



# Hysteresis phenomena of a Macroscopic Fundamental Diagram in freeway networks

Nikolas Geroliminis<sup>a,\*</sup>, Jie Sun<sup>b</sup>

<sup>a</sup> School of Architecture, Civil and Environmental Engineering, Urban Transport Systems Laboratory, Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland

<sup>b</sup> Department of Civil Engineering, University of Minnesota, Twin Cities, USA

## ARTICLE INFO

### Keywords:

Freeways  
Macroscopic Fundamental Diagram  
Traffic hysteresis  
Congestion

## ABSTRACT

Observations of traffic pairs of flow vs. density or occupancy for individual locations in freeways or arterials are usually scattered about an underlying curve. Recent observations from empirical data in arterial networks showed that in some cases by aggregating the highly scattered plots of flow vs. density from individual loop detectors, the scatter almost disappears and well-defined macroscopic relations exist between space-mean network flow and network density. Despite these findings for the existence of well-defined relations with low scatter, these curves should not be universal. In this paper we investigate if well-defined macroscopic relations exist for freeway network systems, by analyzing real data from Minnesota's freeways. We show that freeway network systems not only have curves with high scatter, but they also exhibit hysteresis phenomena, where higher network flows are observed for the same average network density in the onset and lower in the offset of congestion. The mechanisms of traffic hysteresis phenomena at the network level are analyzed in this paper and they have dissimilarities to the causes of the hysteresis phenomena at the micro/meso level. The explanation of the phenomenon is dual. The first reason is that there are different spatial and temporal distributions of congestion for the same level of average density. Another reason is the synchronized occurrence of transitions from individual detectors during the offset of the peak period, with points remain beneath the equilibrium curve. Both the hysteresis phenomenon and its causes are consistently observed for different spatial aggregations of the network.

© 2011 Elsevier Ltd. All rights reserved.

## 1. Introduction

Analysis of traffic flow theory and modeling of vehicular congestion has mainly relied on fundamental laws, inspired from physics using analogies with fluid mechanics, many particles systems, etc. One main difference of physical systems and vehicular traffic is that humans make choices in terms of routes, destinations and driving behavior, which creates additional complexity to the system. While most of the traffic science theories make a clear distinction between free-flow and congested traffic states, empirical analysis of spatio-temporal congestion patterns has revealed additional complexity of traffic states and non-steady state conditions (see for example Muñoz and Daganzo, 2003; Helbing et al., 2009). Thus, the known fundamental diagram (initially observed for a stretch of highway and provide a steady-state relationship between speed, density and flow) is not sufficient to describe the additional complexity of traffic systems and it also contains significant experimental errors in the congested regime (see for example Kerner and Rehborn (1996) for a highway stretch or Geroliminis and Daganzo (2008) for a city street). The assumption of a unified curve suppresses all other traffic states and phase

\* Corresponding author. Tel.: +41 21 69 32481; fax: +41 21 69 35060.

E-mail address: [nikolas.geroliminis@epfl.ch](mailto:nikolas.geroliminis@epfl.ch) (N. Geroliminis).

transitions not belonging on this curve. Thus, as described in Zhang (2003) “phase curve obtained through statistical averaging suppresses finer traffic dynamics represented by the fine structures in the scatter”.

Nevertheless, it was recently observed from empirical data in Downtown Yokohama (Geroliminis and Daganzo, 2008) that by aggregating the highly scattered plots of flow vs. density from individual loop detectors, the scatter almost disappears and a well-defined Macroscopic Fundamental Diagram (MFD) exists between space-mean flow and density.

The first theoretical proposition of a macroscopic relationship between average network flow and density with an optimum accumulation belongs to Godfrey (1969). Earlier studies looked for macro-scale traffic patterns in data of lightly congested real-world networks (Ardekani and Herman, 1987; Olszewski et al., 1995) or in data from simulations with artificial routing rules and static demand (Mahmassani et al., 1987; Mahmassani and Peeta, 1993). However, the data from all these studies were too sparse or not investigated deeply enough to demonstrate the existence of an invariant macroscopic relation for real urban networks. Later, Daganzo (2007) conjectured (as part of an urban traffic dynamics model) that a well-defined relationship between flow and density must arise if a network is “uniformly loaded” in space and demand changes slowly with time.

The empirical verification of its existence with dynamic features is recent (Geroliminis and Daganzo, 2007, 2008). The experiments and simulations in these references suggest that under some conditions this macroscopic relation is indeed a reproducible curve named as the ‘Macroscopic Fundamental Diagram’ (MFD). These references also showed that the MFD is a property of the network infrastructure and control and not of the demand, i.e. space-mean flow is maximum for the same value of vehicle density independently of time-dependent origin–destination tables. If this collective behavior holds for more real networks, transportation managers can now have reliable tools to anticipate the results of “smart” traffic management strategies (based on average network densities and speeds), such as pricing or perimeter control, by observing whether their system is in a state that is producing the desired mobility levels.

Nevertheless, despite these recent findings for the existence of well-defined MFD’s with low scatter, these curves should not be a universal recipe. More real-world experiments are needed to identify the types of networks and demand conditions, for which invariant MFD’s with low scatter exist. In particular, networks with an uneven and inconsistent distribution of congestion may exhibit traffic states that are well below the upper bound of an MFD and much too scattered to line along an MFD. By analyzing real data from a medium-size French city Buisson and Ladier (2009) showed that heterogeneity has a strong impact on the shape/scatter of an MFD. Recent findings from empirical and simulated data for arterial networks (Geroliminis and Sun, 2011; Mazloumian et al., 2010) have identified the spatial distribution of vehicle density in the network as one of the key components that affect the scatter of an MFD and its shape. The second reference observed well defined relations between flow and density when link density variance is constant. In other words, the average network flow is consistently higher when link density variance is low, for the same network density. Recently, Cassidy et al. (2011), observed macro-relations in freeway networks if data come from periods where all lanes on all links of the studied network are either in the congested or the uncongested regime.

In this paper we investigate if well-defined MFD’s exist for freeway network systems, by analyzing real data from Minnesota’s freeways. We show that different freeway subnetworks do not have a well-defined MFD with low scatter. An interesting observation is that the aggregated patterns do not just exhibit some high degree of random scatter. We show that freeway network systems exhibit *consistent* hysteresis phenomena in an MFD plane, where higher network flows are observed for the same average network density in the onset and lower in the offset of congestion. We show that freeway systems are hysteretic, dynamic and path-dependent systems. In a dynamic system with no hysteresis and a well-defined MFD (as for example the arterial network of Yokohama), it is possible to predict the system’s outflow or flow at a specific time, given only its network density at that time. In a system with hysteresis, this is not possible without knowing the system’s current state, and there is no way to know the system’s state without looking at the history of the network density. This means that it is necessary to know the path that the network density followed before reaching its current value, i.e. freeway systems are hysterical systems with lasting memory properties.

Many physical systems experience hysteresis phenomena including electromagnetism, thermodynamics, biology and economics. Two observed traffic flow phenomena appear to challenge the notion of a smooth flow–density curve for *a specific location of a road segment*: One is the sudden flow drop often observed in empirical flow–density phase plots and the other is traffic hysteresis, a phenomenon characterized by different loops in the congested part of speed–density curves. The flow drop, known as “capacity drop” phenomenon can be addressed by the discontinuity between free-flow and congested regions and has been observed by many researchers at downstream of active bottlenecks (e.g. Banks, 1991; Cassidy and Bertini, 1999).

Also, others have unveiled flow or lane changing mechanisms that cause the phenomenon (e.g. Chung and Cassidy, 2004; Laval and Daganzo, 2006). Traffic hysteresis loops have firstly observed by Edie (1965) and Treiterer and Myers (1974) as separation between the acceleration and deceleration curves in speed–density diagrams. Assymmetric theories that explained the mechanisms of hysteresis have been developed by Newell (1965) and Zhang (1999), Yeo and Skabardonis (2009) and others. Nevertheless, traffic hysteresis loop may disappear when we aggregate lane-observed data in 1–5 min intervals (Daganzo, 2002). The same reference attributed the hysteresis phenomenon to lane changing and the non-conservative nature of flow in a single lane. The fact that the hysteresis loop disappears when we aggregate lanes supports this interpretation.

The mechanisms of traffic hysteresis phenomena at the network level are analyzed in this paper and they have dissimilarities to the causes of the hysteresis phenomena at the micro/meso level, as they persist for different types of spatial aggregation. To investigate the causes of the phenomenon, we carefully analyzed the spatial distribution of individual detectors’

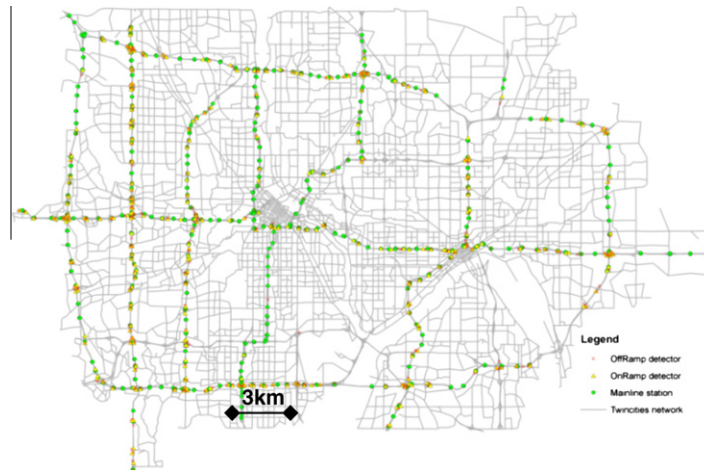


Fig. 1. The selected test site and the detectors locations.

occupancy measurements at different scales of the freeway network. The paper is organized as follows. Section 2 introduces the hysteresis for the whole network and explains the phenomenon by focusing on the spatial analysis of occupancy and flow data. Section 3 investigates the magnitude of the phenomenon for different structure and size subnetworks (compact congested subnetwork and incompact subnetwork with the 100 most congested detectors of the morning peak). This analysis unveils additional features of the phenomenon and gives us insights that the different spatial congestion distribution is not the unique cause for the hysteresis. The new sources are presented in Section 4, which identifies transition phenomena and non-equilibrium points of individual detectors as an additional cause of hysteresis. These transitions occur in many detectors in a synchronized manner and create hysteresis loops even when the spatial distributions are similar. We also analyze the relative magnitude of the two hysteresis sources.

The data utilized in this paper consist of count and occupancy measurements from mainline loop detectors on the freeways that are in the central part of the Twin Cities metropolitan area. The detectors are chosen from a network that consists of several major freeways constrained within I-494 and I-694 ring. Those include I-94, I-394, TH-169, TH-100, and part of I-35W and I-35E that are inside the ring. Detector locations are known through a GIS file. Loop detectors are located on the mainline of freeways approximately every 1/2 mile and on entrance and exit ramps. Loop detectors collect information about volume and occupancy and report the data back to the traffic management center every 30 s via fiber optic lines. An illustration of the study site and the detector locations is shown in Fig. 1.

## 2. Spatial analysis: whole network

### 2.1. Existence of hysteresis loops

We now show the existence of hysteresis phenomenon in a macroscopic diagram between average network flow and average occupancy for the whole study network.

Denote by  $i$  a freeway segment with a mainline detector approximately in the middle; and by  $q_i$  and  $o_i$  the flow and occupancy measured by the corresponding detector in a particular time slice. We use  $\mathbf{D}$  for the set of all mainline detectors in our study area (later we will analyze subsets of  $\mathbf{D}$ ). As is well known, the density at a detector location is  $k_i = \frac{o_i}{l}$ , where  $l$  is the space-mean effective vehicle length, which is about  $l \cong 6$  m. As the mainline detectors are closely placed approximately every 1/2 mile, we will continue our analysis with occupancy measures without converting them to density measures. One could do this by using some of the many existing first or higher order traffic flow models.

We are interested in spatially aggregated patterns of these variables at the network level. To this end, we define the average network flow  $Q = \sum_i \frac{q_i}{N}$  and average network occupancy  $O = \sum_i \frac{o_i}{N}$ , where  $N$  is the number of mainline detectors. (time  $t$  is omitted from all variables). To analyze the effect of spatial heterogeneity, we represent the global variance of the values of occupancy among all links in the network at a given time  $t$  by  $V$ .

Fig. 2a shows the time-series of average flows and occupancies for the day of May 22nd, 2007. Note that the values of  $Q$  vary for the same values of  $O$  during the day (for example for  $O = 12\%$ , the values of  $Q$  are different for the onset and offset of morning and evening peak. This indicates that it might not exist a well-defined Macroscopic Fundamental Diagram for the specific network. To get a clearer picture, we plot  $Q$  vs.  $O$  for three different weekdays, May 22nd, 23rd and 24th, 2007. The results are shown in Fig. 2b. We first observe curves with high scatter, especially for values of occupancy, which maximize flow. The value of maximum flow varies across days and within days for the morning and evening peak. An interesting observation is that this scatter is not random, but the  $Q$ - $O$  curves follow clockwise hysteresis loops. For some days the loops are

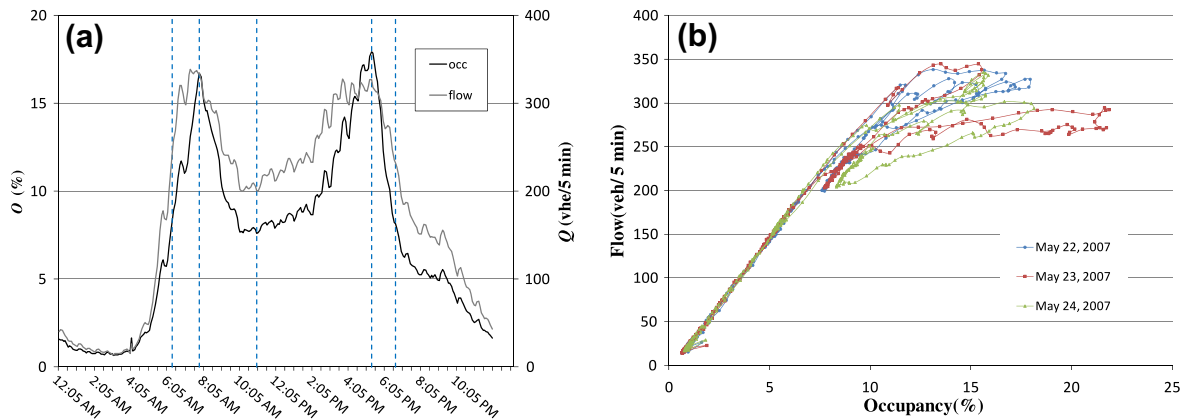


Fig. 2. (a) Time-series of network flow and occupancy for May 22nd, 2007; (b)  $Q$  vs.  $O$  across three different days (5 min data).

similar (for May 22nd the evening loop is internal to the morning loop), while for others (May 24th) the loops follow completely different paths.

To better understand the phenomenon we first look at the spatial distribution of individual detectors' occupancies. Recently, Geroliminis and Sun (2011) further explored the characteristics of arterial MFDs and by analyzing the Yokohama arterial data, identified the following regularity condition for the existence of a well-defined MFD: "If the spatial distribution of link occupancy is the same for two different time intervals with the same average network occupancy then these two time intervals should have the same average network flows, i.e. in an MFD plane these two are points close to each other". Using a similar approach, we investigate if the same regularity condition applies to freeway networks.

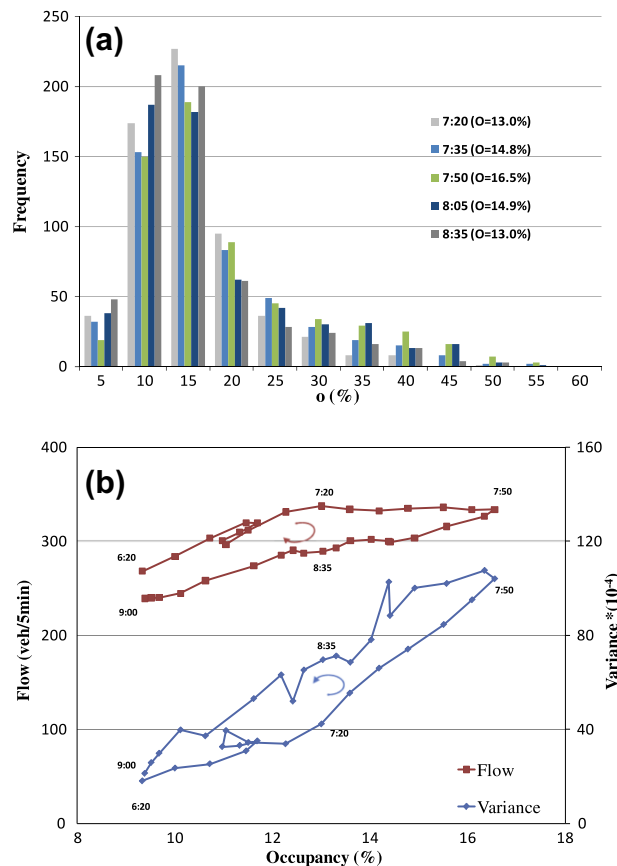
Fig. 3a compares histograms of individual detectors' occupancy for five different times in the onset and offset of morning with similar values of  $O$  (7:50 represents the most congested time of day with  $O = 16.5\%$ ). The other four histograms represent two time pairs with high magnitude of hysteresis loops, for  $O = 13\%$  and  $O = 14.8\%$ . In contrary with the Yokohama network, these distributions are different, especially when flow differences are higher (close to 20%). Note the large differences in the occupancy bins (i) for the most frequent values (occupancy 10–25%) and especially (ii) for values close to the critical occupancy of individual detectors (for the study network it ranges between 10% and 13% for most detectors). It looks that the number of detectors operating at capacity is smaller in the offset of morning congestion and the occupancy distribution spreads away from the mean, which increases its variance.

To verify the above conjecture, Fig. 3b plots in the same graph  $Q$  vs.  $O$  and  $V$  vs.  $O$ . It is clear that hysteresis loops are observed in both graphs, a clockwise loop for  $Q$ – $O$  and an anti-clockwise one for  $V$ – $O$ . Variance is always higher for the same value of  $O$  in the offset of congestion (7:50–9:00). Higher variance is associated with lower network flow. The reason is that flow of individual detectors  $q_i$  decreases in both sides, when moving away from the critical occupancy value that maximizes flow. Thus, when occupancy spreads away from the mean, network flow decreases. We analyzed additional days and we observed similar trends. The consistency of these results deserves a more detailed analysis of the data to provide a deeper explanation for the variance increase.

## 2.2. A deeper look in the spatial heterogeneity

From a mobility standpoint, we are interested in the relationship between network flow and occupancy variance. A better understanding of the phenomenon could provide benefits for freeway users as strategies that decrease occupancy variance, could increase network flow and trip endings. To identify the reason for the variance increase in the offset of congestion, we have a more detailed look at individual detectors' occupancy values. We create occupancy contour plots of the whole network, where at each detector location the occupancy value is shown as the diameter of a circle. The results for six different time instances between 6:00 am and 9:00 am are shown in Fig. 4. These consecutive snapshots show the propagation of congestion in the network.

Note that there are multiple pockets of congestion (active bottlenecks) in different parts of the network, which propagate upstream as time passes. By comparing Fig. 4b with 4f ( $O = 11.5\%$ ) and 4c with 4e ( $O = 13\%$ ), we note that the congestion pockets are in different parts of the network and have different sizes. These pockets are not necessarily interconnected, which is a common characteristic of propagation in arterials (see for example results from Geroliminis and Daganzo (2007, 2008) or Mazlounian et al. (2010)). The main reason is that the aforementioned arterial networks are mainly redundant and congestion propagates in a similar manner as drivers change their routes to identify less congested ones. Redundancy ensures that drivers have many route choices and that most links are on many desirable routes. But, most freeway networks are not redundant and congestion propagates unevenly. Despite the fact that interconnected arterials exist next to some freeways in the Twin Cities network, drivers do not deviate from their freeway route, because the arterial routes are even slower than the congested freeways.



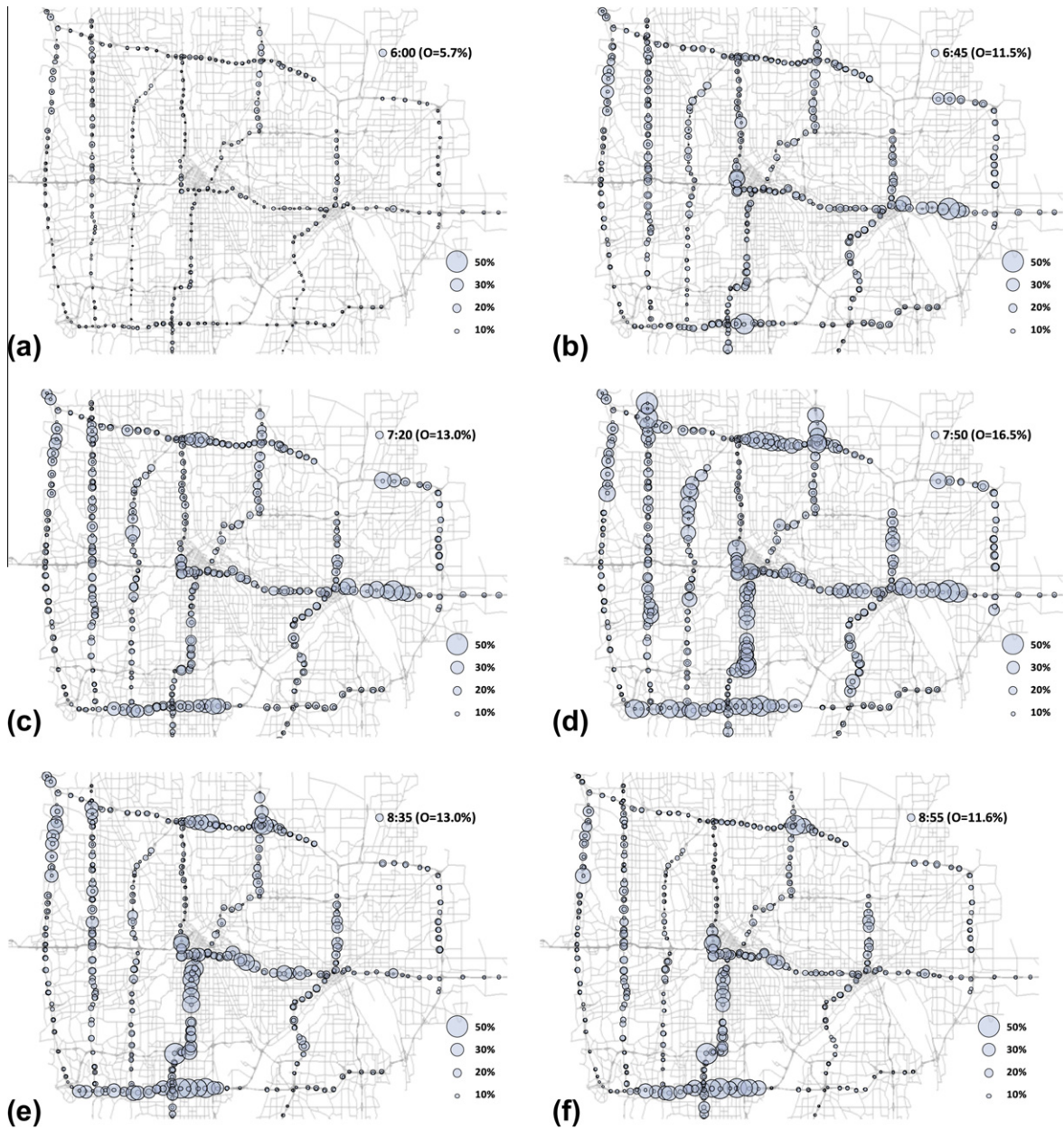
**Fig. 3.** Spatial analysis for the whole network: (a) Histograms of occupancy for five time periods; (b)  $Q$  vs.  $O$  (left axis) and  $V$  vs.  $O$  (right axis) for the morning peak.

By looking at these results our conjecture is that the spatial distribution of origins might be different from that of destinations. To verify this, we utilize the data from the on- and off-ramps. We estimate the input and the output flows of the network during the morning peak (6:00–9:00). Fig. 5 summarizes the spatial distribution of origins and destinations. The area of circles in the graph is analogous to the number of trips generated or terminated. The graph only shows the 25% of detectors, which generated/attracted 65% of the demand. It is clear that the spatial distribution of origins is completely different from that of the destinations. Most trips are generated in the perimeter of the city, while the majority of trips are attracted in the city center. Note also, that origins are more evenly distributed in the network, while destinations are concentrated in a few locations. This explains the fact (shown in the histograms of Fig. 3a) that in the onset of congestion, more detectors operate at the critical occupancy and close to capacity, while in the offset of congestion the distribution is more spread with higher variance.

To clarify this, consider a simple analogy of a queueing system of two servers in parallel with separate queues. In case demand is higher than capacity for both servers, the system will operate at its maximum service rate, as both servers operate at capacity. In a different case, where the same total demand is unevenly distributed between the two servers, the one will operate at capacity, but the other will not as demand is low. Something similar happens in the onset and offset of congestion as described above.

We further enhance in this direction by focusing on two time instances with a high degree of hysteresis, at 7:20 and 8:35, with  $O = 13\%$  and looking at individual detectors. For each detector we estimate the change in flow and occupancy,  $\Delta q_i$  and  $\Delta o_i$  between these times and 7:50, which is the peak time with the highest value of  $O = 16.5\%$ . We also estimate the distribution of  $\Delta o_i$  for 7:20 → 7:50, 7:50 → 8:35 and 7:20 → 8:35. The results are summarized in Fig. 6.

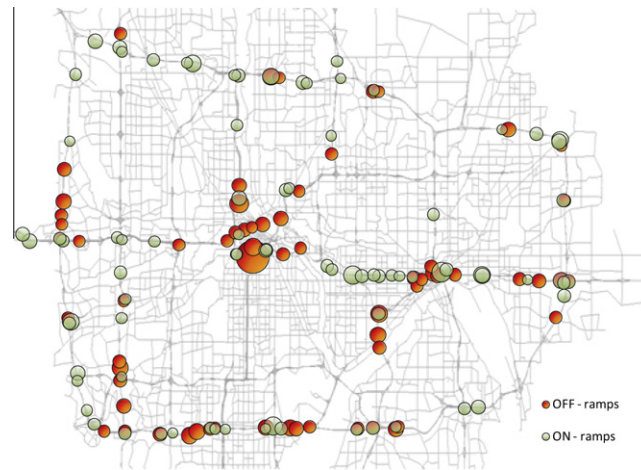
Fig. 6a plots  $\Delta q_i$  vs.  $\Delta o_i$  of individual detectors for 7:20 → 7:50 and 7:50 → 8:35. The average value of  $\Delta o_i$  is 3.5% for the first pair and  $-3.5\%$  for the second pair. Note that for the majority of detectors  $\Delta o_i$  is positive in the onset and negative in the offset of congestion. Note also, that there is a significant number of points lying in the 1st and 3rd quadrant at a slope equal to the free-flow speed. These are mainly points that remained in the uncongested regime during the whole peak period. Let us first look at the 7:20 → 7:50 transitions. Most points in the 4th quadrant (lower right) moved from the uncongested (U) to the congested (C) regime (U → C) or remained at the congested one with higher values of occupancy (C → C) and significant decrease in flow. Nevertheless, there is not a well-defined slope that these movements followed, as points are spread all over



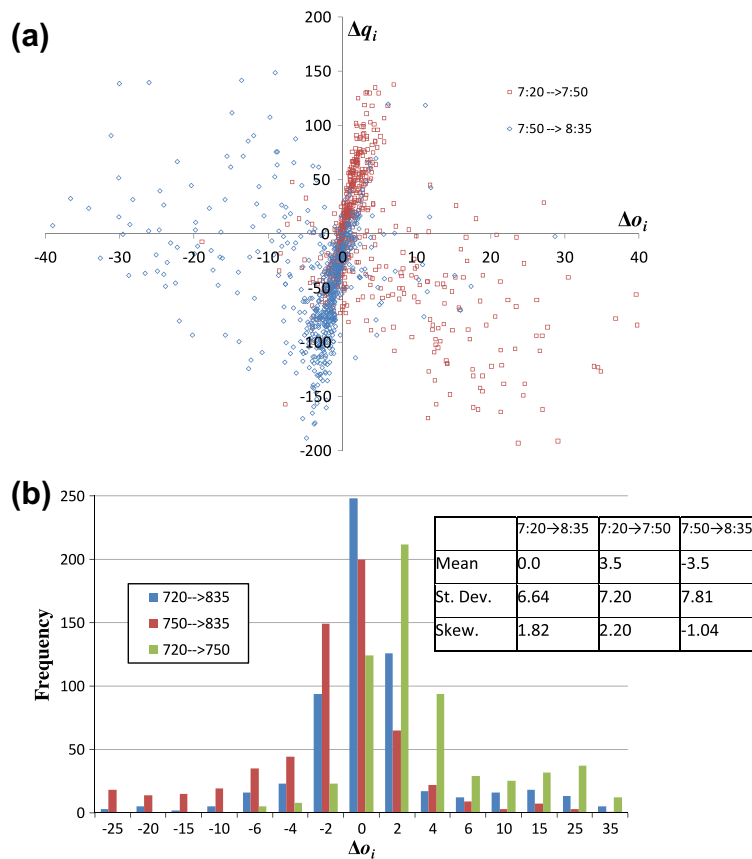
**Fig. 4.** Contour plot of individual detectors' occupancy for six different times of day 5.22.2007 (the occupancy value of each detector is analogous to the diameter of the circle in the specific location). The first graph shows a network snapshot before the starting of the morning peak at 6:20 am. Time 7:50 is when the average occupancy gets the maximum value, while the pairs (7:20, 8:35) and (6:45, 8:55) represent instances with the same value of  $O$  in the onset and offset of morning peak.

the quadrant. In the offset of congestion, the pattern is not exactly the same. There are still many points in the free-flow slope. But, the  $C \rightarrow U$  movements are not all associated with flow increase, as there is a large number of points in the left part of 3rd quadrant and less points in the 2nd quadrant (upper left). This observation provides us with a hint that a careful look in the flow measurements might give us additional insights. This analysis will be done in Section 4.

Another interesting observation, which confirms the above findings, is that the distribution  $\Delta o_i$  of for 7:20  $\rightarrow$  8:35 has a significant positive skewness. Note that the right tail (positive  $\Delta o_i$ ) is longer, while the mass of the distribution is concentrated on the left of the mean, which is zero. Thus, there are many points  $U \rightarrow U$  (belonging to the  $[-2, 0]$  bin) with small decrease in occupancy, but significant decrease in flow, as they are in the free-flow regime. On the other side, there are more points in the right tail of the distribution (moving to higher congestion levels) than in the left tail. The combination of these two characteristics creates a distribution in the offset of congestion with higher variance and less network flow. The table



**Fig. 5.** Input and output flows from the on- and off-ramps during the morning peak (the graph shows only the highest 25% of detectors, that carry more than 60% of the total demand).



**Fig. 6.** Spatial analysis of individual detectors: (a)  $\Delta q_i$  vs.  $\Delta o_i$  for 7:20 → 7:50 and 7:50 → 8:35 time pairs (units  $\Delta q_i$ : veh/5 min,  $\Delta o_i$ : %) ; (b) Histogram of  $\Delta o_i$  for 7:20 → 7:50, 7:50 → 8:35 and 7:20 → 8:35.

attached to Fig. 6b shows the values of the first three moments for the 7:20 → 7:50 (onset), 7:50 → 8:35 (offset) and 7:20 → 8:35 (total) distributions of  $\Delta o_i$ . While the variances are about the same, the onset distribution has two times higher skewness than the offset distribution (with opposite signs of course). In Section 3, we will investigate the existence and magnitude of hysteresis phenomenon for the freeway subnetwork consisted of routes leading in the center of the city.

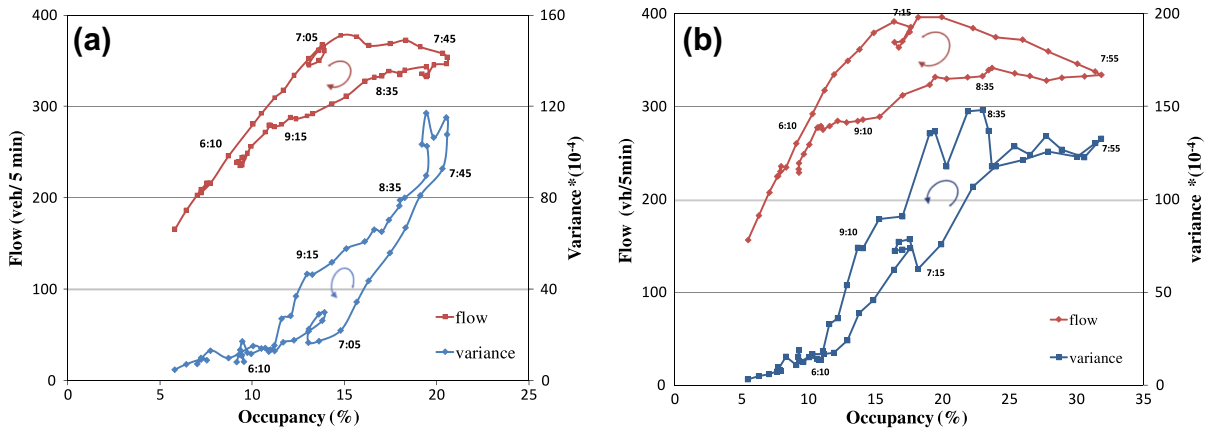


Fig. 7.  $Q$  vs.  $O$  and  $V$  vs.  $O$  (a) Compact congested network; (b) 100 most congested detectors.

### 3. Spatial analysis for different size subnetworks

The analysis of the previous section shows that the freeway system of Twin Cities Metropolitan Area is hysteretic and path-dependent and that it is necessary to know the path that the network density followed before it reached its current value, i.e. this freeway system is a hysterical system with lasting memory properties. The question that arises is if the hysteresis will remain when we look at different subnetworks by analyzing a subset of the detectors. To this end, we focus in two subnetworks:

- A compact congested subnetwork,  $D'$ , consisted of all the medium/high demand routes that lead in the center of the city. The choice of this subnetwork is in accordance with the findings of Fig. 5 (high off-ramps flows close to the center of the city during the morning peak). This part of the network includes approximately 160 detectors.
- An incompact subnetwork,  $D''$ , consisted of the 100 most congested detectors spread all over the network during the morning peak. The two subnetworks have about 50 detectors in common.

Following the same procedure as in Section 2.1, we estimate network averages for flow,  $Q$ , occupancy,  $O$ , and variance of occupancy,  $V$ , every 5 min in  $D'$  and  $D''$ . The results are summarized in Fig. 7. We observe very similar patterns both in the  $Q$ – $O$  and  $V$ – $O$  curves. Clockwise hysteresis loops are observed in  $Q$ – $O$  with higher flows in the onset of congestion and anti-clockwise loops are observed in  $V$ – $O$  with higher variance for lower flow. The magnitude of the hysteresis loops is slightly higher than  $D$  (~10%). Note also that the maximum value of  $O$  for the two subnetworks is higher, as they contain more of the congested detectors (16.5% for  $D$ , 20% for  $D'$  and 31% for  $D''$ ).

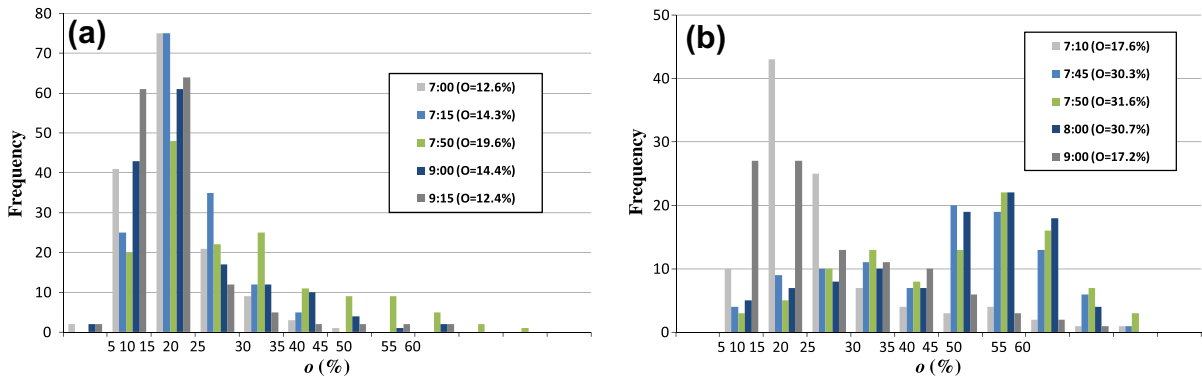
Another interesting observation is that the  $Q$ – $O$  curve for subnetwork  $D''$ , gets the maximum value of flow for  $O=17\%$  and then flow  $Q$  decreases for higher occupancy values in the onset of congestion. This decreasing part of the macroscopic curve was not evident for  $D$  and less evident in  $D'$  (there is no decreasing part in Fig. 3b and an insignificant one with low slope in 7a). Other studies that tried to obtain MFDs for arterial or freeway networks faced with a similar conclusion when the network was heterogeneously loaded. The explanation is simple. When a network consists of large uncongested parts and also has many pockets of congestion, the spatial aggregation of flow and occupancy will produce points below the maximum observed flow of individual detectors and the decreasing part of the individual fundamental diagrams will either weaken or disappear, as it happens for  $D$  and  $D'$ . Note also that the maximum value of network flow,  $Q$ , is 337 vh/5 min for  $D$ , 378 vh/5 min for  $D'$  and 397 vh/5 min for  $D''$ .

Let us also look at the distributions of individual detectors' occupancy for the two subnetworks. Fig. 8 plots these distributions for five time periods, at 7:50 which is the time with the highest occupancy value for both subnetworks and two pairs of same occupancy for the onset and offset of congestion. The compact subnetwork  $D'$  has similar distributions with the whole network  $D$ ; unimodal with the mode close to the mean (compare Fig. 8a with Fig. 3a). The number of detectors operating at capacity is smaller in the offset of morning congestion and the occupancy distribution spreads away from the mean, which increases its variance and decreases the network flow.

Let us now look at the spatial distributions of occupancy for  $D''$  (see Fig. 8b). When network occupancy is low (in the tails of congestion, 7:10 and 9:00) the distributions are unimodal and the number of detectors at capacity is more in the onset of congestion. But, the case is very different for higher values of  $O$ . The distributions are bimodal with both modes in the congested regime and only a few detectors at the uncongested or capacity regime. Note also that for values of  $O$  greater than 23% (between 7:30 am and 8:30 am) the distributions in the onset and offset of congestion have almost the same variance, but hysteresis still exists.<sup>1</sup> This finding cannot be explained with our current analysis, which links the spatial heterogeneity to the

<sup>1</sup> For distributions with the same  $O$  between 7:30 and 8:30, we applied the non-parametric Mann–Whitney  $U$  test and we could not reject the null hypothesis that the two data sets are drawn from the same distribution. This is not the case when  $V$  is different.





**Fig. 8.** Histograms. (a) Compact congested network; (b) 100 most congested detectors. (Histograms are plotted in a sequence of increasing time; earliest on the left, latest on the right).

hysteresis. To this end, we will carefully look at flow characteristics of individual detectors in Section 4 and investigate if there is an additional cause for hysteresis.

#### 4. Non-equilibrium states in individual detectors' measurements

The previous sections identified the spatial heterogeneity of congestion as the main cause of hysteresis phenomenon. Nevertheless, there are still some unexplained characteristics that cannot be attributed to this cause. For example, see (i) that hysteresis still exist when there is no difference in the variances in the onset and offset of congestion for some time instances for  $D'$  and (ii)  $\Delta q_i, \Delta o_i$  the movements from the congested to the uncongested regime in Fig. 6a, as explained in Section 2.2. To further enhance these observations, we first look at the fundamental diagram of individual detectors  $q_i$  vs.  $o_i$ . The results are summarized in Fig. 9.

Fig. 9 shows 5 min  $(q_i, o_i)$  pairs for a representative sample of the detector stations (all lanes together) during the morning peak 6:00–9:00. Detectors are placed approximately 250 ft upstream of the merge location. It is clear that in the offset of congestion the flow-occupancy trajectories follow a path of lower flow in the proximity of critical occupancy. The same pattern has been observed in many detectors that entered the congested regime (see Fig. 9a–d). Many researchers have observed pairs below the “equilibrium” curve during transitions in and out of congestion. Cassidy (1998) explained these pairs as non-stationary points that fall systematically under the well-defined stationary curve. Later, Muñoz and Daganzo (2003) attributed these transitions to non-equilibrium states that arise when the back of the queue passes over the detector.

We have also observed capacity drop phenomena for some active bottlenecks, as shown in Fig. 9e with a sharp change in capacity after the breakdown. The flow–density plot for the bottleneck (using flow as the total output flow at the bottleneck, and density at the upstream mainline detector) of Fig. 9e, shows a 17% capacity drop (from a peak at  $\sim 4200$  vh/h during the uncongested regime to  $\sim 3500$  vh/h during congestion). The aim of this paper is not to explain the cause of transitions or capacity drop at individual locations, but to investigate if these phenomena occur in a synchronized way, which results in a larger magnitude of hysteresis loops in the macroscopic  $Q$ – $O$  relation. Indeed, our analysis shows that this is the case.

We now look again at the change in flow and occupancy of individual detectors for the whole network,  $D$ , between times 7:20 and 8:35, which both have  $O = 13\%$ . Now, instead of the whole set of detectors, we focus on the subset, which during both times have occupancy value close to the critical occupancy (for most of the detectors this value is in the 9–13% range). The pattern is almost identical to the one of Fig. 6a, with points lying in a line with slope equal to free flow speed,  $v_f$ . Then, we extracted from the above data only detectors with  $-1\% < \Delta o_i < 1\%$  and we plotted  $\Delta q_i$  vs.  $\Delta o_i$ , as shown in Fig. 10. Would the same pattern persist when we zoom in this subset of detectors?

The answer is no. While many detectors underlie in the free-flow speed slope, there is a number of detectors beneath this line with a significant drop of flow,  $\Delta q_i < 0$ . These detectors represent locations that either (i) entered the congested regime and returned to the initial occupancy at a transient state, or (ii) a breakdown occurred and the initial high capacity flows cannot be reached. Thus, as the number of these locations is not negligible, we can conclude that there is a “cause-effect” relation between the low flow observations and the hysteresis loops. But, what is the attribute in the hysteresis phenomenon?

To enhance more in this direction, data were aggregated for the onset and offset of congestion and average flows were estimated from all detectors belonging in a specific occupancy range. In other words, we test if detectors with similar occupancy in the onset and offset of congestion, synchronously and consistently exhibit higher flows in the onset than in the offset of congestion. If  $q_{it}, o_{it}$  is the flow and occupancy value of detector  $i$  at interval  $t$ , then we estimate the average flow  $Q_x(o)$  for all time intervals  $t \in x = [onset, offset]$  and for all detectors  $i$  with  $o_{it} \in [o - 1.5\%, o + 1.5\%]$  Fig. 11 shows the results for all occupancy ranges for  $D, D'$ , and  $D''$  and confirms our previous statement. Significant differences (around 50 vh/5 min) are observed for occupancy between 10% and 20%, which is the value range of individual detectors occupancies for which tran-

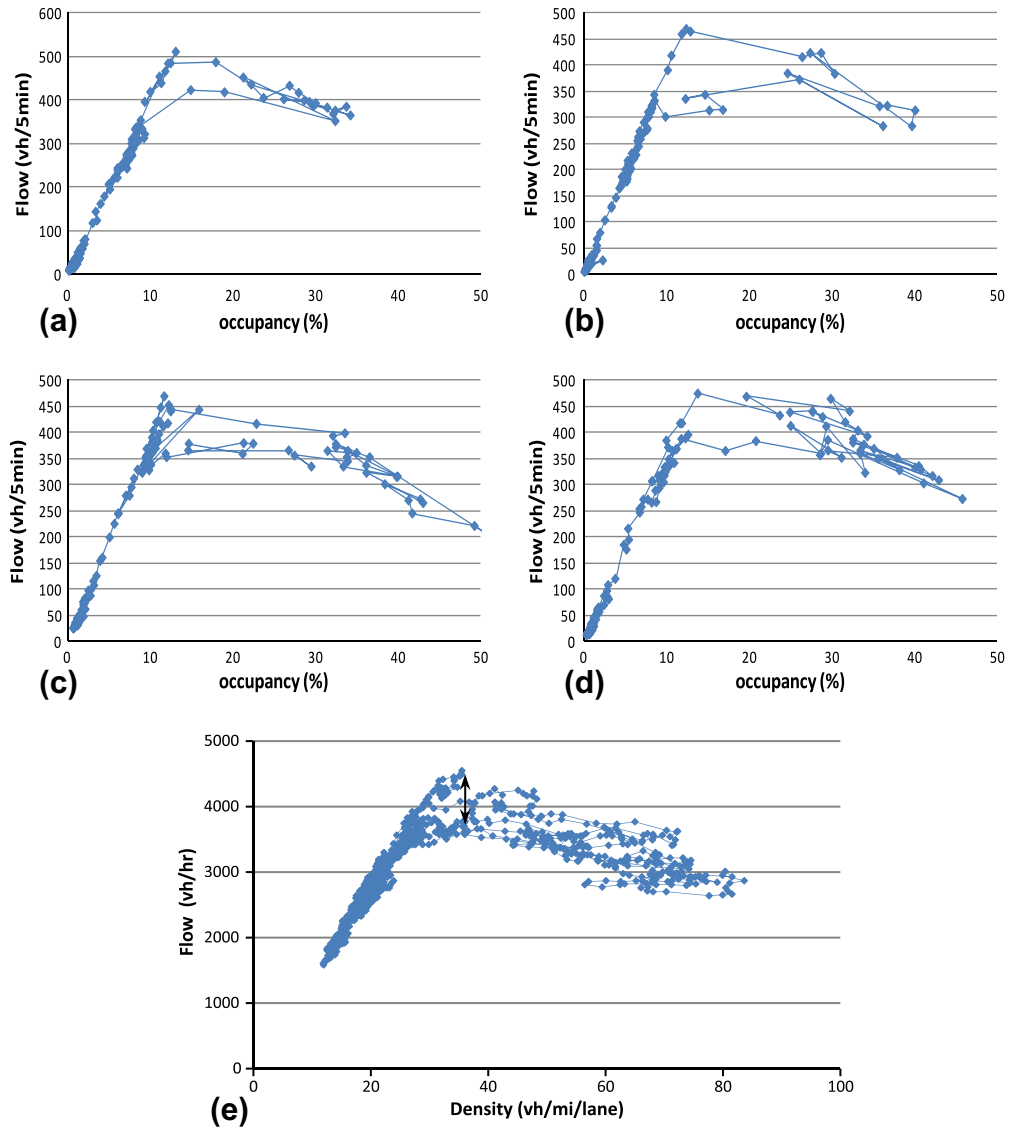


Fig. 9. Individual mainline detectors plots: (a–d) Flow –occupancy pairs for a sample of detectors with transitions; (e) capacity drop for the active bottleneck of Plymouth Ave., at H-169.

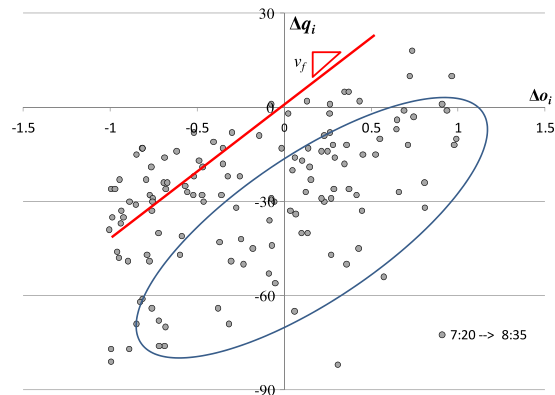


Fig. 10.  $\Delta q_i$  vs.  $\Delta o_i$  between 7:20 and 8:35 for points close to the critical occupancy and with  $-1\% < \Delta o_i < 1\%$ .

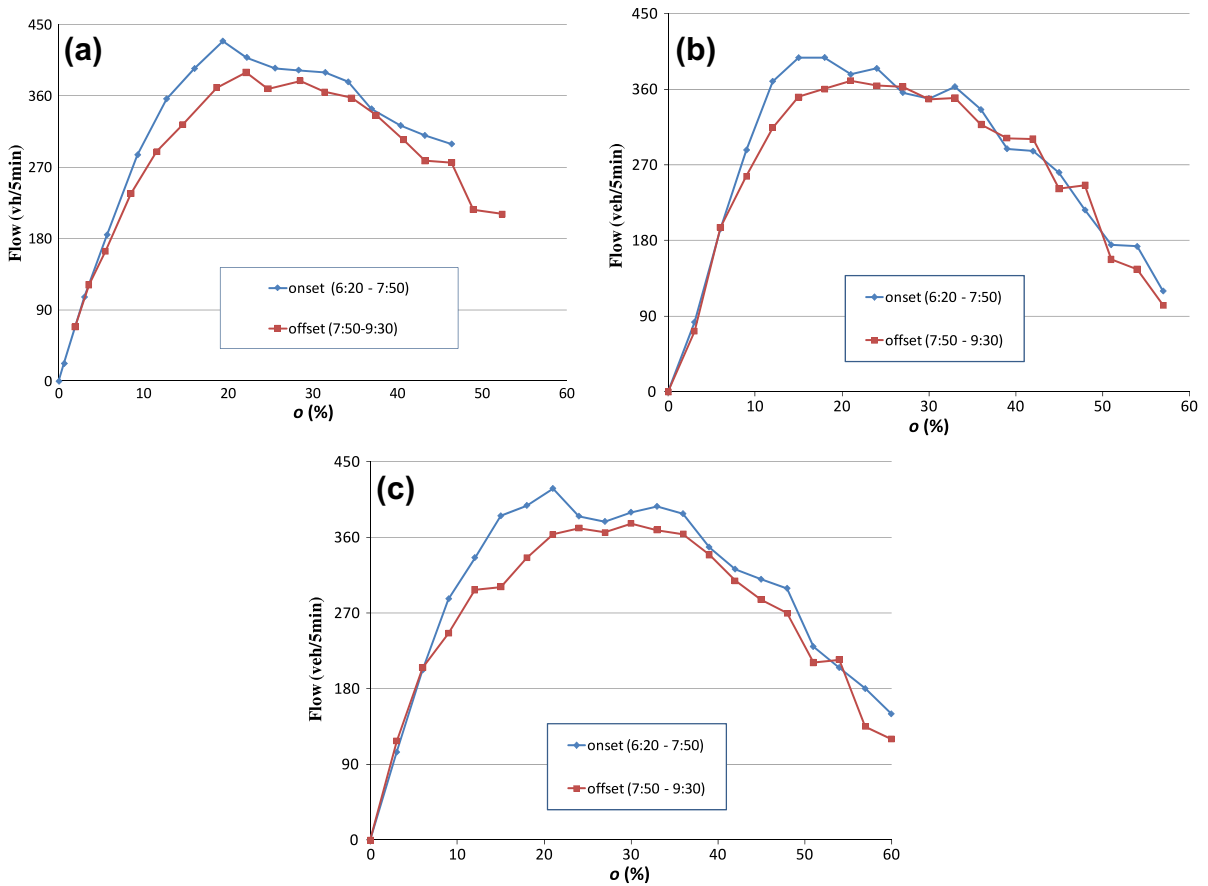


Fig. 11. Average flow of all detectors with occupancy  $\in [o - 1.5\%, o + 1.5\%]$  in the onset and offset of congestion (a) Whole network  $D$ ; (b) Compact subnetwork  $D'$ ; (c) 100 most congested detectors  $D''$ .

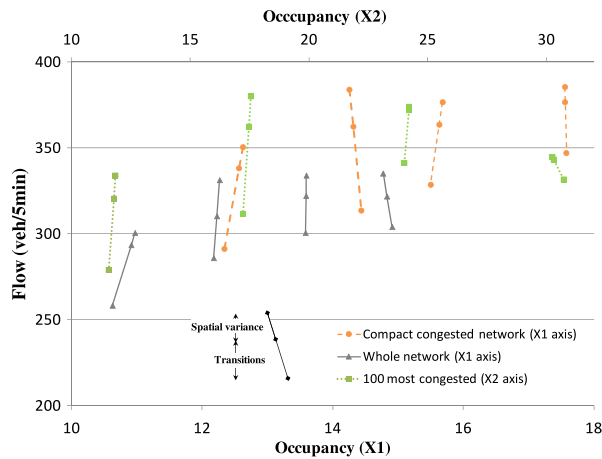


Fig. 12. Hysteresis magnitude for different times and subnetworks attributed to spatial heterogeneity and transitions.

sient states and capacity drops occur. Similar differences are observed if instead of the whole 3 h peak period, we compare data from two time instances, one in the onset and one in the offset of congestion.

We can now quantify the fractions of hysteresis caused by the spatial heterogeneity and the non-equilibrium/transient flows. Denote  $p(o)$  a discrete probability density function (derived from the histograms of Figs. 3a and 8), which shows what fraction of detectors has occupancy in the range  $[o - \epsilon, o + \epsilon]$  and  $q(o)$  the average flow of all detectors in the specified occu-

pancy range. Density function  $p(o)$  can be estimated for a whole period (e.g. the onset or offset of congestion) or for a specific 5 min interval,  $t$ . By definition, we have

$$O_t = \sum_o o \cdot p_t(o), \tag{1a}$$

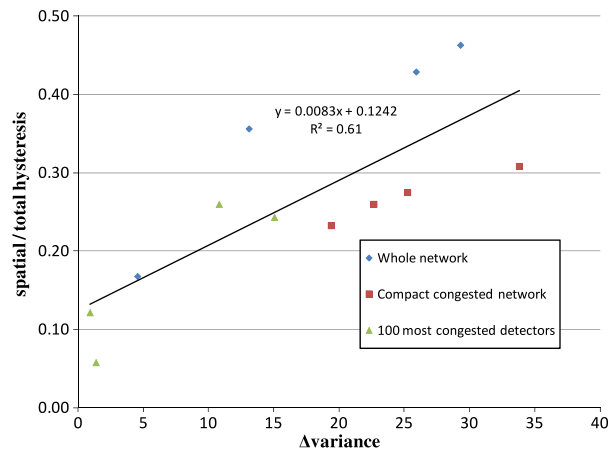
and

$$Q_t = \sum_o p_t(o) \cdot q_t(o). \tag{1b}$$

Out of the two causes of traffic hysteresis, the spatial heterogeneity is related to the first term of RHS of Eq. (1b), while the non-equilibrium/transient flows to the second term of the same equation. If we choose two time periods in the onset and offset of congestion  $t_{on}$ , and  $t_{off}$ , with  $O_{t_{on}} = O_{t_{off}}$ , then to estimate the fraction of hysteresis attributed to the spatial heterogeneity,  $f_s$ , we have to ignore the differences of the flows in the onset and offset of congestion, i.e.  $q_{t_{off}}(o) = q_{t_{on}}(o) \forall o$ .

**Table 1**  
Summary of analysis for attributes of hysteresis.

<i>Whole network</i>								
Time	7:00	9:00	7:15	8:50	7:25	8:25	7:35	8:05
O	10.97	10.63	12.27	12.18	13.58	13.60	14.77	14.91
Q	300.65	258.26	331.25	285.70	334.00	300.61	334.99	303.89
V	32.61	37.18	33.81	63.15	55.39	68.51	74.12	100.07
$\Delta V$	4.58		29.35		13.12		25.95	
H	42.39		45.55		33.38		31.11	
S	7.08		21.08		11.87		13.33	
$f_s$ (%)	16.7		46.3		35.6		42.8	
<i>Compact congested network</i>								
Time	7:00	9:15	7:15	9:00	7:25	8:55	7:35	8:25
O	12.62	12.34	14.26	14.44	15.68	15.50	17.56	17.58
Q	350.43	291.31	383.86	313.37	376.53	328.68	385.32	346.91
V	22.11	44.79	22.32	56.14	42.68	67.96	64.88	84.31
$\Delta V$	22.68		33.82		25.28		19.43	
H	59.13		70.49		47.85		38.41	
S	15.35		21.69		13.13		8.92	
$f_s$ (%)	26.0		30.8		27.4		23.2	
<i>100 Most congested detectors</i>								
Time	6:25	9:30	7:10	9:00	7:30	8:25	7:45	8:00
O	11.86	11.56	17.56	17.22	24.01	24.22	30.26	30.74
Q	333.85	278.94	379.91	311.51	373.75	341.17	344.84	331.70
V	17.89	32.96	79.29	90.12	118.81	120.21	118.75	119.69
$\Delta V$	15.07		10.83		1.40		0.94	
H	54.91		68.40		32.58		13.14	
S	13.34		17.76		1.88		1.59	
$f_s$ (%)	24.3		26.0		5.8		12.1	



**Fig. 13.** Fraction of hysteresis attributed to spatial heterogeneity vs. variance differences ( $f_s$  vs.  $\Delta V$ ).

$$f_s = \frac{\sum_o p_{t_{on}}(o) \cdot q_{t_{off}}(o) - Q_{t_{off}}}{Q_{t_{on}} - Q_{t_{off}}}. \quad (2)$$

Fig. 12 presents the results of the aforementioned analysis for different values of  $O$  for all three networks. The upper part of each interval is the hysteresis attributed to the spatial heterogeneity and the lower part the one attributed to the non-equilibrium flows. Note that the hysteresis because of spatial heterogeneity is larger for  $D$  and in some cases is negligible for  $D'$ . Table 1 shows in details the 12 pairs, with values for time,  $Q$ ,  $O$ ,  $V$ , the variance difference  $\Delta V$ , the hysteresis  $H$ , the attribute to spatial heterogeneity  $S$  and the fraction  $f_s = \frac{S}{H}$ .

By carefully looking at the values of Table 1, it is evident that the magnitude of hysteresis because of spatial heterogeneity is related to the variance differences of the analyzed time intervals. Fig. 13 provides additional evidence for this statement. It plots fraction  $f_s$  vs. the variance differences  $\Delta V$  for all the pairs of Table 1. There is a clear positive trend. When  $\Delta V$  is close to zero  $f_s$  is smaller than 10%, while when  $\Delta V$  is large  $f_s$  can be close to 50%. We also see that the trend is not the same for all subnetworks, which means that the topology and the distribution of congestion play a role in the magnitude and the attributes of the hysteresis.

## 5. Conclusions

The results in this paper show that freeway networks do not have well-defined MFDs between network flow and density, as these networks have topological or control characteristics that are different of arterial networks (non-redundant, no traffic signals, transient states etc.). Despite the fact that high scatter in the MFD for freeways has been conjectured by Daganzo and Geroliminis (2008), the idea that freeway networks are hysteretic and path-dependent at the aggregate level is fresh. We identified two causes for the phenomenon: One attribute is the different degree of spatial heterogeneity in vehicle density in the onset and offset of the peak period. A significant second attribute is the synchronized occurrence of transient periods and capacity drop phenomena in the offset of congestion, which created lower observed flows of individual detectors for values close to the critical occupancy. The magnitude of the hysteresis was analyzed for different subnetworks of the whole network and at different times.

While the results of this paper provide a clearer understanding of the non-existence of well-defined MFDs for freeway networks, additional research is needed in different types of networks to understand how variations in the topology/structure of the network can affect the shape, the scatter and the existence of an MFD. A proper parameterization and description of heterogeneity that would yield good estimates of macroscopic representation of the network is of critical importance. A future study should investigate if the findings for the Twin Cities Metropolitan freeway system persist for other freeway networks. One could utilize data from freeway ring roads with similar distribution of origins and destinations and investigate if hysteresis loops of the same type still occur. Despite the fact that capacity drop is not observed in individual arterial links because of the signals effect, arterial networks can also experience hysteresis loops if there is large heterogeneity in the density distribution (see for example Mazlounian et al., 2010 or Daganzo et al., 2011).

This paper provides additional empirical support that MFD's should not be universally expected. This means that while MFD can be a useful and simple tool to traffic engineers to improve cities' accessibility and mobility levels, a careful analysis is necessary before these control strategies/policies are introduced based on monitoring of aggregated variables.

## References

- Ardekani, S., Herman, R., 1987. Urban network-wide traffic variables and their relations. *Transportation Science* 21 (1), 1–16.
- Banks, J., 1991. Two-capacity phenomenon at freeway bottlenecks: a basis for ramp metering? *Transportation Research Record* 1320, 83–90.
- Buisson, C., Ladier, C., 2009. Exploring the impact of homogeneity of traffic measurements on the existence of macroscopic fundamental diagrams. *Transportation Research Record* 2124, 127–136.
- Cassidy, M.J., 1998. Bivariate relations in nearly stationary highway traffic. *Transportation Research Part B – Methodological* 32 (1), 49–59.
- Cassidy, M.J., Bertini, R.L., 1999. Some traffic features at freeway bottlenecks. *Transportation Research Part B – Methodological* 33 (1), 25–42.
- Cassidy, M., Daganzo, C.F., Jang, K., 2011. Macroscopic fundamental diagrams for freeway networks: theory and observation. In: *The 90th Transportation Research Board Annual Meeting*, Paper No. 11-0657, Washington, DC.
- Chung, K., Cassidy, M.J., 2004. Test of theory of driver behavior on homogeneous freeways. *Transportation Research Record* 1883, 14–20.
- Daganzo, C.F., 2002. A behavioral theory of multi-lane traffic flow. Part I: Long homogeneous freeway sections. *Transportation Research Part B – Methodological* 36 (2), 131–158.
- Daganzo, C.F., 2007. Urban gridlock: macroscopic modeling and mitigation approaches. *Transportation Research Part B – Methodological* 41, 49–62.
- Daganzo, C.F., Geroliminis, N., 2008. An analytical approximation for the macroscopic fundamental diagram of urban traffic. *Transportation Research Part B: Methodological* 42 (9), 771–781.
- Daganzo, C.F., Gayah, V.V., Gonzales, E.J., 2011. Macroscopic relations of urban traffic variables: Bifurcations, multivaluedness and instability. *Transportation Research Part B: Methodological* 45 (1), 278–288.
- Eadie, L.C., 1965. Discussion of traffic stream measurements and definitions. In: *Proceedings of the Second International Symposium on the Theory of Road Traffic Flow*, London, pp. 139–154.
- Geroliminis, N., Daganzo, C.F., 2007. Macroscopic modeling of traffic in cities. In: *Transportation Research Board 86th Annual Meeting*, Washington DC, Paper No. 07-0413.
- Geroliminis, N., Daganzo, C.F., 2008. Existence of urban-scale macroscopic fundamental diagrams: some experimental findings. *Transportation Research Part B: Methodological* 42 (9), 759–770.
- Geroliminis, N., Sun, J., 2011. Properties of a well-defined macroscopic fundamental diagram for urban traffic. *Transportation Research Part B: Methodological* 45, 605–617. doi:10.1016/j.trb.2010.11.004.
- Godfrey, J., 1969. The mechanism of a road network. *Traffic Engineering and Control* 11 (7), 323–327.

- Helbing, D., Treiber, M., Kesting, A., Schönhof, M., 2009. Theoretical vs. empirical classification and prediction of congested traffic states. *European Physical Journal B* 69 (4), 583–598.
- Kerner, B.S., Rehborn, H., 1996. Experimental features and characteristics of traffic jams. *Physical Review E* 53 (2), R1297–R1300.
- Laval, J.A., Daganzo, C.F., 2006. Lane-changing in traffic streams. *Transportation Research Part B: Methodological* 40 (3), 251–264.
- Mahmassani, H.S., Peeta, S., 1993. Network performance under system optimal and user equilibrium dynamic assignments: implications for advanced traveler information systems. *Transportation Research Record* 1408, 83–93.
- Mahmassani, H.S., Williams, J., Herman, R., 1987. Performance of urban traffic networks. In: Gartner, N.H., Wilson, N.H.M. (Eds.), *Transportation and Traffic Theory (Proceedings of the 10th International on Transportation and Traffic Theory Symposium, Cambridge, Massachusetts)*. Elsevier.
- Mazlounian, A., Geroliminis, N., Helbing, D., 2010. The spatial variability of vehicle densities as determinant of urban network capacity. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 368, 4627–4648.
- Muñoz, J.C., Daganzo, C.F., 2003. Structure of the transition zone behind freeway queues. *Transportation Science* 37 (3), 312–329.
- Newell, G.F., 1965. Instability in dense highway traffic: a review. In: *Proceedings of the 2nd International Symposium on the Theory of Road Traffic Flow*, London, UK, pp. 73–83.
- Olszewski, P., Fan, H.S.L., Tan, Y.W., 1995. Area-wide traffic speed-flow model for the Singapore CBD. *Transportation Research Part A: Policy and Practice* 29 (4), 273–281.
- Treiterer, J., Myers, J.A., 1974. The hysteresis phenomenon in traffic flow. In: *Proceedings of the 6th International Symposium on Transportation and Traffic Theory*, pp. 13–38.
- Yeo, H., Skabardonis, A., 2009. Understanding stop-and-go traffic in view of asymmetric traffic theory. In: *Proceedings of the 18th ISTTT, Hong Kong*, pp. 99–116.
- Zhang, H., 1999. A mathematical theory of traffic hysteresis. *Transportation Research Part B: Methodological* 33, 1–23.
- Zhang, H., 2003. On the consistency of a class of traffic flow models. *Transportation Research Part B: Methodological* 37, 101–105.