Disentangling the city traffic rhythms: A longitudinal analysis of MFD patterns over a year

Lukas Ambuhl, Allister Loder, Ludovic Leclercq, Monica Menendez

Transportation Research Part C 126 (2021) 103065

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

Disentangling the city traffic rhythms: A longitudinal analysis of MFD patterns over a year

Lukas Ambühl [a], Allister Loder [a,*], Ludovic Leclercq [b], Monica Menendez [c]

[a] Institute for Transport Planning and Systems, ETH Zurich, Switzerland
[b] Univ. Gustave Eiffel, Université de Lyon, ENTPE, LICIT, Lyon, France
[c] Division of Engineering, NYU Abu Dhabi, United Arab Emirates

ARTICLE INFO

Keywords:
Macroscopic fundamental diagram (MFD)
Clustering
Empirical data
Prediction

ABSTRACT

Urban road transportation performance is the result of a complex interplay between the network supply and the travel demand. Fortunately, the framework around the macroscopic fundamental diagram (MFD) provides an efficient description of network-wide traffic performance. In this paper, we show how temporal patterns of vehicle traffic define the performance of urban road networks. We present two high-resolution traffic datasets covering a year each. We introduce a methodology to quantify the similarity of macroscopic traffic patterns. We do so by using the concepts of the MFD and a dynamic time warping (DTW) based algorithm for time series. This allows us to derive a few representative MFD clusters that capture the essential macroscopic traffic patterns. We then provide an in-depth analysis of traffic heterogeneity in the network which is indicative of the previously found clusters. Thereupon, we define a parsimonious classification approach to predict the expected MFD clusters early in the morning with high accuracy.

1. Introduction

Urban road transportation performance is the result of a complex interplay between the network supply and the travel demand. The infrastructure's upper (theoretical) performance is determined by the network topology and traffic control. The observed network performance, however, also depends on the demand, in particular, the spatial and temporal distribution of vehicle flows. Thus, the observed performance is typically lower than the upper limit (Çolak et al., 2016; Loder et al., 2019; Ji and Geroliminis, 2012).

Fortunately, the framework around the macroscopic fundamental diagram (MFD) provides an efficient description of the ideal network-wide dynamic traffic performance (Geroliminis and Daganzo, 2008; Loder et al., 2019; Mahmassani et al., 1987) that is the prime ingredient to the so-called bathtub models (Vickrey, 2020). Generally, the MFD links the network averages of the fundamental traffic variables density and flow with a smooth and concave relationship. The two variables increase until they reach the critical point of urban traffic. Thereafter, increasing vehicle density in a neighborhood leads to a decrease in vehicle flow. Then, it is said that a neighborhood is congested at the macroscopic level. The network average journey speed results from flow divided by density, i.e., every network has a critical speed that is below free flow speed at which congestion starts. In recent years three distinct definitions of the MFD have emerged as summarized in Fig. 1:

* Corresponding author.
E-mail address: allister.loder@ivt.baug.ethz.ch (A. Loder).

https://doi.org/10.1016/j.trc.2021.103065
Received 4 October 2020; Received in revised form 8 January 2021; Accepted 16 February 2021
Available online 9 March 2021

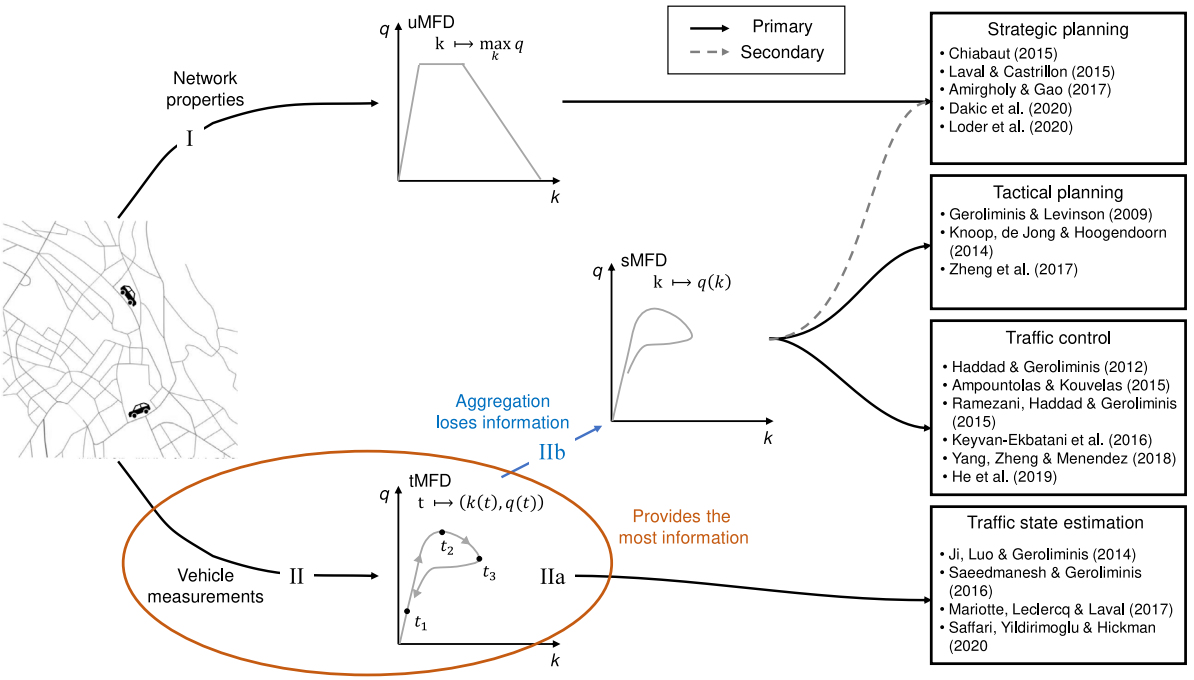


Fig. 1. Existing MFD definitions and corresponding literature examples.

I **MFD as theoretical upper bound MFD (uMFD)**: This is the original notion of the MFD introduced by Daganzo (2007). Here, the MFD is seen as a function of the network topology and traffic control only. It answers the question: "Given an urban road network with its traffic control, what are the maximum feasible traffic states?". In other words, it tries to define the network's maximum traffic capacity under "optimal circumstances". As an example of a brute force approach to find such an upper bound, we can imagine the following experiment: Once the network and traffic control are defined, we run a Monte-Carlo simulation with many different demand patterns and calculate the MFD for each scenario. The scenario (or a combination thereof) with the highest (steady-state) average flow (or travel production) for any average vehicle density (or accumulation) yields the upper bound MFD. These exact demand properties are unlikely to be observed in reality, and thus it is the theoretically feasible upper bound MFD. Even though slightly misleading, it is the common understanding that this upper bound is demand-independent because it is essentially not a description of what is observed, but an upper bound to what could be observed.
Prominent analytical approximations for such upper bound are limited to approaches that require network reduction (or mapping) to a single corridor, e.g. method of cuts (see Daganzo and Geroliminis (2008), Laval and Castrillón (2015), Leclercq et al. (2015)). The inputs to such models are the network topology and traffic control details. Endogenously defined demands, however, are not a required input. The method endogenously solves the demand selection process described by our above brute force approach. This further fosters the notion that the upper bound MFD is independent of demand.
The static perspective offered by the upper bound MFD is useful for urban planning studies or defining the potential of the traffic system.

II **MFD as the observed daily macroscopic traffic performance:** Here, the focus lies on the observed (or realized) shape in the average flow and density plane.
With the first publications (see Geroliminis and Daganzo (2007) and Buisson and Ladier (2009)) on the empirical existence of the MFD, the notion thereof altered as well. Empirical studies showed that relating the measured average traffic density and flow in an urban neighborhood usually yielded, in general, a well-defined curve (but which still differed day after day in shape and evolution). It was therefore natural to call these observed macroscopic traffic states the empirical MFD. Unfortunately, these studies only had access to a few days, so the differences between the days were neither discussed nor further investigated. Later studies showed that the observed daily macroscopic traffic states are (well) below the upper theoretical bound (see Tilg et al. (2020), Ambühl et al. (2020) and Saffari et al. (2020)). Imagine a directed demand in an urban network, where vehicles only travel from South to North, and this happens day after day. It is obvious that the observed MFD will not correspond to the upper bound, because many links in the network (predominantly the ones from North to South) will not be used at their maximum flow. Arguably, observing MFDs close to the uMFD necessitates homogeneously distributed traffic and slow-varying demand that is hard to find in a real, regular network (Daganzo, 2007; Ambühl et al., 2020; Buisson and Ladier, 2009; Loder et al., 2019), because of inefficient network loading and spillbacks due to given demand patterns that inevitably lead to a heterogeneous distribution of traffic flow (Leclercq et al., 2015; Knoop et al., 2015; Geroliminis and Sun, 2011; Mazloumian et al., 2011; Mahmassani et al., 2013). So far, most studies are limited to only a few observation days, so the differences

between days were not further investigated. However, many real-world MFD applications rely on a repeatable and stable MFD, making inter-day variation an important aspect to consider (Haddad and Geroliminis, 2012; Zheng et al., 2012; Yang et al., 2019; Geroliminis et al., 2014). For the observed or realized MFD, it depends on the application, whether the temporal evolution is considered or not, creating subsequently two different definitions: the static and dynamic MFD.

IIa In the static MFD, **sMFD**, the focus lies on the observed (or realized) **shape** of the average flow as a function of density, only. The temporal aspects in the evolution of the macroscopic traffic performance are disregarded. This allows the reduction of the complexity of the observed urban traffic dynamics. Therefore, the sMFD is a very helpful tool for any application that requires a static view of the daily system's performance, e.g. traffic control based on the static critical point (see e.g. Geroliminis and Levinson (2009), Haddad and Geroliminis (2012), Knoop et al. (2014), Zheng and Geroliminis (2016), He et al. (2019), Ramezani et al. (2015), Ampountolas and Kouvelas (2015), Keyvan-Ekbatani et al. (2016) and Yang et al. (2018)).

IIb The dynamic MFD, **dMFD**, a definition introduced by Mahmassani et al. (2013), considers the MFD as a tool to observe, understand, and model the loading and unloading of urban traffic. Tracking average flow and density *over time* allows to clearly define when the traffic system becomes saturated, when it is recovering, etc. This definition also deals with dynamic aspects of the observed MFD's shape, e.g. hystereses and the reason why they might exist (see e.g. Ji et al. (2014), Saeedmanesh and Geroliminis (2014), Mariotte et al. (2017), Saffari et al. (2020) and Gayah and Daganzo (2011)).
It is not astonishing that most applications around the MFD rely on tracking network performance over time. The classic example is perimeter control, where traffic strategies are deployed to prevent the system from becoming macroscopically congested. Similarly, other examples also require short-term traffic prediction. Therefore, the dynamic aspects of the evolution of the macroscopic traffic conditions are paramount.

From Fig. 1 and the above definitions of the MFD it can be concluded that the uMFD uses network information and can be considered demand-independent, while the dMFD and sMFD use vehicle measurements and are demand dependent. Estimating the dMFD always allows to derive the sMFD, but not the other way around. Consequently, analyzing the performance patterns based on the dMFD uses the full information provided by the measurements for applications that require either the dMFD or the sMFD. Until now, the lack of longitudinal empirical data at a large urban scale, as well as a methodology to measure the similarity between empirical dMFDs, prevented a detailed analysis.

Fig. 2 shows the empirical dMFDs over a year for Zurich (430'000 inhabitants) and Lucerne (80'000 inhabitants) in Switzerland, for a region of roughly $10\ \mathrm{km}^2$ each in their respective downtowns. While the overall trends are clear, the dMFDs still exhibit a range of flows for a given density. Highlighting two days in each dMFD emphasizes that there might be substantial variations between different days of the year. Even within the same day, the dMFD does not follow the same path during the loading and unloading phases, i.e. potentially reflecting hysteresis effects. Thus, these dMFDs raise a series of questions that this study aims to answer: How can we establish and measure the repeatability of empirically observed dMFDs? How can we disentangle the differences in the dMFDs shown in Fig. 2? What are the underlying drivers of such differences? How can we predict the expected dMFD shape and its evolution well in advance, i.e. early in the morning?

We approach these research questions in two steps:

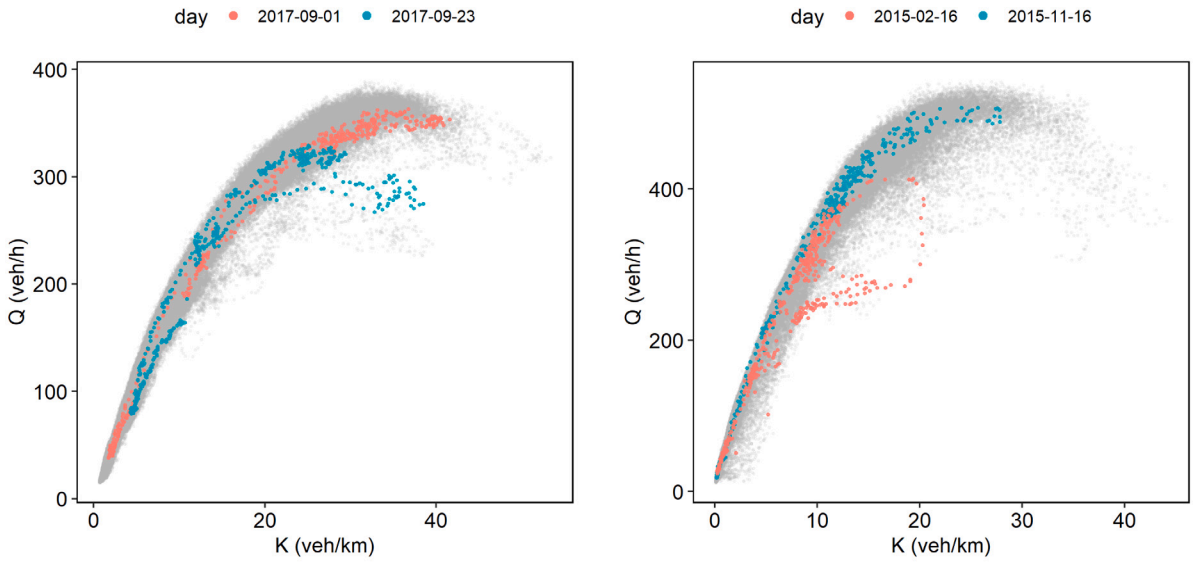
1. In a first step, we use the dMFD definition provided in definition IIb to compare its daily evolution over the course of a year. We then measure the similarity between days and reduce the full-year dataset to a few representative clusters.
2. In a second step, we then explain the differences between these representative clusters as a result of the dynamic loading of the network, measured by the evolution of the spatial heterogeneity in the network as well as the perimeter flows.

This allows us to introduce the following four contributions: (i) we present two very large empirical high-resolution traffic datasets covering a year each; (ii) we introduce a methodology to quantify the similarity across dMFDs and identify clusters across different days ; (iii) we explain the differences between the clusters as a result of the dynamic loading of the network, measured by the spatial heterogeneity in the network, as well as the perimeter flows, which play a key role in the activation of traffic bottlenecks; and (iv) we use a parsimonious classification approach to predict the expected clusters accurately early in the morning.

The remainder of this paper is organized as follows. In Section 2, we present the datasets. Then, in Section 3 we introduce a novel method to measure the similarity across dMFDs. In Section 4, we then derive clusters of dMFDs, which show similar shapes and macroscopic loading/unloading patterns. In Section 5, we investigate the reasons for the differences found between the clusters. In Section 6, we define a surprisingly parsimonious way to accurately predict the expected dMFD cluster as early as possible within a day. In Section 7, we close with the conclusions of this study.

2. Data

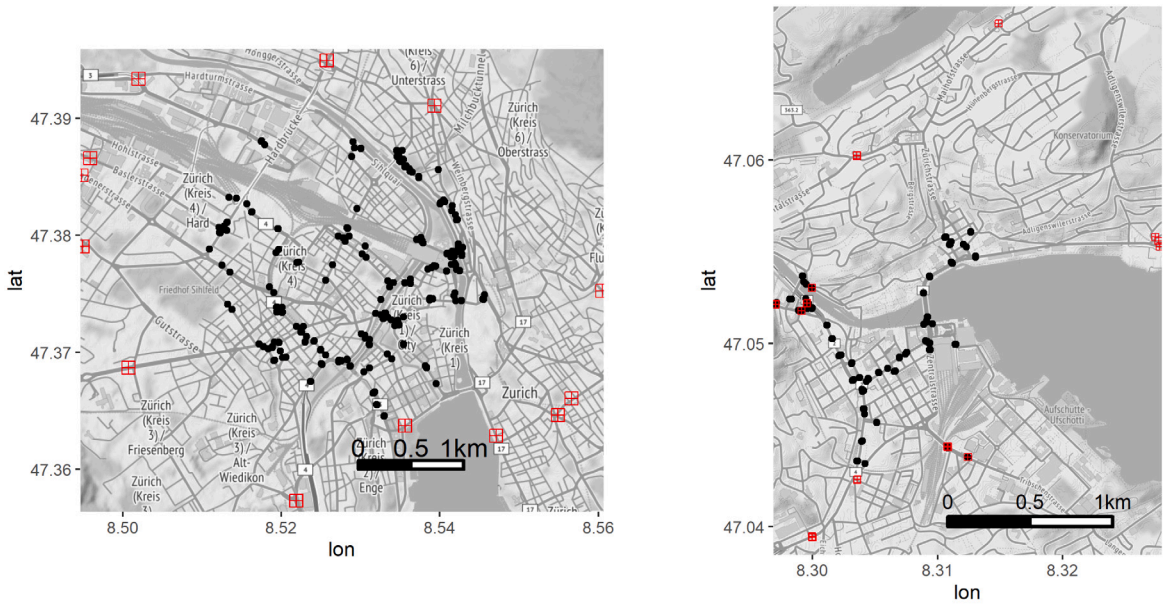
The empirical data from Fig. 2 originates from stationary traffic sensors (386 in Zurich, 131 in Lucerne) and covers a year (365 days) at a resolution of 3 min for each of the two cities. We use an advanced automated outlier detection algorithm based on principal component analysis by Filzmoser et al. (2008) for large multivariate datasets to assure high quality of the data (see also *R* package *mvoutlier*). Missing data points and outliers are then imputed using a k-Nearest Neighbor (kNN) methodology by Kowarik and Templ (2016) (see also *R* package *VIM*). In addition, the time series is smoothed using a TBATS model, a fast exponential smoother known to account for complex trends and seasonality by de Livera et al. (2011) (see also *R* package *forecast*).



(a) Empirically observed macroscopic fundamental diagram (dMFD) for a year-long period in Zurich. We highlight two different days.

(b) Empirically observed macroscopic fundamental diagram (dMFD) for a year-long period in Lucerne. We highlight two different days.

Fig. 2. Empirical dMFDs for Zurich and Lucerne, Switzerland.



(a) Zurich.

(b) Lucerne.

Fig. 3. Loop detectors included in the analysis. Red squares indicate the perimeter loops, whereas black dots represent the loops used to derive dMFDs. The maps are oriented towards the North.

Fig. 3 shows the considered regions and the detectors included in this study. Note that we also highlight detectors outside of the regions. They serve as perimeter detectors (32 in Zurich, 19 in Lucerne) and measure the traffic into and out of our regions.

In Fig. 2, Zurich yields substantially lower flows than Lucerne, because the detectors in Zurich also cover roads with lower priority (recall that Zurich has approximately three times as many detectors as Lucerne). Conversely, in Lucerne, most detectors are located on arterials with relatively high green/red ratios. Interestingly, there are few signs of severe macroscopic congestion in either city. This is due to the adaptive traffic signal control in both cities and a demand-responsive gating of vehicles, especially

in the city of Zurich (see Ambühl et al. (2018b)). Nonetheless, the lowest average network speeds are approximately $4-6$ km/h - roughly walking speed.

The coverage rate of links with a detector in Zurich is roughly 40% and in Lucerne 25% of the main road system (includes primary–tertiary roads according to the definitions by OpenStreetMap). A threshold of 25% randomly distributed detectors allows unbiased estimation of the MFD (Ortigosa et al., 2014; Saffari et al., 2020). Admittedly, the monitored links are not randomly distributed in the network, but they cover mostly the important arterials with recurring congestion. For Zurich, a simulation study has shown that the loop detector distribution yields an unbiased MFD (Ambühl et al., 2016). Thus, they are of prime interest for an intervention. We acknowledge that there might be another measurement bias, insofar that we measure traffic on roads that are monitored, and thus managed differently from roads without loop detectors. Note, we have no reported substantial change (e.g. removal/construction of/on a bridge) to the network during the observation period. Also, there is no evidence of an over-representation of certain clusters in certain months.

3. Measuring similarities of the dMFD

The dMFDs shown in Fig. 2 pool the individual dMFDs across 365 days. They show the length-weighted average of lane-flow and occupancies. To disentangle the daily differences, we first need to understand how the (dis-)similarity between dMFDs can be measured. Similar dMFDs should not only look similar with respect to the average flow and density but also in time, meaning that similar trends in traffic conditions should happen within a narrow time frame. In the following, we introduce a time series based solution which measures not only the difference in magnitude but also the evolution over time. This is especially important as similar dMFDs could potentially benefit from a similar macroscopic control. Thus, identifying the (dis-) similarities between daily dMFDs is not only important to understand the underlying traffic mechanisms, but also from an application point of view.

3.1. Dynamic time warping

In the following, we track the macroscopic traffic performance with two variables. Our approach makes sure that both, the average flow and density, are included in the analysis simultaneously. Given that the dMFD itself (see also Fig. 1) consists of these two variables, we believe it is natural to refer to our approach as MFD-based. Therefore, we consider the dMFD as a *joint time series* of average flow $Q(t)$ and density $K(t)$ for the day.

Thereby, the dMFD time series also captures the dynamics of the network loading and recovery process over time. This perspective is useful to determine a reasonable number of daily dMFD clusters over a year. This would allow reducing the many dMFDs observed to a few representative dMFD clusters. dMFDs contained in the same cluster, therefore, are candidates for the same macroscopic control strategy.

A variety of time series similarity measures exist (Tormene et al., 2009; Bagnall et al., 2017). Given its low complexity and widespread use, we resort to the dynamic time warping (DTW) algorithm, which was first used in speech and handwriting recognition, but has also found applications in other fields, e.g. evolutionary genetics or crowd-sensing (Yuan and Raubal, 2012; Bahlmann and Burkhardt, 2004; Berndt and Clifford, 1994). Existing transportation literature applies DTW for vehicle or pedestrian trajectory matching (Sun et al., 2018; Taylor et al., 2015; Przybyla et al., 2015; Sharma et al., 2018). To the best of our knowledge, this powerful metric has not been applied to the field of the MFD.

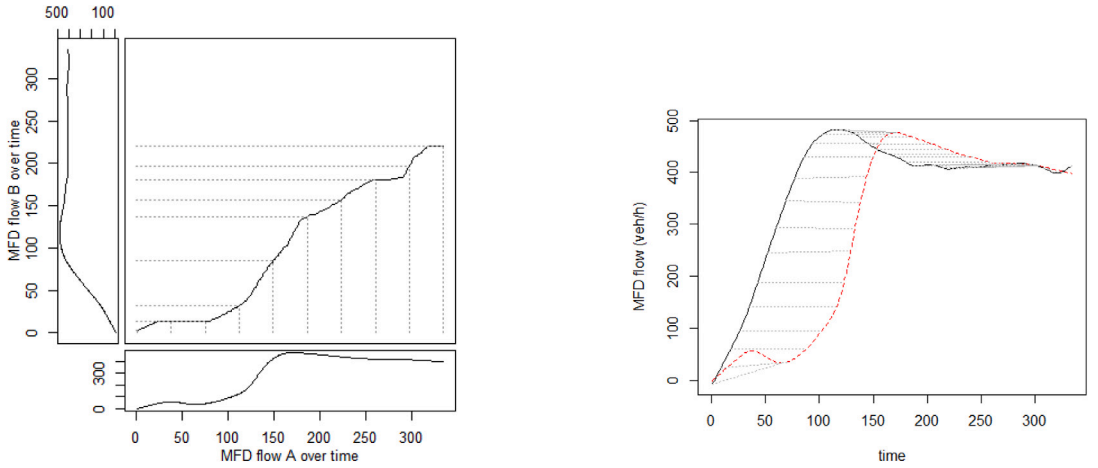
DTW measures the similarity between n-dimensional time series by warping the time axis within a specified time window (Berndt and Clifford, 1994). In contrast to comparing the (euclidean) distance between sequences at the same time, DTW thus deforms the time axis in both $Q(t)$ and $K(t)$ within the allowed limits. The algorithm searches for a warping path that aligns the trajectories of $\{Q_A(t), K_A(t)\}$ and $\{Q_B(t), K_B(t)\}$ so that their distance δ is minimized. The DTW framework is described schematically as follows:

1. Calculate the euclidean distance between the first time interval in MFD A and every time interval in MFD B. Store the minimum distance calculated (so-called time warping).
2. Move to the second interval and repeat 1. Then move to the next interval and repeat.
3. Repeat 1 and 2 but with MFD B as a reference point.
4. The sum of the minimum distances stored in step 1–3 represents the DTW distance.

In mathematical terms, DTW searches for the warping path that minimizes the cumulative distance between sequences $\{q_A(t), k_A(t)\}$ and $\{q_B(t), k_B(t)\}$. This warping path is subject to some restrictions: (i) continuity: the alignment cannot jump; (ii) boundary conditions: the full interval range is covered; (iii) warping window: guarantees that (probably unrelated) intervals very far apart do not get aligned (Tormene et al., 2009).

Fig. 4(a) shows DTW schematically. In this example, we show the process for the univariate case, where the aim is to apply the DTW algorithm on Q only. In the figure, we see two dMFD flows over time, one in the horizontal rectangle (A), and one in the vertical (B). Standard DTW aligns each value of $Q_A(t)$ to a value of $Q_B(t)$ (or vice versa). This alignment is shown in the square in the middle of Fig. 4(a). Note that if we had aligned the two patterns by comparing the dMFD flow values for identical time steps, the alignment would be the square's diagonal.

Measuring the similarity using DTW has two important implications for dMFDs: (i) If dMFD A is shifted by some interval or progresses faster, but otherwise results in a shape similar to dMFD B, A and B are deemed similar (within the limits of the warping window). (ii) It allows us to capture the loading and the unloading of urban traffic networks (i.e. hysteresis). A simpler euclidean distance measure or a functional form estimation would result in ambiguities in the multi-valued dMFD (e.g., hysteresis).



(a) Schematic comparison of two potential dMFD flows, A (horizontal) and B (vertical) over 300 time intervals. The path indicates how the flows are aligned using DTW.

(b) The two dMFD flow sequences compared over time. The red curve corresponds to dMFD flow A. Time is fictitious.

Fig. 4. Schematic overview of dynamic time warping. Time is fictitious.

3.2. Calculating the distance matrix

We apply the above framework to the datasets described in Section 2. For the results to be scale-independent, we normalize the average flow and average density as follows.

$$\tilde{Q}(t) = \frac{Q(t) - Q_{min}}{Q_{max} - Q_{min}}; \qquad \tilde{K}(t) = \frac{K(t) - K_{min}}{K_{max} - K_{min}} \tag{1}$$

Q_{min} and Q_{max} are the global minimum and maximum dMFD flow over all 365 days, respectively. This holds analogously for K. Notice that a similar approach was used by Ortigosa et al. (2014) to quantify differences across dMFDs.

We compute the DTW distance between all days with a warping time window of 3 h, which roughly represents the duration of an extended loading and unloading period. A short sensitivity analysis is presented in the following subsection.

This process yields a distance matrix M_{MFD} of 365x365 in size for each city. The lower the distance between two days, the more similar their dMFDs are deemed. This matrix then serves as an input to the clustering algorithm described in Section 4 aiming to find a set of representative dMFDs. In that section, we will investigate ways to further disentangle the distance matrix in a clustering approach.

3.3. Warping time window

As defined above, we use a warping time window of 3 h, which roughly represents the duration of an extended loading and unloading period. In the following, we investigate the sensitivity of the results with respect to reasonable changes to the time warping window.

We aim at finding an acceptable compromise for the time window size. In a time window too rigid for small (random) delays in the evolution of the observed MFD, we risk overfitting the data. In the extreme case of setting the time window to zero, the DTW distance yields the Euclidean distance. Contrarily, when choosing a large time window, we might disregard important differences between the observed MFDs. Therefore, we chose a time period similar to the duration of the rush-hour, i.e. the duration of an extended loading and unloading period. This way, we allow for certain flexibility in the on- and offset of congestion but do not risk missing a major change in the evolution of the macroscopic traffic performance.

With this in mind, we now investigate how similar the distance matrices become, once the DTW time window changes. For that, we focus on four additional windows, 2 h, 2.5 h, 3.5 h, and 4 h. To that end, we compute the Mantel test statistic (Mantel, 1967), i.e. the original distance matrix with a time window of 3 h and the one of the sensitivity analysis. This test is frequently applied in biology, more specifically in ecology, to compare distances between species, e.g. in terms of genetics, behavior, geographic distribution (Van Schaik et al., 2003; Waddle, 1994). Its interpretation is similar to Pearson's correlation coefficient with a range of $[-1, 1]$.

Table 1 summarizes the results. We can see that the Mantel test are very high, indicating that our approach is robust against reasonable changes in the time window.

Table 1
Sensitivity analysis for the DTW time window. The Mantel test is with respect to the chosen 3 h window. All tests are significant at 0.01.

Window size	2 h	2.5 h	3.5 h	4 h
Mantel test statistics	0.996	0.999	0.999	0.997

Table 2
Agglomeration coefficient for different linkage criteria in hierarchical clustering. The higher the coefficient the higher the strength of the clustering structure.

Dataset	Linkage criteria			
	Single	Average	Complete	Ward
Zurich	0.866	0.959	0.972	0.994
Lucerne	0.781	0.898	0.961	0.991

4. Clustering empirical dMFDs

4.1. Agglomerative hierarchical clustering

To find similar dMFDs, we cluster the distance matrix from Section 3.2. We would like to stress that we investigate a very large longitudinal dataset (a full year). Thus the resulting clusters will reproduce and characterize almost all situations the networks may encounter. There are many different cluster algorithms that are suitable for our problem (see also Jain et al. (1999)). For example, graph-based methods (e.g. Saffari et al. (2020)) or vector quantization methods (e.g. Necula (2015)) were used in the past for clustering traffic patterns. Here, we resort to hierarchical clustering given that it is a simple, fast, standard, and widely used framework to cluster well-defined distance matrices. It is also the standard clustering algorithm for matrices derived from DTW (Sardá-Espinosa, 2019). Moreover, it is one of the very basic clustering techniques, and therefore very accessible, also to practice. The main advantage of the hierarchical clustering compared to other methods is that it only requires a distance matrix, in lieu of the original data point. We choose the more commonly used agglomerative case, where each dMFD is initiated as a cluster on its own (Hastie et al., 2009). The clustering then merges (or agglomerates) similar dMFDs depending on a linkage criterion which specifies the dissimilarity of different cluster sets. This process is repeated until only 1 cluster is left. This successive agglomeration results in a nested association of all dMFDs.

Linkage criteria

For each step in this successive agglomeration, we need to evaluate which clusters should be merged. For this, a multitude of linkage criteria exist (for a complete overview, see also Kaufman and Rousseeuw (2009)). Essentially the criteria determines the distance between two different clusters.

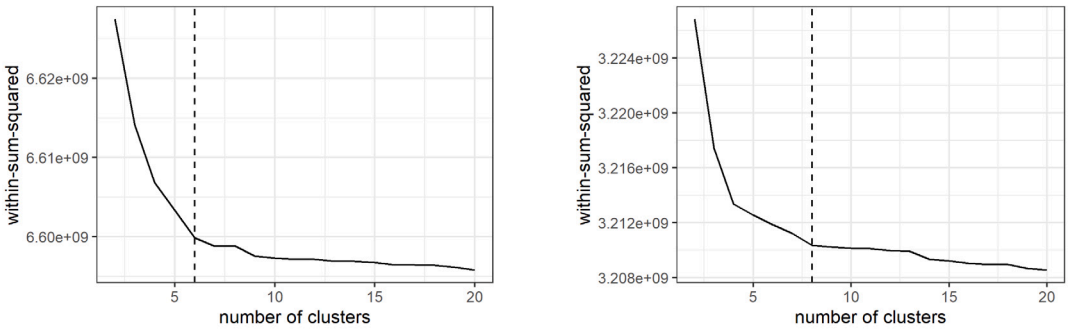
Here, we will analyze four common linkage criteria to decide which one is the most suitable. The *single* linkage defines the cluster dissimilarity as the distance between the closest elements in the clusters. The *average* linkage measures the average pair-wise distance between all elements in the clusters. The *complete* linkage uses the distance between the points that are furthest away between the clusters. *Ward's* criteria finds the pair of clusters that leads to a minimum increase in the total within-cluster variance after merging at each step (Ward, 1963).

The choice of the appropriate linkage criteria can be derived by the agglomeration coefficient, which describes the strength of the clustering structure (Kaufman and Rousseeuw, 2009). For each observation i, we calculate the dissimilarity to the first cluster with which it is merged and then divide this value by the dissimilarity of the merger in the final step of the algorithm. We denote this as $m(i)$. The agglomeration coefficient is then the average of all $1-m(i)$. It is normalized to range between 0 and 1 where higher values indicate a better cluster outcome.

Table 2 shows the agglomeration coefficient for commonly used linkage criteria. It becomes clear that Ward's minimum variance method yields the highest agglomeration coefficient. Thus, for any further analysis, we rely on Ward's linkage criteria.

Number of clusters

Similar to the chosen linkage method, we analyze the within-cluster variance to determine a suitable number of clusters. Fig. 5 shows the evaluation of the variability within each cluster, the within-cluster sum of squares, as a function of the potential number of clusters. The optimal number of clusters is evaluated to be 6 for Zurich and 8 for Lucerne. This number is derived from the visual inspection of the clusters and the "elbow method". The latter defines a reasonable cut-off for the explained variance captured by the clusters (Thorndike, 1953).



(a) Zurich. (b) Lucerne.

Fig. 5. Within-sum-squared as a function of the potential number of clusters. This measures the variability of the observations within each cluster. The dashed line shows the optimal number of clusters.

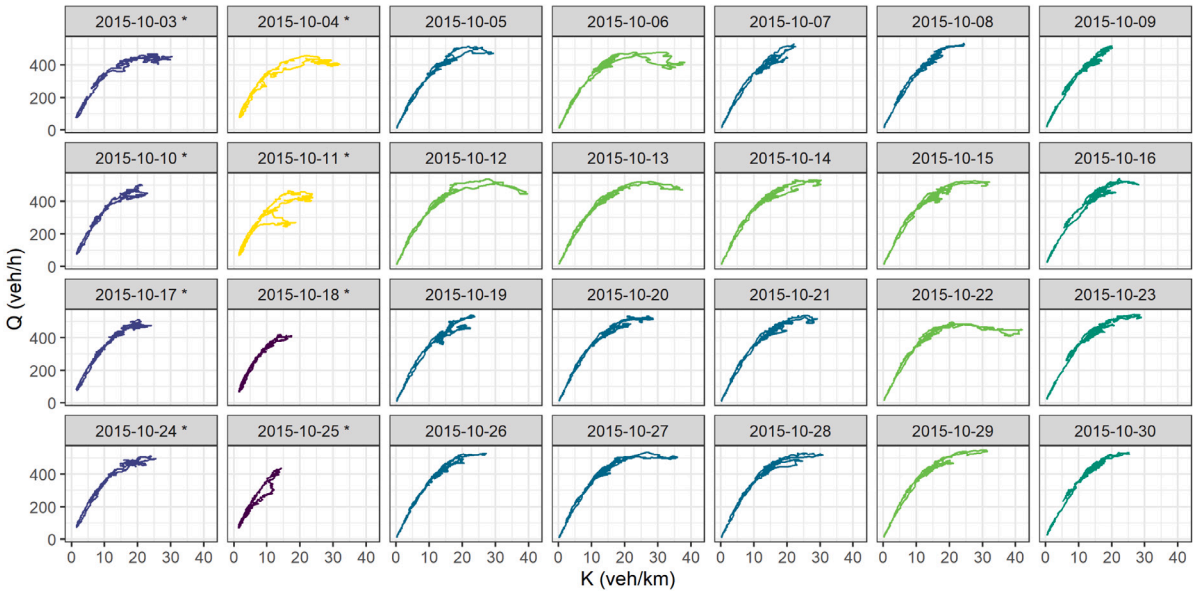


Fig. 6. Daily sMFDs in Lucerne during October 2015 color-clustered. * indicates a weekend day (Saturday and Sunday).

4.2. Clusters

Fig. 6 shows the daily dMFDs for four consecutive weeks in October 2015 for the city of Lucerne. Note that the dMFDs shown are not scaled, but correspond to the actual values of average flow (Q) and density (K). The colors indicate the dMFD cluster. For Lucerne, it is apparent that weekends (denoted by an asterisk) are well distinguished. Although it might be trivial to detect weekdays from weekends and holidays, this clear identification in the data underlines the power of DTW.

Although Fig. 6 shows rather similar sMFDs for weekdays, the clustering reveals differences in the loading and unloading patterns in the dMFD. This implies that the sMFD alone does not inform of the onset of network loading or the duration of congestion, but the resulting dMFD clusters capture these aspects, while at the same time reducing the dimensions of the full dataset to a more manageable scale.

Fig. 7 shows all clusters separately. Each one includes all observed dMFDs belonging to that specific cluster. Note that there is a substantial difference between the maximum flow attained in different clusters for the same city. For example, Lucerne's cluster 5 compared to cluster 6, exhibits a difference of roughly 50 veh/h or 10%. Moreover, for each of the clusters, although we still observe some differences in the dMFD, this one is much lower than that shown in Fig. 2.

Similarly, Fig. 8 shows the clustered average flow and average density. This figure emphasizes the similarity within a cluster and the differences between clusters. Recall that DTW captures not only differences in the magnitude of the observed dMFD but also its evolution over time. These results further validate the chosen approach as well as the number of clusters from a visual perspective. Clustering the observed dMFDs takes into account the loading and unloading profile of a network (within the limits of the DTW warping window).

Table 3
Properties of the medoid cluster.

Cluster	Q_{max} (veh/h)	Hysteresis	Size of hysteresis (veh/h)
Zurich 1	338	no	–
Zurich 2	331	no	–
Zurich 3	371	yes	40
Zurich 4	286	no	–
Zurich 5	369	yes	40
Zurich 6	319	no	–
Lucerne 1	413	no	–
Lucerne 2	448	no	–
Lucerne 3	522	yes	59
Lucerne 4	524	yes	62
Lucerne 5	491	yes	45
Lucerne 6	523	yes	58
Lucerne 7	440	no	–
Lucerne 8	450	yes	56

To further extract the features of each cluster, we estimate each cluster's medoid. Table 3 summarizes the properties of clusters' medoids for Zurich and Lucerne.

From a broader perspective, the relatively low number of clusters identified in a full year confirms seminal findings on the regularity of human mobility (González et al., 2008; Louail et al., 2015; Lopez et al., 2017). We can now confirm this from a macroscopic traffic perspective. For example, most dMFDs on a Saturday are similar to each other. This is the first long-term study confirming that the observed dMFD is indeed a suitable instrument for the macroscopic modeling and control of traffic, as it repeats itself over time. In other words, these clusters are of relevance for macroscopic traffic control, e.g., perimeter control. They allow to potentially pre-determine a control scenario for each cluster found and thereby reduce the control complexity.

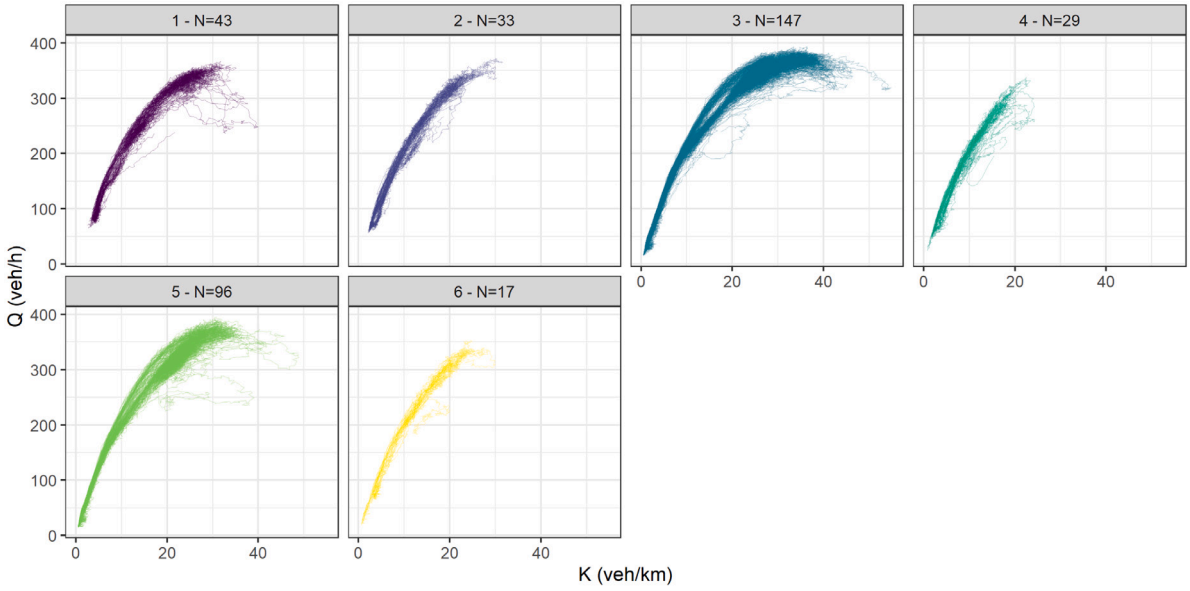
5. Understanding similarities

In the previous sections, we have successfully clustered the dMFDs using DTW similarity. In this section, we will investigate the reasons for the differences across clusters. First, we analyze the heterogeneity in the network and compare it to simulation results from the literature. Second, we compare the heterogeneity patterns and the clusters obtained in Section 4. Third, we include perimeter flows into the analysis to stress better when differences are driven by the demand patterns.

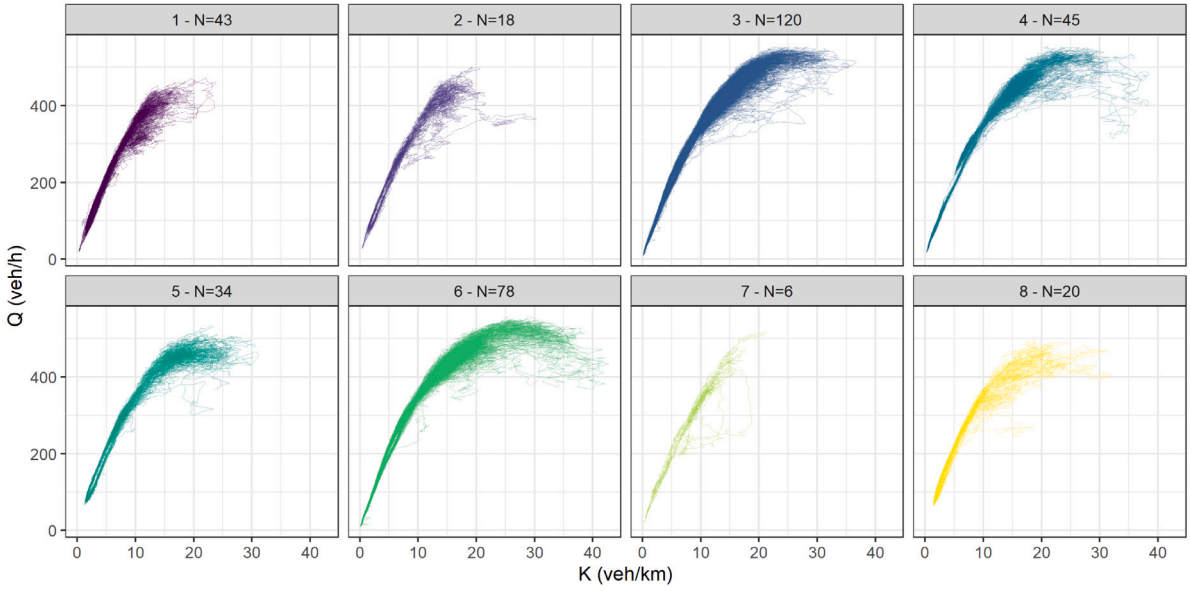
5.1. Empirical evidence on traffic heterogeneity (in density)

Existing literature connects spatial heterogeneity in traffic density and the observed dMFD. Among the first to investigate the effects of traffic heterogeneity was Buisson and Ladier (2009). In a dataset of three days, the authors attributed hysteresis loops in the empirical dMFD from Toulouse to heterogeneity in the distribution of vehicles. Later, simulation studies further investigated the issue. Mazloumian et al. (2011) and Geroliminis and Sun (2011) discovered that the spatial distribution of vehicle density affects the scatter of an observed dMFD and its shape. It is postulated that two time intervals with the same number of vehicles in a network and the same spatial distribution of vehicle densities lead to the same average flow. Similarly, Mahmassani et al. (2013) and Muhlich et al. (2015) verified the effects of the heterogeneity in traffic density on the network flow and the size of the dMFD hysteresis loops. Indeed, for a given average density, lower standard deviations in density lead to higher flows. Knoop et al. (2015) introduced a generalized notion of the MFD, where traffic heterogeneity in density is taken into account. The higher the heterogeneity in density, the lower the average flow. Similarly, Ramezani et al. (2015) look into the relationship between the average flow and the observed heterogeneity. Based on a aggregated modeling approach the authors introduce a hierarchical perimeter flow control that takes into account the neighborhood's heterogeneity. In the search for well-defined empirical dMFDs, numerous studies focused on a partitioning of urban networks which minimizes the overall heterogeneity measured in $sd(k)$ or $var(k)$ (Ji and Geroliminis, 2012; Saeedmanesh and Geroliminis, 2016, 2017; Ambühl et al., 2019).

So far, however, no empirical study has verified the simulation-based findings where heterogeneity in traffic density describes the dMFD's flow accurately. Let us first investigate the heterogeneity in density for the dMFDs covering the whole year shown in Fig. 9. We use the standard deviation in density as the heterogeneity measure and color the dMFDs accordingly. As expected, we observe lower heterogeneities for lower average densities. Thus, it is more interesting to compare the heterogeneities for a given average density. Here, we see that as the heterogeneity increases, the average flow decreases for any given average density. In other words, we reach high average flows, when traffic is well distributed across the network. While these findings confirm previous simulation studies, the figures also reveal another interesting phenomenon: a frontier for average flows — they cannot exceed a certain upper bound. Although a similar graph was previously shown in Ambühl et al. (2018a), the analysis here emphasizes that the existence of this upper bound is reinforced by the unstructured alignment of traffic states below this frontier. Note that this is data from one whole year. Thus, arguably, we have found strong indications for the empirical existence of the *upper* bound of the MFD (as in Ambühl et al. (2018a) and we identify at least one reason why the observed dMFDs fall below it.



(a) Zurich's clusters.

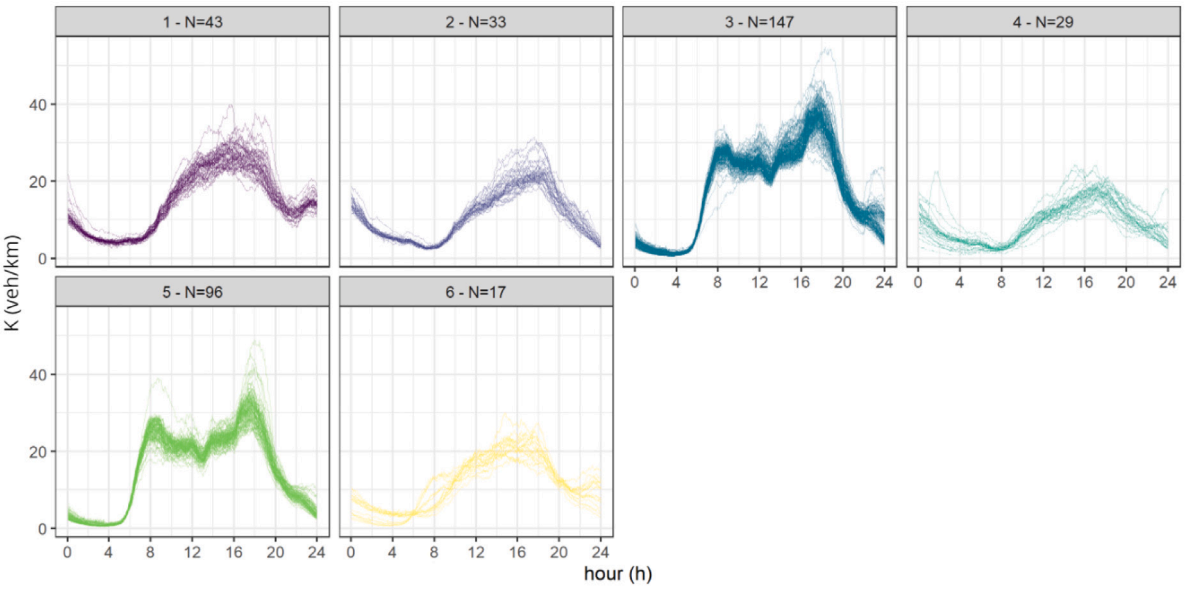


(b) Lucerne's clusters.

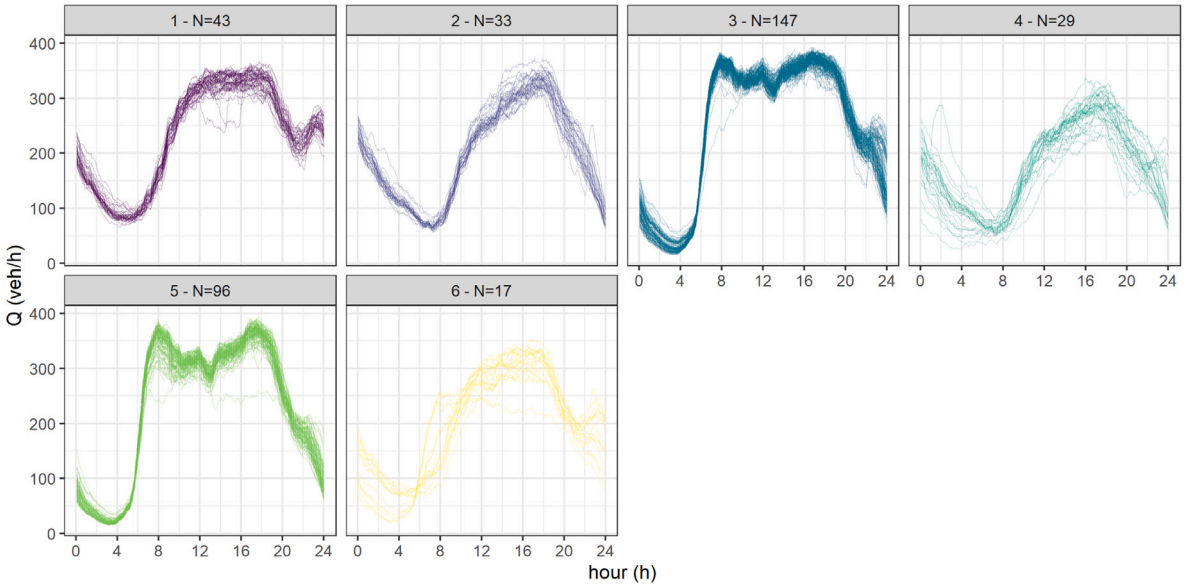
Fig. 7. Observed clusters and their size (N).

Simulation studies by Mazloumian et al. (2011), Mahmassani et al. (2013), and Knoop et al. (2015) have shown the standard deviation in density as a function of the average flow. Thus, we transform Fig. 9 into Fig. 10, to present further empirical evidence confirming the prior simulation findings. Note that the x-axis is now the standard deviation of density as in Mazloumian et al. (2011), Mahmassani et al. (2013), and Knoop et al. (2015) and the contour lines show the isolines of average density. To draw iso-lines of the average density, we bin our data into 1 veh/km, and 10 veh/h. For each bin, we calculate the average standard deviation. For a bin to be taken into account, it needed at least 5 observations.

It becomes clear that as the standard deviation in density increases for a given average density, the resulting average flow decreases. The slope also becomes steeper for higher average densities. In other words, the higher the average density is, the more traffic flow is governed by heterogeneity: a small change in the standard deviation has a drastic effect on the average flow. This means that the higher the observed average density, the more fragile the average flow becomes (i.e., more sensitive to spatial variations in density). Let us think of a perturbation (e.g. accident) as the cause for a change in standard deviation. In very



(a) Zurich's dMFD density.

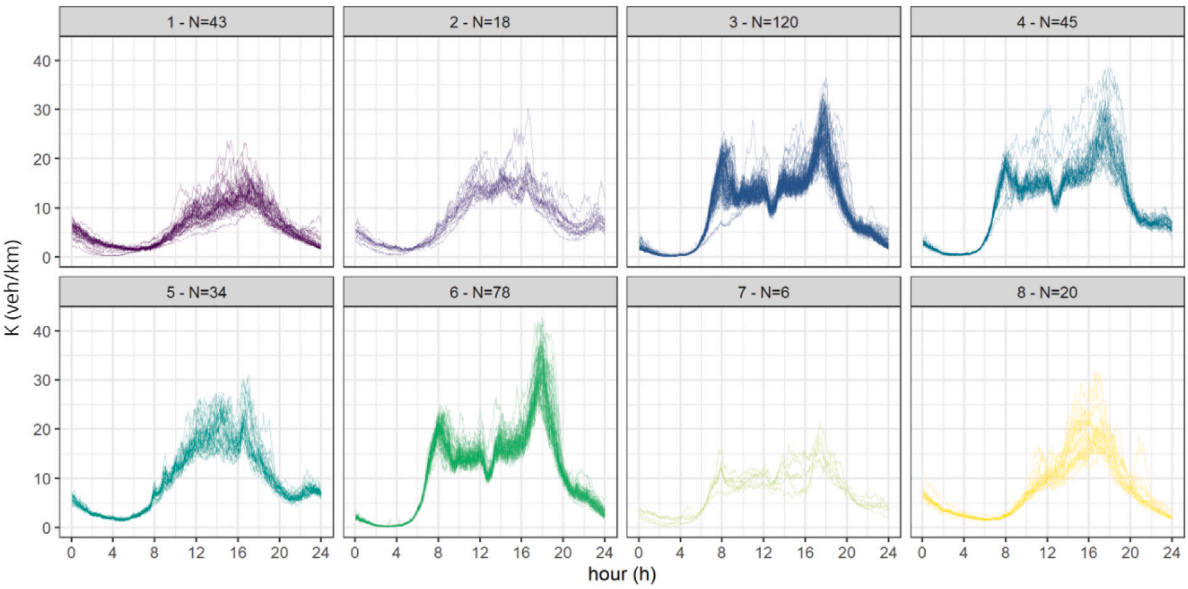


(b) Zurich's dMFD flow.

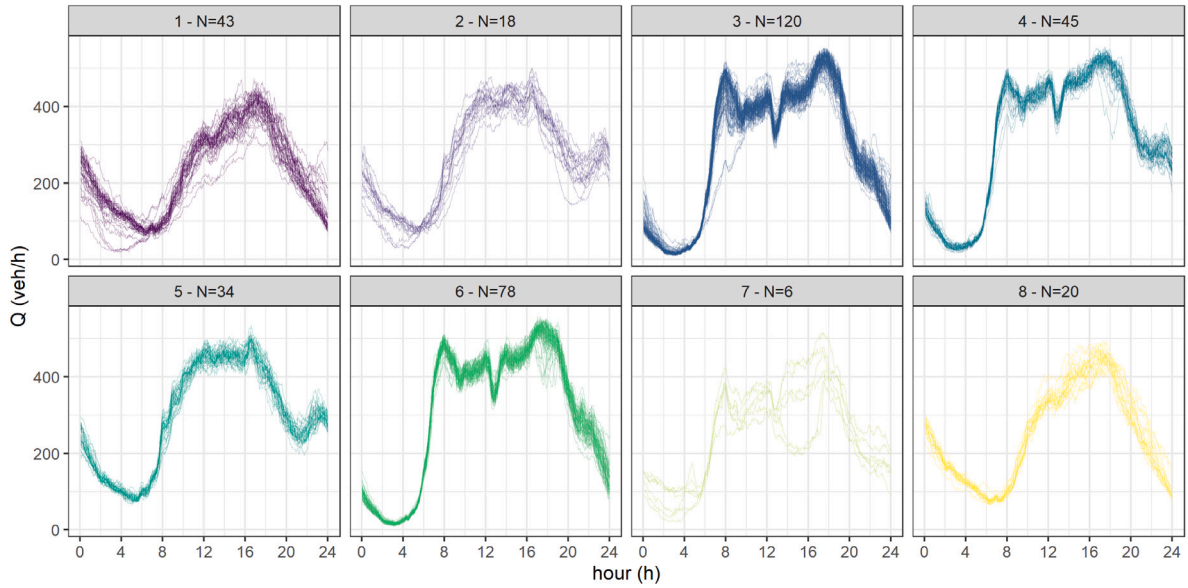
Fig. 8. Observed clusters in dMFD flow and density and their size (N).

uncongested regimes such perturbation does not affect the average flow substantially. Drivers can adapt their routes quickly, and find short detours. In congested regimes, however, this becomes much more complicated. It is hard to find a suitable detour when much of the network is already congested. Thus, average flows react strongly to such a shock. It is clear that this understanding is limited to the ranges of congestion that we observe empirically. It is important to note that we have not measured heavy macroscopic congestion (i.e., states close to gridlock). Hence, we might not be able to extend these conclusions to such situations.

We further differentiate between the loading and the unloading part of the dMFD. We find that generally, the average flow is slightly lower when unloading the network than during the loading phase. This is due to the hysteresis effects, which are discussed in e.g. Gayah and Daganzo (2011) or Paipuri et al. (2019). The slopes, however, follow the same trends independent of the loading or the unloading phase. Thus, the underlying mechanisms are the same.



(c) Lucerne's dMFD density.



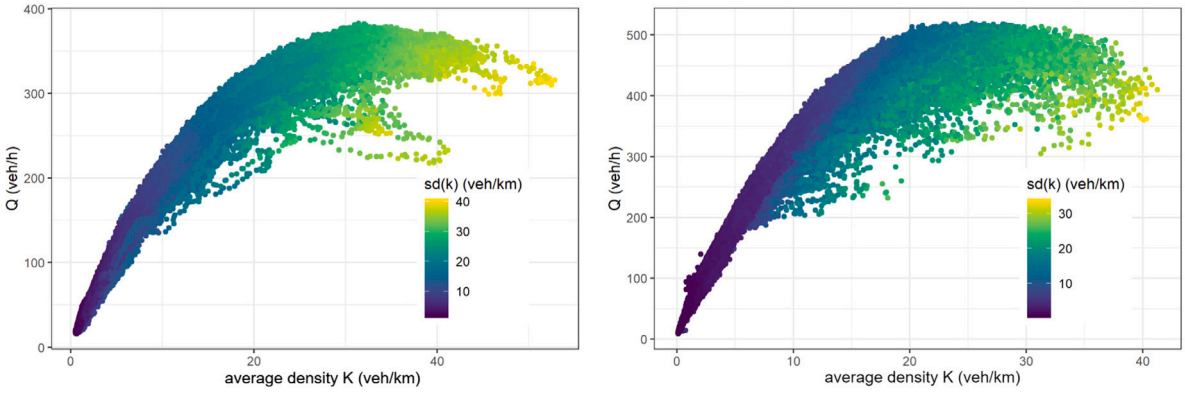
(d) Lucerne's dMFD flow.

Fig. 8. (*continued*).

5.2. Quantitative assessment of the interplay between the heterogeneity and the observed dMFD

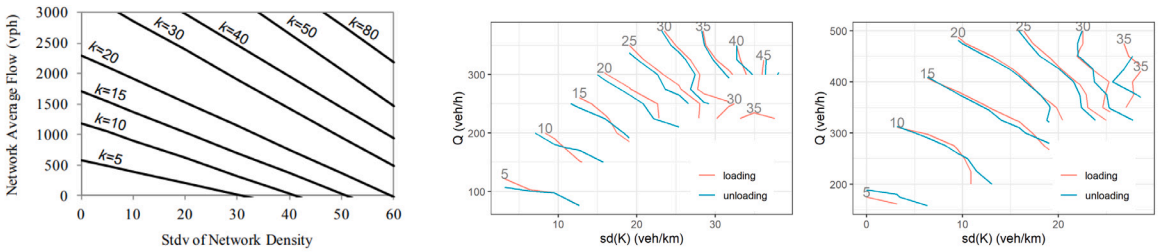
The results in Fig. 10 show a substantial influence of the heterogeneity in density on the dMFD's flow. We expect that the clustering of the evolution of the heterogeneity should yield similar results like the ones based on the dMFD in Section 4. Recall that we clustered our dMFDs with respect to their evolution over time. Unfortunately, time is not represented in Fig. 10, making a comparison with our previously derived clusters non-trivial. Thus, in the following, we resort to a more detailed analysis of traffic heterogeneity over time.

For completeness, we will also include the heterogeneity in flow. We create a joint time series composed of the normalized standard deviation of the vehicle flow and density in the network (normalization analog to Eq. (1)). Identical to the previous analysis in Section 3.2, we apply the DTW similarity measure on this joint time series of heterogeneity. The time-warping window is again 3 h. As before, this yields a distance matrix.



(a) Zurich. (b) Lucerne.

Fig. 9. All dMFDs over a year and the heterogeneity in density measured by the standard deviation (sd(k)) in density.



(a) Heterogeneity in Simulation. Figure from Mahmassani et al. (2013). (b) Observed heterogeneity in Zurich. (c) Observed heterogeneity in Lucerne.

Fig. 10. Heterogeneity in density (sd(k)) and the average flow. The contours describe the isolines for the vehicle density in veh/km.

Table 4
Mantel statistics for the comparison between the dMFD and the heterogeneity measures in flow and density. These statistics show a high structural overlap between the observed patterns.

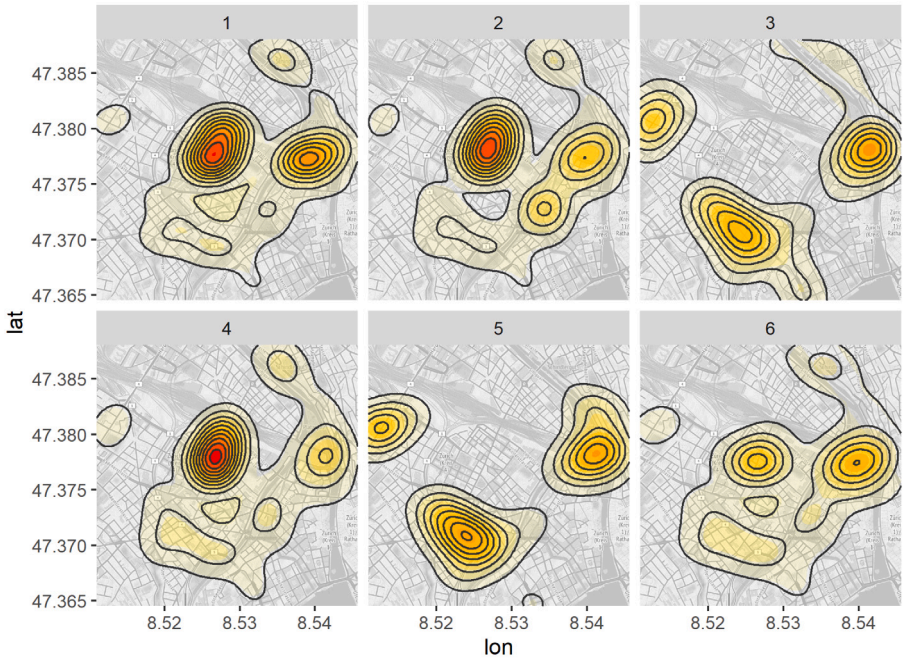
Dataset	Heterogeneity		
	Flow and density	Flow	Density
Zurich	0.987	0.981	0.978
Lucerne	0.972	0.971	0.842

We then compare how similar this new heterogeneity based distance matrix is to the one obtained from the dMFDs in Section 3.2. To that end, we compute the Mantel test statistic, whose interpretation we discussed in Section 3.3.

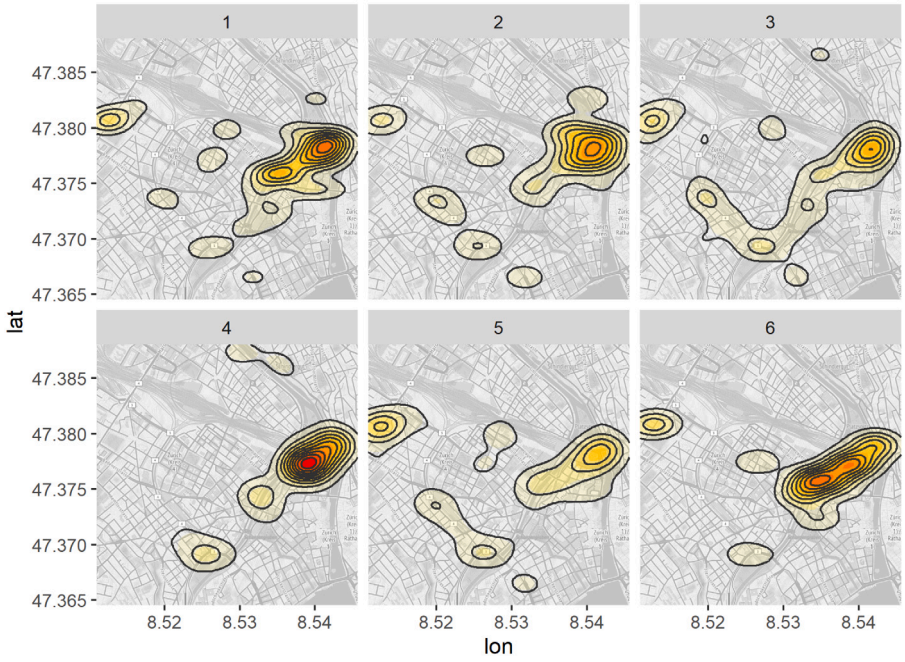
Table 4 presents the Mantel statistics for the cases where the dMFDs' DTW matrix is compared with the DTW distance matrix based on (i) the joint heterogeneity measures (normalized standard deviation of flow and density), (ii) only flow heterogeneity, (iii) only density heterogeneity. Not surprisingly, the joint time series yields the highest Mantel statistic, indicating a very high structural overlap between the two distance matrices. The statistic remains very high ($\geq 0.97 - 0.98$) if single heterogeneity measures are used on their own, except for the case of the density heterogeneity in Lucerne (≥ 0.84). This indicates that the heterogeneity in flow and density, are suitable predictors for the observed dMFD. Interestingly, we find that the heterogeneity in density alone is not the best predictor of the dMFD.

These results imply that if we were to cluster daily traffic patterns using their heterogeneity indicators, we would retrieve almost the same clusters. As a matter of fact, clustering the heterogeneities of flow and density yields the same clusters as before on 95% (Zurich) and 94% (Lucerne) of the days.

These findings here are interesting for three reasons. First, even though we use a dataset from stationary detectors which neither cover the whole network nor the full extent of the roads being monitored, the relationship holds at a higher level. Second, we observe these strong correlations despite the complexity of real traffic with its many transportation modes and interactions. Third, while the heterogeneity in density still yields reasonably high Mantel coefficients, the heterogeneity in flow only is a slightly better predictor throughout the analysis. Interestingly, such heterogeneity in flow has neither been analyzed nor reported in the literature, although in the case of Zurich flows have been used to inform the perimeter control strategies. Note that if the heterogeneity in



(a) Morning peak (04:30-09:00).



(b) Evening peak (15:30-19:00).

Fig. 11. Congestion hotspots during morning and evening peaks for Zurich's 6 clusters.

density describes the distribution of congestion over the network, we may claim that the heterogeneity in flow represents better the demand distribution over the network. In particular, we note that different flow values in the free-flow part of the fundamental diagram lead to close density values. So, in free-flow regimes, it is possible to observe a low heterogeneity in density, but high heterogeneity in flow at the same time.

With the above analysis, we have quantitatively confirmed the interplay between the spatial heterogeneity patterns and system performance. Our results imply that the evolution of the spatial heterogeneity in flow and density predicts the observed traffic performance (or vice versa).

5.3. Impact of the perimeter flow

Perimeter flows into the analyzed neighborhood are informative about the dynamic network loading process. Differences in perimeter flows could lead to differences in the network loading process and thus bottleneck activation. Consequently, differences in perimeter flow patterns could explain the observed differences in the dMFD. In other words, it is reasonable to assume that the interaction between perimeter flows and network topology are the key determinants of the spatial heterogeneity inside a neighborhood. To investigate this hypothesis, we use the detectors in Zurich and Lucerne shown in Fig. 3 that are located on main arterials leading into our previously investigated areas with a buffer of roughly 1 km.

Here, we define the distance matrix based on the flows measured on 32 (Zurich) and 20 (Lucerne) perimeter detectors. To capture differences in the perimeter flow, we refrain from aggregating the flows. Instead, the flow of the 32 detectors in Zurich build a 32-dimensional time series that serves as input to the DTW process defined in Section 3. Therefrom we construct a distance matrix for all 365 days. A similar approach is used for Lucerne.

Now, we can compare the obtained perimeter flow distance matrix and the one based on heterogeneity in flow and density derived in Section 5.2. We find a high Mantel statistic of 0.976 (Zurich) and 0.973 (Lucerne). We claim that the effects measured are the perimeter flows' influence on the spatial traffic heterogeneity of the neighborhood and the observed dMFD. Depending on the perimeter flow patterns, different bottlenecks are activated. To *qualitatively* validate this hypothesis we now analyze the congestion hotspots in the two cities.

Figs. 11 and 12 show the congestion hotspots (i.e., active bottlenecks) in Zurich and Lucerne during the morning and evening peaks. The intensity of a hotspot is defined by the average congestion duration of the detectors within that hotspot. A detector is classified as congested according to its fundamental diagram. We observe that similar clusters, as seen in Fig. 7 and Table 3, reveal similar, yet not identical, hotspot distributions. The morning and evening peaks differ substantially and indicate a clear shift of congestion within the network. This confirms that different bottlenecks are active during different times of the day and their activation depends on the traffic patterns represented by our clusters.

Unfortunately, the Mantel test statistics and the analysis provided in Figs. 11 and 12 do not offer information about the direction of influence (i.e. causality). Therefore, in the following, we investigate the causality between the perimeter flow and the observed heterogeneity.

We analyze the number of congested links in the perimeter as well as inside the neighborhood. It is clear that causality is generally hard to establish with empirical data. Fig. 13 shows that perimeter detectors record congestion with a certain delay compared to the ones inside the dMFD neighborhood. Hence, congestion seems to propagate from the neighborhood towards the perimeter and not the other way around. We deduce that the perimeter detectors affect the neighborhood's heterogeneity and not the other way around. This is also consistent with the fact that traffic congestion propagates backward. Therefore, we find causal evidence that the perimeter flow patterns indeed activate different bottlenecks, which in turn change the shape and evolution of the dMFD.

We have shown that the interaction between perimeter flow and network are the key determinants of the observed dMFD. However, it is worth highlighting that the perimeter flows alone are sufficient to characterize the dMFD clusters (see the high Mantel statistics presented in this section). This is an important finding as it simplifies potential cluster predictions substantially.

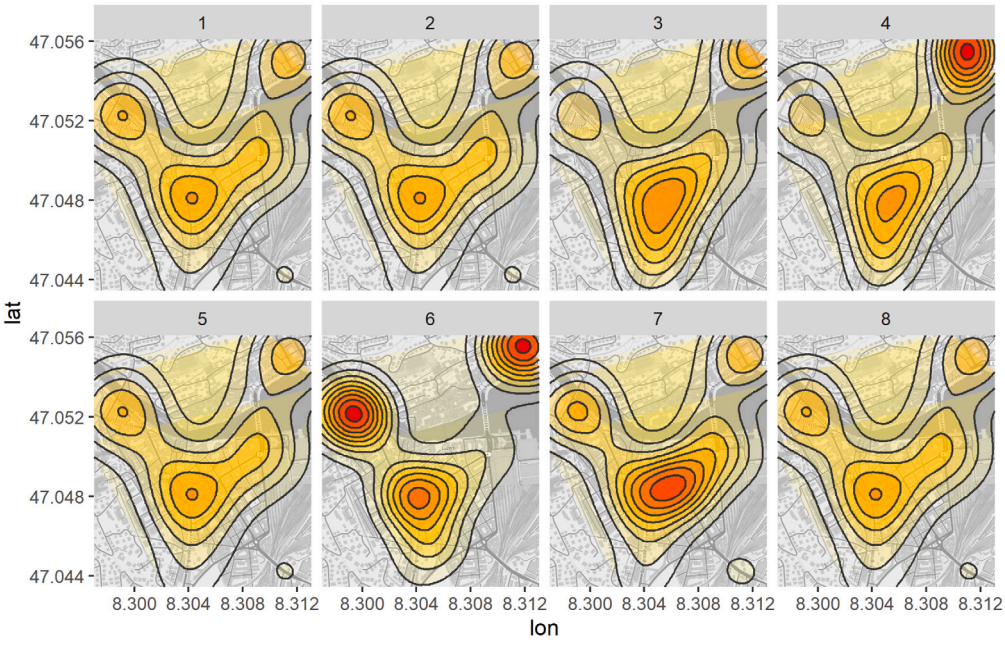
From a broader perspective, our perimeter detectors' flow can be seen as a proxy for the traffic demand pattern. Thus, our findings provide empirical evidence for the influence of traffic demand on the *observed* dMFD. Note that this is fundamentally different, but not contradictory to the theoretical/ideal MFD originally introduced by Daganzo (2007). The latter is the *upper bound* to the network, i.e. the maximum performance of a traffic network under perfect conditions. There is no evidence showing that this one is affected by the traffic demand. The observed dMFD, on the other hand, is.

6. Predicting the observed urban traffic performance

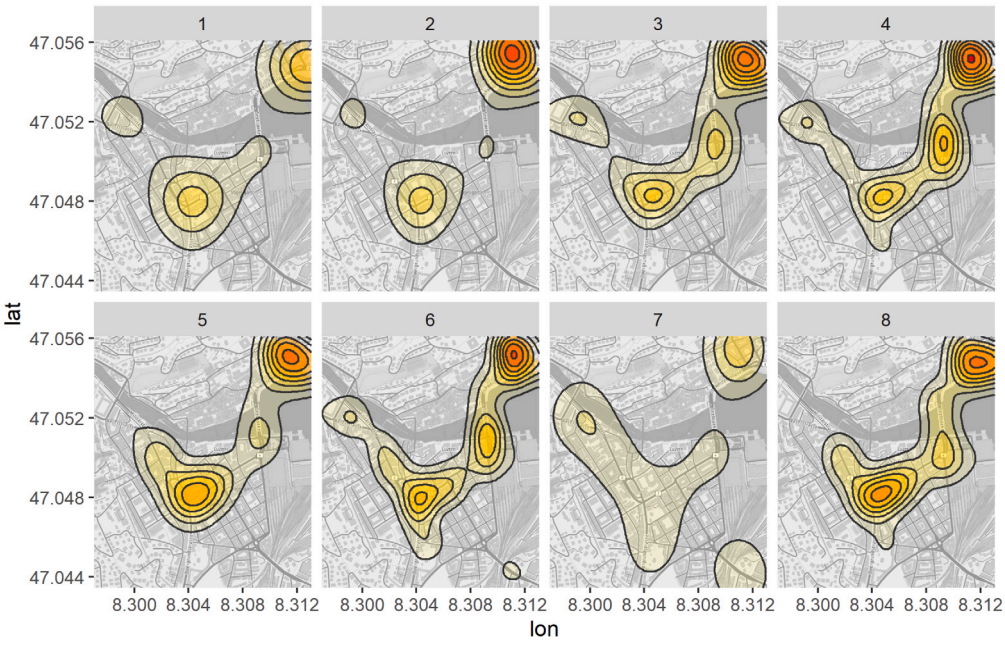
In the previous sections, we found a relatively small number of representative clusters to describe the daily macroscopic traffic patterns. We further showed that the perimeter flows are essentially linked to the observed shape and properties of the dMFD. Consequently, in this section, we are interested in understanding whether monitoring the perimeter loop detectors allows us to identify the correct cluster early in the day. The earlier we can predict the corresponding dMFD cluster, the better cities can deploy the appropriate control. Limiting the observation to a few perimeter loops is reasonable as most cities' traffic monitoring schemes do not monitor all roads. That is to say, for control purposes, it is paramount to build a parsimonious early detection system.

Like most clustering algorithms, our hierarchical clustering method is not able to label *new* data instances. Hence, we resort to the concept of *classification*. The core idea of our prediction routine is to use the clusters derived in Section 4 as labels in a supervised learning model.

To that end, we split our empirical dataset into two. We use a classical 70/30 approach, where 255 days (70%) of our dataset are used to train (or calibrate) our prediction, and the remaining 110 days (30%) are used to test (or validate) the prediction. The following steps describe the approach, where each of the 110 test days is estimated separately.



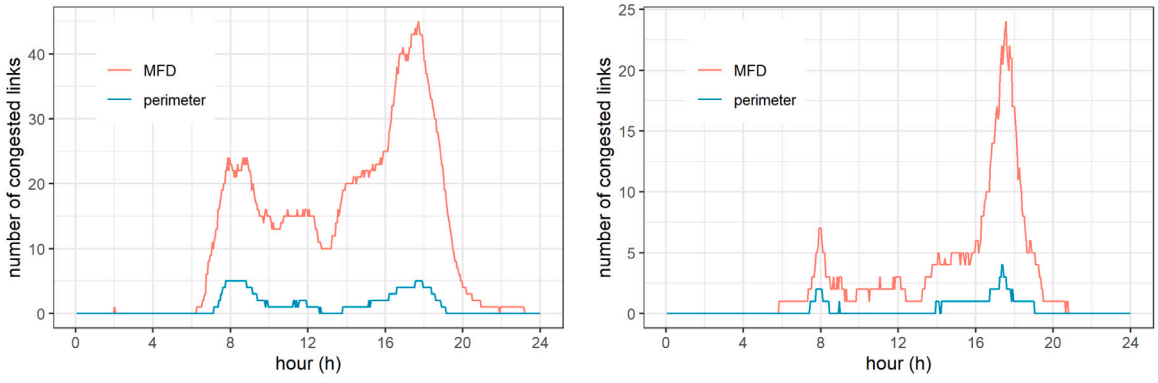
(a) Morning peak (04:30-09:00).



(b) Evening peak (15:30-19:00).

Fig. 12. Congestion hotspots during morning and evening peaks for Lucerne's 8 clusters.

1. Create temporal subsets of the perimeter flow dataset (6 time slots: midnight–05:00; midnight–08:00; midnight–11:00; midnight–14:00; midnight–17:00; midnight–20:00). As in Section 5.3, we refrain from aggregating the perimeter detectors — instead we create a multi-dimensional time series, where each dimension represents a perimeter detector.
2. Run DTW on the datasets of step 1 and calculate the distance matrix across all days.



(a) Zurich. (b) Lucerne.

Fig. 13. Number of congested links in the perimeter and the neighborhood covered by the dMFD. Notice that the perimeter loops exhibit congestion after the dMFD neighborhood itself.

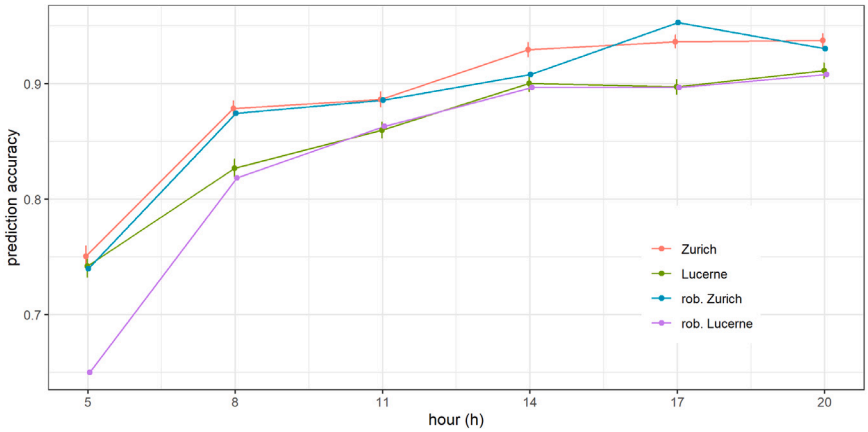


Fig. 14. Accuracy of our cluster prediction.

3. Find the k-nearest neighbor (kNN) (here k = 10) in the training dataset for every test date. Assign the most likely cluster based on the frequency of the nearest clusters for each test day.
4. Compare the results to the original cluster assignment (labels) derived in Section 4.

Remember that the aim is to predict the correct cluster as early as possible. Therefore, to evaluate the prediction accuracy over time, we repeat this approach for different time slots, e.g. midnight–05:00, midnight–08:00, etc (i.e., step 1).

Fig. 14 shows the fraction of correctly determined clusters over time. For example, running the prediction using data until 08:00, results already in a fraction of 88% correctly determined clusters for Zurich. To retrieve the 95th confidence interval of this estimation, we repeat the prediction process by randomly assigning training and test days 100 times. It is shown as a range around the prediction points in Fig. 14. This accuracy can be considered as very high, for two reasons. First, we only use a (small) subset of the data (only perimeter detectors, and temporal limits) to predict. Second, some days are very difficult to predict, e.g., due to an accident, that very much influences the dMFD's trajectory suddenly. In such cases, the cluster found over the whole day might substantially differ from the one found early in the morning and it remains almost impossible to correctly classify such special cases.

The cluster prediction is slightly less efficient for the case of Lucerne in Fig. 14. We can correctly estimate 84% of the clusters by 08:00 with our approach. The overall trend, however, is similar to the one of Zurich. Given the stochastic nature of empirical data, there might be multiple reasons why we can estimate the clusters less accurately. Lucerne offers lower data quality and has a higher number of clusters. Also, the morning patterns in Lucerne might be less decisive for the whole dMFD pattern. This is confirmed by our hotspot analysis in Fig. 12, where the differences in the morning peak are less apparent.

Nonetheless, in most cases, we can predict the correct cluster early in the day. We can thus conjecture that the morning traffic conditions are strongly related to those in the afternoon, especially in Zurich. This is intuitive, as most working trips have an inverse OD pattern for morning and evening. This shows that the morning peak is decisive for much of the day's traffic patterns. In addition to the fact that there are relatively few dMFD clusters during a year, these findings are important for macroscopic control as they allow to identify the prevalent macroscopic traffic pattern early in the day and simplify macroscopic traffic state estimations.

Changes to the network will induce changes to the observed MFD. Our analysis shows that the underlying cause of this is a different bottleneck activation. However, this is not a limitation to our prediction algorithm. On the contrary, in an industrial deployment (i.e. off-the-shelf type of deployment), where the classification is updated (or relaunched) regularly, this change will be picked up naturally by our prediction algorithm. Therefore, we consider the algorithm to be robust by design. To further investigate such claim, we show the robustness of our prediction algorithm by using only the first 6 months to calibrate our prediction method and apply it in the following 3 months. If there were changes to the network in the second half of the year, predicting its daily MFDs based on the first half of the year would be harder. Fig. 14 also includes this sensitivity analysis. We see that the new prediction (rob. Zurich and rob. Lucerne) remains robust. In fact, we do not see a substantially lower prediction accuracy. This implies first that our prediction algorithm also works well with a lower number of days, and second that there were no substantial changes to the network between the first and the second half of the year.

We also tested our framework using a support vector machine (SVM) and a random forest classifier. Interestingly, the resulting accuracy did not vary significantly. Given the simplicity of the kNN classifier, we refrained from showing the other results.

7. Conclusions

The data-driven findings presented in this paper show how temporal patterns of vehicle flows define the performance of urban road networks. The contributions of this paper are five-fold.

First, we present two high-resolution traffic datasets covering a year each. This is one of the largest periods for which high-resolution traffic patterns have been analyzed so far.

Second, we introduce a methodology to quantify the similarity across multiple macroscopic traffic patterns. We do so by using the concepts of the MFD and a DTW based algorithm for time series. This allows us to derive a few (6 for Zurich, 8 for Lucerne) representative dMFD clusters that capture the essential macroscopic traffic features. The clusters clearly differ in the K-Q-space and with respect to the observed maximum flow, the critical density, and the presence of hysteresis (see also Table 3).

Third, we provide an in-depth analysis of traffic heterogeneity in the network. Thereby, we confirm, for the first time, previous simulation results with empirical data, including analysis about the loading and the unloading of the network. We also find evidence for an upper bound as outlined in the early theoretical studies on the dMFD. Further, our analysis reveals that not only heterogeneity in traffic density is of importance, but even more so the heterogeneity in traffic flow. We then show that the evolution of the spatial heterogeneity is a good predictor for the clusters found.

Fourth, using a quantitative approach, we find that perimeter flows are closely linked to the evolution of the spatial heterogeneity in a network. Depending on the perimeter flow distribution, different bottlenecks are activated. Using a congestion hotspot analysis, we further confirm that our clusters indeed yield different traffic states. Coupled with an analysis of the number of congested links in the perimeter and the dMFD neighborhood, we find that perimeter flows affect the observed spatial heterogeneity and are thus also linked to the dMFD clusters found.

Fifth, we use a parsimonious classification approach to predict the expected clusters early in the morning. This indicates that, for the two cities investigated, the morning traffic patterns are closely linked to the evening patterns. It is clear that special cases, like accidents, lower the accuracy of our methodology as they create unpredictable patterns. In reference to our introductory Fig. 1, any application requiring the definitions described in points IIa and IIb will benefit from our approach. As explained previously, a macroscopic control will not only benefit from an accurate estimation of the observed dMFD's shape but also from knowledge about the network loading process, i.e. when and how fast do we reach the maximum flow, whether a hysteresis is expected, etc.

Our findings have practical implications for traffic monitoring and traffic control for cities around the world. We show that the *observed* (or realized) MFD (dMFD or sMFD) is demand-dependent. Thus, applications relying on the *observed* MFD require a thorough online calibration. Furthermore, we find further evidence for the regularity of human mobility by clustering the observed macroscopic traffic conditions. We show that it is enough to monitor the traffic performance of a few roads only to classify daily patterns. This allows not only to reduce monitoring requirements (and costs), but also provides opportunities for enhanced traffic control. We show how the revealed relationships can be used in an early pattern detection for the expected dMFD, providing opportunities to counteract congestion early and apply macroscopic traffic control accordingly. Our framework is flexible and adaptive: if the level of detail is regarded as too high, it is easily possible to reduce the number of clusters chosen. The number of clusters shown in this paper is the one deemed optimal by statistical methods, but it might not necessarily be the one that satisfies the specific application requirements. If changes to the network occur, they will induce changes to the bottleneck activation, and they will automatically be picked up by our prediction algorithm. In an industrial deployment (i.e. off-the-shelf type of deployment), it is clear that the algorithm would automatically update regularly based on the new days observed.

From a broader perspective, our study allows to predict average travel times in a region accurately, and is thus suitable for any models that require such dynamic estimation of macroscopic travel times, e.g., pick-up and drop-off modeling of on-demand mobility.

Nonetheless, our study bears some limitations. First, similar to many empirical studies, and despite our care when dealing with the empirical datasets, there might still be uncaught measurement errors and outliers. Second, causality is difficult to prove with empirical data only. Nonetheless, we show the first evidence for the causality between perimeter flows, spatial traffic heterogeneity, and the dMFD itself. Still, this understanding needs a more thorough investigation. Third, the generality of the findings are also subject to discussion. The methodology applies globally, while the results inferred from data, i.e. the number and characteristics of each cluster, are context specific. In particular, the clustering's reduction of dimensions in the reported order of magnitude is an

effect expected in most cities worldwide. Consequently, cities have to identify *their* clusters with the proposed methodology, which becomes more and more feasible with the increase in data availability, e.g. floating car data.

This study lays the groundwork for future research aiming at deepening the understanding of the effects that demand patterns have onto the dMFD. For example, some dMFDs exhibit clear hysteresis, while others do not. Once such effects are understood, we could potentially determine ways to reduce detrimental hysteresis effects. Similarly, an avenue of research is the analysis of the clear upper bound that we found in our dMFDs. What are the key factors affecting such upper bound, how can we shift traffic towards such upper bound? The purpose of this study is to demonstrate that it is conceptually possible to define a framework capable of disentangling, clustering, and predicting the observed daily MFDs. From an algorithmic perspective, future research could further refine the techniques used in this paper, be it the clustering algorithm or the prediction mechanism.

CRediT authorship contribution statement

Lukas Ambühl: Conceptualization, Data curation, Investigation, Formal analysis, Methodology, Writing. **Allister Loder:** Data curation, Investigation, Methodology, Writing. **Ludovic Leclercq:** Conceptualization, Investigation, Methodology, Supervision, Writing. **Monica Menendez:** Supervision, Writing.

Acknowledgments

L. Ambühl acknowledges the support by the ETH Research Grant, Switzerland ETH-27 16–1 under the project name SPEED. A. Loder acknowledges the support by the ETH Research Grant, Switzerland ETH-04 15-1 under the project name How many are too many? L. Leclercq acknowledges funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 646592 – MAGnUM project). M. Menendez acknowledges support by the NYUAD Center for Interacting Urban Networks (CITIES), United Arab Emirates, funded by Tamkeen under the NYUAD Research Institute Award CG001 and by the Swiss Re Institute under the Quantum Cities™ initiative.

References

Ambühl, L., Loder, A., Bliemer, M.C., Menendez, M., Axhausen, K.W., 2018a. Introducing a re-sampling methodology for the estimation of empirical macroscopic fundamental diagrams. Transp. Res. Rec. 2672 (20), 239–248. http://dx.doi.org/10.1177/0361198118788181.

Ambühl, L., Loder, A., Bliemer, M.C., Menendez, M., Axhausen, K.W., 2020. A functional form with a physical meaning for the macroscopic fundamental diagram. Transp. Res. B 137, 119–132. http://dx.doi.org/10.1016/j.trb.2018.10.013.

Ambühl, L., Loder, A., Menendez, M., Axhausen, K., 2016. Empirical macroscopic fundamental diagrams: New insights from loop detector and floating car data. Paper presented at the 96th Annual Meeting of the Transportation Research Board, Washington, DC, USA.

Ambühl, L., Loder, A., Menendez, M., Axhausen, K.W., 2018b. A case study of zurich's two-layered perimeter control. In: 7th Transport Research Arena. Vienna, http://dx.doi.org/10.3929/ETHZ-B-000206987.

Ambühl, L., Loder, A., Zheng, N., Axhausen, K.W., Menendez, M., 2019. Approximative network partitioning for MFDs from stationary sensor data. Transp. Res. Rec. 2673 (6), 94–103. http://dx.doi.org/10.1177/0361198119843264.

Ampountolas, K., Kouvelas, A., 2015. Real-time estimation of critical values of the macroscopic fundamental diagram for maximum network throughput. Paper presented at the 94th Annual Meeting of the Transportation Research Board, Washington, DC, USA.

Bagnall, A., Lines, J., Bostrom, A., Large, J., Keogh, E., 2017. The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances. Data Min. Knowl. Discov. 31 (3), 606–660. http://dx.doi.org/10.1007/s10618-016-0483-9.

Bahlmann, C., Burkhardt, H., 2004. The writer independent online handwriting recognition system frog on hand and cluster generative statistical dynamic time warping. IEEE Trans. Pattern Anal. Mach. Intell. 26 (3), 299–310. http://dx.doi.org/10.1109/TPAMI.2004.1262308.

Berndt, D.J., Clifford, J., 1994. Using dynamic time warping to find patterns in time series. In: Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining. AAAIWS'94, AAAI Press, pp. 359–370.

Buisson, C., Ladier, C., 2009. Exploring the impact of homogeneity of traffic measurements on the existence of macroscopic fundamental diagrams. Transp. Res. Rec. 2124 (1), 127–136. http://dx.doi.org/10.3141/2124-12.

Çolak, S., Lima, A., González, M.C., 2016. Understanding congested travel in urban areas. Nature Commun. 7, 10793. http://dx.doi.org/10.1038/ncomms10793.

Daganzo, C.F., 2007. Urban gridlock: Macroscopic modeling and mitigation approaches. Transp. Res. B 41 (1), 49–62. http://dx.doi.org/10.1016/j.trb.2006.03.001.

Daganzo, C.F., Geroliminis, N., 2008. An analytical approximation for the macroscopic fundamental diagram of urban traffic. Transp. Res. B 42 (9), 771–781. http://dx.doi.org/10.1016/j.trb.2008.06.008.

de Livera, A.M., Hyndman, R.J., Snyder, R.D., 2011. Forecasting time series with complex seasonal patterns using exponential smoothing. J. Amer. Statist. Assoc. 106 (496), 1513–1527. http://dx.doi.org/10.1198/jasa.2011.tm09771.

Filzmoser, P., Maronna, R., Werner, M., 2008. Outlier identification in high dimensions. Comput. Stat. Data Anal. 52 (3), 1694–1711. http://dx.doi.org/10.1016/j.csda.2007.05.018.

Gayah, V.V., Daganzo, C.F., 2011. Clockwise hysteresis loops in the macroscopic fundamental diagram: An effect of network instability. Transp. Res. B 45 (4), 643–655. http://dx.doi.org/10.1016/j.trb.2010.11.006.

Geroliminis, N., Daganzo, C.F., 2007. Macroscopic modeling of traffic in cities. Paper presented at the 86th Annual Meeting of the Transportation Research Board, Washington, DC, USA.

Geroliminis, N., Daganzo, C.F., 2008. Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. Transp. Res. B 42 (9), 759–770. http://dx.doi.org/10.1016/j.trb.2008.02.002.

Geroliminis, N., Levinson, D.M., 2009. Cordon pricing consistent with the physics of overcrowding. In: Lam, W.H.K., Wong, S.C., Lo, H.K. (Eds.), Transportation and Traffic Theory 2009: Golden Jubilee. Springer, Boston, MA, pp. 219–240. http://dx.doi.org/10.1007/978-1-4419-0820-9_11.

Geroliminis, N., Sun, J., 2011. Properties of a well-defined macroscopic fundamental diagram for urban traffic. Transp. Res. B 45 (3), 605–617. http://dx.doi.org/10.1016/j.trb.2010.11.004,.

Geroliminis, N., Zheng, N., Ampountolas, K., 2014. A three-dimensional macroscopic fundamental diagram for mixed bi-modal urban networks. Transp. Res. C 42, 168–181. http://dx.doi.org/10.1016/j.trc.2014.03.004.

González, M.C., Hidalgo, C.A., Barabási, A.L., 2008. Understanding individual human mobility patterns. Nature 453 (7196), 779–782. http://dx.doi.org/10.1038/nature06958.

Haddad, J., Geroliminis, N., 2012. On the stability of traffic perimeter control in two-region urban cities. Transp. Res. B 46 (9), 1159–1176. http://dx.doi.org/10.1016/j.trb.2012.04.004.
Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning. In: Springer Series in Statistics, vol. 27, http://dx.doi.org/10.1007/b94608.
He, H., Yang, K., Liang, H., Menendez, M., Guler, S.I., 2019. Providing public transport priority in the perimeter of urban networks: A bimodal strategy. Transp. Res. C 107, 171–192. http://dx.doi.org/10.1016/j.trc.2019.08.004.
Jain, A.K., Murty, M.N., Flynn, P.J., 1999. Data clustering: A review. ACM Comput. Surv. 31 (3).
Ji, Y., Geroliminis, N., 2012. On the spatial partitioning of urban transportation networks. Transp. Res. B 46 (10), 1639–1656. http://dx.doi.org/10.1016/j.trb.2012.08.005.
Ji, Y., Luo, J., Geroliminis, N., 2014. Empirical observations of congestion propagation and dynamic partitioning with probe data for large-scale systems. Transp. Res. Rec. 2422 (1), 1–11. http://dx.doi.org/10.3141/2422-01.
Kaufman, L., Rousseeuw, P., 2009. Finding Groups in Data: An Introduction to Cluster Analysis. John Wiley & Sons, http://dx.doi.org/10.1002/9780470316801.
Keyvan-Ekbatani, M., Gao, X., Gayah, V.V., Knoop, V.L., 2016. Examining perimeter gating control of urban traffic networks with locally adaptive traffic signals. In: Traffic and Granular Flow '15. Springer International Publishing, pp. 579–586. http://dx.doi.org/10.1007/978-3-319-33482-0_73.
Knoop, V.L., de Jong, D., Hoogendoorn, S., 2014. The influence of the road layout on the network fundamental diagram. Transp. Res. Rec. 2421 (1), 22–30. http://dx.doi.org/10.3141/2421-03.
Knoop, V.L., Van Lint, H., Hoogendoorn, S.P., 2015. Traffic dynamics: Its impact on the macroscopic fundamental diagram. Physica A 438, 236–250. http://dx.doi.org/10.1016/j.physa.2015.06.016.
Kowarik, A., Templ, M., 2016. Imputation with the R package VIM. J. Stat. Softw. 74 (1), 1–16. http://dx.doi.org/10.18637/jss.v074.i07.
Laval, J.A., Castrillón, F., 2015. Stochastic approximations for the macroscopic fundamental diagram of urban networks. Transp. Res. B 81, 904–916. http://dx.doi.org/10.1016/j.trb.2015.09.002.
Leclercq, L., Parzani, C., Knoop, V.L., Amourette, J., Hoogendoorn, S.P., 2015. Macroscopic traffic dynamics with heterogeneous route patterns. Transp. Res. C 7, 631–650. http://dx.doi.org/10.1016/j.trpro.2015.06.033.
Loder, A., Ambühl, L., Menendez, M., Axhausen, K.W., 2019. Understanding traffic capacity of urban networks. Sci. Rep. 9 (1), 16283. http://dx.doi.org/10.1038/s41598-019-51539-5.
Lopez, C., Leclercq, L., Krishnakumari, P., Chiabaut, N., Van Lint, H., 2017. Revealing the day-to-day regularity of urban congestion patterns with 3D speed maps. Sci. Rep. 7 (1), 14029. http://dx.doi.org/10.1038/s41598-017-14237-8.
Louail, T., Lenormand, M., Picornell, M., Cantú, O.G.a., Herranz, R., Frias-Martinez, E., Ramasco, J.J., Barthelemy, M., 2015. Uncovering the spatial structure of mobility networks. Nature Commun. 6, http://dx.doi.org/10.1038/ncomms7007.
Mahmassani, H.S., Saberi, M., Zockaie, A., 2013. Urban network gridlock: Theory, characteristics, and dynamics. Transp. Res. C 36, 480–497. http://dx.doi.org/10.1016/j.trc.2013.07.002.
Mahmassani, H., Williams, J., Herman, R., 1987. Performance of urban traffic networks. In: Proceedings of the 10th International Symposium on Transportation and Traffic Theory. pp. 1–20.
Mantel, N., 1967. The detection of disease clustering and a generalized regression approach.. Cancer research 27 (2), 209–220.
Mariotte, G., Leclercq, L., Laval, J.A., 2017. Macroscopic urban dynamics: Analytical and numerical comparisons of existing models. Transp. Res. B 101, 245–267. http://dx.doi.org/10.1016/j.trb.2017.04.002.
Mazloumian, A., Geroliminis, N., Helbing, D., 2011. The spatial variability of vehicle densities as determinant of urban network capacity. Phil. Trans. R. Soc. A 368 (1928), 4627–4647. http://dx.doi.org/10.1098/rsta.2010.0099.
Muhlich, N., Gayah, V.V., Menendez, M., 2015. An examination of MFD hysteresis patterns for hierarchical urban street networks using micro-simulation. Transp. Res. Rec. 2491 (2491), 117–126. http://dx.doi.org/10.3141/2491-13.
Necula, E., 2015. Analyzing traffic patterns on street segments based on GPS data using R. Transp. Res. Proc. 10, 276–285. http://dx.doi.org/10.1016/j.trpro.2015.09.077.
Ortigosa, J., Menendez, M., Tapia, H., 2014. Study on the number and location of measurement points for an MFD perimeter control scheme: a case study of Zurich. EURO J. Transp. Logist. 3 (3–4), 245–266. http://dx.doi.org/10.1007/s13676-013-0034-0.
Paipuri, M., Leclercq, L., Krug, J., 2019. Validation of MFD-based models with microscopic simulations on real networks: Importance of production hysteresis and trip lengths estimation. Transp. Res. Rec. 2673 (5), 478–492. http://dx.doi.org/10.1177/0361198119839340.
Przybyla, J., Taylor, J., Jupe, J., Zhou, X., 2015. Estimating risk effects of driving distraction: A dynamic errorable car-following model. Transp. Res. C 50, 117–129. http://dx.doi.org/10.1016/j.trc.2014.07.013.
Ramezani, M., Haddad, J., Geroliminis, N., 2015. Dynamics of heterogeneity in urban networks: Aggregated traffic modeling and hierarchical control. Transp. Res. B 74, 1–19. http://dx.doi.org/10.1016/j.trb.2014.12.010.
Saeedmanesh, M., Geroliminis, N., 2014. Observing MFDs for heterogeneous traffic networks with stop-line loop detector data. In: 16th Swiss Transport Research Conference, Ascona.
Saeedmanesh, M., Geroliminis, N., 2016. Clustering of heterogeneous networks with directional flows based on "Snake" similarities. Transp. Res. B 91 (91), 250–269. http://dx.doi.org/10.1016/j.trb.2016.05.008.
Saeedmanesh, M., Geroliminis, N., 2017. Dynamic clustering and propagation of congestion in heterogeneously congested urban traffic networks. Transp. Res. B 105, 193–211. http://dx.doi.org/10.1016/j.trb.2017.08.021.
Saffari, E., Yildirimoglu, M., Hickman, M., 2020. A methodology for identifying critical links and estimating macroscopic fundamental diagram in large-scale urban networks. Transp. Res. C 119, 102743. http://dx.doi.org/10.1016/j.trc.2020.102743.
Sardá-Espinosa, A., 2019. Time-series clustering in R Using the dtwclust package. R J. 11 (1), http://dx.doi.org/10.32614/rj-2019-023.
Sharma, A., Zheng, Z., Bhaskar, A., 2018. A pattern recognition algorithm for assessing trajectory completeness. Transp. Res. C 96, 432–457. http://dx.doi.org/10.1016/j.trc.2018.09.027.
Sun, L., Gong, Q., Yao, L., Luo, W., Zhang, T., 2018. A dynamic time warping algorithm based analysis of pedestrian shockwaves at bottleneck. J. Adv. Transp. 2018, http://dx.doi.org/10.1155/2018/1269439.
Taylor, J., Zhou, X., Rouphail, N.M., Porter, R.J., 2015. Method for investigating intradriver heterogeneity using vehicle trajectory data: A dynamic time warping approach. Transp. Res. B 73, 59–80. http://dx.doi.org/10.1016/j.trb.2014.12.009.
Thorndike, R.L., 1953. Who belongs in the family? Psychometrika 18 (4), 267–276. http://dx.doi.org/10.1007/BF02289263.
Tilg, G., Amini, S., Busch, F., 2020. Evaluation of analytical approximation methods for the macroscopic fundamental diagram. Transp. Res. C 114, 1–19. http://dx.doi.org/10.1016/j.trc.2020.02.003.
Tormene, P., Giorgino, T., Quaglini, S., Stefanelli, M., 2009. Matching incomplete time series with dynamic time warping: an algorithm and an application to post-stroke rehabilitation. Artif. Intell. Med. 45 (1), 11–34. http://dx.doi.org/10.1016/j.artmed.2008.11.007.
Van Schaik, C.P., Ancrenaz, M., Borgen, G., Galdikas, B., Knott, C.D., Singleton, I., Suzuki, A., Suci Utami, S., Merrill, M., 2003. Orangutan cultures and the evolution of material culture. Science 299 (5603), 102–105. http://dx.doi.org/10.1126/science.1078004.
Vickrey, W., 2020. Congestion in midtown manhattan in relation to marginal cost pricing. Economics of Transportation 21, 100152. http://dx.doi.org/10.1016/j.ecotra.2019.100152.
Waddle, D.M., 1994. Matrix correlation tests support a single origin for modern humans. Nature 368 (6470), 452–454. http://dx.doi.org/10.1038/368452a0.

Ward, J.H., 1963. Hierarchical grouping to optimize an objective function. J. Amer. Statist. Assoc. 58 (301), 236–244. http://dx.doi.org/10.1080/01621459.1963.10500845.

Yang, K., Menendez, M., Zheng, N., 2019. Heterogeneity aware urban traffic control in a connected vehicle environment: A joint framework for congestion pricing and perimeter control. Transp. Res. C 105, 439–455. http://dx.doi.org/10.1016/j.trc.2019.06.007.

Yang, K., Zheng, N., Menendez, M., 2018. Multi-scale perimeter control approach in a connected-vehicle environment. Transp. Res. C 94, 32–49. http://dx.doi.org/10.1016/j.trc.2017.08.014.

Yuan, Y., Raubal, M., 2012. Extracting dynamic urban mobility patterns from mobile phone data. In: Xiao, N., Kwan, M., Goodchild, M., Shekhar, S. (Eds.), Geographic Information Science. GIScience 2012. Springer, Berlin, Heidelberg, pp. 354–367. http://dx.doi.org/10.1007/978-3-642-33024-7_26.

Zheng, N., Geroliminis, N., 2016. Modeling and optimization of multimodal urban networks with limited parking and dynamic pricing. Transp. Res. B 83, 36–58. http://dx.doi.org/10.1016/j.trb.2015.10.008.

Zheng, N., Waraich, R.A., Axhausen, K.W., Geroliminis, N., 2012. A dynamic cordon pricing scheme combining the macroscopic fundamental diagram and an agent-based traffic model. Transp. Res. A 46 (8), 1291–1303. http://dx.doi.org/10.1016/j.tra.2012.05.006.