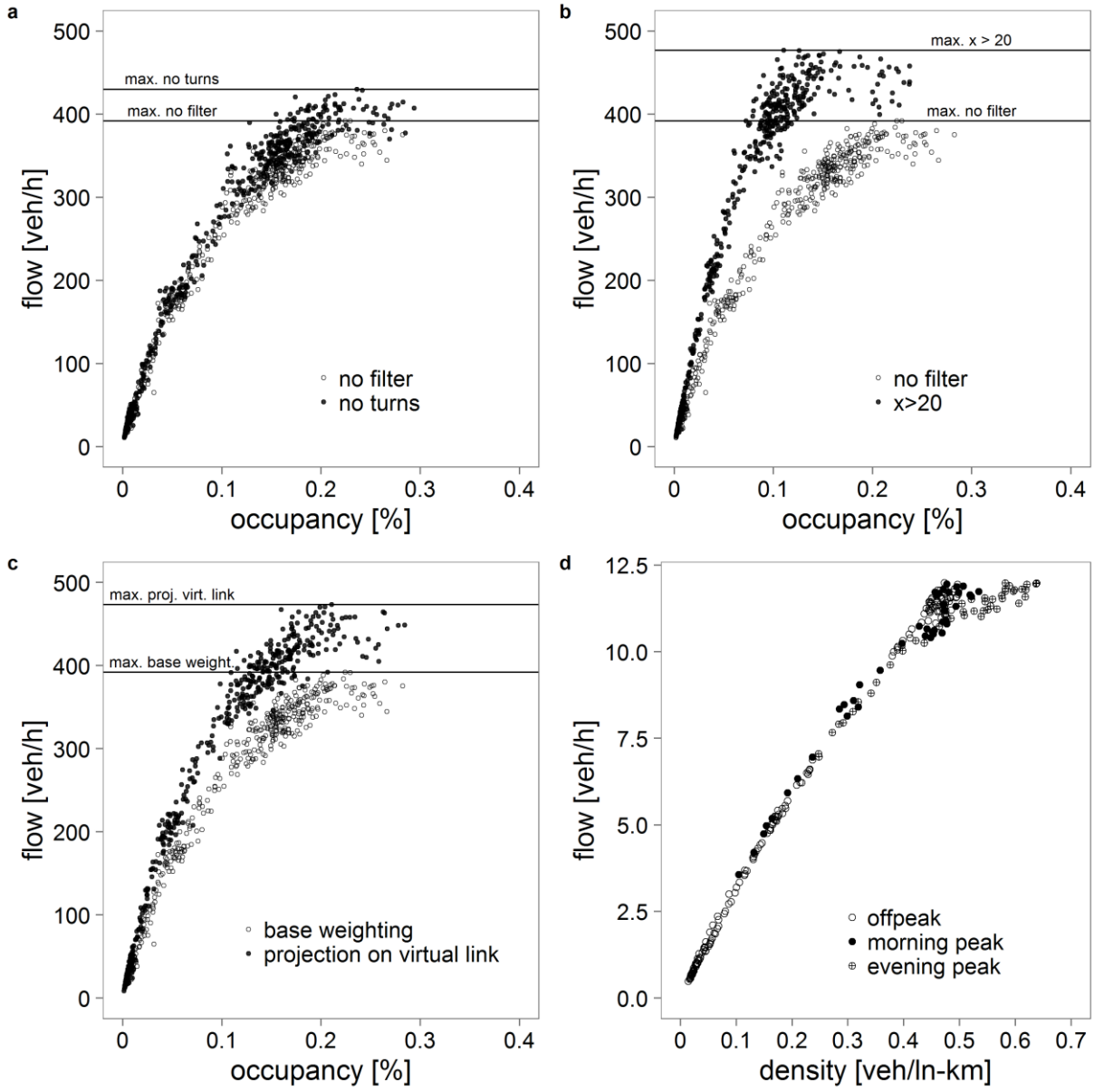

FIGURE 2 LDD and FCD MFDs: (a) LDD filtering turns (b) LDD filtering loop detector position (c) LDD new proposed weighting (d) FCD MFD

MFD BASED ON BOTH SOURCES

Combination of LDD and FCD

In the following, we transform both data sets to common scales and combine both sources in a way that their respective drawbacks are reduced. Then, we validate the transformations used.

As the MFD neither based on LDD nor FCD alone gives absolute numbers for both, average flow or density, we need to transform LDD occupancy to density. Here we use a simple and common scaling based on the spacing effective mean length, *s* (6). FCD speeds and flows need to be transformed with ρ, the ppr *(26)*.

Table 3 Estimation formulas

Source	Flow q	Density k
(i) LDD (eq. 3 and eq. 4 adjusted for *s*)	$\tilde{q} = \hat{q}_{LDD}$	$\tilde{k} = \frac{\hat{o}_{LDD}}{s}$
(ii) FCD (eq. 5 and eq. 6 adjusted for $\hat{\rho}$)	$\tilde{q} = \frac{\hat{q}_{FCD}}{\hat{\rho}}$	$\tilde{k} = \frac{\hat{k}_{FCD}}{\hat{\rho}}$
(iii) Combination of LDD & FCD	$\tilde{q} = \hat{q}_{LDD}$	$\tilde{k} = \frac{\hat{q}_{LDD}}{\hat{v}_{FCD}}$

For LDD, we apply the *projection on a virtual link*. We assume *s*=6.3m *(30)*, which is slightly above the 5.5m used in (*6*), due to the presence of larger vehicles in Zurich compared to Yokohama. For FCD, ρ is estimated by comparing FCD15 to LDD. For each link we divide the number of probes passing a loop detector by the total number of vehicles passing that loop detector, and average such value across all the links. We compare the transformed MFDs to a combined MFD based on the approach by *(15)* to leverage the strength of each source. This approach has the advantage that it needs no transformation. The flow of the combined MFD is calculated from LDD, and the density is calculated by dividing this flow by the FCD speeds.

Table 3 gives an overview of the three approaches, (i) LDD, (ii) FCD, and (iii) combined sources. Figure 3a shows the three approaches, again, for a Monday in the region "City" and Figure 3b in the region "Hard".

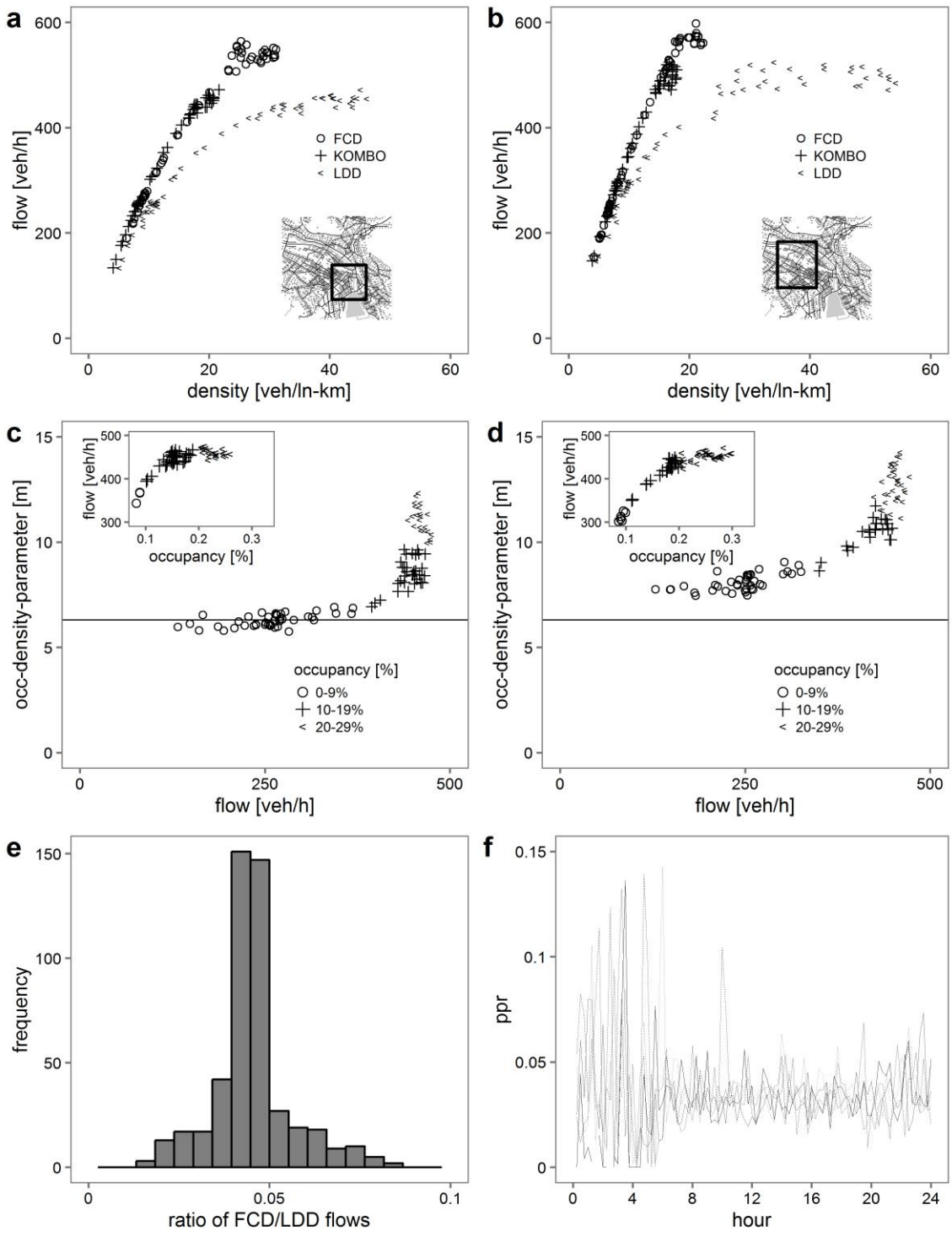

FIGURE 3 MFD based on multiple sources. (a) MFDs for "City", (b) MFDs for Hard, (c) occupancy-density parameter for LDD with x>20, (d) occupancy-density parameter for LDD with projection on virtual link, (e) FCD-LDD ratio of flows, (f) estimated probe penetration rate for region "City" by time of week and day of week.

We observe a similar trend in both regions. LDD shows higher densities for any given flows, even though we use the projection on a virtual link method. Note not all links represented in FCD are also available in LDD. An analysis where only links with both data sources available, still shows this divergence (not shown here for brevity). Thus explanations, other than the spatial differences of the data sources are needed and are discussed below. FCD shows a high consistency with the combination 'Kombo' MFD. Since such a combination increases the accuracy in simulations *(17)*, we can assume that it is more appropriate to use FCD, than LDD if we were to use a single data source, only. Obviously, for the combined MFD the maximum flow is that of the corresponding LDD.

LDD Biases

We observe higher densities for any given flows in LDD compared to FCD and the combined approach. This can be attributed to two reasons: (i) a placement bias, (ii) a link selection bias.

(i) *Placement bias*: We observe that the relative position of loop detectors on links is not uniformly distributed (see Figure 1). To alleviate the effects of this uneven distribution, we apply the projection on a virtual link method. Still, not many loop detectors are located in the middle of the link, although this position would provide important information on traffic states. Thus, the accuracy in the middle of the virtual link is the lowest, whereas it is the highest in the beginning and the end, where we collect information from most of the loop detectors.

FCD measurements, on the other hand, are available throughout the link. Since the speed usually drops close to the traffic light, the mean speed of the entire link is more representative for the middle of the link. Thus, the accuracy of the FCD is highest in the middle of a link. When we compare FCD with LDD, we compare to a large extent measurements with high accuracy in the middle (FCD) versus measurement with low accuracy in the middle (LDD).

(ii) *Link selection bias*: Loop detectors are placed at points of interest, such as in front of traffic signals or on links where congestion is more likely to occur *(14)*. The latter leads to a link selection bias, as we do not measure the average traffic state on the whole

network, but on selected links with higher probability of getting congested. Conversely, FCD is distributed more homogeneously over the network. Thus, when comparing network averages, we observe a lower density for FCD. Again, the difference between the two data sources increases with congestion. With the available LDD, it is non-trivial to correct for this bias, because traffic states on links without loop detectors are unknown. Thus they must be predicted with additional data.

We identify these two biases as the main reasons for a divergence between LDD and FCD MFDs.

Validation of transformation parameters

In the following, we validate the transformation parameters, s and ρ, assuming both data sources provide error-free measurements. We can calculate the transformation parameters correctly, (i) s and (ii) ρ.

(i) Using $\hat{s} = \frac{\hat{o}_{LDD}}{\hat{q}_{LDD}/\hat{v}_{FCD}}$ we can estimate s from both datasets. Figures 3c and 3d show this parameter in relation to MFD flow and occupancy. Figure 3c is based on selecting only loop detectors that are located more than 20m upstream of a traffic signal, and Figure 3d on the *projection on a virtual link*. The added horizontal line corresponds to the space effective mean length of 6.3m – the value used in Figures 3a and 3b and which was based only on average car and detector length *(30)*. We also highlight in the small windows the LDD MFD. Both plots show the same general trend: at low flow levels the parameter is constant and increases with greater flow until strong vertical scatter occurs around critical occupancy. Figure 3c shows for lower flow levels a good agreement with the 6.3m. We argue that detectors located more than 20m upstream of a traffic signal are more likely to measure free flow conditions, even more so at lower flows. Still, with increasing flow the difference between LDD and FCD increases as both biases become more apparent. Figure 3c validates the 6.3m as a rough approximation of the transformation parameters.

(ii) The ppr, $\hat{\rho}$, can be estimated on links with loop detectors installed. If LDD and FCD provided full network coverage, the ppr would be (a) equal to the number of probes

divided by the total number of vehicles. This would be equivalent to (b) the average FCD flow divided by the average LDD flow the ratio of flows, hereafter called ratio of flows, and to (c) the average number of probes on a lane divided by the average number of vehicles passing a loop detector. With neither full spatial nor temporal overlap of FCD5 with LDD, none of the three ratios are equal. Figure 3e shows the histogram of the ratio of flows (b) for region "City" and exhibits a clear peak at 0.04. Figure 3f shows $\hat{\rho}$ (estimated ppr) using (c) for region "City" during five working days. We observe that $\hat{\rho}$ is slightly lower than the average ratio of flows. The variability is high at night, and low during the day – especially during peak hours. At night, not many vehicles circulate. Thus, already one vehicle can represent a ppr of 20%. During daytime, absolute numbers are much higher, and thus the variability is reduced. We suggest to use (c), since (b) is influenced by the potential placement bias discussed before. Notice, this is valid for our type of FCD, and not necessarily for other kinds of FCD. Nevertheless this shows that for a rough approximation, the ppr can be estimated indeed using (c), validating this approach.

NOTABLE FEATURES OF ZURICH'S MFD

In this section we briefly outline two notable features observed in the Zurich empirical MFD. We first find indications of clockwise and counter-clockwise hysteresis loops. Then, the bifurcation seen in Figure 2d is further studied here.

For the hysteresis, we use FCD5, as it gives the 2-year average effects, thereby smoothening noise. In Figure 4a we observe a counter-clockwise hysteresis loop for region "City" and in Figure 4b a clockwise hysteresis loop for region "Hard", both for an average working day. Arguably, the hysteresis is not caused by variations in probe penetration, since during peak hours it can be assumed to be constant (see Figure 3f for region "City"). We attribute the counter-clockwise hysteresis to Zurich's traffic management system that controls signal cycles on arterials (i.e. access control). With a critical accumulation of vehicles reached, the system prevents more cars from entering the city, similarly to a perimeter control scheme *(14)*. This is relevant, as it shows the effectiveness of such a traffic management scheme.

The LDD MFD does not show a hysteresis. However, this might be partially explained due to the fact that the access control system was not working on a regular basis during the LDD observation period because of the construction work mentioned above *(28)*.

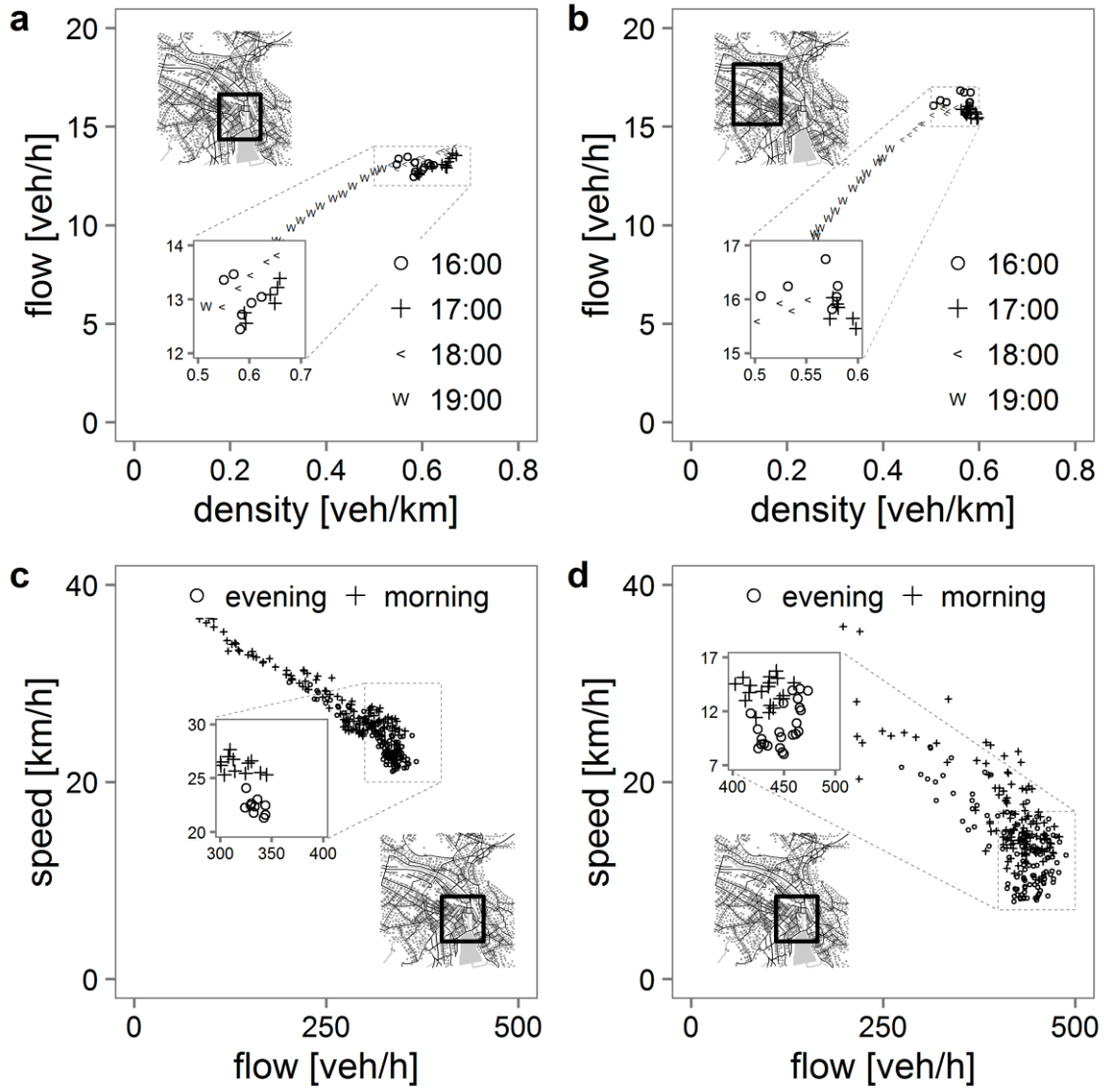

Figure 4 Hysteresis and bifurcation in Zurich: (a) FCD Hysteresis City (b) FCD Hysteresis Hard (c) FCD peak hour speed-flow (d) LDD peak hour speed-flow

We present the macroscopic speed-flow relations in Figure 4c for an average working day in FCD5. The scatter shows a distinction between morning and evening peak – evening speeds are dropping below the morning levels for any given flow. These findings are confirmed in Figure 4d

based on LDD. Similar to differences seen in the on- and off-set of congestion, it seems that filling the city in the morning is different from emptying it in the evening. This is important because it confirm differences between loading and unloading and guides cities to proper management schemes.

CONCLUSIONS

To the authors' knowledge, this is the first study that analyzes jointly LDD *and* FCD empirically to this level of detail in respect to MFD estimation. This allows a deeper understanding of both data sources and discussion on their limitations. The contributions of this paper are threefold, first we point out the limitations of each data source, second we propose new or validate common practice methods that aim at overcoming such limitations by comparing both data sources, and third we combine for the first time empirical data in a way that the effects of such limitations are reduced. These three points are further explained below.

- Loop detectors are (i) usually installed close to traffic signals and (ii) on links with greater congestion probability. We confirm that (i) leads to a *placement bias*, since for a reliable MFD loop detectors must be positioned uniformly within the links across the network. From (ii) results a *link selection bias*, confirming findings in *(14)*. This implies that density and congestion levels are more likely to be overestimated. FCD, on the other hand, faces limitations as well, since ρ is typically unknown a-priori, and a homogeneous spatial distribution of probe vehicles and congestion is not ensured. This implies that FCD is more reliable for average traffic states during daytime and on main roads with good coverage of probe vehicles.
- To overcome LDD limitations as much as possible, we propose the methodology *projection on a virtual link.* This method weights the measurements according to their relative position on a link and their frequency at that position in order to reduce the placement bias.

Comparing both datasets requires appropriate scaling of density (LDD) and flow (FCD). For LDD, this conversion parameter can be obtained a-priori. An ex-post estimate shows that the first parameter is a rough approximation, validating common practice, which uses the space effective mean length of a vehicle as

transformation parameter. For FCD we show that the ppr can be estimated with the average vehicle count data on links covered by both data sources.

- We have shown that well-defined and reproducible MFDs exist for each source separately. However, such single-source MFDs differ somewhat, due to the limitations mentioned above, due to noise, and due to temporal and spatial differences in the data sources. We have empirically shown that a combination of the two data sets following an approach by Leclercq et al. *(15)* leads to a well-defined MFD and we state that such combination of LDD and FCD reduces key drawbacks of each data source.

Although the presented MFDs do not show a congested branch, we do observe congestion at link level in some areas. One approach to overcome this issue is partitioning the network, *(e.g. 11)*, another one might be developing a selective MFD that includes only certain links. Future research is needed to understand how to better represent these very local congestion inhomogeneities, as further partitioning of the network can yield very small areas, ultimately leading to fundamental diagrams rather than MFDs. On the other hand, link selection might lead to non-representative MFDs.

To summarize, each data source exhibits a well-defined and reproducible MFD, but they differ from each other. This can be traced back to the limitations of the sources themselves, namely placement bias, link selection bias, and inappropriate transformation parameters. A combination of LDD flows and FCD speeds partly eliminates key drawbacks of the two data sources. At the moment, research is undergoing to further mitigate the problems arising when using both data sources simultaneously; a preliminary study *(17)* has shown that applying a data fusion algorithm increases the accuracy of the MFD estimation.

ACKNOWLEDGMENTS

This work was supported by ETH Research Grants ETH-04 15-1 and ETH-27 16-1. We would like to thank the City of Zurich and TomTom for providing loop detector data and floating car data, respectively.

REFERENCES

1. Smeed, R. J. *The traffic problem in towns*. BOOK. Manchester : Statistical Society, 1961.
2. Smeed, R. J. Traffic Studies and Urban Congestion. *Journal of Transport Economics and Policy*, Vol. 2, No. 1, 1968, pp. 33–70.
3. Venables, A. J. Evaluating Urban Transport Improvements Cost–Benefit Analysis in the Presence of Agglomeration and Income Taxation. *Journal of Transport Economics and Policy*, Vol. 41, No. 2, 2007, pp. 173–188.
4. Bettencourt, L. M. a. The origins of scaling in cities. *Science*, Vol. 340, No. 6139, 2013, pp. 1438–1441.
5. Mahmassani, H., J. Williams, and R. Herman. Performance of urban traffic networks. *Proceedings of the 10th International Symposium on Transportation and Traffic Theory*, 1987, pp. 1–20.
6. Geroliminis, N., and C. F. Daganzo. Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transportation Research Part B: Methodological*, Vol. 42, No. 9, 2008, pp. 759–770.
7. Knoop, V., S. Hoogendoorn, and J. Van Lint. Routing Strategies Based on Macroscopic Fundamental Diagram. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2315, No. 1, 2012, pp. 1–10.
8. Zheng, N., R. A. Waraich, K. W. Axhausen, and N. Geroliminis. A dynamic cordon pricing scheme combining the Macroscopic Fundamental Diagram and an agent-based traffic model. *Transportation Research Part A: Policy and Practice*, Vol. 46, No. 8, 2012, pp. 1291–1303.
9. Zheng, N., and N. Geroliminis. Modeling and optimization of multimodal urban networks with limited parking and dynamic pricing. *Transportation Research Part B: Methodological*, Vol. 83, 2016, pp. 36–58.
10. Zheng, N., and N. Geroliminis. On the distribution of urban road space for multimodal congested networks. *Transportation Research Part B: Methodological*, Vol. 57, 2013, pp. 326–341.
11. Ji, Y., and N. Geroliminis. On the spatial partitioning of urban transportation networks. *Transportation Research Part B: Methodological*, Vol. 46, No. 10, 2012, pp. 1639–1656.
12. Leclercq, L., C. Parzani, V. L. Knoop, J. Amourette, and S. P. Hoogendoorn. Macroscopic traffic dynamics with heterogeneous route patterns. *Transportation Research Part C: Emerging Technologies*, Vol. 59, 2015, pp. 292–307.
13. Buisson, C., and C. Ladier. Exploring the Impact of Homogeneity of Traffic Measurements on the Existence of Macroscopic Fundamental Diagrams. *TRANSPORTATION RESEARCH RECORD*, No. 2124, 2009, pp. 127–136.
14. Ortigosa, J., M. Menendez, and H. Tapia. Study on the number and location of measurement points for an MFD perimeter control scheme: a case study of Zurich. *EURO Journal on Transportation and Logistics*, Vol. 3, No. 3–4, 2014, pp. 245–266.
15. Leclercq, L., N. Chiabaut, and B. B. Trinquier. Macroscopic Fundamental Diagrams: A cross-comparison of estimation methods. *Transportation Research Part B: Methodological*, Vol. 62, 2014, pp. 1–12.
16. Courbon, T., and L. Leclercq. Cross-comparison of macroscopic fundamental diagram estimation methods. *Procedia - Social and Behavioral Sciences*, Vol. 20, 2011, pp. 417–426.
17. Ambühl, L., and M. Menendez. Data Fusion Algorithm for Macroscopic Fundamental

Diagram Estimation. *Transportation Research Part C: Emerging Technologies*, No. in press, 2016.
18. Hall, F., V. Hurdle, and J. Banks. Synthesis of recent work on the nature of speed-flow and flow-occupancy (or density) relationships on freeways. *Transportation Research Record*, No. 1365, 1992, pp. 12–18.
19. Neumann, T. Efficient queue length detection at traffic signals using probe vehicle data and data fusion. *ITS 2009 (16th World Congress)*, 2009, pp. 1–12.
20. Qian, G., J. Lee, and E. Chung. Algorithm for Queue Estimation with Loop Detector of Time Occupancy in Off-Ramps on Signalized Motorways. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2278, No. 2278, 2012, pp. 50–56.
21. Tsubota, T., A. Bhaskar, and E. Chung. Macroscopic Fundamental Diagram for Brisbane, Australia Empirical Findings on Network Partitioning and Incident Detection. *TRANSPORTATION RESEARCH RECORD*, No. 2421, 2014, pp. 12–21.
22. Brouwer, J. Measuring real-time traffic data quality based on Floating Car Data. 2014, pp. 1–8.
23. Bazzani, a, B. Giorgini, R. Gallotti, L. Giovannini, M. Marchioni, and S. Rambaldi. Towards Congestion Detection in Transportation Networks Using GPS Data. *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on*, No. May 2010, 2011, pp. 1455–1459.
24. Tsubota, T., A. Bhaskar, and E. Chung. Empirical evaluation of brisbane macroscopic fundamental diagram. *Australasian Transport Research Forum 2013 Proceedings*, Vol. 2013, No. October, 2013, pp. 2–4.
25. Du, J., H. Rakha, and V. V. Gayah. Deriving macroscopic fundamental diagrams from probe data: Issues and proposed solutions. *Transportation Research Part C: Emerging Technologies*, Vol. 66, 2016, pp. 136–149.
26. Gayah, V., and V. Dixit. Using Mobile Probe Data and the Macroscopic Fundamental Diagram to Estimate Network Densities. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2390, No. 1, 2013, pp. 76–86.
27. Beibei, J., H. van Zuylen, and L. Shoufeng. Determining the Macroscopic Fundamental Diagram on the Basis of Mixed and Incomplete Traffic Data. *TRB 95th Annual Meeting Compendium of Papers*, 2016.
28. Stadt Zürich - Dienstabteilung Verkehr (DAV). Loop detector data in Zurich.
29. Ge, Q., and M. Menendez. Sensitivity Analysis for Calibrating VISSIM in Modeling the Zurich Network. *12th Swiss Transport Research Conference*, 2012, p. 19.
30. AKP Verkehrsingenieure AG. *Forschungsprojekt VSS 2011/203 Geometrie des Fahrzeugparks der Schweiz*. 2016.
31. Ji, Y., J. Luo, and N. Geroliminis. Empirical Observations of Congestion Propagation and Dynamic Partitioning with Probe Data for Large-Scale Systems. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2422, 2014, pp. 1–11.
32. Wang, P. F., K. Wada, T. Akamatsu, and Y. Hara. An Empirical Analysis of Macroscopic Fundamental Diagrams for Sendai Road Networks. *Interdisciplinary Information Sciences*, Vol. 21, No. 1, 2015, pp. 49–61.
33. Ampountolas, K., and A. Kouvelas. Real-Time Estimation Of Critical Values Of The Macroscopic Fundamental Diagram For Maximum Network Throughput. 2015, pp. 1–15.