Network inefficiency - Empirical findings for six European cities

Lisa S. Hamm*
Chair of Traffic Engineering and Control, TU Munich
DE-80333, Germany
Email: lisa.hamm@tum.de

Allister Loder
Chair of Traffic Engineering and Control, TU Munich
DE-80333, Germany
Email: allister.loder@tum.de

Gabriel Tilg
Chair of Traffic Engineering and Control, TU Munich
DE-80333, Germany
Email: gabriel.tilg@tum.de

Monica Menendez
Division of Engineering, NYU Abu Dhabi, UAE
Email: monica.menendez@nyu.edu

Klaus Bogenberger
Chair of Traffic Engineering and Control, TU Munich
DE-80333, Germany
Email: klaus.bogenberger@tum.de

* Corresponding author

Word count: 5646 words + 3 table(s) × 250 = 7146 words

Submitted: December 6, 2021

Paper re-submitted for review at the Transportation Research Record

ABSTRACT

When planning road networks, inhomogeneous traffic conditions and the effects of multi-modal interactions are often neglected. This can lead to a substantial overestimation of network capacities. Empirical macroscopic fundamental diagrams or volume delay relationships show considerable scatter, reflecting a reduction in network performance and an inefficient use of infrastructure. The implication is that the external costs of vehicular (car) traffic get underestimated, when planning traffic capacities and speeds based on optimal rather than on real estimates. In this paper, we contribute with an explorative and empirical approach to analyze network inefficiency and quantify its drivers. We propose to measure network efficiency by introducing the idea of excess delays for the macroscopic fundamental diagram. We define excess delays as the difference between the observed speed and the optimal network speed at a given density. We apply the concept on traffic data sets of six European cities that differ in the data collection method and use quantile regression methods for analysis. We find that excess delays are present in every data set and increase with the road network's traffic load. We further confirm the intuition that traffic signal control, network loading, and multimodality influence the level of network inefficiency. The excess delay formula allows quantifying this information in a simple way and provides additional insights apart from the standard MFD model. The approach supports planners to obtain better real-world and less optimistic speed predictions for traffic analyses and suggests shifting urban transport to more spatial and temporal efficient modes.

Keywords: macroscopic fundamental diagram, multi-modal, network efficiency, urban congestion, policy

INTRODUCTION

We plan our road networks based on guidelines assuming normal and homogeneous traffic conditions, not accounting for multi-modal interactions, and therefore an overestimation of capacities. This results in rather best-case estimates of road traffic. Cost-benefit analyses may only consider these best-case estimates, but lack information on some of the factors that can negatively affect network performance, which, in turn, potentially alter planning decisions. Such negative effects are observed everywhere in the unpredictable reality: empirical macroscopic fundamental diagrams or volume delay relationships show considerable scatter, implying a reduction in performance and inefficient use of infrastructure (1, 2, 3, 4, 5). In urban road networks, the literature suggests at least three sources contributing to inefficient infrastructure use: (i) interaction effects between different vehicles types, (ii) traffic dynamics, and (iii) traffic control strategies. The implication is that when cities plan and manage traffic capacities and speeds based on an optimal rather than on the real estimates, the external costs of vehicular (car) traffic get underestimated.

In this paper, we contribute with an explorative and empirical approach to analyze network inefficiency and quantify its drivers. We propose to measure network efficiency by introducing the idea of excess delays for the macroscopic fundamental diagram (MFD). Excess delays add up to inherent delays of traffic. The latter ones are already described by the MFD (6, 7) and fundamental diagram. In this paper, we define the original MFD as *idealized*, i.e. as the maximum flow for each density independently from demand. In contrast, the *observed* MFD is what we observe from empirical or simulation data, and includes multi-modal interactions and demand-related effects. Excess delays are the difference between both MFDs for a given density. More specifically, for a certain density, we compare the speed, measured in units of pace, derived from the idealized and the observed MFD. The excess delay approach allows quantifying the effects of signal control, network loading, and multimodality on urban traffic in a simple way. Furthermore, it provides additional insights such as the possibility to facilitate the modeling of hysteresis patterns in the MFD.

Based on this procedure, we can measure excess delays for four real-world loop detector data sets, one drone data set, and one simulation-based data set. We find that excess delays are present in every data set and that they increase with the road network's traffic load. We find also that there is a difference between the network loading and unloading dynamics, and that there exists an intuitive influence of traffic signal control and multimodality on network inefficiency. Interestingly, the estimates of all six sources are comparable, even though the data sets differ in the collection method and the underlying network sizes, suggesting the global applicability of the quantitative results of this analysis. The proposed approach of network inefficiency and excess delays helps planners and decision-makers to obtain better real-world and less optimistic speed predictions for their particular analysis.

The remainder of this paper is organized as follows. In the next chapter, we will briefly stress on MFDs as a traffic analysis tool, define excess delays and outline the resampling approach that we use to generate MFDs. In the following chapter, we will present three explorative approaches to fit an optimal speed curve to the resampled data sets. Thereafter, we will describe the empirical data sets of five cities and the simulation data set. Then, we will present the results of the analysis, and finish with a conclusion on our findings and policy implications.

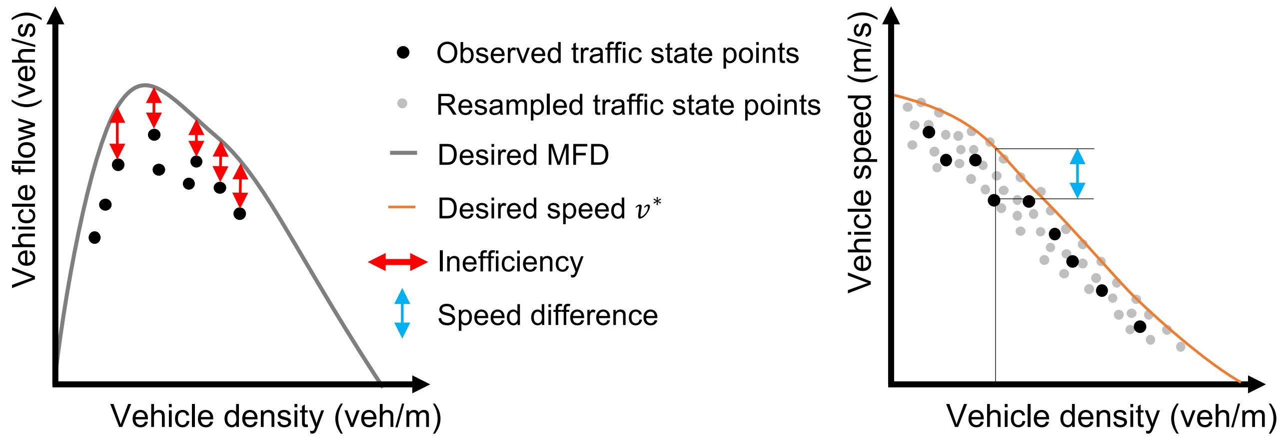

FIGURE 1 : Network inefficiency and estimating excess delays using the MFD.

METHOD

We quantify urban network inefficiency by the measure of *excess delays* based on the formulation of the MFD. As mentioned above, we define excess delays as those that exist in addition to the delays which can be derived by the idealized MFD for a given density. In the following subsections we discuss each building block in the process to calculate excess delays from MFD data.

The macroscopic fundamental diagram

The MFD describes the relationship between vehicle accumulation (density) and average traffic speed (or flow) in an urban road network (6). The general shape of the MFD can be seen in Figure 1. One can distinguish between the idealized or upper MFD (6, 8, 9) and the observed MFD (10, 11), both in the flow-density and in the speed-density relationship. The upper MFD represents the upper envelope to all possible states that are observed in the MFD. We define the optimal (desired) relationship between vehicle density, flow, and speed as the idealized MFD, which can be related to the social optimum. However, the desired speed-density relationship is rarely reached in reality and therefore difficult to measure. Delays in addition to the delays defined by the MFD then always occur when an observed data point does not match the desired speed-density relationship. In this paper, we will fit an upper speed MFD and derive an optimal speed-density relation by using a resampling approach (12). The resampling method is expected to result in less biased upper bound estimates with more supporting data (time span, experimental variation). If the underlying data is biased, e.g., only exhibiting one loading pattern, it could be that the resampled upper bound is biased in such a way that the excess delay estimation is less reliable in the sequel analysis. The MFD literature suggests that additional or *excess* delays occur for three main reasons:

1. **Multimodal vehicle interactions:** So far, most MFD literature focused on car traffic, although in the last years, interactions between different modes are receiving increasing attention: For example, bi-modal interactions, i.e. between cars and buses (13, 14, 15) or cars and pedestrians (16), but also tri-modal interactions between cars, buses, and bicycles (17). As the latter show, interactions between different modes have different effects on the overall pace of the vehicles compared to cases where only unimodal interactions are considered.

Multimodal interaction effects come on top of the delays that occur due to network loading and unloading (increasing and decreasing vehicle densities).

2. **Network loading processes and hysteresis:** Research has shown that the onset and offset of congestion leads to different density distributions from a spatial perspective (18). This is reflected in the MFD, where for a given density the observed flows during unloading of the network, i.e. the offset of congestion, are usually lower than during the loading phase. While the corresponding traffic dynamics can be explained by the FD on the link level, the upper MFD curve fails to do so.

3. **Traffic signal control:** In some circumstances, the urban traffic controller may exert additional red times for car traffic, e.g. to protect a certain perimeter from overcrowding (19) or to prioritize public transport (20). In either case, general traffic experiences additional delays that are in excess to those that are reflected in the desired MFD.

Another contributor to excess delays could be the network topology and supply characteristics, such as speed limits, lane widths, number of intersections, or its structure. For example, in a city with a high number of links with speed limits below the city-wide speed limit, the mean speed could be lower than average on the other links and therefore result in "artificial" excess delays. For single analyses focusing on one city, this effect may be negligible. When comparing two different cities with substantial differences in the network topology, we suggest controlling for these effects in the delay model. This limitation will be further investigated in our future research, e.g., by simulation experiments that analyze the interaction effects of network structure and loading on excess delays in particular.

Network inefficiency based on excess delays

Figure 1 shows where network inefficiency can be seen in the MFD. We define inefficiency as the *gap* between the upper or desired MFD, and the observed traffic states in the flow-density representation of the MFD. In the speed-density relationship, the gap can be directly translated into an additional delay. Here, we express delay in the units of additional time per unit distance (s/m). We quantify these additional delays in the following sequence. See Table 1 for the model specification.

1. **Estimate the re-sampled MFD:** To approximate the smooth upper bound as good as possible that may correspond to the upper or desired speed MFD, we apply the re-sampling method proposed by Ambühl et al. (*12*) on the aggregated data. The authors propose to apply sampling of representative subsets to generate the resampled data set. This procedure aligns the data to more homogeneous distributions of network flow and density. Additionally, it results in a smooth upper bound or boundary between observed and not observed traffic states.

2. **Estimate the observed speed MFD:** The observed MFD from which v_{obs} is derived and for which the delays γ are being calculated is estimated using the methods described in Leclercq et al. (*21*) depending on the data source.

3. **Estimate the desired speed MFD v^*:** We estimate the upper or desired speed MFD, $v^*(k)$, using the data from the re-sampled speed MFD and three different estimation methods to

TABLE 1 : Model specification

Variable	Meaning
k	Average network density in vehicles per meter
v_{obs}	Average journey speed observed in the network in meters per second
$v*$	Desired journey speed in the network in meters per second
γ	Excess delays in seconds per meter
L	Network loading indicator. Equals to one when network is in the loading state (increasing vehicle density), zero otherwise. Unitless
Parameter	Meaning
λ	MFD smoothing parameter (see (9)). Unitless
ß$_0$	Intercept of the excess delay model in seconds per meter
ß$_k$	Effect of traffic density on excess delays in seconds per vehicle
ß$_L$	Effect of network loading on excess delays in seconds per meter

test for sensitivity of the relationships. More specifically, we fit in a quantile regression in the 99th percentile using (i) the functional form for the MFD proposed by Ambühl et al. (*9*) with the smoothing parameter λ; (ii) the 99th percentile of the speed distribution at density bins and (iii) the exponential function proposed by Underwood (*22*) $v^*(k) = exp(log(c_0) + log(c_1 * k))$.

4. **Estimate excess delays:** Finally, we calculate excess delays by $\gamma = 1/v_{obs} - 1/v^*$, where v^* is evaluated at the same density as observed for v_{obs}.

DATA

To enable an extended empirical comparison of excess delays and network inefficiency, we use traffic data from six European cities: Athens, Innsbruck, London, Lucerne, Paris, and Zurich. The cities are diverse, with large differences in e.g. surface areas, population size, network size, traffic densities, and average speeds. We chose two very large cities (> 1 million inhabitants), two mid-size cities (> 400.000 inhabitants) and two smaller cities (> 50.000 inhabitants) to test the method for different network scales. Also, the data collection methods for the data sets differ. Table 2 shows an overview of the six data sets.

We first aggregate the data based on 2-minute time intervals for the Athens data set, and on 5-minute intervals for all other data sets, except for Paris. We chose a shorter time interval for Athens to generate a larger database. The aggregated data sets were used as base data for the resampling approach. Here, the number of randomly generated subsamples and the fraction size had to be specified. We chose a size of 100 subsamples and fraction sizes between 0.2 and 0.25. In other words, we resampled 20 to 25 percent of the aggregated data for each subsample (see Table 2).

TABLE 2 : Overview of the traffic data sets from six European cities and the resampling(*) parameters

Data set	Athens	Innsbruck	London	Lucerne	Paris	Zurich
Population size (mil.)	0.664	0.131	8.961	0.082	2.161	0.403
Network size (km^2)	1.3	33	160	5	24	15
Detection method	drone	simulation	loop det.	loop det.	loop det.	loop det.
Time window	4 days	4 hours	22 days	365 days	335 days	365 days
Aggregation interval (seconds)*	120	300	300	300	3600	300
No. of subsamples*	100	100	20	100	20	100
Fraction size (in %)*	0.25	0.25	0.2	0.25	0.2	0.25

Loop detector data: Lucerne, London, Paris, Zurich

The data of Lucerne, Zurich, London, and Paris is part of the UTD19 dataset (23, 24). The large-scale traffic data was assembled through stationary loop detectors in the city areas. Loop detectors measure the occupancy, i.e. the time fraction that a vehicle occupies a detector, and the traffic count, i.e., the number of vehicles passing the detector, for a fixed time interval. In Lucerne and Zurich, data was collected over a time period of a year. In London and Paris, data was collected over 22 and 335 days, respectively. We selected all observations during the daytime, between 6 a.m. and 8 p.m. The Paris data set is aggregated on a time interval of an hour, as the loop detector only generates observations in this frequency. The other three data sets are aggregated in intervals of 5 minutes. The loop detector data differs from the pNEUMA data set and Innsbruck simulation data with the former covering the morning hours only and the latter simulating four hours of weekday traffic.

Drone data: Athens

The observations of the data set pNEUMA (25) were collected by ten drones flying over the central business district of Athens, Greece, during the morning hours. The data contains the latitude and longitude values for different vehicle types, namely car, bus, motorcycle, medium and heavy vehicles, and taxis over time fractions of 0.04 seconds. The number of observations for pNEUMA is smaller compared to the detector-based UTD-datasets, as the measurement period comprised 4 weekdays only. Because the observations of some drone flights indicated measurement errors, the preprocessing was extensive. For example, we removed the eighth drone flight and the observations of the last 2 minutes of every drone flight, as the reported speeds and densities indicate that there might be measurements errors. To derive density values from the data set, we assumed a network length of 100 km according to the official data set description. Finally, we aggregated the observations in 2-minute intervals.

Simulation data: Innsbruck

The Innsbruck data was generated with a microscopic traffic simulation using SUMO (26). The network was retrieved from OpenStreeMaps (27). For all included traffic behavior models, the default parameters were chosen. Thus, the simulation is not calibrated. However, for the current study, such a calibration is not essential as we investigate traffic flow dynamics and their effects on the network level. Such mechanisms are included in the simulation due to the physical modeling of driving behavior.

The origin-destination patterns were randomly generated. The loading curve has a trapezoidal shape, i.e. after a step-wise loading, the maximum demand was kept for some time intervals, until the demand was decreased again in a step-wise manner. Such loading curves are commonly applied to mimic a rush-hour including its onset and offset. The loading and unloading phase lasted for 1.5 [h] each, the plateau phase for 1 [h], which results in a total simulation time of $T = 4$ [h].

Vehicles follow the shortest path from their origin to their destination. In the simulation, we apply a quasi-dynamic traffic assignment, where the shortest paths are updated for all vehicles considering the current traffic states in the network for every 2 minutes. In other words, all vehicles can adapt their route before they reach their destination if such a change is of advantage. Such an assignment represents a reasonable trade-off between realism and computational cost (28).

To allow the analysis of the impact of signal control on the network inefficiency, we varied the signal control parameters. We assume that all traffic signals follow a fixed-time control logic, have a common cycle length, have the same green-to-cycle ratio of 0.5, and no offsets apply. Three different cases with a cycle length of 60, 90, and 120 [s] were investigated. Previous research has shown that offsets do not have a major impact on the resulting MFD (29).

RESULTS

The resampling method proved successful to build the upper bound of the MFD for all six data sets. As expected, we obtain a decreasing non-linear speed-density relation for all cities (see Figure 2). We see that the range of vehicle density is higher for London, Lucerne, Paris, and Zurich than for Innsbruck and Athens. Not surprisingly, the larger data sets show less scatter than the smaller ones (Athens, Innsbruck). For the latter, the resampling method proves especially useful as it generates a more reliable database. In Figure 2, we obtain for every estimation method (Ambühl (9), the percentile approach, and Underwood (22)) the optimal speed curves v^* for all six data sets. We find that all three fitting methods for v^* obtain comparable relationships. Eventually, differences can be explained by the estimation methods, e.g. a limited flexibility due to functional assumptions in the Underwood and Ambühl case. With substantial observations and scatter in the uncongested regime and less in the congested regime, these two functions, which weigh each point equally in the estimation, have to balance these differences. This leads to a relationship that appears to be below the upper MFD in the uncongested regime, while better describing the upper MFD in the congested regime. This has implications for the estimation of excess delays that can be less reliable in the uncongested regime.

Based on the fitted optimal speed curves from Figure 2, we examined the relationship between excess delays and densities. The kernel density estimates of delays on the right-hand side of Figure 3 show the frequency of the delay value range, calculated with Underwoods' method. As expected, Athens and Innsbruck show more variance in the excess delay distribution as the database is small. The UTD-data sets approach a Gaussian bell shape, especially Lucerne and Zurich, being the largest data sets. Athens, Innsbruck, and Zurich have a mean excess delay of approximately

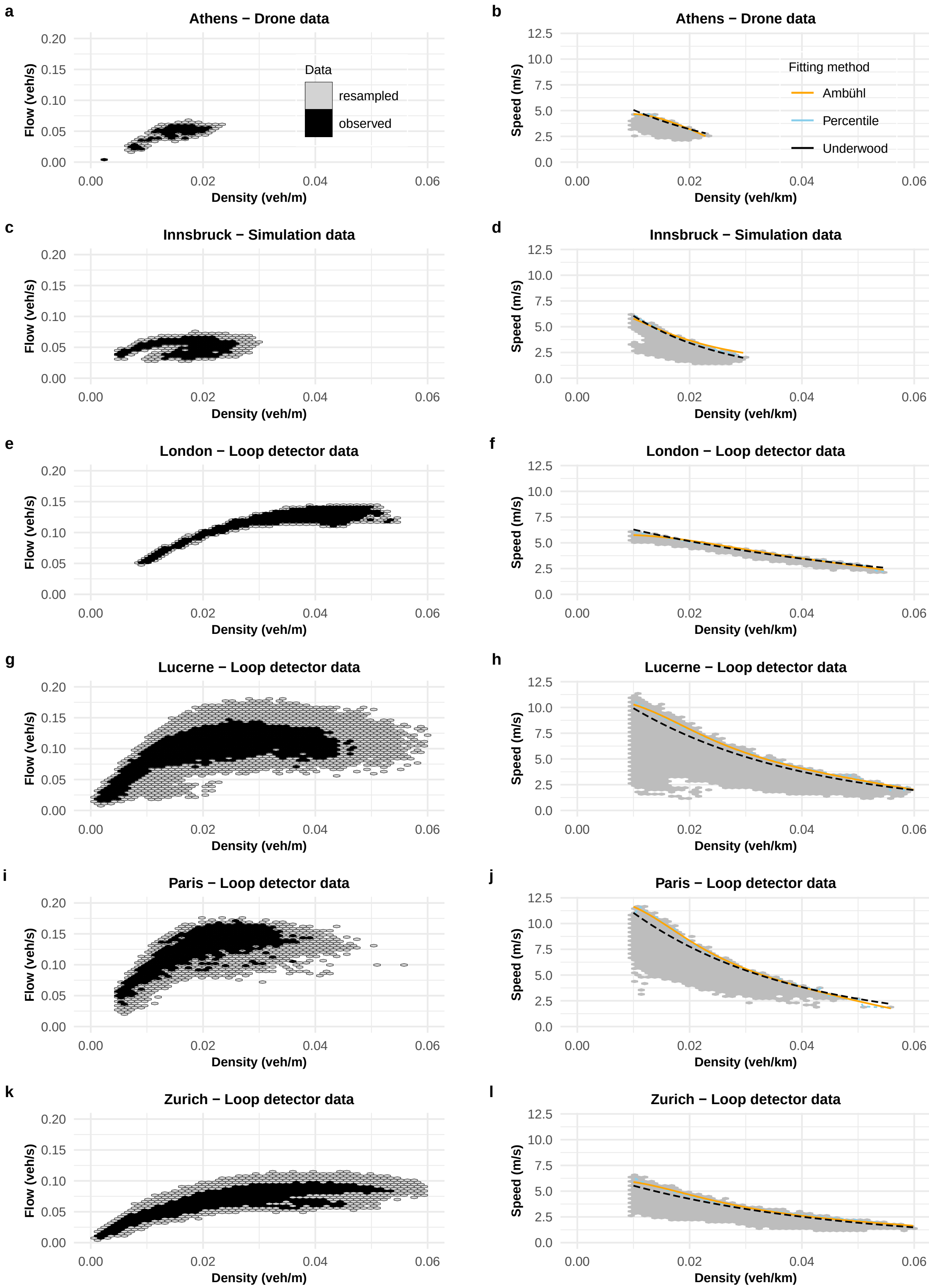

FIGURE 2 : Desired flow-density (left) and speed-density (right) MFDs and three fitting methods for v^* for six cities.

0.05 to 0.06 seconds per meter. London, Lucerne, and Paris have a mean excess delay of approximately 0.02 to 0.03 seconds per meter. As can be seen in Figure 3, we obtain positive relationships between higher vehicle densities and excess delays, especially in higher density ranges, for Innsbruck, London, Lucerne, and Zurich. This indicates that v^* is less likely obtained, the more density is observed in an urban road network. The plotted results indicate that the estimation methods do not differ considerably.

Network loading

To test the influence of network loading and unloading process on the development of excess delays, we derive from the data an indicator variable L that equals one if the network is loading and zero otherwise. We assume that the network is in a loading state when the difference in density of two consecutive intervals is positive, after applying a three-interval moving average on the density to reduce the noise in the data. We then estimated a linear regression to analyze the effect of density k and slope L on excess delays γ. We estimate the linear model as given in Eqn. 1, where β_0, β_k and β_L are parameters to be estimated using ordinary least squares. ε represents the error terms that are assumed to be normally distributed.

$$\gamma = \beta_0 + \beta_k \cdot k + \beta_L \cdot L + \varepsilon \tag{1}$$

As the excess delay values and relationships that result from the three v^* fitting approaches look very similar (see Figure 2), we present the results only for the Underwood model in Table 3. The results for the other three models do not alter the findings. Generally, we find that the model formulated in Eqn. 1 explains substantial variance found in Athens, Innsbruck, London, and Lucerne. The low R^2 in Paris and Zurich suggests that the model does not well describe the data for these cities. Potentially, the value for Paris results from the temporal aggregation at one-hour intervals instead of 2-5 minute intervals, where many of the dynamic effects might be averaged out. In future research, we will investigate further the factors of the distribution of excess delays in Paris and Zurich.

For Athens, Innsbruck, London, Lucerne, and Zurich we find positive and statistically significant effects of vehicle density on excess delays. However, their effect sizes differ by one order magnitude. Future research has to investigate why this effect is so substantially different between the shown networks. Potential reasons are data bias, network topology, traffic control, etc. In addition, Figure 4 for Innsbruck suggests that the linear model in Eqn. 1 is falsely specified as the loading and unloading effect and their interaction are clearly not linear. This means that the model formulation from Eqn. 1 might not capture all underlying mechanisms, which could be a reason for the alternate effect direction found in Paris. Here, the effect estimation regarding the loading part results in $\beta_k = 0.377$ ($p < 0.01$), which supports the findings related to the other five cities. The effect corresponds to increasing excess delays with traffic density. More specifically, we observe an increase of the estimates around 1 s/m for every 0.01 veh/m increase. Note that this effect is in addition to the speed reduction already captured in the MFD. Nevertheless, the findings from Paris indicate that an analysis of the differences could reveal further insights.

The six cities cover different spatial scales to understand the behavior of the loading indicator for different city sizes. The indicator variable for loading is negative and statistically significant in Athens, Innsbruck, Lucerne, and Zurich, while Zurich and Lucerne report an effect of one order of magnitude less than Athens and Innsbruck. It means that during the network loading, fewer ex-

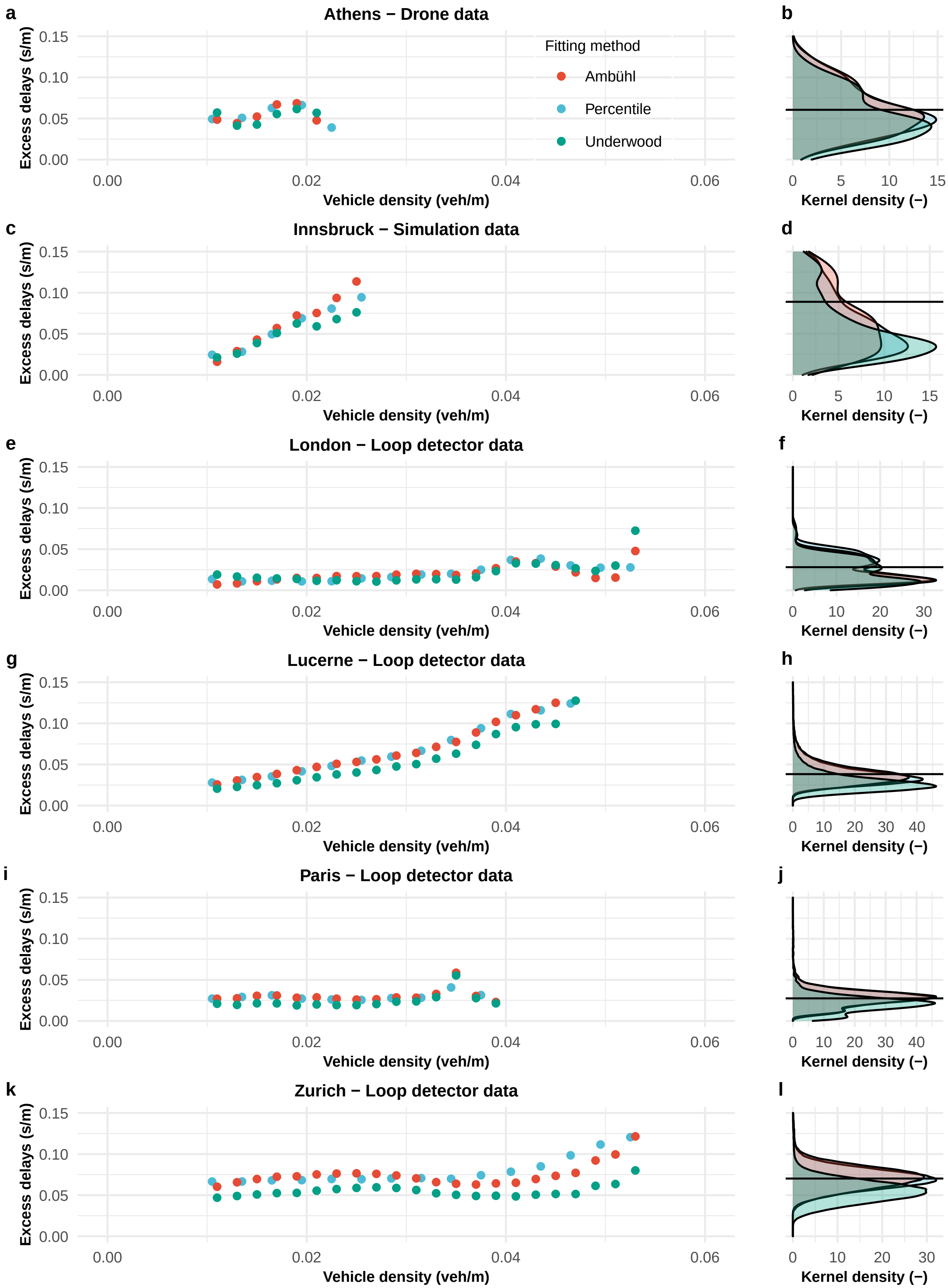

FIGURE 3 : Relationship between vehicle density and excess delays (left), and kernel densities (right), for each fitting method.

TABLE 3 : Model estimation results for excess delays.

Dataset	Effect of density	Effect of loading	R^2	Density range (veh/m)
Athens	2.96 ($p < 0.01$)	-0.038 ($p < 0.01$)	0.26	0.01 to 0.02
Innsbruck	3.53 ($p < 0.01$)	-0.090 ($p < 0.01$)	0.53	0.01 to 0.03
London	1.13 ($p < 0.01$)	0.001 ($p < 0.01$)	0.19	> 0.02
Lucerne	2.30 ($p < 0.01$)	-0.003 ($p < 0.01$)	0.30	> 0.02
Paris	-0.24 ($p < 0.05$)	0.001 ($p < 0.01$)	0.006	> 0.02
Zurich	0.48 (p <. 0.01)	-0.002 ($p < 0.01$)	0.03	> 0.02

cess delays are present compared to the unloading, e.g. supporting the development of hysteresis in the MFD. The positive effect in London and Paris deserves more attention, especially from an econometric perspective. In the model specification from Eqn. 1, there are no control variables included. Consequently, any factor that could contribute to excess delays and that correlates either with traffic density or the loading indicator variable is partially included in the estimated effect. As the model is estimated for the entire urban area in London and Paris, this could be an increase in bus services during the loading phase that increases excess delays (30), a gating traffic control scheme that increases red phases for inbound traffic to protect the urban core from gridlock which adds waiting time and thus increases excess delays (19), or any origin-destination-based effect. Consequently, future research should improve the estimates with more detailed model formulations.

Traffic control

The simulation of Innsbruck enables to vary signal control parameters and study corresponding effects on the excess delay. For this purpose, we investigate three different scenarios, where the cycle length of all signals equals 60, 90, and 120 seconds. Figure 4 shows the resulting scatter plot for the loading and unloading part of the MFD as well as a local polynomial regression fitting (loss) of R's ggplot package to investigate the trend in the data.

For all three scenarios, we observe the already revealed positive relationship between vehicle density and excess delays in the loading part. In the unloading part, we see that excess delays are decreasing with vehicle density, hinting towards a clockwise hysteresis. Importantly, the data suggest that differences in the relationship between different traffic signal settings and excess delays exist. The influence seems to be nonlinear concerning the cycle length as the trend lines do not appear in ascending or descending order, but in the sequence 90, 120, and 60 seconds. This confirms that traffic control indeed has not only an effect on inherent delays already included in the MFD, but indeed also on excess delays as suggested in this paper. The impact of cycle times seems to be larger in the unloading part compared to the loading part as it can be seen in Figure 4. However, in future research, we will investigate this relationship more extensively using more simulation scenarios.

Note that the apparent relationship between excess delay and density in Figure 4 in the unloading part suggests that simple mathematical modeling of hysteresis effects is possible. This will be further explored in future research.

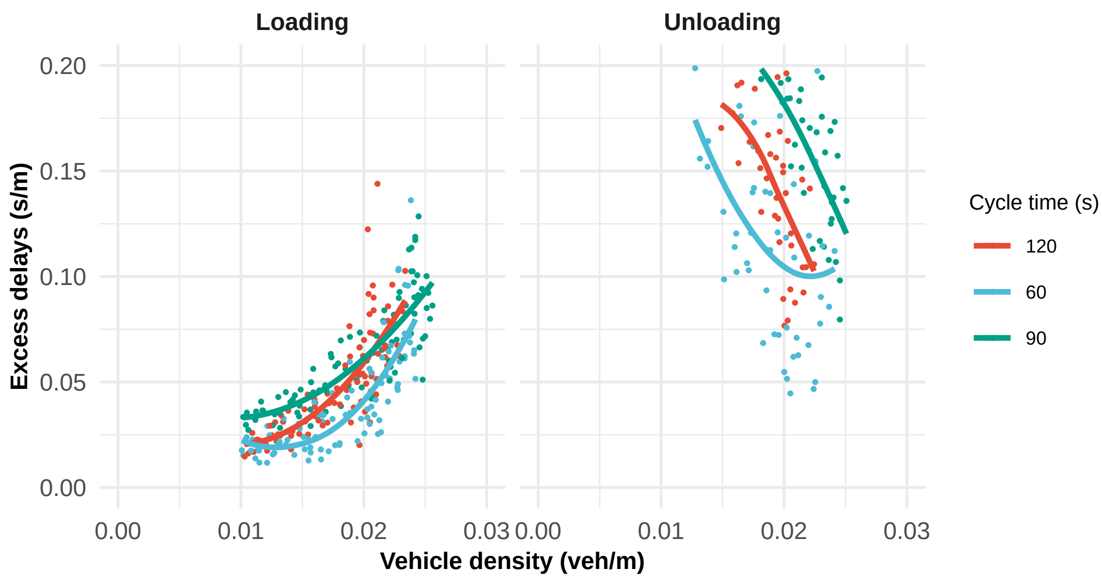

FIGURE 4 : Influence of cycle lengths on excess delays using the Innsbruck simulation data.

Multimodal traffic in Athens

We further investigate the variation of excess delays in Athens by distinguishing between different vehicle types. We use the pNEUMA data here, as it is the only data set allowing such in-depth multimodal analyses. Note that there exists already a paper on multi-modal interactions at space-mean network speed working with the pNEUMA data set (31). The authors use regression models of the space-mean speed on the vehicle accumulations of the multimodal traffic. The core distinction of the analysis presented in this paper is that we do not use a speed-accumulation relationship but an excess delay formulation, i.e., additional time delay in s/m in relation to the maximum speed at a given density. In each time interval, we compute the share of taxis, large vehicles (labeled as 'large and medium-sized vehicles' in the original data), buses, and motorcycles. Figure 5 shows the resulting scatter plots. For the share of taxis, large vehicles and motorcycles we find a positive relationship with excess delays, while for the share of buses we find a negative relationship. This seems perhaps surprising given that the 3D-MFD assumes negative interaction costs between cars and buses (13). To explore the multivariate nature of the data, we estimate a linear model of excess delays as a function of the taxi, large vehicle, bus, and motorcycle share. We find statistically significant (1 % level) marginal effects of taxi share of 0.39 (s/m), and truck share of 0.62 (s/m). This means that when the taxi share increases by ten percentage points, excess delays increase by 0.04 (s/m). This can be translated to an additional delay of 3.3 minutes for a typical journey with a length of 5 km per 10 % taxi share increase. When the large vehicle share increases by ten percentage points, excess delays increase by 0.06 (s/m). For buses, the model estimate of -1.37 (s/m) (statistically significant at 1 % level) confirms the relationship from Figure 5. This counter-intuitive relationship may result from the limited sample size and experimental variation: The share of buses correlates strongly negatively with vehicle density. In other words, the share of buses only increases as a consequence of an overall decreasing vehicle density (fixed timetable). Thus, this variable approximates more high and low demand traffic states and less the impact of buses. This makes the revealed estimate reasonable. Nevertheless, this finding emphasizes that an

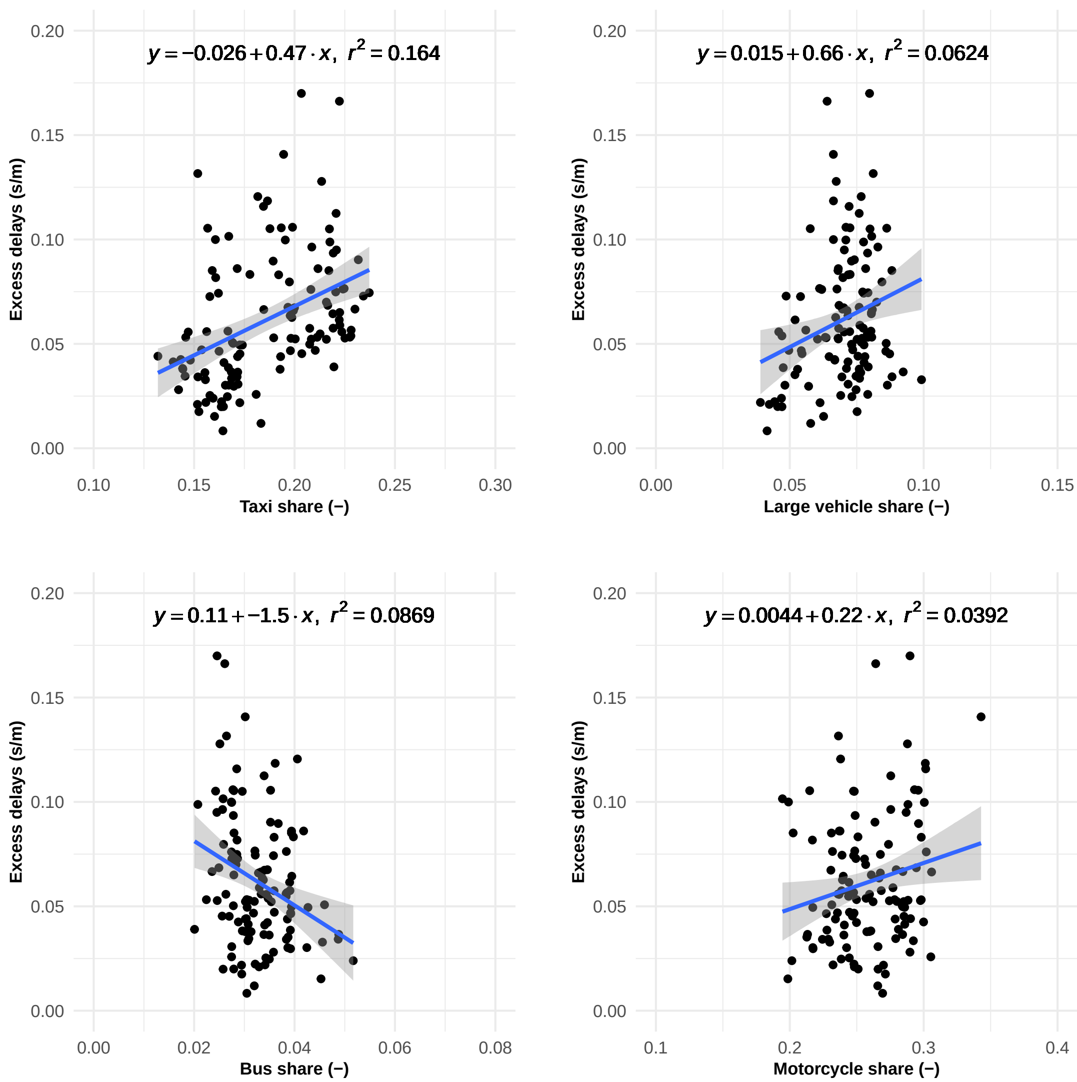

FIGURE 5 : The influence of multimodal traffic on excess delays in Athens.

important variable is omitted in the present model formulation. The model does not reveal a statistically significant effect of motorcycles, i.e. the relationship found in Figure 5 is not supported. Overall, the model has a goodness of fit of $R^2 = 0.27$, when controlling for potential outliers it increases to $R^2 = 0.37$.

Multi-modal traffic occurs in all cities including the resulting interaction effects. In larger cities such as Paris or London, the flows of bicycles and scooters are clearly observable and are quite likely contributing to excess delays. To account for them in the excess delay formulation, many observations and sufficient experimental variation are required (high/low volumes of cars and high/low volumes of bicycles or scooters) to reveal the interaction effects shown in field experiments (17, 32). Another modeling challenge is the violation of traffic regulations, which might impede the observation and estimation of effects. In the case of the pNEUMA data set, the experimental variation was limited, leading to a high correlation of densities and the expected effects could not be revealed. In cities with loop detector data, one approach to control for the impact of bicycles/scooters on excess delays would be to use bicycle counts from permanent counting locations as a proxy. Unfortunately, such data was not available to the authors.

CONCLUSION

In this paper, we showed that urban road networks experience substantial inefficiencies as seen in the presence of excess delays. We defined excess delays as the difference between the optimal and observed pace. Using five empirical data sets (loop detector and drone data) from European cities and one simulation data set, we observed network inefficiencies in every city. Even though the extent of excess delays differs across cities, their general effects and evolution are highly similar. This supports the applicability of the method for other cities. We further investigated causes for the emergence of excess delays, which would make them predictable: (i) network loading, causing inherent delays produced by increasing density (ii) signal control, which we showed for different cycle lengths in Innsbruck and (iii) multimodal interaction effects between different vehicle types, supported by data from multimodal traffic data set of Athens. Regarding (i), the results of the delay-density relation for the Innsbruck data suggest that our approach might simplify the mathematical modeling of hysteresis effects.

With this paper, we not only contribute to improved modeling of the evolution of congestion in cities at the network level but also to a more realistic capacity planning for urban road networks. We show that there exists an optimal, achievable speed curve for large, medium-sized, and small cities. We also demonstrate that the proposed method applies to different forms of data sets - loop detector data, drone data, and simulation data. The inefficiency of excess delays can be measured easily by the proposed methods and only requires average speed and density values for a given area and time period. To find out which factors affect excess delays, it would be suitable to compare the measured excess delays of a specific area in a city for a given time period with a simulation of this respective scenario. Then, measures could be derived to reduce the effects on excess delays and therefore minimize speed drop.

The practical implications of this paper primarily concern the applications of the proposed approach. First, it helps to identify and quantify factors on excess delays in a city more conveniently compared to modeling speeds directly - either using empirical data or simulation experiments. Then, measures (design features, traffic management) could be derived to reduce excess delays and therefore minimize speed and accessibility losses either by scenario analysis or by a cross-sectional analysis of several cities. Second, applications of the proposed network inefficiency approach can be possible everywhere in the field of network-wide traffic management where an improved capacity and speed estimate is valued for improved traffic and economic outcomes, e.g., road pricing or perimeter control. Third, the proposed method can be applied as a performance indicator to assess the impact of time and space allocation in an urban network: In those areas or hours with a high share of excess delays, traffic could be allocated to other roads or shifted to other time periods, e.g., through intelligent passenger information systems. As the optimal speed v^* is almost never reached in real-world scenarios, network planners might create more space and capacity to obtain the optimal speed for cars. This is difficult to realize, e.g. due to limited space in urban areas and the risk of induced traffic. Multimodal system improvement in combination with setting up mode-independent accessibility values (for example, by measuring the minimum required transport speed or maximum acceptable travel time) which have to be met by a traffic system, could be a possible solution.

This paper introduces and discusses the idea of measuring network inefficiency by the concept of excess delays. Therefore, there are limitations to the study and opportunities for future research. First, we did not consider at the present stage the influence of structural network effects, e.g. speed limits or network design, on the evolution of excess delays. As the MFD is governed by network topology, one could argue that it influences excess delays too. Second, the fitting of v^* to identify the upper MFD can be improved as the relationships do not perfectly match the resampled upper MFD, e.g. by weighting observations. Only when we can correctly describe the upper MFD, we can retain unbiased excess delay estimates that are important for further modeling. This also requires a throughout data filtering and unbiased MFD estimation prior to the derivation of excess delays. Once unbiased excess delays estimates are retrieved, we can follow on the first evidence present on its driving factors (network loading, signal control, and multimodality) to improve the estimates and, using more extensive experiments (empirical data, simulation), obtain global validity of these estimates.

In closing, describing network inefficiency by excess delays seems to be promising because it makes the former predictable. As effect sizes are similar, too, across cities in our study, we are convinced that the revealed effects can be found in every city. Last, we consider that using drone data for calibrating a city's multimodal excess delays effects is promising, as it allows quantifying otherwise unobserved factors.

ACKNOWLEDGEMENTS

Lisa S. Hamm acknowledges support from the German Federal Ministry of Transport and Digital Infrastructure (BMVI) for the funding of the project TEMPUS (Test Field Munich - Pilot Test Urban Automated Road Traffic). Allister Loder acknowledges support from the German Federal Ministry of Transport and Digital Infrastructure (BMVI) for the funding of the project KIVI (Artificial Intelligence in Ingolstadt's Transportation System), grant no. 45KI05A011. Gabriel Tilg acknowledges support from the German Federal Ministry of Transport and Digital Infrastructure (BMVI) for the funding of the project LSS (Capacity increase of urban networks). Monica Menendez acknowledges the support from the NYUAD Center for Interacting Urban Networks (CITIES), funded by Tamkeen under the NYUAD Research Institute Award CG001 and by the Swiss Re Institute under the Quantum Cities™ initiative. As a data source, the pNEUMA data set was used: open-traffic.epfl.ch

AUTHOR CONTRIBUTIONS

The authors confirm contribution to the paper as follows: L. Hamm, A. Loder contributed to the study conception and design, and model design. L. Hamm, A. Loder, G. Tilg, M. Menendez contributed to the draft manuscript preparation, and the results analysis and interpretation. All authors reviewed the results and approved the final version of the manuscript.

References

1. Bai, L., S. Wong, P. Xu, A. H. Chow, and W. H. Lam, Calibration of stochastic link-based fundamental diagram with explicit consideration of speed heterogeneity. *Transportation Research Part B: Methodological*, Vol. 150, 2021, pp. 524–539.
2. Buisson, C. and C. Ladier, Exploring the impact of homogeneity of traffic measurements on the existence of macroscopic fundamental diagrams. *Transportation Research Record*, Vol. 2124, No. 1, 2009, pp. 127–136.
3. Mazloumian, A., N. Geroliminis, and D. Helbing, The spatial variability of vehicle densities as determinant of urban network capacity. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, Vol. 368, No. 1928, 2010, pp. 4627–4647.
4. Geroliminis, N. and J. Sun, Properties of a well-defined macroscopic fundamental diagram for urban traffic. *Transportation Research Part B: Methodological*, Vol. 45, No. 3, 2011, pp. 605–617.
5. Knoop, V. L., H. Van Lint, and S. P. Hoogendoorn, Traffic dynamics: Its impact on the macroscopic fundamental diagram. *Physica A: Statistical Mechanics and its Applications*, Vol. 438, 2015, pp. 236–250.
6. Daganzo, C. F., Urban gridlock: Macroscopic modeling and mitigation approaches. *Transportation Research Part B: Methodological*, Vol. 41, No. 1, 2007, pp. 49–62.
7. Geroliminis, N. and C. F. Daganzo, Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transportation Research Part B: Methodological*, Vol. 42, No. 9, 2008, pp. 759–770.
8. Daganzo, C. F., L. J. Lehe, and J. Argote-Cabanero, Adaptive offsets for signalized streets. *Transportation Research Part B: Methodological*, Vol. 117, 2018, pp. 926–934.
9. Ambühl, L., A. Loder, M. C. Bliemer, M. Menendez, and K. W. Axhausen, A functional form with a physical meaning for the macroscopic fundamental diagram. *Transportation Research Part B: Methodological*, Vol. 137, 2020, pp. 119–132.
10. Ambühl, L., A. Loder, L. Leclercq, and M. Menendez, Disentangling the city traffic rhythms: A longitudinal analysis of MFD patterns over a year. *Transportation Research Part C: Emerging Technologies*, Vol. 126, 2021, p. 103065.
11. Tilg, G., L. Ambühl, S. Batista, M. Menendez, L. Leclercq, and F. Busch, Semi-analytical estimation of macroscopic fundamental diagrams: from corridors to networks. In *100th Annual Meeting of the TRB*, 2021.
12. Ambühl, L., A. Loder, M. Bliemer, M. Menendez, and K. Axhausen, Introducing a resampling methodology for the estimation of empirical macroscopic fundamental diagrams. *Transportation Research Record*, Vol. 2672, No. 20, 2018.
13. Geroliminis, N., N. Zheng, and K. Ampountolas, A three-dimensional macroscopic fundamental diagram for mixed bi-modal urban networks. *Transportation Research Part C: Emerging Technologies*, Vol. 42, 2014, pp. 168–181.
14. Dakic, I., L. Ambühl, O. Schümperlin, and M. Menendez, On the modeling of passenger

mobility for stochastic bi-modal urban corridors. *Transportation Research Part C: Emerging Technologies*, Vol. 113, 2020, pp. 146–163.

15. Tilg, G., Z. Ul Abedin, S. Amini, and F. Busch, Simulation-based design of urban bi-modal transport systems. *Frontiers in Future Transportation*, Vol. 1, 2020.
16. Daganzo, C. F. and V. L. Knoop, Traffic flow on pedestrianized streets. *Transportation Research Part B: Methodological*, Vol. 86, 2016, pp. 211–222.
17. Loder, A., L. Bressan, M. J. Wierbos, H. Becker, A. Emmonds, M. Obee, V. L. Knoop, M. Menendez, K. W. Axhausen, and A. Loder, How Many Cars in the City Are Too Many ? Towards Finding the Optimal Modal Split for a Multi-Modal Urban Road Network. *Frontiers in Future Transportation*, Vol. 2, No. May, 2021.
18. Gayah, V. V. and C. F. Daganzo, Clockwise hysteresis loops in the macroscopic fundamental diagram: An effect of network instability. *Transportation Research Part B: Methodological*, Vol. 45, 2011, pp. 643–655.
19. Geroliminis, N., J. Haddad, and M. Ramezani, Optimal Perimeter Control for Two Urban Regions With Macroscopic Fundamental Diagrams: A Model Predictive Approach. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 14, No. 1, 2013, pp. 348–359.
20. Guler, S. I. and M. Menendez, Analytical formulation and empirical evaluation of pre-signals for bus priority. *Transportation Research Part B: Methodological*, Vol. 64, 2014, pp. 41–53.
21. Leclercq, L., N. Chiabaut, and B. Trinquier, Macroscopic fundamental diagrams: A cross-comparison of estimation methods. *Transportation Research Part B: Methodological*, Vol. 62, 2014, pp. 1–12.
22. Underwood, R. T., Speed, Volume and density relationships. In *Quality and Theory of Traffic Flow*, Pennsylvania State University, University Park, 1961, pp. 141–188.
23. Loder, A., L. Ambühl, M. Menendez, and K. W. Axhausen, *UTD19. Understanding traffic capacity of urban networks*. Zurich, 2020-08.
24. Loder, A., L. Ambühl, M. Menendez, and K. W. Axhausen, Understanding traffic capacity of urban networks. *Scientific Reports*, Vol. 9, No. 1, 2019, p. 16283.
25. Barmpounakis, E. and N. Geroliminis, On the new era of urban traffic monitoring with massive drone data: The pNEUMA large-scale field experiment. *Transportation Research Part C: Emerging Technologies*, Vol. 111, No. November 2019, 2020, pp. 50–71.
26. Lopez, P. A., M. Behrisch, L. Bieker-Walz, J. Erdmann, Y.-P. Flötteröd, R. Hilbrich, L. Lücken, J. Rummel, P. Wagner, and E. Wießner, Microscopic traffic simulation using SUMO. In *The 21st IEEE International Conference on Intelligent Transportation Systems*, IEEE, 2018.
27. OpenStreetMap contributors, *Innsbruck dump retrieved from https://planet.osm.org* , 2020.
28. Mühlich, N., V. V. Gayah, and M. Menendez, Use of Microsimulation for Examination of Macroscopic Fundamental Diagram Hysteresis Patterns for Hierarchical Urban Street Networks. *Transportation Research Record*, Vol. 2491, No. 1, 2015, pp. 117–126.
29. Girault, J.-T., V. V. Gayah, I. Guler, and M. Menendez, Exploratory analysis of signal coordination impacts on macroscopic fundamental diagram. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2560, No. 2560, 2016, pp. 36–46.
30. Loder, A., L. Ambühl, M. Menendez, and K. W. Axhausen, Empirics of multi-modal traffic networks – Using the 3D macroscopic fundamental diagram. *Transportation Research Part C: Emerging Technologies*, Vol. 82, 2017, pp. 88–101.
31. Paipuri, M., E. Barmpounakis, N. Geroliminis, and L. Leclercq, Empirical observations of

multi-modal network-level models: Insights from the pNEUMA experiment. *Transportation Research Part C: Emerging Technologies*, Vol. 131, 2021, p. 103300.

32. Wierbos, M. J., V. L. Knoop, F. S. Hänseler, and S. P. Hoogendoorn, A macroscopic flow model for mixed bicycle–car traffic. *Transportmetrica A: Transport Science*, Vol. 17, No. 3, 2021, pp. 340–355.