 doi:10.20944/preprints202212.0178.v1

Effects of Loop Detector Position on the Macroscopic Fundamental Diagram

Garyoung Lee[a], Zijian Ding[a], Jorge Laval[a,*]

[a]*School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, United States*

Abstract

Loop detectors are probably the widest-used technology for traffic state estimation. Previous research has shown that loop detector positions within the link significantly affect the estimation of the macroscopic fundamental diagram (MFD) of a given network. This paper examines the biases produced by the positioning of loop detectors on the MFD, using both analytical and simulation methods, as well as empirical data from UTD-19. We confirm earlier results that a uniform distribution of loop detector positions reduces the bias. For non-uniform distribution, we found that: (i) the subsets of the MFD by the loop detector position help estimate whether the loop detector MFD will have a bias; (ii) if the detectors in the network are positioned more downstream with a larger variation, the loop detector MFD is more likely to have a discrepancy in position subsets of the MFD; (iii) a lower ratio of link length to green signal time elevates the MFD bias as well. This research opens the possibility for the bias of MFD induced by the loop detector data to be corrected by only using itself.

Keywords: Macroscopic Fundamental Diagram, Loop Detector Data, Traffic Simulation

1. Introduction

The modeling of traffic flow dynamics in large urban networks has proven challenging over the years (Greenshields et al., 1935, Smeed, 1967, Mahmassani et al., 1984, Geroliminis and Daganzo, 2008, Daganzo and Geroliminis, 2008, Daganzo, 2005, Geroliminis and Boyacı, 2012, Laval and Castrillón, 2015, Geroliminis and Sun, 2011, Mazloumian et al., 2010, Leclercq et al., 2015, Zheng and Geroliminis, 2013, Buisson and Ladier, 2009, Yildirimoglu et al., 2015, Ding et al., 2017, Gayah et al., 2014, Huang et al., 2018, Ambühl and Menendez, 2016, Courbon and Leclercq, 2011, Leclercq et al., 2014, Ambühl et al., 2017, Aghamohammadi and Laval, 2022, Loder et al., 2019). An important branch of the efforts to control congestion is aggregated modeling. After Greenshields et al. (1935) observed for the first time the fundamental diagram of a single uninterrupted link, researchers took a profound interest in the aggregated relationship between average flow and density in entire urban signalized networks (Smeed, 1967, Mahmassani et al., 1984). The encapsulation of network traffic states into two variables is known as the Macroscopic Fundamental Diagram (MFD). Geroliminis and Daganzo (2008), and their empirical study in Yokohama, Japan demonstrated that the MFD is a convincing model to describe a network-level traffic performance. When aggregated at a network level, a high scatter of average flow and density from individual loop detectors nearly vanished and the points gather along the MFD curve.

Analytical, empirical, and simulation studies have been conducted to observe the MFD. Contemporaneous with Geroliminis and Daganzo (2008), Daganzo and Geroliminis (2008) presented the method of cuts (MoC) using variational theory (Daganzo, 2005) in a homogeneous signalized corridor, which sets the upper bound for the MFD. Stemmed from the literature, Geroliminis and Boyacı (2012) applied variational theory to parallel corridors with weak heterogeneity. Considering a strong heterogeneity of the real-world, Laval and Castrillón (2015) proposed the stochastic MoC (SMoC) to handle networks with different block lengths and signal timings.

Numerous studies have verified that the MFD is applicable to other cities or arbitrary networks. In contrast to the findings from Geroliminis and Daganzo (2008) that the MFD is independent of demand, later researchers challenged

*Corresponding author: jorge.laval@ce.gatech.edu
[1]790 Atlantic Drive, Atlanta, GA 30332

that the finding is only apposite for homogeneous networks with low congestion levels. Otherwise, the MFD reveals a trapezoidal shape and the hysteresis phenomenon (Geroliminis and Sun, 2011). A host of posterior literature suggested that the MFD shape is dependent on demand (Mazloumian et al., 2010, Leclercq et al., 2015), network topology and heterogeneity (Zheng and Geroliminis, 2013, Geroliminis and Boyacı, 2012, Buisson and Ladier, 2009), routing strategy (Yildirimoglu et al., 2015, Ding et al., 2017), and signal control schemes (Gayah et al., 2014, Huang et al., 2018) (Zhang et al., 2020).

While various exogenous influential factors of the MFD have been exhibited, the endogenous factor–the bias induced by the nature of empirical data–has not been discussed to a comparable extent. Loop detector data is arguably the most prevailing empirical source for exploring various facets of the MFD. Some studies use probe data, but even this is usually fused with loop detector data (Ji et al., 2014, Ambühl and Menendez, 2016, Du et al., 2016, An et al., 2020, Saffari et al., 2022). Although the loop detector is compelling as it measures traffic flow at all times, many have faced limitations in solely exploiting it reliably for various problems (Kong et al., 2009, Kim et al., 2020, Min et al., 2022). Particularly, its installation position along the link is found to be critical to accurately represent the traffic flow.

Buisson and Ladier (2009) first realized that the position of the loop detectors within the links plays a substantial role in defining the MFD shape. They split the measurements from the detectors in Toulouse, France into three subsets according to the physical distance to the downstream traffic signal. Closer to the signal, the free-flow branch of MFD showed a lower slope. The overestimation of the queue stood as the rationale. Using simulation, Courbon and Leclercq (2011) compared three positionings– constant, uniformly distributed, and normally distributed–of virtual detectors on a corridor with an identical block length. Although the constant distance setting displayed the largest bias, the detectors farther from the downstream signal reproduced the free-flow conditions well and the closer ones reproduced the queues well at the cost of the lower slope of the free-flow branch. Uniform distribution of the detectors showed the most accurate fit to the MoC. Leclercq et al. (2014) proved that the uniform distribution of detectors is also the best strategy for the homogeneous network to reproduce accurate traffic state. Ambühl et al. (2017) leveraged this finding to explain not only the discrepancy in the MFD drawn by the loop detector and floating car data from Zurich, Switzerland, but also the decreased average occupancy when the detectors closer to the downstream signal were excluded. They pointed out that the loop detector bias owes to the non-uniform placement and the link selection: the detectors are mostly placed at the beginning or the end of the link, and they are installed in certain links to control traffic signals and congestion.

Although previous literature showed how the loop detector positions influence the shape of MFD and why the bias happens, still there are important gaps to be filled. First, two different biases induced by the nature of detectors are distinguished here: (i) the bias between the link MFD and the loop detector (LD)-MFD (henceforth, LD bias) and (ii) the bias between position-based subsets of LD-MFD (henceforth, subset bias). Note that the link MFD can be thought of as the "ground-truth" as it gives Edie's generalized traffic state definitions (Edie et al., 1963). The subset bias refers to MFDs using detectors that belong to a particular location subset: upstream, midstream or downstream within the link. Notice that Courbon and Leclercq (2011) and Buisson and Ladier (2009) concluded that the detector position affects the MFD without distinguishing these two biases although what they measured turned out to be LD bias and subset bias, respectively. Second, identifying network characteristics that contribute to each bias is required since the literature used only a single network when explaining the existence of bias. Third, even though the uniform distribution of the positions is proved to be the best strategy (Courbon and Leclercq, 2011, Leclercq et al., 2014), we need a further investigation on how an arbitrary distribution of detector positions affects the MFD. Lastly, as mentioned in Ambühl et al. (2017), the variability of block lengths and the spatial density across the network should be taken into consideration.

To bridge these gaps, we (i) analytically investigate the condition and the extent of the LD bias and subset bias occurrence in a corridor, (ii) empirically analyze the characteristics of loop detector position that generate subset bias, and (iii) simulate the impact of different network topology on the bias. These three objectives are addressed in the remainder of the paper.

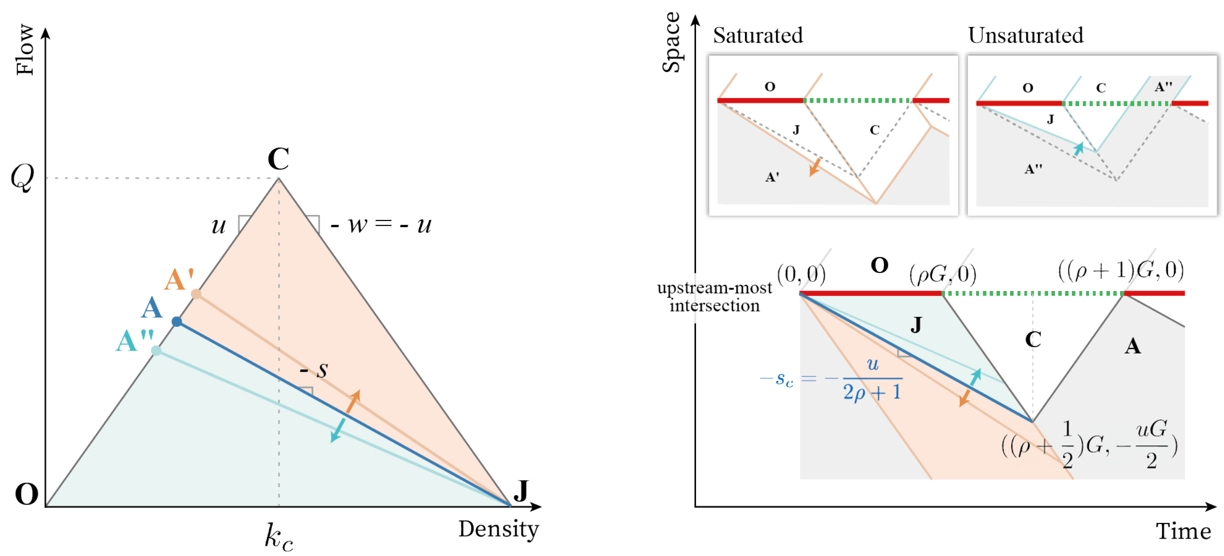
2. Analytical Corridor Approximation

Recall that the LD bias refers to the bias between the link MFD and the LD-MFD and the subset bias refers to the bias between position-based subsets of LD-MFD. In this section, we assume a homogeneous corridor that obeys a symmetric triangular fundamental diagram (FD) to analyze LD bias and subset bias. As customary to simplify the analysis, we will use isosceles fundamental diagrams (free-flow speed = wave speed) since one obtains the same solutions using a general triangular FD (Laval and Castrillón, 2015, Laval and Chilukuri, 2016).

The corridor to be analyzed consists of N links with an identical block length of l. The traffic signal on all intersections is fixed with green time, G, and red time, R, with no offset. The symmetric triangular FD has a free-flow speed of u, a critical density of k_c, and a capacity of Q. As shown in Fig. 1(a) the queue initially grows at a shock wave speed s in the upstream-most intersection, depicted as a state **A**, and clears at a wave speed w. The traffic state of zero flow with zero density, *i.e.*, a void, is depicted as a state **O**. The traffic state of the capacity and the jam density are each denoted as state **C** and state **J**, respectively. The variables and constants used are summarized in Table 1. Note that the symbols with asterisks are constants, and otherwise, variables.

Table 1: Descriptions of constants and variables

Traffic states		Physical attributes of the network	
O	Void state*	l	Link length
A	Current state	N	The number of links*
C	Capacity state*	N_u	The number of links with loop detector located upstream of critical position
J	Jam state*	N_d	The number of links with loop detector located downstream of critical position
Fundamental Diagram		Signal Setting	
u	Free-flow speed*	R	Red signal time
w	Wave speed*	G	Green signal time
s	Shock wave speed	n	Ceiling of time in unit of cycle length for 1st vehicle in green arrives at next intersection
k_c	Critical density*	Dimensionless parameter	
Q	Capacity*	λ	Link length to critical link length ratio ($= \frac{l}{l_c}$)
		ρ	Red time to green time ratio ($= \frac{R}{G}$)



(a) (b)

Fig. 1. Initial conditions: (a) A fundamental diagram with different shock waves; (b) Time-space diagrams of the saturated and unsaturated condition

 doi:10.20944/preprints202212.0178.v1

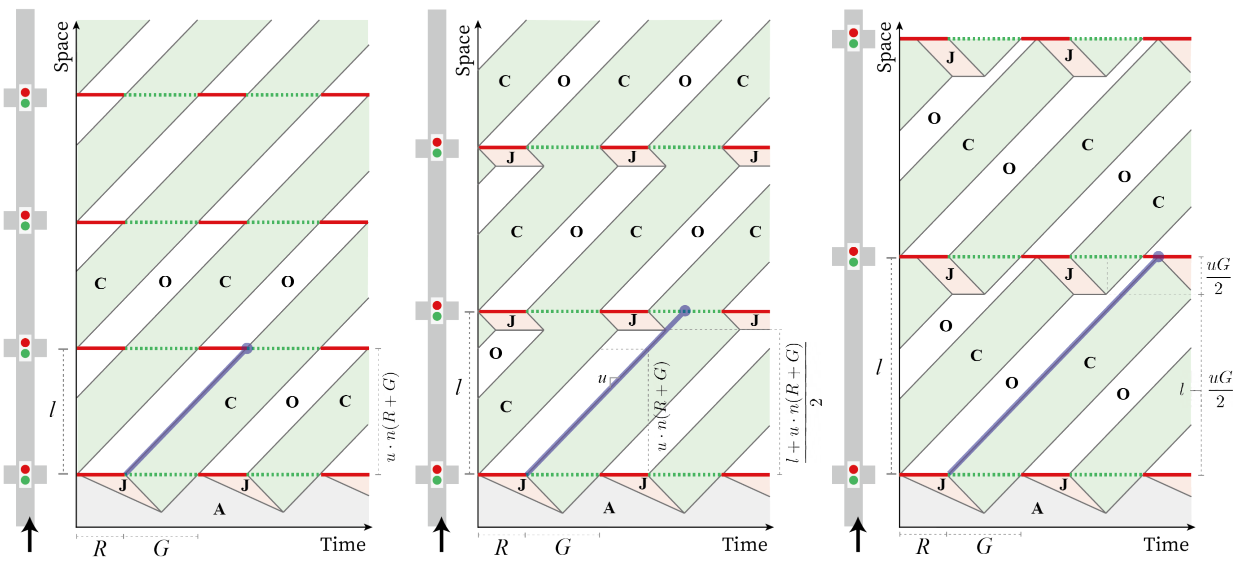
Here, it is important to introduce two dimensionless parameters, the mean block length to critical length ratio, λ, and the mean red signal time to mean green signal time, ρ, which significantly influence the MFD shape according to Laval and Castrillón (2015). The critical length, l^*, corresponds to the minimum block length that prevents spillbacks. Then λ is expressed as:

$$\lambda = \frac{l}{l^*} = \frac{l}{G/(\frac{1}{u}+\frac{1}{w})} = \frac{l}{G/\frac{2}{u}} = \frac{2l}{uG} \tag{1}$$

Under the given settings, all possible patterns of the time-space diagram can be categorized according to four variables: (i) s, (ii) λ, (iii) ρ, and (iv) n. First of all, the shock wave speed becomes the primary determinant. It decides whether the initial condition influences the downstream links. The critical state **A** and the corresponding critical shock wave, s_c, are obtained where the queue clearance wave intersects the end of the green phase (Fig. 1(b)). Thus, the critical shock wave speed is the slope between the origin and the queue dissipation point, $s_c = \frac{u}{2\rho+1}$. If the shock wave speed is steeper than the critical shock wave speed (state $\mathbf{A}'$), the first queue clearance wave is obstructed by the queue of the next red phase. This conveys that the queue accumulated in the red signal phase loses a chance to completely vanish before the next cycle. On the other hand, if the shock wave speed is lower (state $\mathbf{A}''$), the queue clearance wave traverses through the green phase, causing the initial state $\mathbf{A}''$ to spread to the downstream link. Hereafter, the initial condition that exceeds or equals the critical shock wave speed is considered a saturated initial condition, otherwise an unsaturated.

Under saturated conditions, the time-space diagram of the upstream-most intersection is duplicated at downstream links, while unsaturated initial conditions do not provide this guarantee. The uncertainty of the repetition ascribes to the spread of the initial state to the downstream links. This causes some cases of unsaturated initial condition to not be expressed in a closed form. Hence, we only address the saturated condition in the following.

Fig. 2 depicts all possible types of time-space diagrams in the saturated initial condition. It is confirmed for all three cases that the traffic state patterns recur throughout the corridor. Importantly, the existence of the jam state and the coverage of the void state distinguish three cases. As indicated by a blue line, the difference originates from the time it takes for the first vehicle in a green phase to reach the next intersection.



(a) Case 1 **(b)** Case 2 **(c)** Case 3

Fig. 2. Three types of time-space diagram at the saturated initial condition

A method to derive the MFD from the time-space diagram is explained in the following. Note that only the downstream links of the upstream-most intersection are considered. We assume N detectors are installed in each of the N links, and the aggregation interval is a multiple of a cycle length. This enables setting the aggregation interval simply as a cycle length. In order to obtain the LD-MFD, firstly, draw a horizontal line on a time-space diagram at the position of a detector by the amount of the aggregation interval. Second, calculate the weighted average of the traffic state, *i.e.*, a pair of density-flow, using the time proportion of each state as a weight. This corresponds to the FD of a detector, which is then a linear combination of traffic states **O**, **C**, and **J**. Lastly, for every time interval measured, average the density-flow pair of all detectors to obtain the LD-MFD (Equation 2). While the LD-MFD uses the proportion of time as a weight, the link MFD can be obtained by using the proportion of the area of each traffic state throughout the link as a weight. The link MFD in the saturated initial condition equals to the intersection point where the stationary cut and the forward cut of MoC.

$$q_{LD} = \frac{\Sigma_i q_i l_i}{\Sigma_i l_i}, \quad k_{LD} = \frac{\Sigma_i k_i l_i}{\Sigma_i l_i} \tag{2}$$

where q_{LD} is the average flow, k_{LD} is the average density, and q_i and k_i are the flow and density measured by the loop detector i of installed link length of l_i.

The MFD for each case through the above process is summarized in Table 2. We now disentangle the constraints and MFD formulae; i.e., how λ, ρ, and n act as the keys to distinguish these three.

2.1. Case 1: No queues

According to Fig. 2(a), the first vehicle of the green phase never encounters the red phase and so do all other vehicles. Notice that only the void and the capacity state exist. This is only possible when the time for the vehicle to complete passing the link is equal to a multiple of a cycle length. This constraint is simplified using definitions of λ and ρ as follows:

$$\frac{l}{u} = n(R+G) \tag{3a}$$

$$\lambda = 2n(\rho+1) \tag{3b}$$

Since the traffic states are homogeneous along the link, MFD is not subject to any bias no matter where the detector is positioned. In any of the positions during a cycle $R+G$, the void state **O** and the capacity state **C** are measured by the amount of the red time R and the green time G, respectively. Thus, the FD of all loop detectors correspond to $\frac{\rho}{1+\rho} \cdot \mathbf{O} + \frac{1}{1+\rho} \cdot \mathbf{C} + 0 \cdot \mathbf{J}$, so do the MFD.

2.2. Case 2: Jam exists & Voids are finite

Compared to Case 1, Fig. 2(b) bares the jam state for a certain length. The jam accumulates because some vehicles departed at a previous intersection cannot pass the intersection ahead being blocked by the red time. This happens when the first vehicle of the green time was able to pass the next intersection without stopping but was not the foremost vehicle passed during its green time. That is, the time for the first vehicle to arrive at the next intersection is a multiple of a cycle length plus a partial or a full amount of green time. This condition is formulated as below:

$$n(R+G) < \frac{l}{u} \leq n(R+G) + G \tag{4a}$$

$$2n(\rho+1) < \lambda \leq 2n(\rho+1) + 2 \tag{4b}$$

As traffic states are non-homogeneous along the link, identifying the critical position that turns the void state into the jam state is necessary. Separating the travel time into a multiple of a cycle length and the remainder gives us the length of the jam state. The distance vehicle traveled during n-multiple of cycle length is $u \cdot n(R+G)$. The remaining distance of $l - u \cdot n(R+G)$ is bisected by the void state and the jam state due to the symmetry of the FD. Then, the length of jam state and the void state are $(l - u \cdot n(R+G))/2$ and $(l + u \cdot n(R+G))/2$, respectively. If the loop detector is installed downstream of the critical position, the jam state **J** and the capacity state **C** will be each measured for the red time R and the green time G during an aggregation interval $(R+G)$. On the other hand, if the detector is installed at upstream of the critical position, the red time and the green time are each occupied by the void state **O** and the capacity state

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 December 2022 doi:10.20944/preprints202212.0178.v1

C. With specifying that N_d and N_u detectors are installed at each downstream and upstream of critical position, the LD-MFD will have a linear combination of $\frac{\rho}{1+\rho} \cdot \frac{N_u}{N_u+N_d} \cdot \mathbf{O} + \frac{1}{1+\rho} \cdot \mathbf{C} + \frac{\rho}{1+\rho} \cdot \frac{N_d}{N_u+N_d} \cdot \mathbf{J}$.

Here, the bias of LD-MFD is unavoidable unless the number of detectors in each position label is proportional to its length: *i.e.*, $N_d \propto (l - u \cdot n(R+G))/2$ and $N_u \propto (l + u \cdot n(R+G))/2$. Namely, the link MFD, which is theoretically a weighted average of each traffic state's area, can be simply calculated by substituting corresponding lengths to the number of detectors. Similar to Buisson and Ladier (2009), partitioning the loop detectors by their position gives us the position-based subsets of the MFD. For example, the downstream subset can be obtained by substituting $N_d = 1$ and $N_u = 0$ to the LD-MFD. The corresponding MFD expressions can be found in Table 2.

2.3. Case 3: Jam exists & Voids are infinite

Compared to Case 2, jam is accumulated for a shorter period of time and the void fills the gap (Fig. 2(c)). The red time is already initiated before the first vehicle of the green time arrives at the next intersection. The amount of red time elapsed before the jam accumulation remains as the void. This is only available when the link travel time of the vehicle equals a multiple of a cycle length plus a full green time plus a partial red time, as below.

$$n(R+G) + G < \frac{l}{u} < n(R+G) + G + R \tag{5a}$$

$$2n(\rho+1) + 2 < \lambda < 2(n+1)(\rho+1) \tag{5b}$$

The critical position is obtained by the relationship between the free-flow speed u and the queue clearance duration G. Using that the jam always dissipates exactly before the red phase begins, the spatial length of the jam state is $u \cdot G/2$. The loop detectors located downstream of the critical position will measure the void state **O** for time $\frac{l}{u} - n(R+G) - G$, the capacity state **C** for green time G, and the jam state **J** for the rest of a cycle. At the upstream of the critical position, the void state **O** and the capacity state **C** are observed by the amount of red time R and the green time G, respectively. With the assumption of the number of detectors N_d and N_u, the weighted average gives LD-MFD. The link MFD and the position-based subsets are calculated likewise to Case 2 and one can refer to Table 2.

2.4. Discussions on Subset bias and LD bias

Fig. 3 illustrates MFD realizations of the corridors with different network parameters. The FD is drawn at the back for comparison with the MFDs. Case 1 shows that neither the LD bias nor the subset bias resides in MFD since the corridor has an ideal signal setting that the queue never forms. Regardless of n and λ, the MFD of a corridor always lies exactly on the free-flow branch of FD and only moves along the branch depending on the ρ. Specifically, the yellow-green dot displayed a much slower critical shock wave speed and a smaller average density than the blue dot due to its higher ρ value. This means that insofar as two different corridors have the same ρ value, their MFDs stand identical regardless of different n.

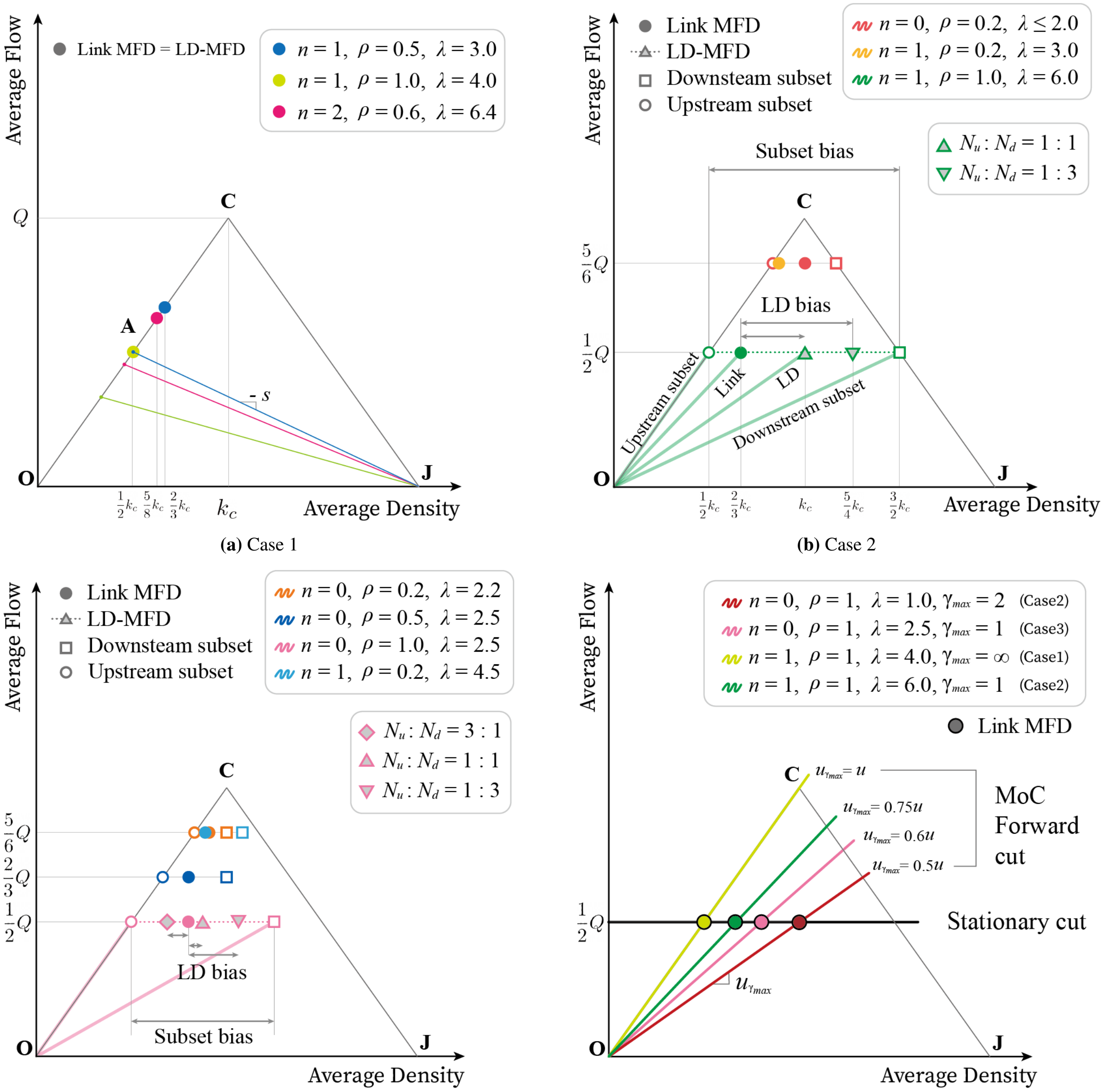
In contrast to Case 1, the difference between MFDs are discernible in Fig. 3(b) and 3(c). The link MFD, LD-MFD, upstream subset, and downstream subset are labeled with a solid circle, gray-filled symbols, open circle, and open square, respectively. The dotted line shows the possible range between the upstream and downstream subsets to which the LD-MFD can fall on. Slopes that connect the origin and the MFD are considered free-flow branches of the corresponding MFD. In the saturated initial condition, the bias takes place only in average density values. The average flow is always $\frac{1}{1+\rho}Q$, which aligns with the stationary cut of MoC (Daganzo and Geroliminis, 2008). Computations reveal that the link MFD corresponds to the point where the stationary cut and the forward cut intersect (Fig. 3(d)).

2.4.1. Subset bias

We can identify the subset bias through the comparison between position-based subsets, *i.e.*, the difference between the downstream subset and the upstream subset. It is prominent that the downstream subset underestimates the average free-flow speed while the upstream subset overestimates it. As the slope of the upstream subset is always fixed with the free-flow speed of FD, u, the downstream subset holds the key to determining the subset bias.

In Case 2, the downstream subsets fall exactly on the congestion branch of FD regardless of the parameters. The magnitude of the subset bias, which is $\frac{2\rho}{1+\rho}k_c$ with respect to the average density, is determined only by the value of ρ. Larger ρ decreases the maximum average flow and increases the subset bias. This means that the subset bias is inevitable unless ρ is negligibly small. Insofar as two corridors have the same ρ, the position-based subsets are identical despite different link MFDs.

In Case 3, the downstream subsets do not lie on the congestion branch. As the subset bias with respect to the average density equals $(2(n+1) - \frac{\lambda}{1+\rho})k_c$, the subset bias gets smaller when λ approximates the right-hand side of the constraint of Case 3 (Equation 5b). This is because the corridor becomes more free-of-congestion with the approximation of λ. We can also notice that larger ρ does not guarantee a larger subset bias.



(a) Case 1 **(b)** Case 2 **(c)** Case 3 **(d)** Comparison with Method of Cuts

Fig. 3. The representation of the LD bias and the subset bias (Note: Link MFD is always identical to the intersection of the MoC forward cut and stationary cut)

Table 2: Summarize of link MFD, LD-MFD, and position-based subsets

Case 1: $\lambda = 2n(\rho+1), \quad n = 0, 1, 2, \cdots$			The proportion of time during a cycle		
LD position	Length	Detector count	$\mathbf{O}(0,0)$	$\mathbf{C}(k_c, Q)$	$\mathbf{J}(2k_c, 0)$
-	l	N	$\frac{R}{R+G} = \frac{\rho}{1+\rho}$	$\frac{G}{R+G} = \frac{1}{1+\rho}$	0
MFD	Linear combination	$\frac{\rho}{1+\rho}\cdot\mathbf{O} + \frac{1}{1+\rho}\cdot\mathbf{C} + 0\cdot\mathbf{J}$			
(LD = Link)	Coordinate	$\left(\frac{1}{1+\rho}k_c, \quad \frac{1}{1+\rho}Q\right)$			
Case 2: $2n(\rho+1) < \lambda \le 2n(\rho+1)+2, \quad n = 0, 1, 2, \cdots$			The proportion of time during a cycle		
LD position	Length	Detector count	$\mathbf{O}(0,0)$	$\mathbf{C}(k_c, Q)$	$\mathbf{J}(2k_c, 0)$
Downstream of critical position	$\frac{l-u\cdot n(R+G)}{2} = \frac{uG}{2}(\frac{\lambda}{2} - (1+\rho)n)$	N_d	0	$\frac{G}{R+G} = \frac{1}{1+\rho}$	$\frac{R}{R+G} = \frac{\rho}{1+\rho}$
Upstream of critical position	$\frac{l+u\cdot n(R+G)}{2} = \frac{uG}{2}(\frac{\lambda}{2} + (1+\rho)n)$	N_u	$\frac{R}{R+G} = \frac{\rho}{1+\rho}$	$\frac{G}{R+G} = \frac{1}{1+\rho}$	0
LD-MFD	Linear combination	$\frac{\rho}{1+\rho}\cdot\frac{N_u}{N_u+N_d}\cdot\mathbf{O} + \frac{1}{1+\rho}\cdot\mathbf{C} + \frac{\rho}{1+\rho}\cdot\frac{N_d}{N_u+N_d}\cdot\mathbf{J}$			
	Coordinate	$\left(\frac{1}{1+\rho}(1+2\rho\frac{N_d}{N_u+N_d})k_c, \quad \frac{1}{1+\rho}Q\right)$			
Link MFD	Linear combination	$\frac{\rho}{1+\rho}\cdot(\frac{1}{2}+\frac{1+\rho}{\lambda}n)\cdot\mathbf{O} + \frac{1}{1+\rho}\cdot\mathbf{C} + \frac{\rho}{1+\rho}\cdot(\frac{1}{2}-\frac{1+\rho}{\lambda}n)\cdot\mathbf{J}$			
	Coordinate	$\left((1-\frac{2\rho}{\lambda}n)k_c, \quad \frac{1}{1+\rho}Q\right)$			
Position-based subset LD-MFD	Coordinate	Downstream: $\left(\frac{1+2\rho}{1+\rho}k_c, \frac{1}{1+\rho}Q\right)$ Upstream: $\left(\frac{1}{1+\rho}k_c, \frac{1}{1+\rho}Q\right)$			
Case 3: $2n(\rho+1)+2 < \lambda < 2(n+1)(\rho+1), \quad n = 0, 1, 2, \cdots$			The proportion of time during a cycle		
LD position	Length	Detector count	$\mathbf{O}(0,0)$	$\mathbf{C}(k_c, Q)$	$\mathbf{J}(2k_c, 0)$
Downstream of critical position	$\frac{uG}{2} = \frac{l}{\lambda}$	N_d	$\frac{\frac{l}{u}-n(R+G)-G}{R+G} = \frac{\lambda-2}{2(1+\rho)} - n$	$\frac{G}{R+G} = \frac{1}{1+\rho}$	$\frac{(n+1)(R+G)-\frac{l}{u}}{R+G} = n+1-\frac{\lambda}{2(\rho+1)}$
Upstream of critical position	$l - \frac{uG}{2} = l(1-\frac{1}{\lambda})$	N_u	$\frac{R}{(R+G)} = \frac{\rho}{1+\rho}$	$\frac{G}{R+G} = \frac{1}{1+\rho}$	0
LD-MFD	Linear combination	$\frac{(\frac{\lambda-2}{2(1+\rho)}-n)\cdot N_d+\frac{\rho}{1+\rho}\cdot N_u}{N_u+N_d}\cdot\mathbf{O} + \frac{1}{1+\rho}\cdot\mathbf{C} + (n+1-\frac{\lambda}{2(1+\rho)})\cdot\frac{N_d}{N_u+N_d}\cdot\mathbf{J}$			
	Coordinate	$\left((\frac{1}{1+\rho}+(n+1-\frac{\lambda}{2(1+\rho)})\frac{2N_d}{N_u+N_d})k_c, \quad \frac{1}{1+\rho}Q\right)$			
Link MFD	Linear combination	$(\frac{1+2\rho}{2(1+\rho)} - \frac{n+1}{\lambda})\cdot\mathbf{O} + \frac{1}{1+\rho}\cdot\mathbf{C} + (n+1-\frac{1}{2(1+\rho)})\cdot\mathbf{J}$			
	Coordinate	$\left(\frac{2(n+1)}{\lambda}k_c, \quad \frac{1}{1+\rho}Q\right)$			
Position-based subset LD-MFD	Coordinate	Downstream: $\left((\frac{1-\lambda}{1+\rho}+2(n+1))k_c, \frac{1}{1+\rho}Q\right)$ Upstream: $\left(\frac{1}{1+\rho}k_c, \frac{1}{1+\rho}Q\right)$			

2.4.2. LD bias

LD bias is clearly recognizable by the green and pink symbols in each Fig. 3(b) and 3(c). LD-MFDs can take place anywhere between the upstream and downstream subsets, depending on the distribution of detectors. The link MFD is generally closer to the upstream subset than the downstream subset under this setting because the length of the jam state is shorter than half of the link length. LD bias will converge to zero when the number of detectors of each position is proportional to its length: *e.g.*, $N_d \propto (l - u \cdot n(R + G))/2$ and $N_u \propto (l + u \cdot n(R + G))/2$ for Case 2. This validates the findings from Courbon and Leclercq (2011) that the uniform distribution of the detectors across the corridor can best emulate the MoC.

Since the range to which LD-MFDs and link MFDs can exist is equal to subset bias, the estimation of subset bias helps conjecture the maximum amount of LD bias. In terms of Case 2, the size of the range of LD-MFD with respect to the average density is $\frac{2\rho}{1+\rho}k_c$. The range increases as ρ increases, which might lead to a higher LD bias. When having the same ρ value, the possible range of LD-MFD is shorter in Case 3 than in Case 2. This is because the LD bias of Case 3 is affected by λ, ρ, and n. Especially for Case 3, the possible ranges of LD-MFD of two corridors with equal n and λ are proportional to the rate of ρ.

2.4.3. Remarks

Under the saturated initial condition, the subset bias is inevitable unless the traffic signal system (i) is perfect that never forms a queue (Case 1) or (ii) has a negligibly small red time portion under Case 2, or (iii) satisfies diminutive $2(n+1) - \lambda/(1+\rho)$ under Case 3. The LD bias occurs unless the signal timing is perfectly programmed or the detector positions are uniform. When LD bias is directly immeasurable, the subset bias can be used to estimate the maximum amount of LD bias.

3. Empirical Data Analysis

We recognized previously that some cases have no bias regardless of loop detector positions. When a bias was noticeable, the LD bias and subset bias were clearly distinguished. We now validate the applicability of the findings in the empirical data.

Courtesy of the UTD-19 dataset provided by Loder et al. (2019), the loop detector data for 40 cities worldwide are easily accessible. Aghamohammadi and Laval (2022) found the discrepancy of the MFD parameters in the dataset and raised the existence of bias from the loop detector position and its coverage. As mentioned earlier, no city can avoid LD bias unless loop detectors are uniformly distributed or the signal timing is perfectly planned, which is not expected in the real world. Unfortunately, LD bias cannot be accurately estimated without complete signal information. However, the subset bias, which can be solely obtained by the loop detector data, enables predicting the possible range of LD bias. This conveys that determining the factors affecting the extent of the subset bias will help understand LD bias. In this respect, we aim to investigate specific characteristics of the loop detector position that induces the subset bias.

Emulating the processing criteria from Aghamohammadi and Laval (2022), 10 cities (Paris, Bolton, Birmingham, Groningen, Innsbruck, Manchester, Melbourne, Rotterdam, Torino, and Utrecht) are ruled out due to the long aggregation intervals and incomplete occupancy measurements. Bordeaux and Constance are also excluded due to the incomplete loop detector information. Hereafter, the characteristics of loop detectors are explored for the remaining 28 cities.

The loop detector installation information of each city is described in Table 3. The number of time intervals can be viewed as the temporal sample size. The number of detectors shows the sample size to compute average network flows and occupancies for a given temporal point. The area is calculated with a convex-hull method that encircles the edge of LD-installed links. As the number of detectors and the area covered differ by city, the spatial density of loop detectors is calculated by dividing the former by the latter. The length distribution of LD-installed links is represented with a mean, standard deviation, skewness, and kurtosis, and so does the distance distribution from the detector to the downstream traffic signal. The distribution of the relative position of the detector, which is the distance divided by the link length, is also presented for normalization.

Table 3: Descriptive statistics related to loop detector in 28 cities

City	No. of time interval	No. of detector	Area covered (km^2)	Detector Density ($\#/km^2$)	Length of LD-installed links (m)				Distance from LD to downstr. signal (m)				Relative Position of LD (%)			
					mean	std	skew	kurt	mean	std	skew	kurt	mean	std	skew	kurt
Augsburg	5757	445	29.3	15.2	219	167	1.7	3.8	37	36	5.6	38.3	24	19	1.3	1.1
Basel	2016	60	2.0	30.7	214	127	1.5	2.4	74	57	5.1	31.3	41	22	0.7	-0.6
Bern	2016	441	20.6	21.4	256	216	1.2	0.9	33	74	8.1	81.5	17	23	1.9	3.2
Bremen	6720	462	100.3	4.6	324	241	1.5	4.0	152	208	2.1	4.4	45	34	0.3	-1.6
Cagliari	24000	80	25.5	3.1	554	304	0.6	0.5	283	231	0.7	-0.6	49	29	0.1	-1.3
Darmstadt	17873	204	29.8	6.8	312	239	1.6	2.6	202	206	1.81	4.8	61	33	-0.6	-1.3
Essen	8159	36	11.4	3.2	419	180	0.9	0.4	270	160	1.3	2.3	66	26	-0.7	-0.8
Frankfurt	288	73	1.8	39.6	216	141	1.0	1.3	79	101	1.9	3.2	33	25	0.6	-1.2
Graz	2880	246	13.0	18.9	229	169	2.3	5.8	94	97	3.3	14.1	47	29	0.3	-1.4
Hamburg	50142	325	18.6	17.5	218	178	2.0	5.3	66	114	4.7	27.8	32	27	1.0	0.0
Kassel	1171	422	54.4	7.8	291	236	1.6	3.0	77	150	4.0	17.8	28	27	1.3	0.5
London	6454	4736	190.1	25.0	218	182	2.6	10.5	124	122	4.4	32.6	64	28	-0.6	-0.9
Los Angeles	2879	1643	55.6	29.6	214	122	1.5	3.0	78	30	4.6	54.4	45	21	0.3	-1.1
Luzern	175116	135	7.2	18.7	172	114	1.6	2.9	82	91	2.6	7.6	49	30	0.1	-1.3
Madrid	4560	977	35.1	27.9	182	101	2.0	8.6	95	53	2.5	13.9	58	24	-0.4	-0.9
Marseille	14400	146	45.4	3.2	205	117	1.6	3.3	158	102	1.3	2.0	77	22	-1.6	1.7
Munich	288	280	138.9	2.0	390	314	2.1	5.7	146	201	2.2	5.7	33	27	0.8	-0.6
Santander	2239	190	9.4	20.3	246	210	2.1	4.5	119	164	3.1	11.7	48	34	0.1	-1.6
Speyer	6720	136	8.4	16.2	302	214	1.0	0.3	31	22	0.0	-1.0	16	18	2.0	3.6
Strasbourg	9349	138	27.6	5.0	281	150	1.5	3.4	156	106	1.5	4.3	56	25	-0.4	-0.8
Stuttgart	2304	20	46.6	0.4	969	569	0.2	-1.1	662	408	0.2	-1.3	70	20	-0.4	-0.1
Taipei	6620	353	41.5	8.5	213	86	1.7	5.3	107	62	1.9	7.8	50	17	0.0	-0.7
Tokyo	17857	228	23.5	9.7	165	115	3.8	19.1	117	82	4.3	28.6	73	18	-1.6	2.6
Toronto	5856	163	5.4	30.2	241	118	1.8	7.1	172	109	2.6	11.8	70	17	-1.1	0.4
Toulouse	3360	455	87.9	5.2	265	178	1.9	4.0	224	163	2.2	5.7	85	17	-2.3	5.2
Vilnius	481	444	15.2	29.2	238	160	1.7	3.4	82	114	3.1	10.4	38	31	0.9	-0.6
Wolfsburg	6720	104	25.2	4.1	540	383	0.6	-1.0	271	335	1.5	1.5	46	34	0.0	-1.7
Zurich	3359	1012	65.5	15.5	255	228	2.6	10.3	68	116	4.1	23.0	31	30	0.9	-0.6
Average	13914	499	40.54	14.98	298				145				48			
Std	33168	902	43.29	11.01	163				124				18			

To consider the different link length distributions by city, a relative position along the link is used as the classification criteria. The lower the relative position, the closer the loop detector is to the downstream signal. The histogram of the relative position for each city can be found in Fig. 4. The distribution highly differs by cities; *e.g.*, right-skewed, left-skewed, inverted bell, etc. In particular, the loop detectors are located mostly downstream in Augsburg, whereas Toulouse has most installations upstream.

The detectors with a relative position of less than 33 percent are labeled downstream detectors, greater than 67 percent as upstream detectors, and otherwise considered midstream detectors. Fig. 5 illustrates the spatial distribution of loop detectors for each city. The detector density is visibly comparable with the link color. In Augsburg, where 445 detectors cover 29.3 square kilometers with a detector density of 15.2 detectors per square kilometer. In contrast, Toulouse with a detector density of 5.2 per square kilometer is much sparser and with loop detectors positioned mostly upstream.

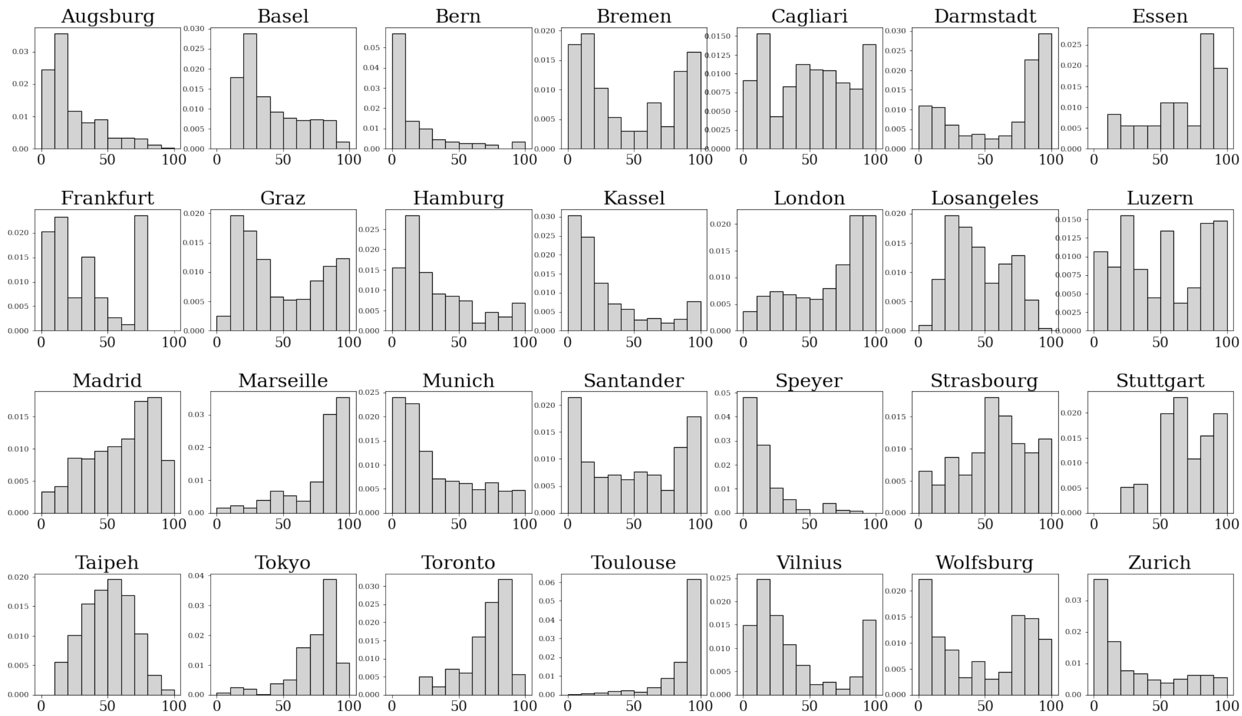


Fig. 4. Histogram of the relative position of loop detectors

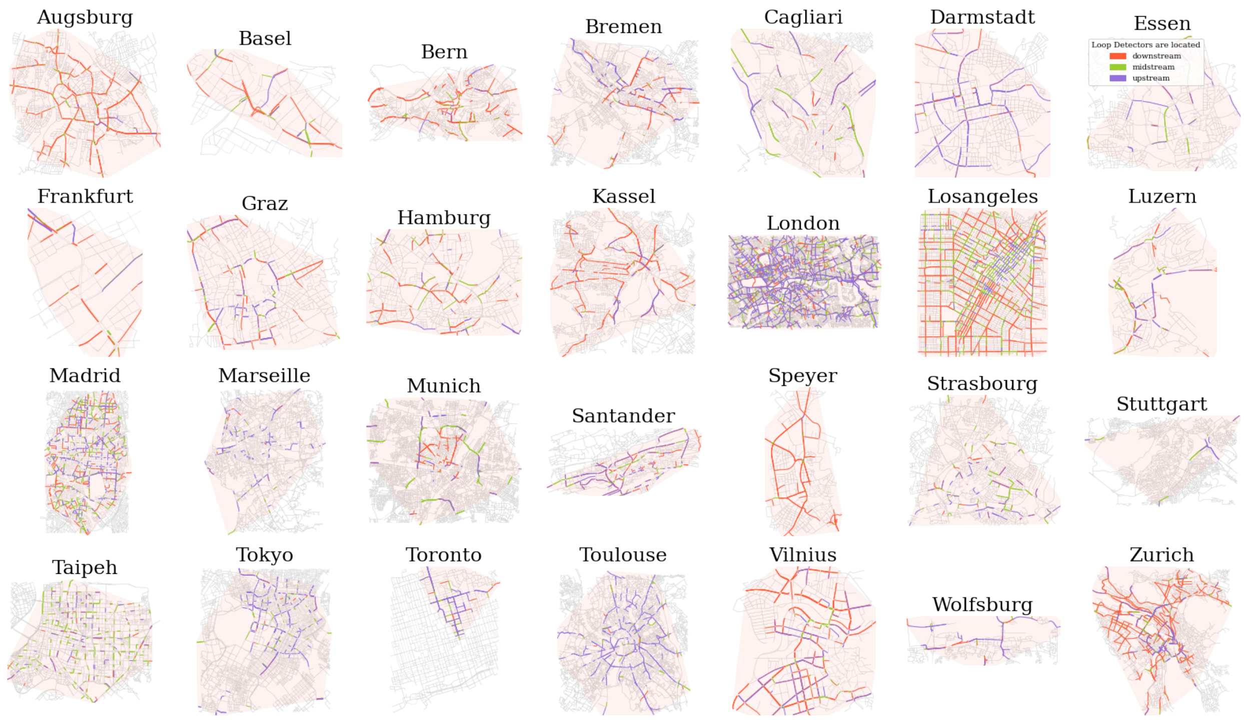
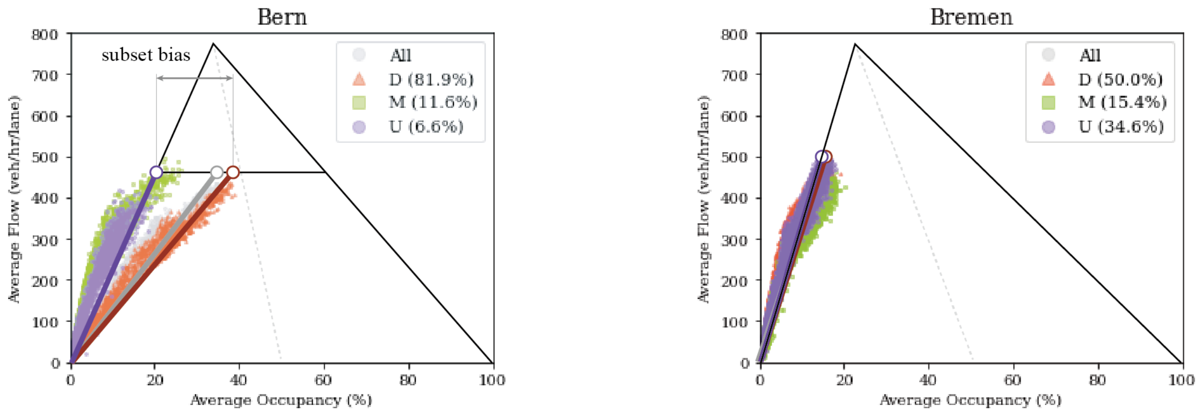


Fig. 5. Links colored by the loop detector position: street maps (Boeing, 2017) (gray), the link with LD located downstream (orange), midstream (green), upstream (purple), and convex-hull area covered by LD (light-coral)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 December 2022 doi:10.20944/preprints202212.0178.v1

As the variations of the relative position distribution and the spatial density are discovered, we now draw MFDs. It is conspicuous that some cities have different slopes of a free-flow branch by the position of loop detectors, while others present the same slope (Fig. 6 and Fig. 7). For example, in Bern, the downstream subset MFD has a relatively low slope compared to the midstream and the upstream. In Bremen, however, the subset MFDs are all overlapped regardless of positions. This aligns well with the observations from the analytical approach that the subset bias is not always extant.



(a) **(b)**

Fig. 6. Representation of subset bias: (a) Bern - subset bias observed; (b) Bremen - no subset bias observed.

However, the expectation that heterogeneous complex networks would always retain subset bias, is contradicted. This also tackles the contentions of the past literature that the downstream loop detectors tend to overestimate average occupancy. The evidence does not direct towards that the city would not have had traffic jams. Rather, we attribute it to the distribution of loop detector positions. The necessity to identify the determinants of the subset bias arises here.

According to Fig. 7, the position-based subsets of 28 cities can be classified into three: scattered, biased, and unbiased. Cities with a scattered MFD are the resultant of the insufficient number of detectors in certain positions such as downstream located loop detectors in Marseille. Augsburg, Darmstadt, Speyer, Stuttgart, Tokyo, Toronto, and Toulouse are classified as scattered. Cities with subset bias show different slopes by their position (Bern, Cagliari, Frankfurt, Graz, Hamburg, Kassel, London, Madrid, Munich, Santander, Vilnius, Wolfsburg, and Zurich). Otherwise, cities with no subset bias show overlapping shapes by subsets.

With all that classified, a simple logistic regression is performed to find variables that might explain the subset bias. Note that cities classified as scattered are excluded from the analysis since a high scatter hampers determining the subset bias. Logistic regression is a probabilistic statistical classification model applied when the dependent variable is categorical (Freedman, 2009, Boateng and Abaye, 2019). Although it is typically used to predict a binary outcome, here we utilize it to determine the significant variables that affect the subset bias. The dependent variable, Y, corresponds to the label for the MFD shape: 13 cities of biased ($Y = 1$) and 7 cities of unbiased ($Y = 0$). The logistic regression in an aggregated form is expressed with a natural logarithm of odds when there are n independent variables of $X_1, X_2, \cdots, X_n$ with $(n+1)$ coefficients of $\beta_0, \beta_1, \cdots, \beta_n$:

$$\ln\left(\frac{P(Y)}{1-P(Y)}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n \quad (6)$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 December 2022 doi:10.20944/preprints202212.0178.v1

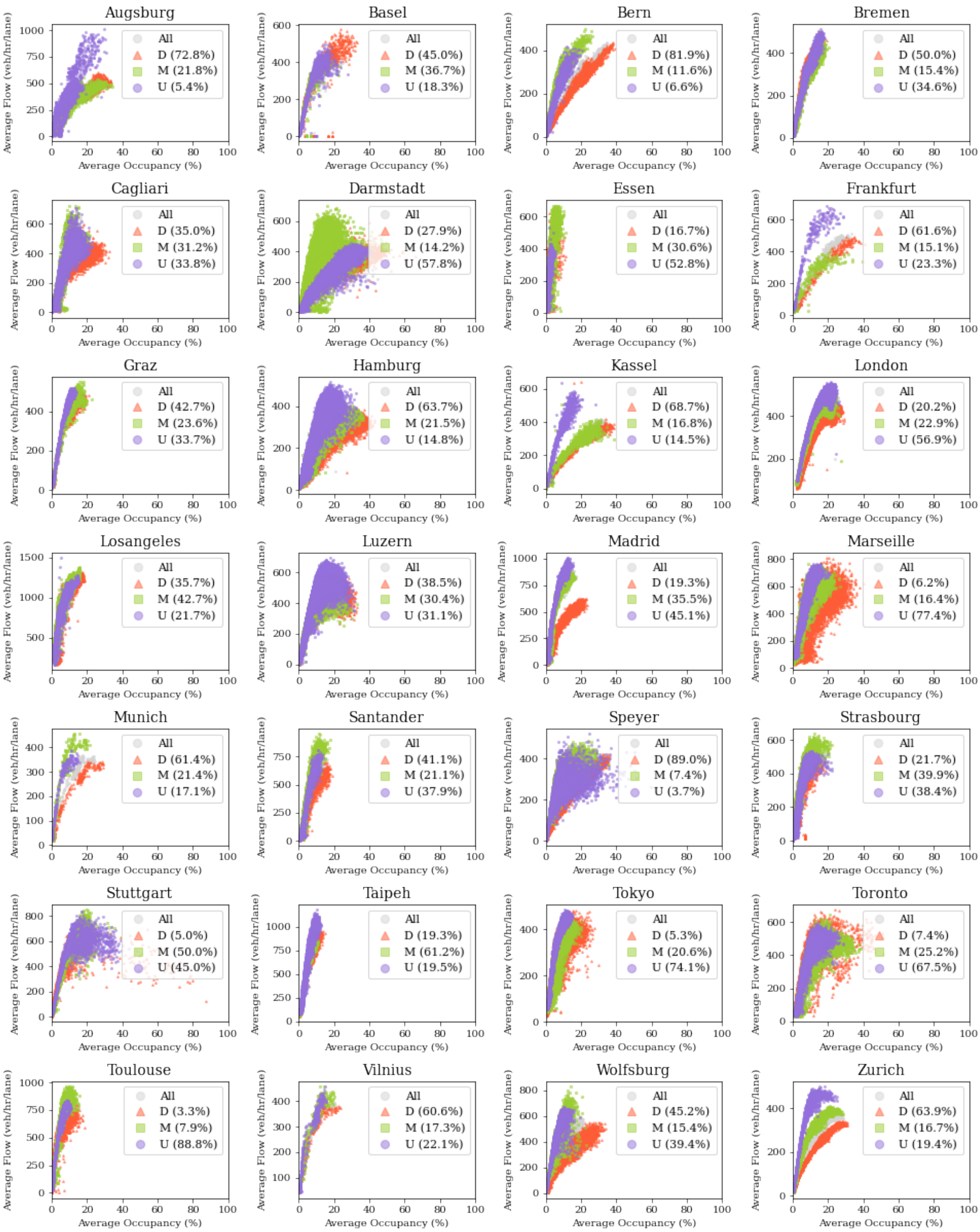


Fig. 7. MFD of all loop detectors, downstream subset (D), midstream subset (M), and upstream subset (U) for 28 cities

Here, the model aims to estimate the coefficients of independent variables. The odds are the probability that the outcome occurs to not occur. The odds ratio indicates how much odds are to be changed by the 1-unit increase of the independent variable, X_i, and it equals to e^{β_i}. The model evaluation is done by a likelihood-ratio (LLR) test which shows how strong a relationship between all of the independent variables and a dependent variable is. The null hypothesis is that all coefficients are equal to zero. The p-value of this test will determine whether the null hypothesis of the LLR test is rejected.

We consider all columns in Table 3 as the candidates for independent variables. Since the relative position (columns 12-14) is obtained by dividing the distance to the downstream signal (columns 9-11) by the link length (columns 6-8), we discard columns from the sixth to the eleventh in order to avoid collinearity. Similarly, as the detector density (column 5) is obtained by dividing the number of detectors (column 3) by the area covered (column 4), we drop columns 3 and 4. Then we execute the logistic regression with the remaining candidates and then apply backward step-wise elimination until the independent variables in the model are all at least 5% significant (Chowdhury and Turin, 2020). Table 4 summarizes the process.

Since each of the p-values is below 0.05 in the final step, the variables of mean and the standard deviation of relative position are statistically significant at the 95% confidence level. Also, the LLR p-value of 0.014 states that the proposed model was more effective than the null model. Based on the coefficients, the odds ratio is obtained as $e^{-0.087} = 0.92$ and $e^{0.170} = 1.19$ for the mean and the standard deviation, respectively. A 1% increase in the mean of relative position results in an 8% decrease in the probability of position-based subsets having different slopes. Similarly, a 1% increase in the standard deviation results in a 19% increase in the probability of having the subset bias. In other words, if the loop detectors are mostly located downstream and have a large variation, subset MFDs of each position are more likely to have different free-flow branch slopes. Although it is conjectured in many articles that the empirical MFD cannot accurately represent entire networks unless the loop detectors are sufficiently installed (Aghamohammadi and Laval, 2022, Ambühl et al., 2017), their spatial density was not a significant factor that decides the subset bias.

Table 4: Backward Elimination of Logistic Regression model

Model 1	Pseudo R-squared:	0.422	Log-Likelihood:	-7.4863	LLR p-value:	0.027
Variables	Coefficient	Std. Err.	t-statistics	p-value	Lower CI	Upper CI
Mean of Relative Position	-0.260	0.130	-1.998	0.046	-0.515	-0.005
Std. Dev. of Relative Position	0.572	0.288	1.984	0.047	0.007	1.136
Skewness of Relative Position	-5.023	3.123	-1.609	0.108	-11.144	1.098
Kurtosis of Relative Position	3.604	2.710	1.330	0.184	-1.708	8.916
Detector Density	0.104	0.072	1.447	0.148	-0.037	0.244
Model 2	Pseudo R-squared:	0.289	Log-Likelihood:	-9.2124	LLR p-value:	0.058
Variables	Coefficient	Std. Err.	t-statistics	p-value	Lower CI	Upper CI
Mean of Relative Position	-0.164	0.1	-1.671	0.095	-0.356	0.028
Std. Dev. of Relative Position	0.278	0.161	1.721	0.085	-0.039	0.594
Skewness of Relative Position	-1.702	1.843	-0.923	0.356	-5.314	1.910
Detector Density	0.064	0.063	1.020	0.308	-0.059	0.187
Model 3	Pseudo R-squared:	0.253	Log-Likelihood:	-9.6762	LLR p-value:	0.038
Variables	Coefficient	Std. Err.	t-statistics	p-value	Lower CI	Upper CI
Mean of Relative Position	-0.093	0.046	-2.051	0.040	-0.183	-0.004
Std. Dev. of Relative Position	0.161	0.077	2.109	0.035	0.011	0.311
Detector Density	0.033	0.048	0.693	0.488	-0.061	0.128
Model 4	Pseudo R-squared:	0.233	Log-Likelihood:	-9.9300	LLR p-value:	0.014
Variables	Coefficient	Std. Err.	t-statistics	p-value	Lower CI	Upper CI
Mean of Relative Position	-0.087	0.042	-2.055	0.040	-0.170	-0.004
Std. Dev. of Relative Position	0.170	0.075	2.279	0.023	0.024	0.317

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 December 2022 doi:10.20944/preprints202212.0178.v1

3.1. Insights from the Analytical Approach

LD bias cannot be obtained directly from loop-detector data since no complete information is accessible for typical cities. As noticed in the analytical approach, nevertheless, the subset bias provides the possible range of LD bias. That is, the upstream and downstream subsets offer the upper and lower bound for the link MFD. For example, the link MFD of Bern will have average occupancy between 20% and 40% at an average flow of 400 veh/hr. Still, the bound may be too wide to shed significant insights. We even cannot conclude whether the cities with a subset bias possess LD bias or not. However, it is meaningful that the cities with no subset bias are likely to have no LD bias.

We have proved the presumption from Aghamohammadi and Laval (2022) that the discrepancy between the MFD parameters might account for the loop detector position. Since the cuts are the upper bound that hems all points in, the bounds would be overestimated if the LD-MFD lies on the left of the link MFD, and otherwise cause underestimation. We can expect that the parameter estimation of the cities with no subset bias would reveal a consistent result.

4. Simulation

So far we have shown in the analytical part that the network parameters such as λ and ρ have a significant impact on the corridor MFD. In the empirical analysis, it is observed that the distribution of the loop detector position itself influences the subset bias in the network MFD. In this regard, we examine how the network parameter, λ, affects the bias in the network MFDs. We run network simulation experiments in the microscopic traffic simulation platform SUMO (Lopez et al., 2018) to verify and extend the previously described analysis results.

4.1. Simulation Settings

Simulations are carried out on a 10×10 grid network. All links have one lane per direction. Each intersection is controlled by a two-phase traffic signal without a protected left turn. Traffic signals for all internal nodes share constant parameters: the cycle length of $C = 90$ s, the green phase of $G = 45$ s, and the offset of $\theta = 0$ s. We adopt a triangular fundamental diagram. The traffic flow parameters are given by: the free-flow speed of $u = 54$ km/h, the wave speed of $w = 16.6$ km/h, the capacity of $Q = 1944$ veh/h, the jam density of $k_j = 153$ veh/km, and the critical density of $k_c = 36$ veh/km. With the above parameters fixed, four networks with different λ are constructed. For λ as 0.5, 1, 2, and 4, the link length corresponds to 80 m, 160 m, 320 m, and 640 m, respectively.

The trips are generated to promote homogeneous density distribution throughout the network by assigning identical demand at all interior links. The routes are randomly assigned with a turning probability of one-third each for turning left, right, or going straight. The U-turns are executed at the boundary links. Vehicles are allowed to leave the network at all links. Free-flow branch of MFD was reproduced by allowing vehicles to take all links as their destination, whereas the destination is restricted to a single boundary link to emulate the congested state.

Three loop detectors are set in every link each downstream, midstream, and upstream. We label the positions by the criterion which divides a link into thirds, as we used in our empirical data analysis. The loop detector is positioned randomly within the position range: *e.g.*, the downstream detector selects anywhere between the relative position of 0% and 33%. This positioning will reduce LD bias and help observe the impact of λ on the subset bias. The aggregation interval is set to 180 s.

4.2. Simulation Results

We have taken both signal and network settings into account figuring out the influential factor of the MFD shape, by changing λ. Fig. 8 illustrates the MFDs of four networks with different λ. The average flow and the density are each normalized by the value of Q and k_j, respectively. Note that the free-flow speed and the wave speed of FD are each 4.3 and 1.3 when normalized. The forward cuts of the MoC and the SMoC calculated through the parameters of each network are also superimposed.

We can verify that the random positioning within the position range does not hinder the accurate representation of the link MFD since the link MFD and LD-MFD aligns well on all four figures. Also, it is observed that the link MFD is well bounded by the forward cuts of the MoC and the SMoC. As discussed in both previous sections, free-flow branch slopes are high by order of upstream, midstream, and downstream subsets. In view of a congested branch, the downstream subset overestimates its slope and the upstream subset underestimates it. Hence, from the viewpoint of subset bias, the downstream and upstream subsets show each left-skew and right-skew distribution from the link

MFD. We have explicitly shown that the upstream subset is not bounded by the forward cut. This empowers our anticipation mentioned in 3.1. that estimating network parameters using LD-MFD will likely suppose longer blocks than the original if the network has detectors installed mostly upstream. The midstream subsets of all four figures are considerably close to the upstream subset, meaning that the jam state is not likely to propagate beyond the third downstream point. Following are the main observations as lambda increases.

- The maximum average flow increases since the long blocks are not prone to spillbacks.
- It is observed that free-flow branches of other MFDs including the link MFD approach that of the upstream subset. The free-flow branch of the upstream subset also gets steeper and closer to the free-flow speed u_f. This implies that the duration of the traffic jam is canceled out by the long block length.
- Larger λ assures the smaller subset bias. While the free-flow branches highly differ by the subsets in shorter blocks, these become overlapped as λ increases. Recalling that we have stated that the possible range of LD bias is equivalent to the subset bias, longer blocks have less chance to show bias in the LD-MFD.
- This is also analytically demonstrated by the slope of MoC and SMoC forward cut (Table 5). The average speed of the observer passing the origin increases when λ gets larger. Namely, the difference between the free-flow slope of the upstream subset and the slope of forward cuts becomes smaller. For long blocks, $u_{\gamma max}$ approximates the free-flow speed u_f. However, in short blocks, $u_{\gamma_{max}}$ is just half of the free-flow speed.

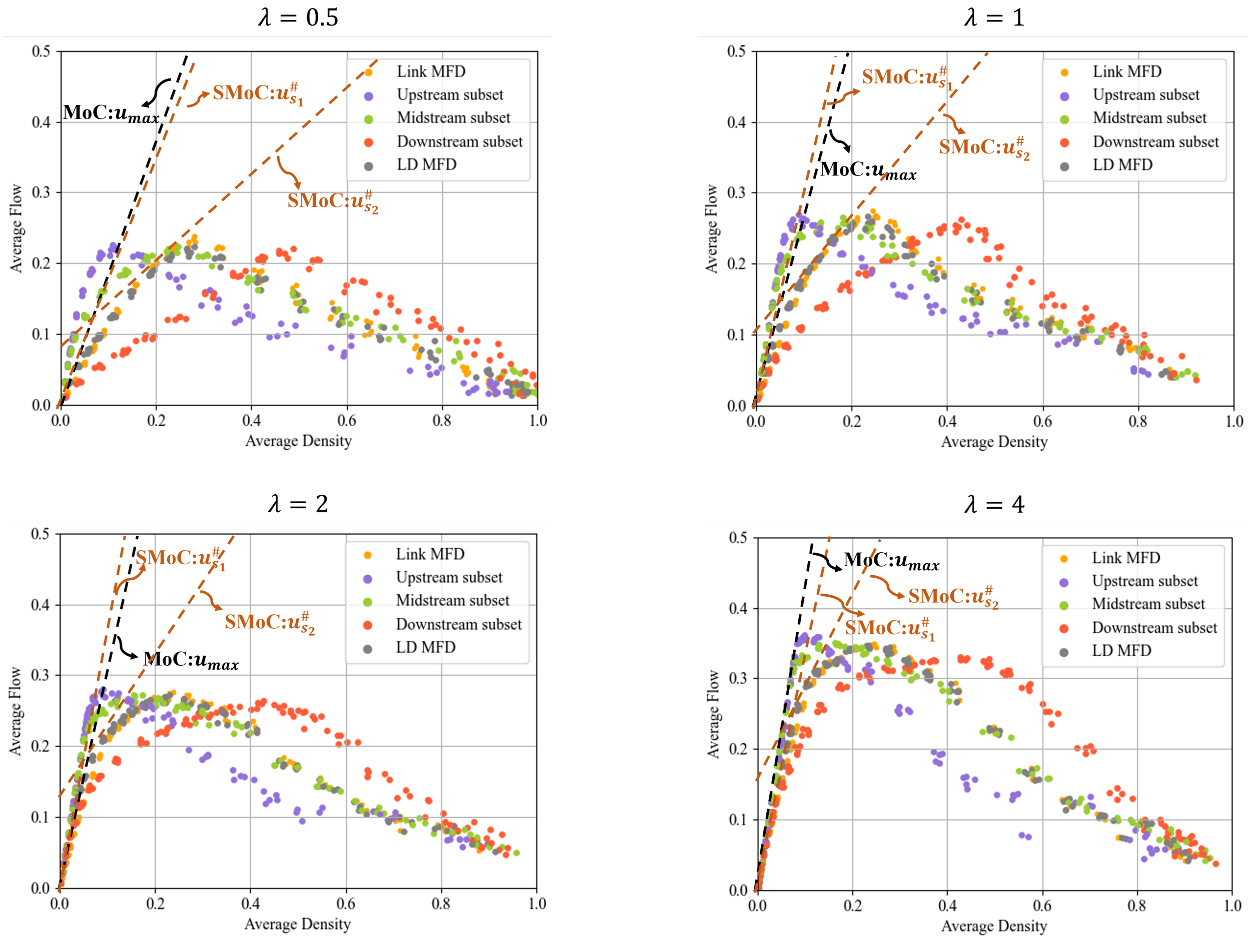


Fig. 8. The link MFD, the LD-MFD, and the position-based subsets for different values of λ

Table 5: The forward cut parameters of the MoC and the SMoC. Note that γ_{max} and $u_{\gamma_{max}}$ are the number of blocks that the observer with free-flow speed can pass without stopping and the average speed of the observer in the forward cut passing the origin, respectively. $u^{\#}_{s_1}$ and $u^{\#}_{s_2}$ are the average speed of the observer in SMoC, each corresponds to the steepest forward cut and the next steepest forward cut. Also note that $c = (1+\delta^2)\rho^2/(\lambda(1+\rho))$.

Variables		$\lambda = 0.5$	$\lambda = 1$	$\lambda = 2$	$\lambda = 4$
MoC	γ_{max}	9	5	3	2
	$u_{\gamma_{max}}$ (km/h)	**27**	**32.04**	**38.16**	**51.84**
	$u_{\gamma_{max}}$ (normalized)	2.15	2.55	3.04	4.13
SMoC	c	1	0.5	0.25	0.125
	$u^{\#}_{s_1}$	2	2.67	3.2	3.55
	$u^{\#}_{s_2}$	0.6	0.83	1.125	1.42

5. Conclusion

This paper presented an analysis of the impacts of loop detector position within the link on the resulting empirical MFDs. Previous research has been based either on empirical data or simulation. Here, we have (i) added an analytical approach based on kinematic wave theory that enables the explanation of these impacts in the corridor MFD, (ii) postulated a logistic regression model based on empirical data to predict the occurrence of bias on a given network, and (iii) revealed that the network parameter λ plays a key role in the bias magnitude.

In the analytical approach, the symmetric triangular diagram on a corridor was used to envision possible biases induced by the nature of loop detectors. Subject to the saturated initial condition constrained by the shock wave speed, the time-space diagram was classified according to the values of λ, ρ, and n. Formulae of the link MFD, LD-MFD, and the position-based subsets were cataloged. Several visualizations of MFDs and biases were presented. Under an ideal signalization, neither LD bias nor subset bias is apparent. If the network is programmed for the first vehicle of the green time to arrive at the next intersection during the green time (e.g., $\lambda < 1$), the subset bias only depends on ρ. If the first vehicle arrives during the red time of the next intersection, the subset bias is subject to λ, ρ, and n. It was proved that the LD bias is inevitable unless the signal is programmed perfectly or red time is negligible or the loop detectors are uniformly distributed. Also, the possible range of LD-MFD can be obtained by the subset bias formulas presented here. The analytical approach can be improved by considering an offset as an additional variable and identifying MFDs in unsaturated initial conditions.

By analyzing the loop detector data of 28 cities provided by UTD-19, we observed a wide variation in loop detector installation parameters such as coverage area or mean link length. When MFDs were subset, some cities showed different free-flow slopes for each subset, as predicted by the theory for a corridor, while others overlapped and therefore exhibit no subset bias. Our logistic regression results strongly suggest that this phenomenon can be explained partially by the mean and standard deviation of loop-detector positions alone. This indicates that the subset bias can be reduced by favoring placements upstream and with low variance. Of course, the other models in Table 4 are also instructive as they exhibit the intuitively correct coefficient sign. In particular, the skewness appears as a strong factor, confirming earlier results that the uniform distribution tends to minimize subset bias.

Our simulation results indicated that the overlapping phenomenon described in the above paragraph can be explained by the network parameter λ. For short-block networks ($\lambda < 1$) we have seen that the different branches do not overlap and that they tend to overlap for very long-block networks ($\lambda \gg 1$). Although the empirical data does not include signal settings to verify these hypotheses, this finding highlights the importance of parameterizing urban networks according to their λ-value.

Finally, the results of this paper strongly indicate that a correction method can be devised to improve the estimation of the link MFD using loop detector data. Overlapping position-based subset MFDs like Fig. 6(b) are fortunate to not require correction. However, when the subset bias is observed (e.g., Fig. 6(a)), LD-MFD may not accurately represent link MFD, which requires a correction method. As we have seen, both the topology of the network such as λ, and the distribution of positions should play a key role. This is currently being investigated by the authors.

Acknowledgments

This research was partially funded by NSF Awards #1932451 and by the STRIDE University Transportation Center.

References

Aghamohammadi, R., Laval, J.A., 2022. Parameter estimation of the macroscopic fundamental diagram: A maximum likelihood approach. Transportation Research Part C: Emerging Technologies 140, 103678.

Ambühl, L., Loder, A., Menendez, M., Axhausen, K.W., 2017. Empirical macroscopic fundamental diagrams: New insights from loop detector and floating car data, in: TRB 96th Annual Meeting Compendium of Papers, Transportation Research Board. pp. 17–03331.

Ambühl, L., Menendez, M., 2016. Data fusion algorithm for macroscopic fundamental diagram estimation. Transportation Research Part C: Emerging Technologies 71, 184–197.

An, K., Hu, X., Chen, X., 2020. Traffic network partitioning for hierarchical macroscopic fundamental diagram applications based on fusion of gps probe and loop detector data. arXiv preprint arXiv:2011.09075 .

Boateng, E.Y., Abaye, D.A., 2019. A review of the logistic regression model with emphasis on medical research. Journal of data analysis and information processing 7, 190–207.

Boeing, G., 2017. Osmnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. Computers, Environment and Urban Systems 65, 126–139.

Buisson, C., Ladier, C., 2009. Exploring the impact of homogeneity of traffic measurements on the existence of macroscopic fundamental diagrams. Transportation Research Record 2124, 127–136.

Chowdhury, M.Z.I., Turin, T.C., 2020. Variable selection strategies and its importance in clinical prediction modelling. Family medicine and community health 8.

Courbon, T., Leclercq, L., 2011. Cross-comparison of macroscopic fundamental diagram estimation methods. Procedia-Social and Behavioral Sciences 20, 417–426.

Daganzo, C.F., 2005. A variational formulation of kinematic waves: Solution methods. Transportation Research Part B: Methodological 39, 934–950.

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

Ding, H., Guo, F., Zheng, X., Zhang, W., 2017. Traffic guidance–perimeter control coupled method for the congestion in a macro network. Transportation Research Part C: Emerging Technologies 81, 300–316.

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.

Edie, L.C., et al., 1963. Discussion of traffic stream measurements and definitions. Port of New York Authority New York.

Freedman, D.A., 2009. Statistical models: theory and practice. cambridge university press.

Gayah, V.V., Gao, X.S., Nagle, A.S., 2014. On the impacts of locally adaptive signal control on urban network stability and the macroscopic fundamental diagram. Transportation Research Part B: Methodological 70, 255–268.

Geroliminis, N., Boyacı, B., 2012. The effect of variability of urban systems characteristics in the network capacity. Transportation Research Part B: Methodological 46, 1607–1623.

Geroliminis, N., Daganzo, C.F., 2008. Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. Transportation Research Part B: Methodological 42, 759–770. doi:10.1016/j.trb.2008.02.002.

Geroliminis, N., Sun, J., 2011. Hysteresis phenomena of a macroscopic fundamental diagram in freeway networks. Procedia-Social and Behavioral Sciences 17, 213–228.

Greenshields, B.D., Bibbins, J.R., Channing, W., Miller, H.H., 1935. A study of traffic capacity.

Huang, J., Hu, M.B., Jiang, R., Li, M., 2018. Effect of pre-signals in a manhattan-like urban traffic network. Physica A: Statistical Mechanics and its Applications 503, 71–85.

Ji, Y., Luo, J., Geroliminis, N., 2014. Empirical observations of congestion propagation and dynamic partitioning with probe data for large-scale systems. Transportation Research Record 2422, 1–11.

Kim, E.J., Kim, D.K., Kho, S.Y., Chung, K., 2020. Spatiotemporal filtering method for detecting kinematic waves in a connected environment. PloS one 15, e0244329.

Kong, Q.J., Li, Z., Chen, Y., Liu, Y., 2009. An approach to urban traffic state estimation by fusing multisource information. IEEE Transactions on Intelligent Transportation Systems 10, 499–511.

Laval, J.A., Castrillón, F., 2015. Stochastic approximations for the macroscopic fundamental diagram of urban networks. Transportation Research Procedia 7, 615–630.

Laval, J.A., Chilukuri, B.R., 2016. Symmetries in the kinematic wave model and a parameter-free representation of traffic flow. Transportation Research Part B: Methodological 89, 168–177.

Leclercq, L., Chiabaut, N., Trinquier, B., 2014. Macroscopic fundamental diagrams: A cross-comparison of estimation methods. Transportation Research Part B: Methodological 62, 1–12.

Leclercq, L., Parzani, C., Knoop, V.L., Amourette, J., Hoogendoorn, S.P., 2015. Macroscopic traffic dynamics with heterogeneous route patterns. Transportation Research Procedia 7, 631–650.

Loder, A., Ambühl, L., Menendez, M., Axhausen, K.W., 2019. Understanding traffic capacity of urban networks. Scientific reports 9, 1–10.

Lopez, P.A., Behrisch, M., Bieker-Walz, L., Erdmann, J., Flötteröd, Y.P., Hilbrich, R., Lücken, L., Rummel, J., Wagner, P., Wiessner, E., 2018. Microscopic traffic simulation using sumo, in: 2018 21st International Conference on Intelligent Transportation Systems (ITSC), pp. 2575–2582. doi:10.1109/ITSC.2018.8569938.

Mahmassani, H.S., Williams, J.C., Herman, R., 1984. Investigation of network-level traffic flow relationships: some simulation results. Transportation Research Record 971, 121–130.
Mazloumian, A., Geroliminis, N., Helbing, D., 2010. The spatial variability of vehicle densities as determinant of urban network capacity. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 368, 4627–4647. doi:10.1098/rsta.2010.0099.
Min, J.H., Ham, S.W., Kim, D.K., Lee, E.H., 2022. Deep multimodal learning for traffic speed estimation combining dedicated short-range communication and vehicle detection system data. Transportation Research Record , 03611981221130026.
Saffari, E., Yildirimoglu, M., Hickman, M., 2022. Data fusion for estimating macroscopic fundamental diagram in large-scale urban networks. Transportation Research Part C: Emerging Technologies 137, 103555.
Smeed, R.J., 1967. The road capacity of city centers. Highway Research Record .
Yildirimoglu, M., Ramezani, M., Geroliminis, N., 2015. Equilibrium analysis and route guidance in large-scale networks with mfd dynamics. Transportation Research Procedia 9, 185–204.
Zhang, L., Yuan, Z., Yang, L., Liu, Z., 2020. Recent developments in traffic flow modeling using macroscopic fundamental diagram. Transport Reviews 40, 529–550. doi:10.1080/01441647.2020.1743918.
Zheng, N., Geroliminis, N., 2013. On the distribution of urban road space for multimodal congested networks. Procedia-Social and Behavioral Sciences 80, 119–138.