Upscaling Macroscopic Fundamental Diagram Estimation From the Equipped to Full Networks

Nandan Maiti[1], Manon Seppecher[1], and Ludovic Leclercq[1]

[1]Univ. Eiffel, ENTPE, LICIT-ECO7, Lyon, 69500, France

Short summary

Accurately estimating traffic variables across unequipped portions of a network remains a significant challenge due to the limited amount of sensors-equipped links, such as loop detectors and probe vehicles. A common approach is to apply uniform scaling, treating unequipped links as equivalent to equipped ones, which leads to a strong bias in MFD estimation. Two main approaches are proposed: (1) Hierarchical Network Scaling and (2) Variogram-based data imputation. The hierarchical scaling method categorizes the network into several clusters according to spatial and functional characteristics, applying tailored scaling factors to each category. The variogram-based imputation leverages spatial correlations to estimate traffic variables for unequipped links, capturing spatial dependencies in urban road networks. Validation results show that the hierarchical scaling approach yields the most accurate estimates, demonstrating reliable performance with as little as 5% uniform detector coverage, while the variogram-based method provides strong results with over 10% detector coverage.
Keywords: Network scaling, Macroscopic fundamental diagrams, Spatial imputation, Loop detectors, Equipped networks.

1 Introduction

The Macroscopic Fundamental Diagrams (MFDs) has emerged as a critical tool in network traffic management, providing a relationship between aggregate traffic variables such as flow, density, and speed across an entire urban road network (Daganzo & Geroliminis, 2008). This concept offers valuable insights for network traffic control (Keyvan-Ekbatani et al., 2015; Ampountolas et al., 2017), particularly in perimeter control strategies (Aboudolas & Geroliminis, 2013; Mariotte & Leclercq, 2019; Jiang & Keyvan-Ekbatani, 2023), where traffic inflow and outflow are managed to optimize network-wide performances. By monitoring traffic at a macroscopic level, the MFDs enable decision-makers to regulate traffic in real-time, improve congestion management, and enhance the overall efficiency of urban transportation systems.

Despite its potential, estimating reliable MFDs for a city network requires extensive empirical data. Loop detector device (LDD), the most commonly used data source for MFDs estimation, collects information from fixed sensors embedded in road infrastructure (Buisson & Ladier, 2009; Geroliminis & Sun, 2011a; Keyvan-Ekbatani et al., 2012; Saberi & Mahmassani, 2012; Aboudolas & Geroliminis, 2013; Keyvan-Ekbatani et al., 2013; L. Ambühl et al., 2021; Lee et al., 2023; Mousavizadeh & Keyvan-Ekbatani, 2024). However, LDD faces limitations, such as positional biases in speed estimation (Leclercq et al., 2014; Maiti, 2024), and uneven distribution of sensors in a network (Lee et al., 2023), leading to incomplete coverage. To address these limitations, floating-car devices (FCD), or probe is often used alongside LDD to enhance spatial coverage, primarily for improving spatial speed estimation (Geroliminis & Daganzo, 2008; Gayah & Daganzo, 2011; Geroliminis & Sun, 2011b; Mahmassani et al., 2013; Tsubota et al., 2014; L. Ambühl & Menendez, 2016; Yang Beibei et al., 2016; Du et al., 2016; Saeedmanesh & Geroliminis, 2016; L. . Ambühl et al., 2017; Dakic & Menendez, 2018; Mariotte, Leclercq, et al., 2020; Mariotte, Paipuri, & Leclercq, 2020). However, FCD has its own challenges, including the unknown and variable penetration rate of vehicles equipped with GPS in time and space. Estimating network-wide traffic flows based on the assumption of homogeneous FCD penetration introduces uncertainty in the flow estimation, particularly when coverage is sparse (Leclercq et al., 2014; Shim et al., 2019; Fu et al., 2020).

The limitations of these data sources underscore a broader issue: the network-wide traffic data used for MFD estimation is typically incomplete. LDD and FCD are often available only for

certain links in the network, resulting in what is known as an 'equipped network MFD', an MFD that only reflects the links where sensors or FCD are deployed (Mariotte, Leclercq, et al., 2020). Consequently, this equipped MFD may not be representative of the entire network. This is critical for simulation studies as the demand is usually given for the full region, and estimating capacity and other variables on the equipped network only creates a mismatch with trips that could totally or partially happen in the non-equipped network.

To scale the equipped network MFD to the full city network, previous studies have attempted to partition the network into homogeneous areas and apply a scaling factor that adjusts for the ratio between the total length of links in the network and the length of the equipped links, e.g., Mariotte, Leclercq, et al. (2020). While this approach can offer some level of approximation, it suffers from significant limitations. The partitioning of the network based on equipped network data can lead to erroneous estimations of homogeneous areas, as the unequipped links may exhibit different traffic characteristics (Jiang et al., 2023; Jiang & Keyvan-Ekbatani, 2023; Saeedmanesh & Geroliminis, 2016, 2017; Gu & Saberi, 2019; Johari et al., 2023). Furthermore, applying a uniform scaling factor across areas with varying types of links, such as arterials and local roads, may lead to inaccurate flow estimates.

There is a need for a more advanced approach to spatial scaling of the MFD. To overcome the challenges associated with uniform scaling, a new methodology that accounts for the link hierarchy within a network is proposed. This method involves developing separate scaling factors for different hierarchical levels of the network, such as major arterials, collectors, and local streets. By accounting for the differences in traffic behavior at various link levels, this approach offers a more accurate estimation of network flow variables, allowing for accurate scaling of MFD from sensor-equipped portions of the network to the entire city.

2 Methodology

The proposed methodology estimates traffic variables for an entire network based on partially sensor-equipped data using hierarchical network scaling and variogram-based imputation approaches.

Network Variables Estimation and Scaling Issues

Uniform Scaling or Baseline

Let's assume a traffic network having links (i) of n numbers ($|i| = n$), with a different hierarchy based on serviceability, defined as link class as t. Among all the links, some have loop detectors (LDs) count as j. Therefore, the flow in the links of LDs can be presented as $\{q_j \mid j \in i\}$. The lengths of the links in the network are represented by l_i. The overall network flow ($\hat{q}_N$) can be expressed by total travel distance (TTD) as network 'production' by all vehicles over the time-space region, shown as follows in (1):

$$\hat{q}_N = \frac{\sum_{i=1}^{n} TTD_i}{\sum_{i=1}^{n} l_i \Delta t} \tag{1}$$

The TTD is the sum of equipped network TTD ($TTD_{j,eq}$) and non-equipped network TTD ($TTD_{i \notin i,neq}$)

$$TTD = \sum_{i=1}^{n} TTD_i = \sum_{\forall j \in \{eq\}} TTD_{j,eq} + \sum_{\forall i \in \{neq\}} TTD_{i,neq} \tag{2}$$

The first term of (2) can be calculated from the LDs.

$$\sum_{\forall j \in \{eq\}} TTD_{j,eq} = \sum_{\forall j \in \{eq\}} q_j l_j \tag{3}$$

The second term of (2) can be directly estimated after the imputation method. Otherwise, we know that the expectation relation in (4a) holds for all situations. Note that the covariance term

in (4a) can be disregarded as it was proven negligible in all our data analyses.

$$E(q_i l_i) = E(q_i)E(l_i) + Cov(q_i, l_i) \tag{4a}$$

$$\frac{1}{p}\sum_{i\in\{neq\}}(q_i l_i) = \left(\frac{1}{p}\sum_{i\in\{neq\}} q_i\right)\frac{1}{p}\sum_{i\in\{neq\}} l_i \tag{4b}$$

$$\sum_{i\in\{neq\}}(q_i l_i) = \bar{q}_{neq}\sum_{i\in\{neq\}} l_i \tag{4c}$$

Here, $\bar{q}_{neq}$ is the mean flow in the non-equipped network, and p is the total number of non-equipped links ($|i \notin j| = p$). Also, assume m is the total number of equipped links ($|j| = m$) in the network with mean flow estimated as $\bar{q}_{eq}$ from LDs. If we assume the non-equipped network shares the same average flow and covariance as the equipped, then we can calculate the second term of (2), as (5).

$$\sum_{\forall i\in\{neq\}}(q_i l_i) = \bar{q}_{neq}\sum_{\forall i\in\{neq\}} l_i \tag{5a}$$

$$= \left(\frac{1}{m}\sum_{j\in\{eq\}} q_j\right)\sum_{\forall i\in\{neq\}} l_i \tag{5b}$$

$$= \bar{q}_{eq}\sum_{\forall i\in\{neq\}} l_i \tag{5c}$$

So, the TTD can be expressed as in (6).

$$TTD = \bar{q}_{eq} \times (\sum_{\forall i\in\{neq\}} l_i + \sum_{\forall j\in\{eq\}} l_j) \tag{6a}$$

$$= \bar{q}_{eq} \times l_{net} \tag{6b}$$

Here l_{net} is the total network length, including equipped and non-equipped links. The assumption, $\bar{q}_{neq} = \bar{q}_{eq}$ forms the foundation of the uniform scaling factor, where all variables are scaled by the factor of network length covered.
Unfortunately, this assumption is often invalid, and equipped links are not a randomly selected subset of the full network. We propose a hierarchical scaling approach that will cluster the equipped network into groups that more likely resemble parts of the non-equipped network.

Hierarchical Network Scaling

The general idea in the hierarchical network scaling method is that we want to approach $\bar{q}_{neq} = \bar{q}_{eq}$ at the cluster level. Similar to (5), we can use the $\bar{q}_{eq,t}$ to estimate the $\bar{q}_{neq,t}$, where t represents link clusters. One option is to use the network hierarchy to define the clusters. Thus, we propose the following scaling method for the hierarchical network approach by disregarding covariance term in (7).

$$\sum_{\forall t} TTD_{eq,t} = \sum_{\forall j,t} q_{j,t} l_{j,t} \tag{7}$$

$$E(q_{j,t} l_{j,t}) = E(q_{j,t})E(l_{j,t}) \tag{8}$$

$$\frac{1}{m}\sum_{\forall t}\sum_{j=1}^{m} q_{j,t} l_{j,t} = \sum_{\forall t}(\frac{1}{m}\sum_{j=1}^{m} q_{j,t})\frac{1}{m}\sum_{\forall t}\sum_{j=1}^{m} l_{j,t} \tag{9}$$

$$\sum_{\forall j,t} q_{j,t} l_{j,t} = \sum_{\forall t}(\bar{q}_{eq,t})\sum_{\forall j,t} l_{j,t} \tag{10}$$

Similarly, for non-equipped networks:

$$\sum_{\forall t} TTD_{neq,t} = \sum_{\forall i \in \{neq\}} q_{i,t} l_{i,t} \tag{11}$$

$$E(q_{i,t} l_{i,t} | \forall i \notin j, t) = E(q_{i,t} | \forall i \notin j, t) E(l_{i,t} | \forall i \notin j, t) \tag{12}$$

$$\sum_{\forall t} \sum_{i=1}^{p} q_{i,t} l_{i,t} = \sum_{\forall t} (\bar{q}_{neq,t}) \sum_{\forall t} \sum_{i=1}^{p} l_{i,t} \tag{13}$$

$$\sum_{\forall i \notin j,t} q_{i,t} l_{i,t} = \sum_{\forall t} (\bar{q}_{neq,t}) \sum_{\forall i \notin j,t} l_{i,t} \tag{14}$$

Assuming the mean flow for a non-equipped network is the same as an equipped network for a unique hierarchy, i.e., $(\bar{q}_{neq,t} = \bar{q}_{eq,t})$. Therefore, by applying (10) to (14), we get the following estimation of TTD for a non-equipped network.

$$\sum_{\forall i \notin j,t} q_{i,t} l_{i,t} = \frac{\sum_{\forall j,t} q_{j,t} l_{j,t}}{\sum_{\forall j,t} l_{j,t}} \sum_{\forall i \notin j,t} l_{i,t} \tag{15}$$

$$= \sum_{\forall j,t} q_{j,t} l_{j,t} \frac{\sum_{\forall i \notin j,t} l_{i,t}}{\sum_{\forall j,t} l_{j,t}} \tag{16}$$

Thus, the total network flow can be estimated from the equipped network information as follows:

$$\hat{q} = \frac{\sum_{\forall t} TTD}{\sum_{\forall i,t} l_{i,t} \Delta t} \tag{17}$$

$$= \frac{\sum_{\forall t} TTD_{eq} + \sum_{\forall t} TTD_{neq}}{\sum_{\forall i,t} l_{i,t} \Delta t} \tag{18}$$

$$= \frac{\sum_{\forall j,t} q_{j,t} l_{j,t} + (\sum_{\forall j,t} q_{j,t} l_{j,t} \frac{\sum_{\forall i \notin j,t} l_{i,t}}{\sum_{\forall j,t} l_{j,t}})}{\sum_{\forall i,t} l_{i,t} \Delta t} \tag{19}$$

Similarly, we can also derive the network average density from the equipped network. The network average density can be expressed by total travel time (TTT) as the 'accumulation' spent by all vehicles in the network over the time-space domain. Similar to the TTD in (2), we can estimate TTT in the equipped network as $TTT_{eq} = \sum_{\forall j} k_j l_j$. Local densities at LD-level (k_j) need to be corrected since the LDs suffer from location-biased and systematic errors. This study corrected LD-level density estimation as per the methods mentioned in Maiti (2024). Therefore, the total network density can be formulated from the corrected LD in the equipped network as follows:

$$\hat{k} = \frac{\sum_{\forall t} TTT}{\sum_{\forall i,t} l_{i,t} \Delta t} \tag{20}$$

$$= \frac{\sum_{\forall t} TTT_{eq} + \sum_{\forall t} TTT_{neq}}{\sum_{\forall i,t} l_{i,t} \Delta t} \tag{21}$$

$$= \frac{\sum_{\forall j,t} k_{j,t} l_{j,t} + (\sum_{\forall j,t} k_{j,t} l_{j,t} \frac{\sum_{\forall i \notin j,t} l_{i,t}}{\sum_{\forall j,t} l_{j,t}})}{\sum_{\forall i,t} l_{i,t} \Delta t} \tag{22}$$

The variables $q_{j,t}$ and $k_{j,t}$ represent flow and density in the equipped network of t link hierarchy, estimated at the detector level on a link.

Spatial Imputation: Variogram

Instead of scaling observations from the equipped network to represent the non-equipped network, an alternative approach is to impute flow and density values for all links and subsequently calculate the full MFD. To achieve this, we propose using a spatial variogram-based method for the imputation process.

Let the traffic flow data for certain links in a network be given as $\{(s_i, q_i)\}$, where $s_i = (x_i, y_i)$ represents the geographical coordinates (latitude and longitude) of the i-th known link, and q_i is the observed traffic flow at that link. The task is to estimate traffic flow $q(s_0)$ at unknown locations s_0 by leveraging spatial interpolation techniques based on a variogram-based model.

The variogram measures the spatial dependence of a random variable, in this case, traffic flow, across a network. The variogram $\gamma(h)$ represents how traffic flow differences are expected to change with increasing distance (h) between two locations. Unlike the traditional variogram model's spatial distance, in this study, the distance between two locations is measured by the shortest path distance along the road network. For any two points s_i and s_j separated by distance $h = \|s_i - s_j\|$, the variogram is defined as:

$$\gamma(h) = \frac{1}{2} E\left[(q(s_i) - q(s_j))^2\right] \tag{23}$$

Here, E denotes the expectation, and h is the shortest path distance between s_i and s_j.
The variogram is often approximated from data using the empirical variogram:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{s_i, s_j : \|s_i - s_j\| = h} (q_i - q_j)^2 \tag{24}$$

where $N(h)$ is the number of pairs of points separated by distance h. This empirical variogram helps identify the spatial structure in the traffic flow data. The variogram $\gamma(h)$ is typically modeled using one of several functional forms (e.g., spherical, exponential, Gaussian) based on the empirical semivariances derived from the data. For instance, the spherical variogram is expressed as:

$$\gamma(r) = \begin{cases} C_0 + C\left(\frac{3r}{2a} - \frac{r^3}{2a^3}\right), & \text{if } r \leq a \\ C_0 + C, & \text{if } r > a \end{cases} \tag{25}$$

where a is the range, C_0 is the nugget, and C is the sill, serve as constants for the model.

Spatial Interpolation: Kriging System Setup

The predicted traffic flow $\hat{q}(s_0)$ at location s_0 is given by:

$$\hat{q}(s_0) = \bar{q}(s_0) + \epsilon(s_0) \tag{26}$$

The deterministic component $\bar{q}(s)$ represents the mean traffic flow, and the stochastic component $\epsilon(s)$ represents random fluctuations in traffic flow. The random component can be defined as a weighted multiplication of the deviation of the observed traffic flow at s_i from the mean s_0. The kriging weights w_i are applied to these deviations, meaning that locations closer to s_0 (with higher spatial correlation) will have a greater influence on the prediction. This spatially weighted adjustment ensures that the predicted traffic flow at s_0 not only considers the global mean $\bar{q}$ but also incorporates the local variations (the deviations from the mean) based on the observed traffic flows at the nearby locations.

$$\hat{q}(s_0) = \bar{q} + \sum_{i=1}^{n} w_i \left(q(s_i) - \bar{q}\right) \tag{27}$$

Since $\bar{q}$ is assumed to be constant, the estimator simplifies to:

$$\hat{q}(s_0) = \sum_{i=1}^{n} w_i q(s_i) + \bar{q}\left(1 - \sum_{i=1}^{n} w_i\right) \tag{28}$$

For an unbiased estimator, the weights w_i must satisfy the constraint:

$$\sum_{i=1}^{n} w_i = 1 \tag{29}$$

Therefore, the predicted traffic flow at s_0 is given by:

$$\hat{q}(s_0) = \sum_{i=1}^{n} w_i q(s_i) \tag{30}$$

The expected value of the predicted traffic flow is equal to the expected value of the true traffic flow:

$$E\left[\hat{q}(s_0)\right] = \bar{q} = E\left[q(s_0)\right] \tag{31}$$

The kriging weights w_i are determined by solving the kriging system of $n+1$ equations for the weights w_i and Lagrange multiplier μ is:

$$\begin{pmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1n} & 1 \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2n} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma_{n1} & \gamma_{n2} & \cdots & \gamma_{nn} & 1 \\ 1 & 1 & \dots & 1 & 0 \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \\ \mu \end{pmatrix} = \begin{pmatrix} \gamma_{10} \\ \gamma_{20} \\ \vdots \\ \gamma_{n0} \\ 1 \end{pmatrix} \tag{32}$$

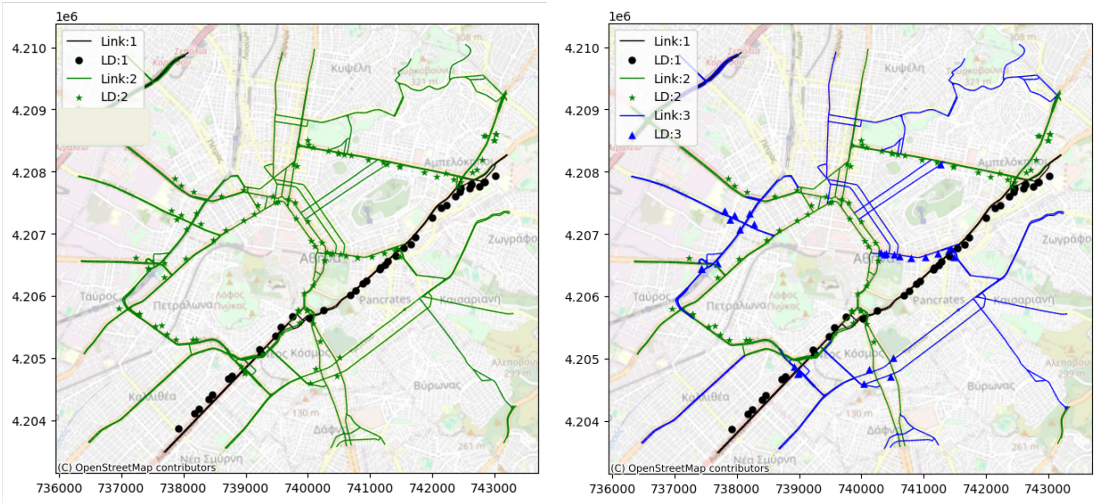
Where γ_{ij} is the semivariance between known points s_i and s_j, γ_{i0} represents the semivariance between s_i and s_0.

3 Data

This section describes the empirical data used in this study, focusing on LDD collected under varying levels of sensor deployment. Additionally, it outlines the link hierarchy classification, which forms the basis for the hierarchical scaling methodology proposed in this study.

Data Description

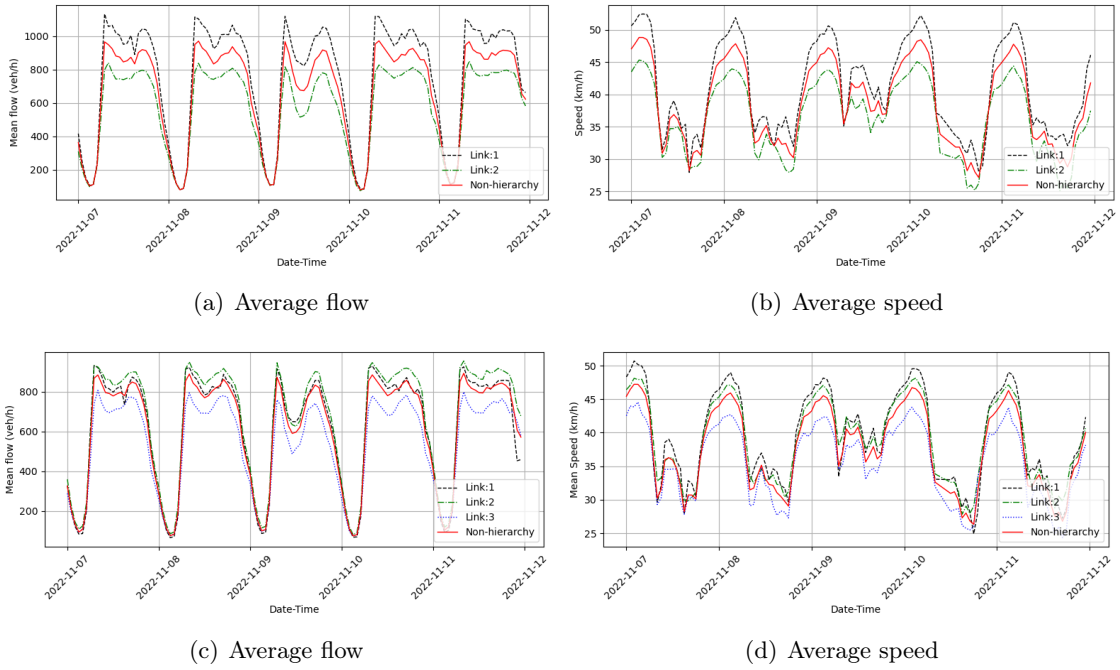
The study focuses on the road network of downtown Athens, Greece. The data for this study was collected from loop detectors over a weekday period from November 7th to 11th, 2022, covering a 24-hour time span each day.
The links in the road network were classified in two ways (see Figures 1). The first classification follows a three-hierarchy (3-H), where roads are categorized into three groups: Link-1, corresponding to the most critical roads, and Link-3, the least important major roads. The second classification follows a two-tier hierarchy (2-H), where the network is divided into two types: Link-1, representing the most important roads, and Link-2, which includes the remaining major roads. We can see in Figure 2 that the average speed and flow in each group are significantly different, so using a common flow average from the equipped network to estimate the average flow on the non-equipped network leads to strong bias.



(a) 2-Link Hierarchies

(b) 3-Link Hierarchies

Figure 1: Network fully equipped with LDD and features hierarchical links, including (a) two types link, (b) three types link



(a) Average flow

(b) Average speed

(c) Average flow

(d) Average speed

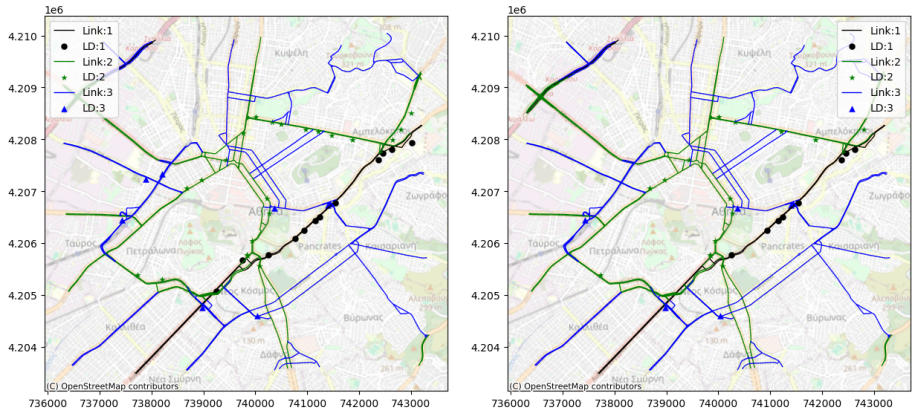
Figure 2: Comparison of average network flow ((a) two-hierarchy, (c) three-hierarchy) and speed ((b) two-hierarchy, (d) three-hierarchy) for different network configurations.

The study focuses on a road network comprising 2,456 links, spanning various hierarchical levels within a 150 km area. A fully equipped network is defined as one where every link is equipped with at least one LD. However, only 142 of these links currently have LDs installed. Therefore, the LDs percentage in the network is approximately 5.78%. To address this limitation, a partially equipped network was constructed, utilizing the data from these 142 links to estimate MFDs for validation purposes. For comparison, the fully equipped network was conceptualized as including the 142 LDs distributed across their respective links, as illustrated in Figures 1 (a, b). To evaluate our methodology, we also created partially equipped networks by randomly removing LDs from various links equally from each hierarchy in both the 2-H and 3-H hierarchical networks. Figure 3 illustrates the distribution of LDs in these partially equipped networks.
In this study, the proposed approach was validated using two distinct network types: the 2-H network, the 3-H network, and a baseline non-hierarchical network for comparison.

4 Results

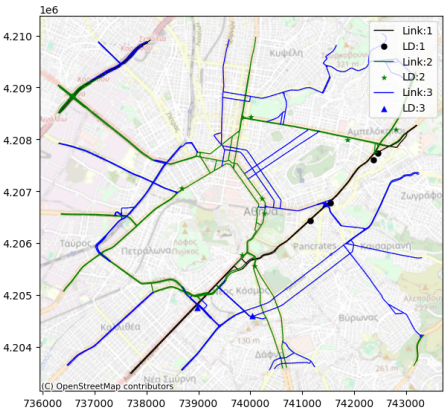
Comparison of proposed methods

The results presented in Figures 4, along with the RMSE values in Table 1, provide a detailed comparison of the proposed hierarchical network scaling method, the variogram-based approach and the baseline non-hierarchical network scaling. The scatter plots in Figure 4 further illustrate these trends, showing the spread of estimated flow values relative to actual flow values from a fully equipped network. As LD coverage decreases, the spread around the ideal $x = y$ line widens, particularly for the non-hierarchical method, which exhibits greater deviations from actual values. Conversely, the 3-Hierarchy and variogram approaches demonstrate tighter clustering around the ideal line, especially at lower LD coverages, reaffirming their superior accuracy under moderate to high LD densities.
These results underscore the robustness of hierarchical network scaling, particularly in sparse LD conditions where unstructured methods like the Variogram may falter. The 3-H method, in particular, shows resilience across all coverage levels, while the Variogram method performs strongly at 10% and higher LD coverage, demonstrating its potential for accurate network-wide estimations when sufficient spatial data is available. The baseline approach, applied under the lower usual case of 5% LD coverage, yields a network-average flow estimation with an RMSE of 234 veh/h. This value is notably four times higher than the error achieved using the proposed three-level hierarchical (3-H) network scaling method. These findings underscore the importance

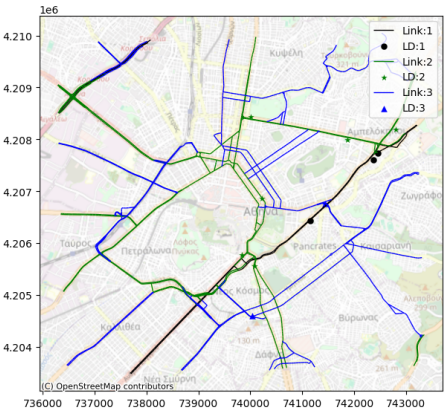


(a) 30% LD of Figure 1 network

(b) 20% LD of Figure 1 network



(c) 10% LD of Figure 1 network



(d) 5% LD of Figure 1 network

Figure 3: Partially loop detectors equipped networks (a) 30% LD equipped network, (b) 20% LD equipped network, (c) 10% LD equipped network, and (d) 5 % LD equipped network

of adopting a hierarchical network approach to improve the accuracy of traffic variable estimation at the network level.

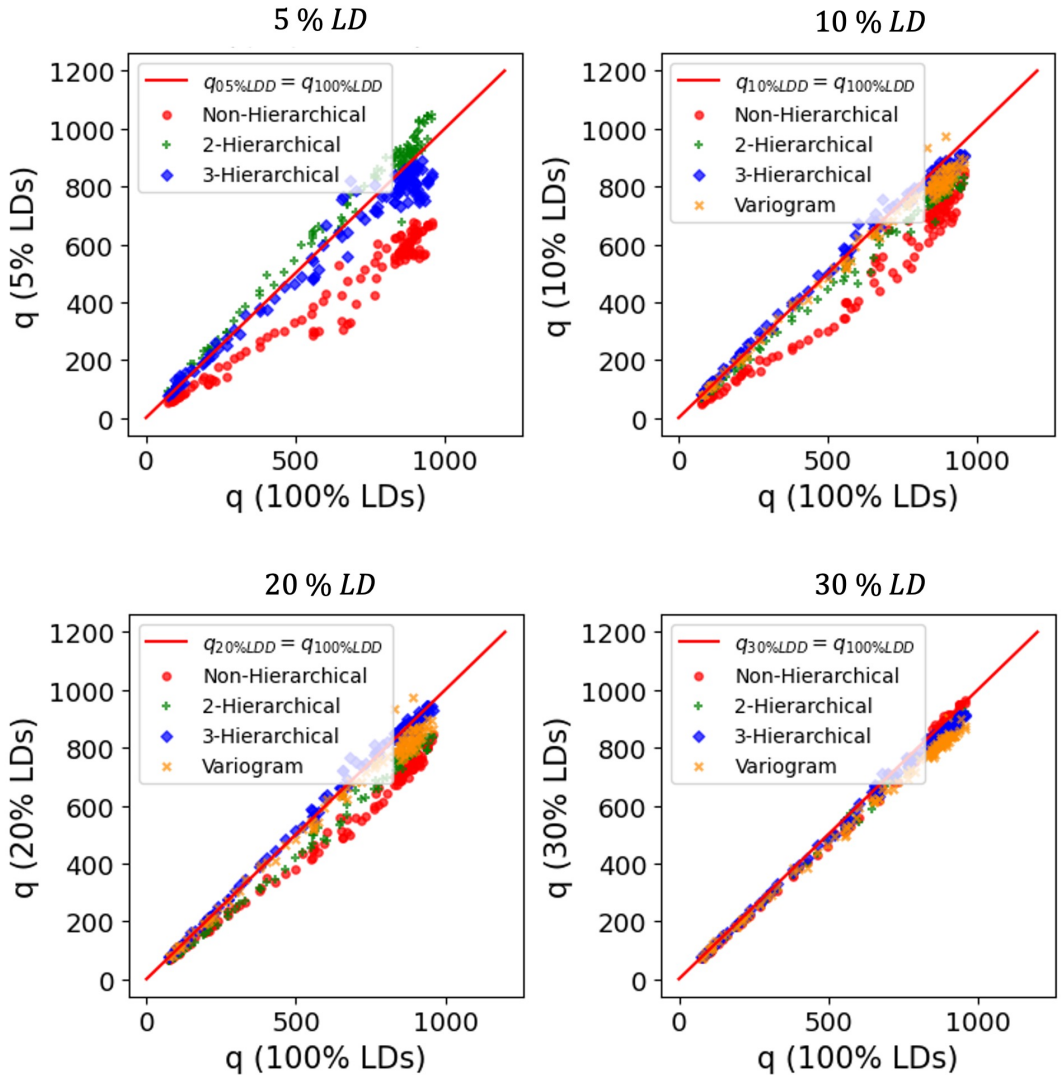


Figure 4: Comparison of actual network flow based on fully equipped network information (100% LD) with estimated flow derived from a partially equipped network. The x-axis represents the actual flow, while the y-axis denotes the estimated flow. From left to right, the subfigures illustrate comparisons for 5%, 10%, 20%, and 30% LD.

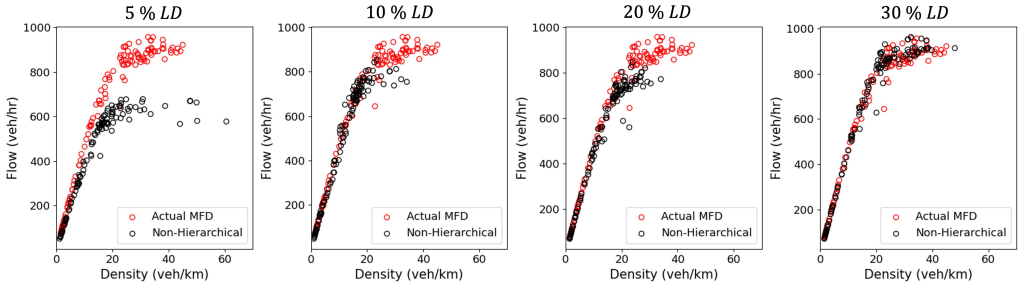
Table 1: Comparison of root mean square error (RMSE) of network flow, estimated using different network scaling methods

Day	Network (% LD)	3-H	2-H	Non-hierarchy	Variogram
day1	30	**12**	30	24	12
	20	**12**	85	128	17
	10	**26**	90	158	43
	5	**65**	110	226	-
day2	30	**13**	29	20	15
	20	**14**	83	127	29
	10	**26**	87	148	45
	5	**53**	108	232	-
day3	30	**25**	24	20	42
	20	**27**	70	133	27
	10	**28**	83	142	42
	5	**51**	120	260	-
day4	30	**12**	37	17	25
	20	**13**	94	119	15
	10	**31**	98	144	39
	5	**64**	117	224	-
day5	30	**13**	42	16	31
	20	**16**	94	110	34
	10	**35**	95	145	39
	5	**48**	115	229	-

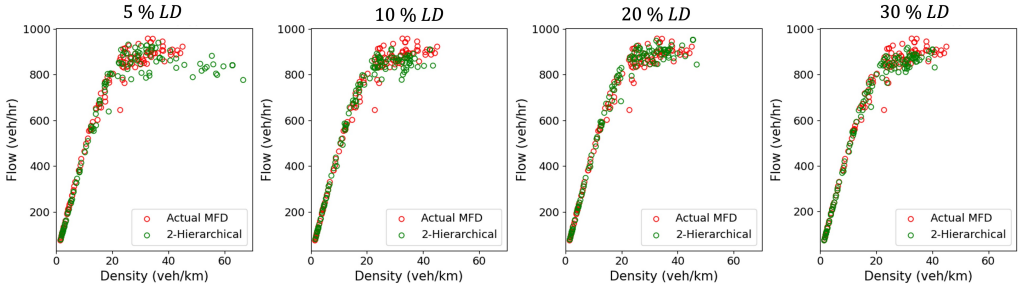
Comparing MFDs

This section demonstrates network MFDs estimated from the partially equipped network using the proposed scaling methodologies and compares them with the actual MFDs from the fully equipped network. In section 2.1, we proposed the network flow ($\hat{q}$) and density ($\hat{k}$) using the hierarchical scaling method, and section 2.2 described $\hat{q}$ estimation using the variogram. Similarly, we can estimate $\hat{k}$ using variogram methods by estimating detector-level density.

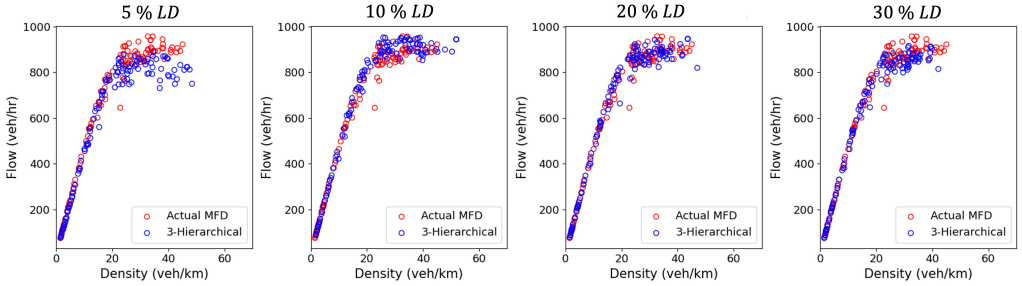
The estimated MFDs derived from the proposed methodologies across varying LD coverage percentages demonstrate clear differences in accuracy among the methods employed, as summarized in Figure 5 and Table 2. The 5% LD scenario represents the most challenging condition for MFD estimation due to sparse detector data. Among the methods, the three-hierarchy (3-H) statistical scaling achieves the lowest RMSE of 48.9 veh/h, significantly outperforming the two-hierarchy (2-H) scaling (51.8 veh/h) and the non-hierarchical approach (175.5 veh/h). The high coefficient of determination ($R^2 = 0.97$) for the 3-H method further highlights its robustness in low-data scenarios, maintaining strong alignment with the actual MFD. For 10% LD, the 3-H scaling method again delivers superior results with an RMSE of 45.3 veh/h, slightly better than the 2-H method (45.7 veh/h), variogram (55.2 veh/h), and markedly better than the non-hierarchical approach (137.6 veh/h). At 20% LD, the differences between the hierarchical methods narrow further, with 3-H achieving 36.5 veh/h RMSE compared to 37.3 veh/h for 2-H. Both methods significantly outperform the non-hierarchical method (133.5 veh/h), demonstrating the effectiveness of incorporating spatial structure in flow estimation. At the highest detector coverage of 30%, all methods perform well due to the availability of more extensive data.



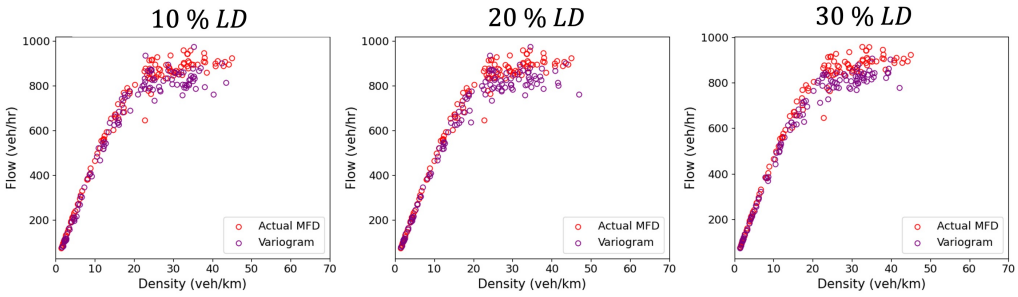
(a) Statistical scaling in non-hierarchical network



(b) Statistical scaling in two-hierarchical network



(c) Statistical scaling in three-hierarchical network



(d) Variogram scaling in non-hierarchical network

Figure 5: Comparison of estimated MFDs from partially equipped networks: (a, b, c) represent 5%, 10%, 20%, and 30% LD from left to right, and (d) shows 10%, 20%, and 30% LD from left to right, compared with the actual MFD derived from a fully equipped (100% LD) network.

Table 2: Comparison of estimated and actual MFDs

Method	Network Types	% LD	RMSE (veh/h)	R^2
Statistical	Non-hierarchy	30	39.9	0.98
		20	133.5	0.79
		10	137.6	0.78
		5	175.5	0.64
	Three-hierarchy	30	**35.9**	0.98
		20	**36.5**	0.98
		10	**45.3**	0.97
		5	**48.9**	0.97
	Two-hierarchy	30	37.1	0.98
		20	37.3	0.98
		10	45.7	0.97
		5	51.8	0.96
Variogram	Non-hierarchy	30	51.5	0.97
		20	54.6	0.96
		10	55.2	0.96

Using the hierarchical network clustering methods, we estimate the full network MFD, as illustrated in Figure 6, and compare these results with the baseline method. It is evident from the figure that the baseline method underestimates the MFD, particularly in the critical density regime. However, in the free-flow regime, both approaches yield comparable results. Recalling the network LD coverage as 5.68%, we can refer to the MFDs in Figure 5 (a) for 5% LD; the same pattern is observed here.

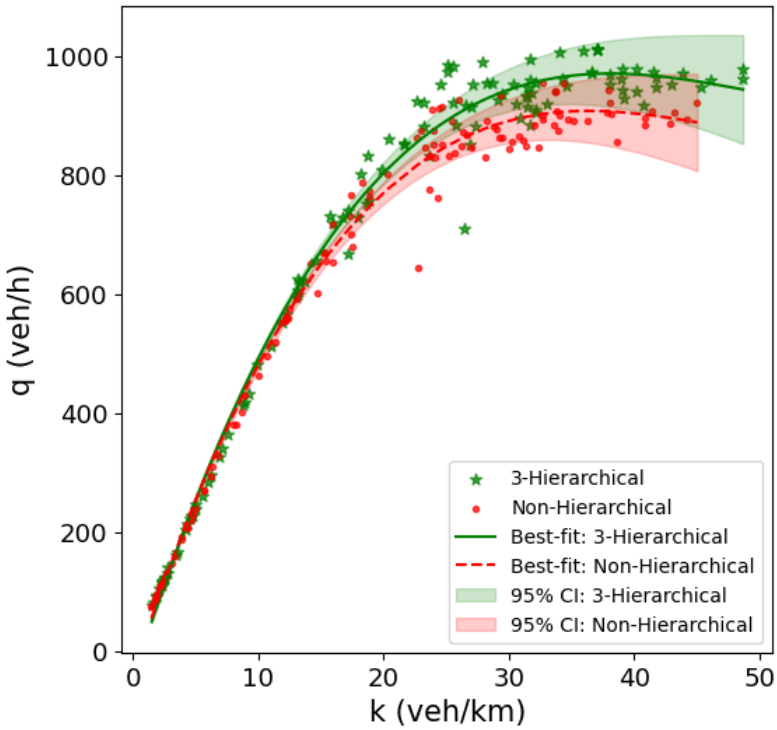


Figure 6: Comparing the estimated MFD using a hierarchical network cluster with three hierarchies to the baseline non-hierarchical method for the entire network consisting of 2456 links.

5 Conclusion

This study presents a novel methodology for estimating network-wide traffic characteristics in partially sensor-equipped urban networks. By integrating hierarchical network scaling and geospatial imputation techniques, the framework provides accurate estimations of MFDs, even in networks with sparse detector coverage. The results underscore the critical role of hierarchical scaling in improving network traffic variables and MFD estimation accuracy, particularly under sparse data conditions. The 3-H method consistently outperforms (lowest RMSE and highest R^2 values) the 2-H and variogram approaches across all LD levels, with its advantage most pronounced at low LD coverage (5% and 10%). This is likely due to the additional spatial structure incorporated in the hierarchical scaling, which allows it to better capture localized variations in flow and density. Interestingly, the variogram approach demonstrates competitive performance for LD coverage levels of 10% or more with small fluctuations in high flow regimes, but it was incompetent at lower (5%) LD due to insufficient data for spatial imputation. This limitation highlights the trade-offs between different methods in terms of data requirements and estimation capabilities. Therefore, this study bridges the gap between partial sensor data and full-network traffic estimation for cities with limited sensor coverage. Overall, the findings emphasize the importance of leveraging network hierarchical frameworks for MFD estimation, particularly in scenarios where data availability is limited.

Acknowledgement

This research has received funding from the European Union's Horizon Europe research and innovation program under Grant Agreement no. 101103808 (ACUMEN). The authors would like to thank Prof. Eleni I. Vlahogianni and her team from NTUA Athens for their support in providing traffic data.

References

Aboudolas, K., & Geroliminis, N. (2013). Perimeter and boundary flow control in multi-reservoir heterogeneous networks. *Transportation Research Part B: Methodological*, *55*, 265–281. doi: 10.1016/j.trb.2013.07.003

Ambühl, L., Loder, A., Leclercq, L., & Menendez, M. (2021, 5). Disentangling the city traffic rhythms: A longitudinal analysis of MFD patterns over a year. *Transportation Research Part C: Emerging Technologies*, *126*. doi: 10.1016/j.trc.2021.103065

Ambühl, L. ., Loder, A. ., Menendez, M. ., & Axhausen, K. W. (2017). Empirical Macroscopic Fundamental Diagrams New insights from loop detector and floating car data. Retrieved from `https://doi.org/10.3929/ethz-b-000167171` doi: 10.3929/ethz-b-000167171

Ambühl, L., & Menendez, M. (2016). Data fusion algorithm for macroscopic fundamental diagram estimation. *Transportation Research Part C: Emerging Technologies*, *71*, 184–197. doi: 10.1016/j.trc.2016.07.013

Ampountolas, K., Zheng, N., & Geroliminis, N. (2017, 10). Macroscopic modelling and robust control of bi-modal multi-region urban road networks. *Transportation Research Part B: Methodological*, *104*, 616–637. doi: 10.1016/j.trb.2017.05.007

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. doi: 10.3141/2124-12

Daganzo, C. F., & Geroliminis, N. (2008). An analytical approximation for the macroscopic fundamental diagram of urban traffic. *Transportation Research Part B: Methodological*, *42*(9), 771–781. doi: 10.1016/j.trb.2008.06.008

Dakic, I., & Menendez, M. (2018, 6). On the use of Lagrangian observations from public transport and probe vehicles to estimate car space-mean speeds in bi-modal urban networks. *Transportation Research Part C: Emerging Technologies*, *91*, 317–334. doi: 10.1016/j.trc.2018.04.004

Du, J., Rakha, H., & Gayah, V. V. (2016, 5). Deriving macroscopic fundamental diagrams from probe data: Issues and proposed solutions. *Transportation Research Part C: Emerging Technologies*, *66*, 136–149. doi: 10.1016/j.trc.2015.08.015

Fu, H., Wang, Y., Tang, X., Zheng, N., & Geroliminis, N. (2020, 9). Empirical analysis of large-scale multimodal traffic with multi-sensor data. *Transportation Research Part C: Emerging Technologies*, *118*. doi: 10.1016/j.trc.2020.102725

Gayah, V. V., & Daganzo, C. F. (2011). Clockwise hysteresis loops in the Macroscopic Fundamental Diagram: An effect of network instability. *Transportation Research Part B: Methodological*, *45*(4), 643–655. doi: 10.1016/j.trb.2010.11.006

Geroliminis, N., & Daganzo, C. F. (2008). Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transportation Research Part B: Methodological*, *42*(9), 759–770. doi: 10.1016/j.trb.2008.02.002

Geroliminis, N., & Sun, J. (2011a). Hysteresis phenomena of a macroscopic fundamental diagram in freeway networks. In *Procedia - social and behavioral sciences* (Vol. 17, pp. 213–228). Elsevier Ltd. doi: 10.1016/j.sbspro.2011.04.515

Geroliminis, N., & Sun, J. (2011b). Properties of a well-defined macroscopic fundamental diagram for urban traffic. *Transportation Research Part B: Methodological*, *45*(3), 605–617. Retrieved from `http://dx.doi.org/10.1016/j.trb.2010.11.004` doi: 10.1016/j.trb.2010.11.004

Gu, Z., & Saberi, M. (2019, 12). A bi-partitioning approach to congestion pattern recognition in a congested monocentric city. *Transportation Research Part C: Emerging Technologies*, *109*, 305–320. doi: 10.1016/j.trc.2019.10.016

Jiang, S., & Keyvan-Ekbatani, M. (2023, 9). Hybrid perimeter control with real-time partitions in heterogeneous urban networks: An integration of deep learning and MPC. *Transportation Research Part C: Emerging Technologies*, *154*. doi: 10.1016/j.trc.2023.104240

Jiang, S., Keyvan-Ekbatani, M., & Ngoduy, D. (2023, 3). Partitioning of urban networks with polycentric congestion pattern for traffic management policies: Identifying protected networks. *Computer-Aided Civil and Infrastructure Engineering*, *38*(4), 508–527. doi: 10.1111/mice.12895

Johari, M., Jiang, S., Keyvan-Ekbatani, M., & Ngoduy, D. (2023). Mode differentiation in partitioning of mixed bi-modal urban networks. *Transportmetrica B*, *11*(1), 463–485. doi: 10.1080/21680566.2022.2089271

Keyvan-Ekbatani, M., Kouvelas, A., Papamichail, I., & Papageorgiou, M. (2012). Exploiting the fundamental diagram of urban networks for feedback-based gating. *Transportation Research Part B: Methodological*, *46*(10), 1393–1403. doi: 10.1016/j.trb.2012.06.008

Keyvan-Ekbatani, M., Papageorgiou, M., & Papamichail, I. (2013, 8). Urban congestion gating control based on reduced operational network fundamental diagrams. *Transportation Research Part C: Emerging Technologies*, *33*, 74–87. doi: 10.1016/j.trc.2013.04.010

Keyvan-Ekbatani, M., Yildirimoglu, M., Geroliminis, N., & Papageorgiou, M. (2015, 8). Multiple concentric gating traffic control in large-scale urban networks. *IEEE Transactions on Intelligent Transportation Systems*, *16*(4), 2141–2154. doi: 10.1109/TITS.2015.2399303

Leclercq, L., Chiabaut, N., & Trinquier, B. (2014). Macroscopic Fundamental Diagrams: A cross-comparison of estimation methods. *Transportation Research Part B: Methodological*, *62*, 1–12. Retrieved from `http://dx.doi.org/10.1016/j.trb.2014.01.007` doi: 10.1016/j.trb.2014.01.007

Lee, G., Ding, Z., & Laval, J. (2023, 9). Effects of loop detector position on the macroscopic fundamental diagram. *Transportation Research Part C: Emerging Technologies*, *154*. doi: 10.1016/j.trc.2023.104239

Mahmassani, H., Hou, T., & Saberi, M. (2013, 12). Connecting networkwide travel time reliability and the network fundamental diagram of traffic flow. *Transportation Research Record*(2391), 80–91. doi: 10.3141/2391-08

Maiti, L., Nandan; Leclercq. (2024). *Correcting Loop Detector Positional Bias When Estimating Macroscopic Fundamental Diagrams Using Link Speed Data Samples.* Lyon.

Mariotte, G., & Leclercq, L. (2019, 12). Heterogeneous perimeter flow distributions and MFD-based traffic simulation. *Transportmetrica B*, *7*(1), 1378–1401. doi: 10.1080/21680566.2019.1627954

Mariotte, G., Leclercq, L., Batista, S. F., Krug, J., & Paipuri, M. (2020, 6). Calibration and validation of multi-reservoir MFD models: A case study in Lyon. *Transportation Research Part B: Methodological*, *136*, 62–86. doi: 10.1016/j.trb.2020.03.006

Mariotte, G., Paipuri, M., & Leclercq, L. (2020). Dynamics of Flow Merging and Diverging in MFD-Based Systems: Validation vs. Microsimulation. *Frontiers in Future Transportation*, *1*. doi: 10.3389/ffutr.2020.604088

Mousavizadeh, O., & Keyvan-Ekbatani, M. (2024, 4). On the important features for a well-shaped reduced network MFD estimation during network loading and recovery. *Transportation Research Part C: Emerging Technologies*, *161*. doi: 10.1016/j.trc.2024.104539

Saberi, M., & Mahmassani, H. (2012, 12). Exploring properties of networkwide flow-density relations in a freeway network. *Transportation Research Record*(2315), 153–163. doi: 10.3141/2315-16

Saeedmanesh, M., & Geroliminis, N. (2016, 9). Clustering of heterogeneous networks with directional flows based on "Snake" similarities. *Transportation Research Part B: Methodological*, *91*, 250–269. doi: 10.1016/j.trb.2016.05.008

Saeedmanesh, M., & Geroliminis, N. (2017, 11). Dynamic clustering and propagation of congestion in heterogeneously congested urban traffic networks. *Transportation Research Part B: Methodological*, *105*, 193–211. doi: 10.1016/j.trb.2017.08.021

Shim, J., Yeo, J., Lee, S., Hamdar, S. H., & Jang, K. (2019, 5). Empirical evaluation of influential factors on bifurcation in macroscopic fundamental diagrams. *Transportation Research Part C: Emerging Technologies*, *102*, 509–520. doi: 10.1016/j.trc.2019.03.005

Tsubota, T., Bhaskar, A., & Chung, E. (2014, 1). Macroscopic Fundamental Diagram for Brisbane, Australia. *Transportation Research Record: Journal of the Transportation Research Board*, *2421*(1), 12–21. doi: 10.3141/2421-02

Yang Beibei, J., van Zuylen, H. J., & Shoufeng, L. (2016). Determining the Macroscopic Fundamental Diagram on the Basis of Mixed and Incomplete Traffic Data. In *Trb 2016 annual meeting* (pp. 1–13).