

Properties of a well-defined macroscopic fundamental diagram for urban traffic

Nikolas Geroliminis^{a,*}, Jie Sun^b

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

^b Department of Civil Engineering, University of Minnesota, Twin Cities, USA

ARTICLE INFO

Article history:

Received 3 January 2010

Received in revised form 6 November 2010

Accepted 6 November 2010

Keywords:

Congestion

Urban traffic

Macroscopic modeling

Arterials

Hysteresis

ABSTRACT

A field experiment in Yokohama (Japan) revealed that a macroscopic fundamental diagram (MFD) linking space-mean flow, density and speed exists on a large urban area. It was observed that when the highly scattered plots of flow vs. density from individual fixed detectors were aggregated the scatter nearly disappeared and points grouped along a well defined curve. Despite these and other recent findings for the existence of well-defined MFDs for urban areas, these MFDs should not be universally expected. In this paper we investigate what are the properties that a network should satisfy, so that an MFD with low scatter exists. We show that the spatial distribution of vehicle density in the network is one of the key components that affect the scatter of an MFD and its shape. We also propose an analytical derivation of the spatial distribution of congestion that considers correlation between adjacent links. We investigate the scatter of an MFD in terms of errors in the probability density function of spatial link occupancy and errors of individual links' fundamental diagram (FD). Later, using real data from detectors for an urban arterial and a freeway network we validate the proposed derivations and we show that an MFD is not well defined in freeway networks as hysteresis effects are present. The datasets in this paper consist of flow and occupancy measures from 500 fixed sensors in the Yokohama downtown area in Japan and 600 loop detectors in the Twin Cities Metropolitan Area Freeway network in Minnesota, USA.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

This paper is to search for the properties of the macroscopic modeling of traffic flow under congested conditions. The understanding of these properties should be the first step to macroscopic modeling-based alternative traffic management schemes, such as pricing and perimeter control to mitigate congestion and improve mobility in urban areas.

It has been recently proposed and tested in Daganzo (2007) and Geroliminis and Daganzo (2007, 2008), that traffic in large urban regions can be modeled dynamically at an aggregate level, if the neighborhoods are uniformly congested. The last two papers showed, using a micro-simulation of the San Francisco Business district and a field experiment in downtown Yokohama (Japan), (i) that urban neighborhoods approximately exhibit a “macroscopic fundamental diagram” (MFD) relating the number of vehicles (accumulation) in the neighborhood to the neighborhood's average speed (or flow) and (ii) there is a robust linear relation between the neighborhood's average flow and its total outflow (rate vehicles reach their destinations). The experiment used a combination of fixed detectors and floating vehicle probes as sensors. It was also observed that when the somewhat chaotic scatter-plots of speed vs. density from individual fixed detectors were aggregated for a

* Corresponding author. Tel.: +41 21 69 32481.

E-mail addresses: nikolas.geroliminis@epfl.ch (N. Geroliminis), sunxx308@umn.edu (J. Sun).

downtown region around 10 km², the scatter nearly disappeared and points grouped neatly along a smoothly declining curve. The same references also showed that (a) the MFD is a property of the network itself (infrastructure and control) and not of the demand, i.e. the MFD should have a well-defined maximum and remain invariant when the demand changes both with the time-of-day and across days; (b) the space-mean flow, is maximum for the same value of density of vehicles or average speed, independent of the origin–destination tables; (c) the average trip length for the study region is about constant with time, i.e. the total outflow vs. density curve is a scaled up version of the average network flow vs. density curve; and (d) the MFD can be estimated accurately using existing monitoring technologies (e.g. detector data, GPS, etc.).

The idea of an MFD with an optimum accumulation is quite old and belongs to [Godfrey \(1969\)](#), but the verification of its existence with dynamic features is recent ([Geroliminis and Daganzo, 2007, 2008](#)). Earlier works had looked for MFD patterns in data from lightly congested real-world networks ([Godfrey, 1969](#); [Ardekani and Herman, 1987](#); [Olszewski et al., 1995](#)) or in data from simulations with artificial routing rules and static demands ([Mahmassani et al., 1987](#); [Williams et al., 1987](#); [Mahmassani and Peeta, 1993](#)) but the data from all these studies were too sparse or not investigated deeply enough to demonstrate that an invariant MFD could dynamically arise in the real world. Although [Geroliminis and Daganzo \(2007, 2008\)](#) have shown that this can indeed happen, more real-world experiments should shed light on the types of networks and demand conditions for which invariant MFD's arise. Invariance is important because knowledge of a neighborhood's MFD allows decision-makers to use demand-side policies to improve mobility. However, to evaluate changes to the network (e.g., re-timing the traffic signals or allocating the percentage of streets devoted to public transit) one needs to know how the MFD is affected by these changes. To this end, [Daganzo and Geroliminis \(2008\)](#) explored the connection between network structure and a network's MFD for urban neighborhoods with cars controlled by traffic signals. [Helbing \(2009\)](#) has also derived analytical theories for an MFD, using a utilization-based approach.

Despite these recent findings for the existence of well-defined MFDs for urban areas, these MFDs should not be universally expected. The framework of an MFD was proposed under the assumption that congestion is evenly distributed, i.e. the areas with an MFD should be (roughly) homogeneously loaded ([Daganzo, 2007](#); [Geroliminis and Daganzo, 2007, 2008](#)). A well-defined MFD was observed in the last two references, but the homogeneity assumption was not deeply explored. In this paper we provide a clearer understanding of the spatial distribution of vehicles in a network and how these affect the scatter and the existence of an MFD. This paper demonstrates with real data that a strict homogeneity of traffic states on the network, as initially conjectured by [Daganzo \(2007\)](#), is not necessary to observe a well-defined MFD.

To this end we define as Aggregated Traffic Relationship (ATR), a relationship between average traffic variables (like speed, density, occupancy or flow) for a specific network that not necessarily have a well defined shape to be called an MFD. In particular, networks with an uneven and inconsistent distribution of congestion may exhibit traffic states that are well below the upper bound of an ATR and much too scattered to line along an MFD. An inconsistent distribution of congestion is typical of freeway networks with multiple recurrent and non-recurrent bottlenecks, and of large urban areas with multiple congested sub-centers. Even networks that satisfy some of the regularity conditions pointed above may exhibit significant scatter on their MFDs because of a rapidly changing demand, e.g. during an emergency evacuation an MFD may not exist for an urban network which exhibits an MFD under normal conditions.

The effect of heterogeneity in ATRs or MFDs has been experimentally examined to some aspects with real data from a medium-size French city by [Buisson and Ladier \(2009\)](#). The authors examined differences between the arterial and freeway network, impact of the distance between the loop detector and the traffic signal within the urban network, differences between normal and non-recurring conditions (during a strike of truck drivers). They showed that heterogeneity has a strong impact on the shape/scatter of an MFD, that may turn not to be an MFD in some cases, e.g. for a freeway network. In parallel with our work, [Daganzo et al. \(2011\)](#) investigated bifurcations and instability issues in an MFD for a two-ring network.

In this paper we investigate what are the properties that a network should satisfy, so that an MFD with low scatter exists. The remainder of this paper is organized as follows: Firstly, we show that the spatial distribution of car density in the network is one of the key components that affect the scatter of an MFD and its shape. We also propose an analytical derivation of the spatial distribution that considers correlation between adjacent links. We investigate the scatter of an MFD in terms of errors in probability density function of spatial occupancy and errors of individual links' fundamental diagram (FD). Later,

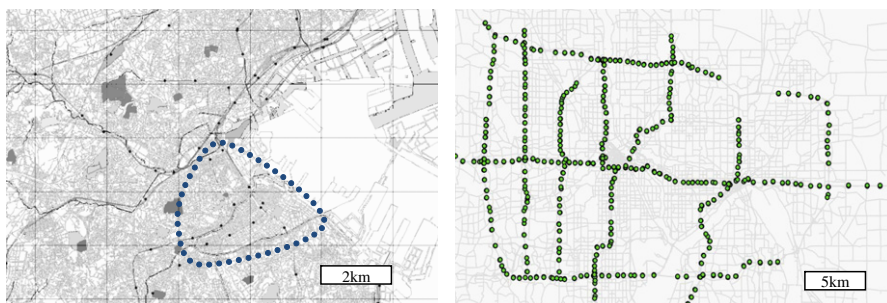


Fig. 1. Test sites: (left) Yokohama downtown arterial network; (right) Minneapolis, Twin Cities freeway network.

using real data from detectors for an urban arterial and a freeway network we validate the proposed derivations and we show that an MFD is not well defined in freeway networks as hysteresis effects are present.

The datasets in this paper consist of flow and occupancy measures from 500 fixed sensors in the Yokohama downtown area and 600 loop detectors in the Twin Cities Metropolitan Area Freeway network (Minnesota, USA). Fig. 1 illustrates the test sites used in the paper.

2. Properties of a well-defined MFD

A macroscopic fundamental diagram (MFD) linking space-mean flow, density and speed has been observed in Yokohama, Japan through a field experiment (Geroliminis and Daganzo, 2008). The relation between space-mean flows on the whole network and the trip completion rates was also explored there, but not how the spatial variability of link vehicle density affects the shape of the MFD or even its existence. In this section, we first explore the spatial distribution of individual detector occupancy for different times of day, which is a metric of the spatial distribution of vehicles in the network.

2.1. Occupancy distributions

The framework of an MFD was proposed under the assumption that congestion is evenly distributed, i.e. the homogeneity assumption of the MFD (Daganzo, 2007; Geroliminis and Daganzo, 2007, 2008). A well-defined MFD was observed in the last two references, but the homogeneity assumption was not deeply explored. In this section, we try to shed some light to this direction and develop a metric to test the range of the spatial heterogeneity of congestion for a well-defined MFD.

The data of a typical workday (12/14/2001) is chosen, which is consistent with the previous test in Yokohama. Our raw data is in form of volume and occupancy, which come directly from loop detectors and some obvious errors/malfunctions (e.g. data with occupancy over 100 or stay 0 for all times) are eliminated before analyzing the data. One can see from the time series plot of average network occupancy, O , for the study site of Fig. 1a, that the network is almost empty at the beginning of the day with average occupancy close to zero. There are two dense peak times during the day: morning peak and evening peak, when the average occupancy would reach as high as almost 50%.

Let us index the road lane section between two intersections as d , and let q^d and o^d donate the flow and occupancy measured by the detector during a time slice on this specific road lane section. It is known that the density k in the proximity of a detector can be given by $\frac{o^d}{s}$, where s is the effective vehicle length, which can be considered as a parameter of a specific network. The weighted average (using link lengths) of occupancy could be space-mean for the network if the detectors happen to be at representative locations within each link. This should happen automatically for flows on time slices large compared with a signal cycle because on this time scale link flows are roughly the same regardless of where they are measured within a link. But obviously, the same is not necessarily true for density. We continue our analysis using occupancy measures, as they are directly available.¹

Let O be the average network occupancy in the proximity of all detectors during time interval t (comparable in size with one signal cycle), and Q be the average flow of all detectors during the same time interval (interval t is omitted from the notation for simplicity purposes). We have to note that Q and O in our analysis are unweighted averages, which may contradict the generalized definitions of network flow and density according to (Edie, 1963), where weighted averages using the length of individual links are estimated. Actually, Geroliminis and Daganzo (2008) showed for the Yokohama test site, that both the weighted and unweighted averages produce a well-defined MFD with similar scatter. Thus, we focus our analysis in a simpler definition.

We now test the following conjecture: 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 flows, i.e. in an MFD plane these two are points close to each other. Mathematically speaking, if we estimate the probability density function $d_r(t)$, of all individual detectors' at a specific time t , in region r , and the average network flow and occupancy, $Q(t)$ and $O(t)$, a well-defined MFD exists in r when for all t_1, t_2 we have:

$$\{Q(t_1) = Q(t_2) \text{ and } O(t_1) = O(t_2)\} \iff d_r(t_1) \sim d_r(t_2).$$

To test the conjecture for a wide range of traffic conditions from uncongested to heavily congested, we choose an average occupancy of 15%, 25%, 35% and 45%. We look at all times during a day that O is close to these values, e.g. for $O = 35\%$ the time points that the dashed line intersects with the occupancy time series in Fig. 2. For each average occupancy level, we collect five time intervals whose differences in O are less than 1% and we plot the histograms for each average occupancy group (shown in Fig. 3). We discretize the full occupancy range from 0% to 100% in 23 groups; one group for empty links with zero occupancy and 22 groups of equal range. The value 22 is the maximum number of vehicles for a link $N = k_{jam} \cdot l$, where the average link length and the jam density for the study site are estimated to be $l = 154$ m and $k_{jam} = 0.14$ veh/m (Daganzo and Geroliminis, 2008).

¹ Using mid-block detector data Geroliminis and Daganzo (2008) found that the weighted and unweighted values of average network occupancy are similar for the city of Yokohama and that $\frac{Q}{s}$ provides a decent approximation of the average network density (errors less than 15%).

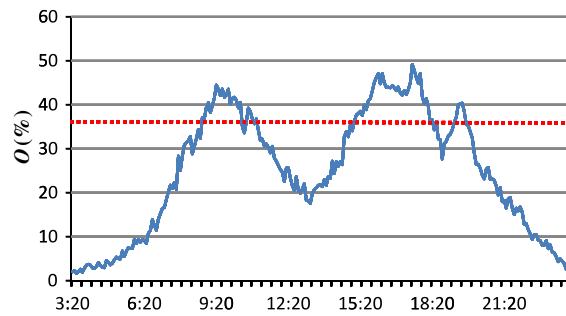


Fig. 2. Time series of average network occupancy for the day of December 14, 2001 in downtown Yokohama.

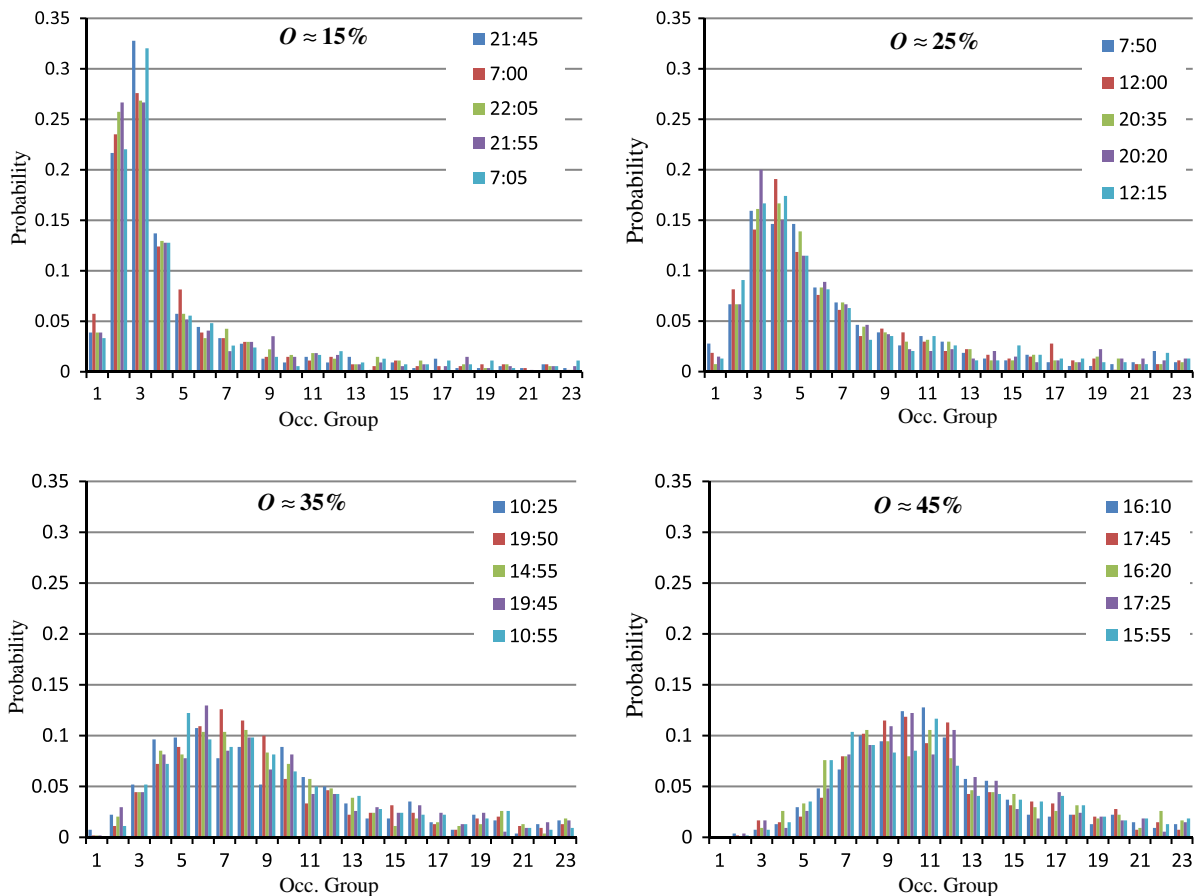


Fig. 3. Histograms for different values of average occupancy in the arterial network, O (Occupancy group #1 represents empty links ($O = 0$), groups #2–#23 have a range of occupancy equal to $100/22 = 4.55\%$).

It is very clear from these graphs that congestion is not evenly distributed. Some parts of the network are more congested than others. As more parts of the network become congested, the distribution of detector occupancy is shifted to the right. Note that the distributions of occupancy with similar average occupancy are similar although they come from different time intervals.

2.2. Statistical analysis

Some statistical tools are needed for our investigation into the property of the individual detector occupancy distributions. The two major statistical tests we are using in this paper are the Chi-square test and the Mann–Whitney U test.

The kind of Chi-square test we are using in this paper is called Pearson's chi-square test. It is used to test whether the sample distribution of a categorical variable is different from a population with a hypothesized distribution or not and

Table 1

Results from statistical tests for occupancy distributions: (a) Mann–Whitney test and (b) Chi-square test.

Test	Time	O	p-Value	Conclusion
1	17:45 17:25	44.93 44.98	0.967	Not significantly different at $\alpha = 0.05$
2	19:50 10:55	35.21 35.61	0.967	
3	12:00 20:20	25.28 25.52	0.727	Not significantly different at $\alpha = 0.05$
4	7:00 21:55	15.20 15.86	0.827	

Test	Degrees of freedom	Chi-square statistics	Critical value ($Pr \leq 0.05$)	p-Value	Conclusion
1	88	84.32	110.90	0.591	Not significantly different at $\alpha = 0.05$
2	88	80.69	110.90	0.697	
3	88	96.77	110.90	0.245	
4	88	99.13	110.90	0.196	

the frequency distribution of the events observed in a sample is approximately chi-square. The chi-square goodness-of-fit test is applied to binned data (i.e., data put into classes) and for each bin a metric for the difference between the frequencies of the two distributions is estimated. The sum across all bins is used to evaluate the similarity of the distributions.

Mann–Whitney U test is a non-parametric test used to see whether two samples are coming from the same distributions or not. The null hypothesis is that the two samples are from a single population and have equal probability distributions. This test is the same as performing a two-sample *t*-test after ranking over the samples. Samples are first ranked and a statistic U is calculated based the rank of observations from samples in the series. For large samples, U approximately follows normal distribution, thus, Z score can be calculated to obtain *p*-value from normal distribution.

The results are summarized in Table 1. In Mann–Whitney test, two samples can be compared at one time. The null hypothesis of the test is that the two samples are drawn from a single population. For each average occupancy level, we choose two different time intervals whose difference in average occupancy is mild (neither the biggest nor the smallest in the group). All the tests give large *p*-values, which indicates that samples within one group are not significantly differently.

As there are five points in every group, a Mann–Whitney U test for each pair in the group is a tedious task. Instead, Chi-square test can take several populations into account simultaneously. In our settings, the null hypothesis of the test is that the distribution of the occupancy is the same at all times with the same average occupancy. The possible outcomes of a link have vehicles range from 0 to 22, i.e. there are 23 different outcomes. Thus, we can divide the total range of occupancy into 23 categories. Instead of probabilities, frequency of each group is used, as Chi-square test is designed for categorical variables that should take integer values.

From Table 1b we draw the same conclusion as we did with the Mann–Whitney U test. We showed that evenly-distributed congestion is not a necessary condition for a well-defined MFD and relaxed the homogeneity assumption first presented in Daganzo (2007). Thus, we conclude that the occupancy distribution for different time intervals with similar average occupancy is similar in a well-defined MFD. This conclusion is still a little different from our initial conjecture, as there are some small errors within the distribution of each average occupancy group. The analysis of errors will be investigated in Section 4.

An interesting observation is that the variance of individual detector occupancy, which is a measurement of spatial inhomogeneity of the congestion level, has a well defined relationship with the average network occupancy, which expresses the average congestion level. Note that the variance for Yokohama is the same for the onset and offset of morning and evening peak if the average network occupancies are the same, i.e. periods with significantly different *O–D* tables. Thus, this observation supports our initial conjecture that a network exhibits a well-defined MFD if the spatial distribution of congestion is similar for times when average network occupancy is about the same. Fig. 4 illustrates the above relationship, by plotting the variance of individual detector occupancy vs. the average occupancy every 5 min for a whole day. The red² curve plots the same variables if it is assumed that all locations in the network have the same probability (equal to the average occupancy) to be occupied by a vehicle. This observation will be utilized in Section 3 to develop an analytical approximation for the spatial distribution of vehicles in a network with a well-defined MFD.

2.3. Does a well-defined MFD exist for freeway networks?

We now test the existence of a well-defined MFD for a freeway system that has different properties than an arterial network, e.g. merges instead of signals, different speeds, less route choices, etc. We choose a typical workday for the Twin Cities freeway network in Minnesota, where flow and occupancy data are available from a large number of loop detectors. We first

² For interpretation of color in Figs. 1–10, the reader is referred to the web version of this article.

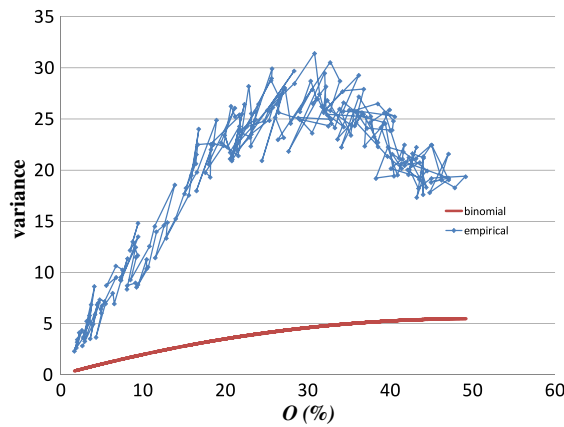


Fig. 4. Variance of individual detector occupancy vs. average network occupancy: (a) empirical data and (b) assuming binomial distribution of congestion.

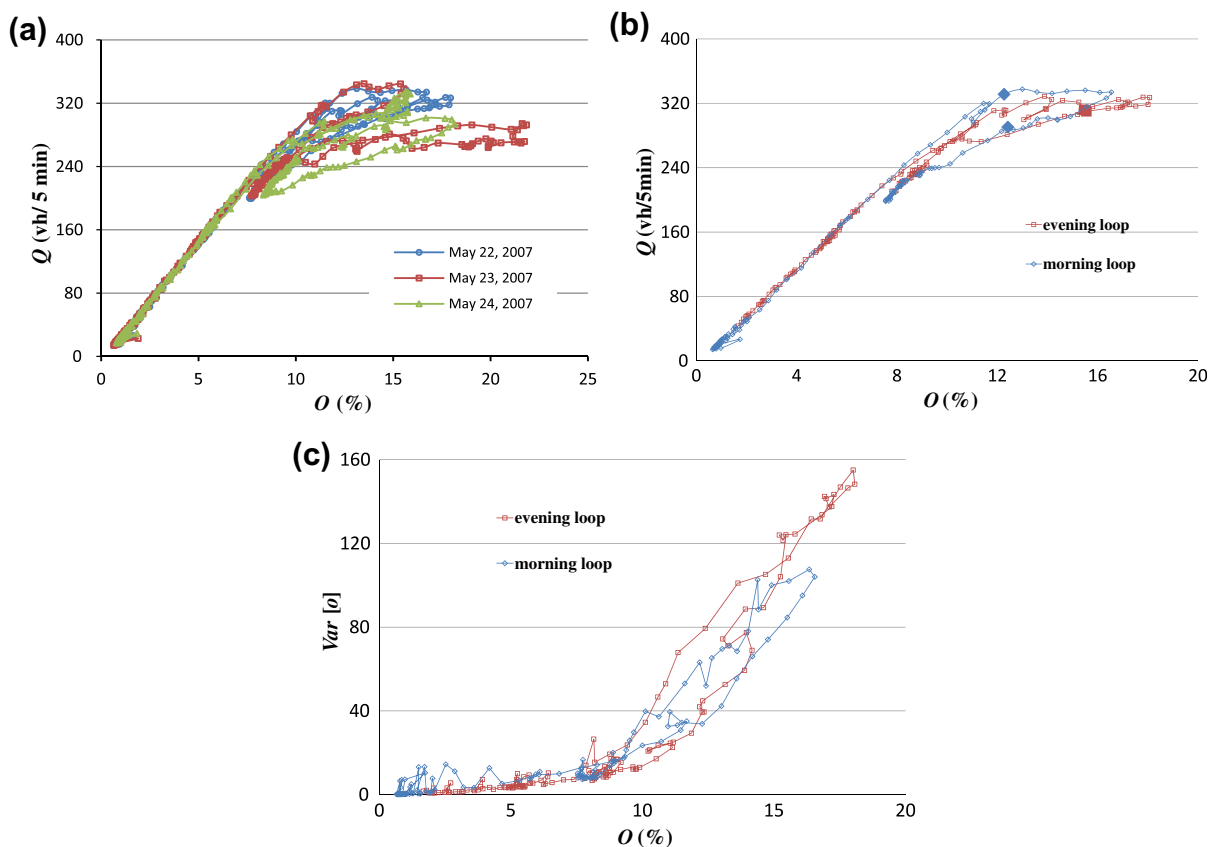


Fig. 5. ATR for freeway network: (a) Q vs. O for three consecutive days; (b) comparison of points with and without hysteresis for May 22nd, 2007; (c) variance of individual detector occupancy, $\text{Var}[o]$ vs. O for May 22nd, 2007.

plot the average network occupancy O , vs. average network flow Q , every 5 min for three consecutive days, May 22nd to 24th, 2007.

Fig. 5a summarizes the results. A non well-defined MFD is noticeable. For the uncongested part we still cannot distinguish one day from the others. For the congested part, however, strong variations and hysteresis phenomena appear (i.e. the evolutions of traffic states follow different paths during the onset and offset of morning and evening peak). We further investigate the Aggregated Traffic Relationship (ATR) for the freeway network, by estimating the variance of individual detector occupancy for every 5 min. Fig. 5c plots the estimated variance vs. the average network occupancy, O , for May 22nd, 2007. It

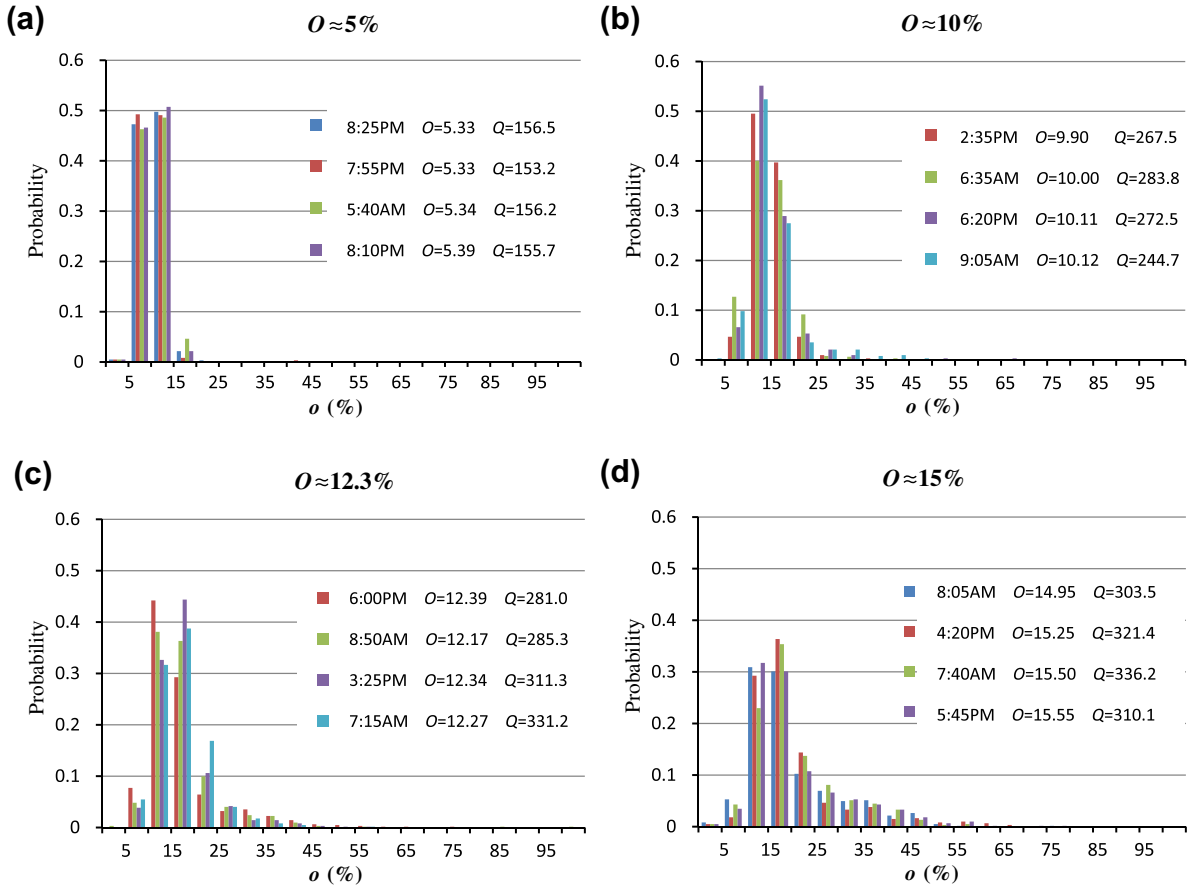


Fig. 6. Histograms for different values of average occupancy, O , for the freeway network.

is clear that different loops are formed in this relationship similar with the hysteresis loops of the Q vs. O relationship. By comparing Figs. 5c and 4, we see that the arterial and the freeway network have different macroscopic behavior, as for similar level of congestion the spatial distribution of congestion is significantly different for the freeway network.

To enhance more in this direction, we estimate empirical histograms of individual detector occupancy for times of day with similar average network occupancy, O , shown in Fig. 6 (similar to Fig. 2). The figure also shows the values of Q and O for each time. Note how different the distributions are when flows are different. Compare for example the distributions at times 3:25 pm vs. 6:00 pm and 7:15am vs. 8:50am in Fig. 6c. These times represent the onset and offset of congestion in the evening and morning peak. Note also that the differences are significant for occupancy range $O = 10$ –15% which is the critical value of occupancy (value for which flow is maximum) for most detectors. This observation provides a hint for an explanation of hysteresis, shown in Fig. 5b and c. For the same value of O , flow is lower when variance of detector occupancy, $\text{var}[o]$, is higher. The reason is that the distribution spreads away from the mean value and there are more detectors in the congested or uncongested regime, but less detectors in the critical occupancy range, which maximizes detector flows. Further investigation is needed to justify the reasons for this change in variance. A deeper analysis and explanation of the phenomenon will be presented in a follow-up publication (work in progress).

To verify the statistical significance of the above findings, we compare two different pairs using the Mann–Whitney statistical test; (i) two points with similar average network flows (large blue points in Fig. 5b) and (ii) two points with different average network flows (large red points in Fig. 5b). The p -value for the first test is 0.847 and we cannot conclude that the two samples come from different populations. In contrast to the first test, the p -value for the second test is very small (0.007); the test is significant even at $\alpha = 0.01$ level. This means the two samples come from different distribution. These results show that freeway networks do not behave like arterial networks.

3. An analytical approximation for the spatial distribution of vehicles in an arterial network

We are interested in the distribution of vehicles over space, conditional on the total number of vehicles in the network, S . In other words, we would like to model the spatial distribution of link density at the network level. We now focus in the distribution of the number of cars in a link of the network. If we assume that (i) the number of vehicles in each link is

uncorrelated with the number in all other links over the network and (ii) the probability that a randomly selected cell in the network contains a car is the same, then the number of vehicles in each link should follow a binomial distribution with parameter p . Parameter p should be equal to S/NL , where NL is the total number of network cells $NL = k_{jam} \times l \times L$, and L is the total number of links.³ We use the average occupancy as the success probability of the binomial parameter to test our occupancy distribution from the experiment.

In general, if the random variable X follows the binomial distribution with parameters N and p , we define $X \sim B(N, p)$. The probability of getting exactly x successes in N trials is given by the probability mass function.

$$\Pr(X = x) = B(x; N, p) = \binom{N}{x} p^x (1 - p)^{N-x} \quad (1)$$

The number of trials for this experiment was estimated as the maximum number of vehicles that fit in one link, $N = l \cdot k_{jam}$. As explained before, this leads to $N = 22vhs$ and the probability of success equals to the average network occupancy, $p = O$. It is found out that none of the histograms in Fig. 3 follows a binomial distribution by applying a Chi-square test. This may be due to correlation between adjacent links, as when a downstream link becomes congested, the density of corresponding upstream links might be affected.

Further analysis of more data from the Yokohama experiment shows that the spatial distribution of occupancy does not follow a binomial distribution, as the spatial variance of occupancy is much higher than the one of a Binomial distribution with probability of success equal to the average occupancy, $p = O$. Fig. 4 shows the variance of individual detector occupancy as average network occupancy, O , changes for a whole day. The figure contains an estimation using real data for December 14, 2001 in downtown Yokohama and the estimated variance if a binomial distribution is assumed with $p = O$. Note the significant difference, especially as the level of congestion increases (variance 4–5 times higher).

In an arterial network, it is quite reasonable to assume that the immediate upstream link can be affected by a downstream link. Mathematically speaking, we assume that the probability of having a car in the cell of one link is correlated with the number of vehicles of its downstream link. Furthermore, in case there are some strong turning movements or ending point like parking ramps, the two links may be uncorrelated and the probability is equal to the a priori probability S/NL . We present here a model to estimate the spatial distribution of car density by introducing the necessary notation.

As mentioned above, the main reason of this difference with the binomial assumption is that traffic congestion in a link has some correlation with successive links, while a binomial distribution assumes zero correlation from one link to another. To capture this type of spatial correlation and congestion propagation from one link to another we introduce a probability mixture model.

Consider $K + 1$ successive links in an arterial network and X_k is a random variable denoting the number of vehicles in link $k = 0, 1, \dots, K$ (zero represents the most upstream link and K the most downstream). Then, it is assumed that X_k follows a binomial distribution with probability of success $p_k = \frac{x_{k-1}}{N}$, where x_{k-1} is the number of vehicles in link $k - 1$

$$\Pr(X_k = x_k | X_{k-1} = x_{k-1}) = B\left(x_k; N, \frac{x_{k-1}}{N}\right). \quad (2)$$

Using the law of total expectation for the unconditional probability $\Pr(X_k = x_k)$ we get

$$\Pr(X_k = x_k) = \sum_{i=0}^N B\left(x_{k-1}; N, \frac{i}{N}\right) \cdot \Pr(X_{k-1} = i). \quad (3)$$

Eq. (3) provides the probability density function (pdf) of X_k if X_0 is known. It is not difficult to prove (tedious, but straightforward) that the limit pdf of X_k as $k \rightarrow \infty$ if $X_0 \sim B(N, p)$, is taking values 0 or N with

$$\lim_{k \rightarrow \infty} \Pr(X_k = N) = p \quad \text{and} \quad \lim_{k \rightarrow \infty} \Pr(X_k = 0) = 1 - p. \quad (4)$$

Fig. 7 shows the probability density function of X_k for $k = 1, 2, \dots, 10$ for $p = \frac{7}{22}$.

Given that an arterial network is two-dimensional, each link has more than one successive links and receives flow usually from 1 to 3 links depending on the topology of the network. To define which is the k th link in the network one can use a trajectory that travels and visits all links in the network only once. There are many different ways to do this, i.e. moving randomly or moving first north to south and then east to west, etc. Because of different turning percentages in each link there is some probability that correlation between link k and $k - 1$ is much smaller than the one described by the model in Eq. (3). For this reason we add one more parameter to the model; this is probability π , that links k and $k - 1$ are independent and $\Pr(X_k = x_k | X_{k-1} = x_{k-1}) = \Pr(X_k = x_k) = B(x_k; N, O)$.

One should expect that probability π should be higher for uncongested conditions (less spatial correlation), as trip generation or turning movements can significantly affect the number of vehicles in low density links. In congested networks, one expects that the aforementioned effects are less intense, as links contain a large number of vehicles. Real data can provide additional insights in this direction and π values can be estimated based on observation. The Yokohama data do not allow to perform this investigation because (as explained in Geroliminis and Daganzo (2008)) the loop detector locations are not

³ The exact distribution is hypergeometric, as if each vehicle randomly chooses without replacement an available position on the network. Hypergeometric converges to binomial for high numbers of vehicles and links.

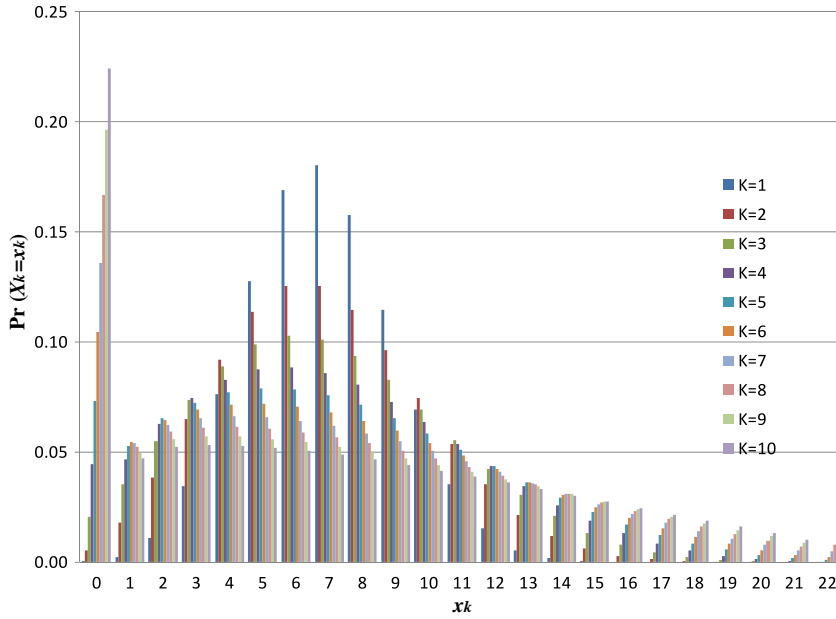


Fig. 7. Probability density function of the proposed model derived from Eqs. (1)–(3) for $p = 7/22$.

assigned to a digital map and there is no information which detectors are close to each other. For this reason, a statistical estimation of π value will be performed here.

To estimate now the spatial distribution of cars in the network we need to know what is the probability that a specific link is of type X_k , i.e. is correlated with $k - 1$ upstream links. This is not difficult to estimate given the modification described above. For a specific link in the network there is probability $1 - \pi$ that this link is correlated with the previous one and probability π that it is of type X_0 . Thus, the type of link follows a geometric distribution with probability of success $1 - \pi$, i.e. the number of failures until the first success.

$$\Pr(X_k = X_i) = (1 - \pi)^i \pi. \quad (5)$$

Thus, using a mixture of Eqs. (3) and (5) we can estimate which is the spatial probability density function of link occupancy.

Fig. 8 compares the proposed analytical model (T) with the empirical data from Yokohama (E) and the binomial distribution (B) for different values of average occupancy, $O = 0.09, 0.18, 0.27, 0.32, 0.37$ and 0.46 , which is approximately equal with $O = \frac{2}{22}, \frac{4}{22}, \frac{6}{22}, \frac{7}{22}, \frac{8}{22}$ and $\frac{10}{22}$, respectively. Parameter π has been calibrated to maximize goodness of fit (minimum average square error).

Table 2 summarizes the results of the analysis. It also contains the average speed in the network, \bar{v} , to illustrate the level of congestion. Note that the theoretical model developed here provides a much better fit than the binomial distribution and a good estimator of spatial variance of link occupancy, especially when the network becomes congested. Table 2b contains the results of the statistical analysis to compare the empirical distribution vs. the proposed model (T) and the binomial distribution. Two commonly used tests were performed to identify, if the proposed model and the empirical data are drawn from the same distribution: (i) a Chi-square test (as in Section 2) and (ii) a Smirnov–Kolmogorov test. This test estimates the maximum vertical distance between the under consideration cumulative distribution functions, D , and compares it with a threshold value, D_{cr} , which depends on the number of data points and a level of significance, α . Both tests conclude that the proposed model (T) is not significantly different than the empirical data, while the binomial model is rejected. Values for D , D_{cr} of the Smirnov–Kolmogorov test and p -value of the Chi-square test are summarized in Table 2b.

4. Analysis of errors for a macroscopic fundamental diagram

This section provides an approximate estimation of the variance of the macroscopic fundamental diagram in terms of $Q(O)$, where O is the average occupancy of all M detectors in the network during time interval t (comparable in size with one signal cycle), and Q is the average flow of all detectors during the same time interval (interval t is omitted from the notation for simplicity purposes). Variables Q and O in our analysis are unweighted averages. If we follow Edie's definition, the approach should be similar with the one we present here, but much more tedious.

Let us repeat the notation of a previous section; q^d and o^d is the flow and occupancy of an individual detector d during a time interval. The relationship between q^d and o^d (expressed as fundamental diagram) exhibits high scatter, especially for

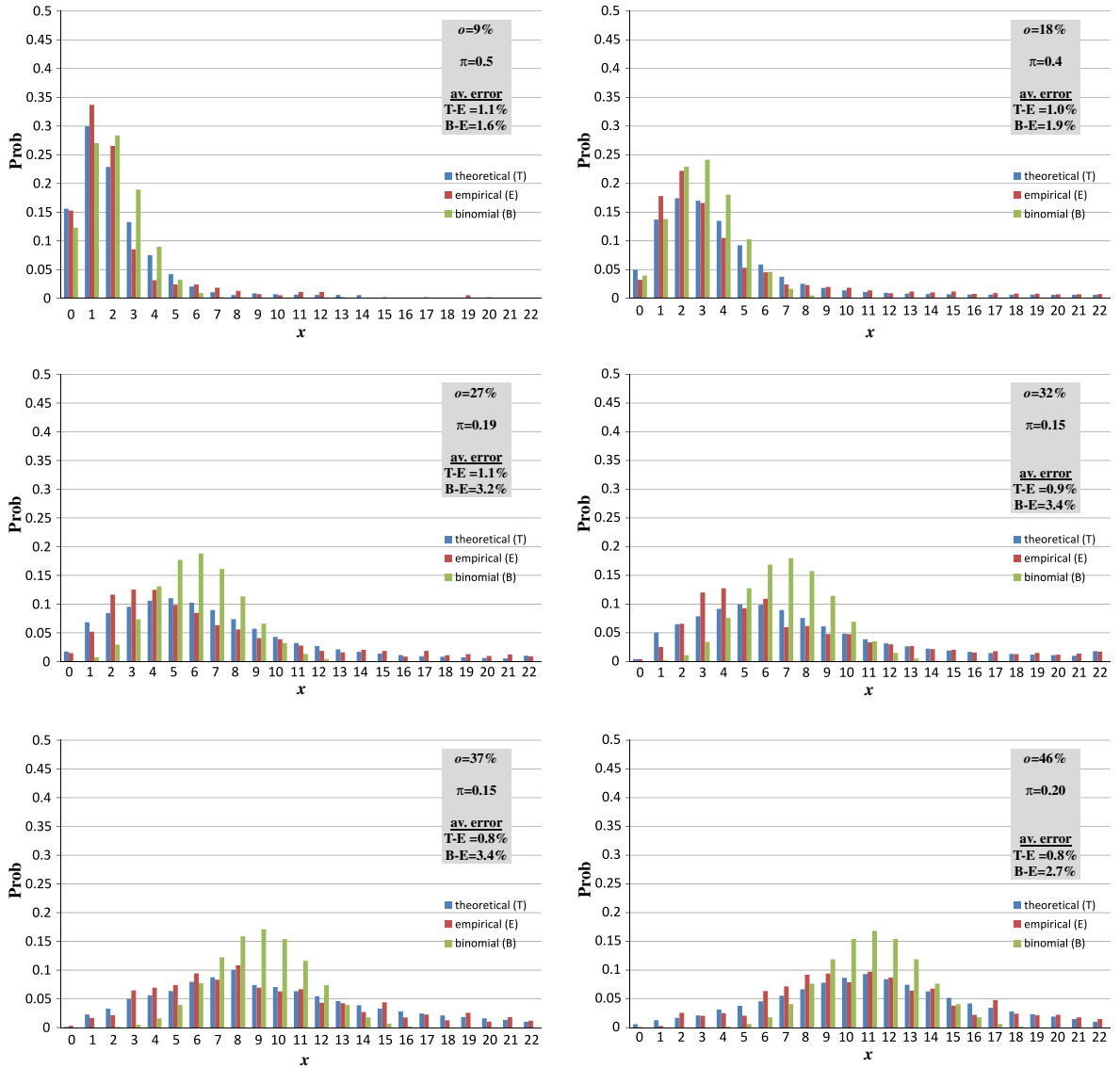


Fig. 8. Distribution comparison between empirical data, proposed model and binomial distribution for different values of average network occupancy, O .

congested conditions. If $q^d(o)$ represents an average (or best fit) fundamental diagram for the individual detector we can write $q^d = q^d(o) + \varepsilon^d(o)$, where $\varepsilon^d(o)$ is an error, which depends on occupancy o^d according to the analysis for real data. If we now look at all individual detectors by omitting their label d (without any type of spatial aggregation or averaging) we may fit a fundamental diagram for all individual detectors $q = q(o) + \varepsilon(o)$. Denote $p(o)$ the continuous probability density function, which shows what fraction of detectors has occupancy o and we can write:

$$Q(O) \cong \int_0^{100\%} p(o) \cdot q(o) do \cong \sum_{i=0}^N \tilde{p}(o_i) \cdot \tilde{q}(o_i), \quad (6)$$

where $\tilde{p}(o_i)$ is the discrete version of the continuous probability density function $p(o)$ and $\tilde{q}(o_i)$ is the average flow of all detectors in occupancy group i , $i = 0, 1, \dots, N$.

We now investigate the variance of $Q(O)$. Note that this error/variance can be high (even higher than the individual detector errors) if there is some type of significant correlation (not random errors) between some of the aforementioned variables. For example (i) if individual detectors exhibit significantly different individual fundamental diagrams $q^d(o)$ in different parts of the network or (ii) if congestion level is not homogeneously distributed among the network for different times of day with similar average occupancy O or (iii) if $q^d(o)$ is not well defined and exhibits hysteresis phenomena (like in freeways).

Table 2

Analytical approximation of spatial distribution of occupancy (a) summary of results of the proposed model (T), the empirical data (E) and binomial distribution (B) and (b) statistical analysis for T vs. E and B vs. E (average error, Chi-square and Smirnov–Kolmogorov test).

O	\bar{v} (km/h)	Variance			π	
		B	T	E		
0.09	24	1.82	6.24	8.47	0.50	
0.18	19	3.29	18.16	20.58	0.40	
0.27	13	4.36	23.61	22.06	0.19	
0.32	11	4.77	25.32	24.33	0.15	
0.37	9	5.32	24.07	23.26	0.16	
0.46	6	5.50	20.84	18.35	0.20	
\bar{O}	χ^2 test (p -value)		S-K (D value) $D_{cr} = 8.6\%$ ($\alpha = 0.05$) (%)		Average error (%)	
	$B-E$	$T-E$	$B-E$	$T-E$	$B-E$	$T-E$
0.09	0	0.698	9.7	7.1	1.6	1.1
0.18	0	0.403	17.5	7.1	1.9	1.0
0.27	0	0.330	19.7	6.2	3.2	1.1
0.32	0	0.642	21.9	5.6	3.4	0.9
0.37	0	0.493	26.2	4.2	3.4	0.9
0.46	0	0.474	15.9	4.5	2.4	1.0

Denote $V[\cdot]$ the variance of a random variable. Let $\tilde{p}(o_i)$ and $\tilde{q}(o_i)$ have random errors $\varepsilon_{\tilde{p}}(o_i)$ and $\varepsilon_{\tilde{q}}(o_i)$ respectively, which are independent of each other. If the random errors of their product is independent of all other products for different i 's, then the variance of average network flow for a given average network occupancy O , $Q(O)$, is

$$V[Q(O)] \cong \sum_{i=0}^N V[\tilde{p}(o_i) \cdot \tilde{q}(o_i)]. \quad (7)$$

The estimation of the variance of the product of two random variables is not an easy task in statistical analysis. But, if these two variables are independent, the variance has a closed form expression (Goodman, 1960),

$$V[XY] = E^2[X] \cdot V[Y] + E^2[Y] \cdot V[X] + V[X] \cdot V[Y]. \quad (8)$$

For physical reasons, there seems no dependence between variables $\tilde{p}(o_i)$ and $\tilde{q}(o_i)$. One variable expresses the fraction of links/detectors with occupancy close to o_i while the other describes the average flow of the same group of links/detectors. For simplicity, we use \tilde{p}_i for $E[\tilde{p}(o_i)]$ and \tilde{q}_i for $E[\tilde{q}(o_i)]$. Combining Eqs. (7) and (8) we get

$$V[Q(O)] \cong \sum_{i=0}^N \tilde{p}_i^2 V[\varepsilon_{\tilde{q}}(o_i)] + \tilde{q}_i^2 V[\varepsilon_{\tilde{p}}(o_i)] + V[\varepsilon_{\tilde{p}}(o_i)] V[\varepsilon_{\tilde{q}}(o_i)]. \quad (9)$$

Also, if $\varepsilon_{\tilde{q}}(o_i)$ is the random error of the flow for all detectors for occupancy group i (close to o_i), using the central limit theorem, a good approximation for the variance of their average $\tilde{q}(o_i)$ is $V[\varepsilon_{\tilde{q}}(o_i)] \cong \frac{V[\varepsilon_{\tilde{q}}(o_i)]}{M\tilde{p}_i}$ (this is the average of approximately $M\tilde{p}_i$ detectors). After some manipulations Eq. (9) becomes

$$V[Q(O)] \cong \sum_{i=0}^N \frac{1}{M} V[\varepsilon_{\tilde{q}}(o_i)] \left(\tilde{p}_i + \frac{V[\varepsilon_{\tilde{p}}(o_i)]}{\tilde{p}_i} \right) + \tilde{q}_i^2 V[\varepsilon_{\tilde{p}}(o_i)]. \quad (10)$$

We now apply Eqs. (6)–(10) to estimate the variance of the macroscopic fundamental diagram for the site of Yokohama. Although some of the assumptions of this section may not be met in the real network, the application still illustrates how the method may work in a real-world application, where the input data include some error.

Firstly, Fig. 9a and b summarizes the estimated mean and variance of $\tilde{p}(o_i)$ for different occupancy groups o_i for $i = 0, \dots, N$ for network average occupancy $o = 15\%$, 25% , 35% and 45% . Note that the variance of the probability density function $V[\varepsilon_{\tilde{p}}(o_i)]$, is maximum for the occupancy group which is very close to the mode (the most frequent value) of $\tilde{p}(o_i)$; for example group #6 for $O = 35\%$.

Secondly, Fig. 9c shows the coefficient of variation, CoV, (dimensionless quantity: standard deviation divided by the mean) of individual detector flow for different values of occupancies. Note that the CoV is almost constant at the value of 0.2, for $o > 5\%$. This is an interesting finding because by assuming that flow's CoV is constant, one can derive a relationship between the level of spatial heterogeneity (as this expressed by the variance and mean of $\tilde{p}(o_i)$) and the variance of an MFD. This relationship can provide useful hints about the type of cities that exhibit a well-defined MFD. We did not categorize the CoV for different network occupancy O levels, as we do not expect that this complication could improve the estimation and has no physical explanation (individual links are expected to produce similar q vs. o pairs independent of the aggregated level

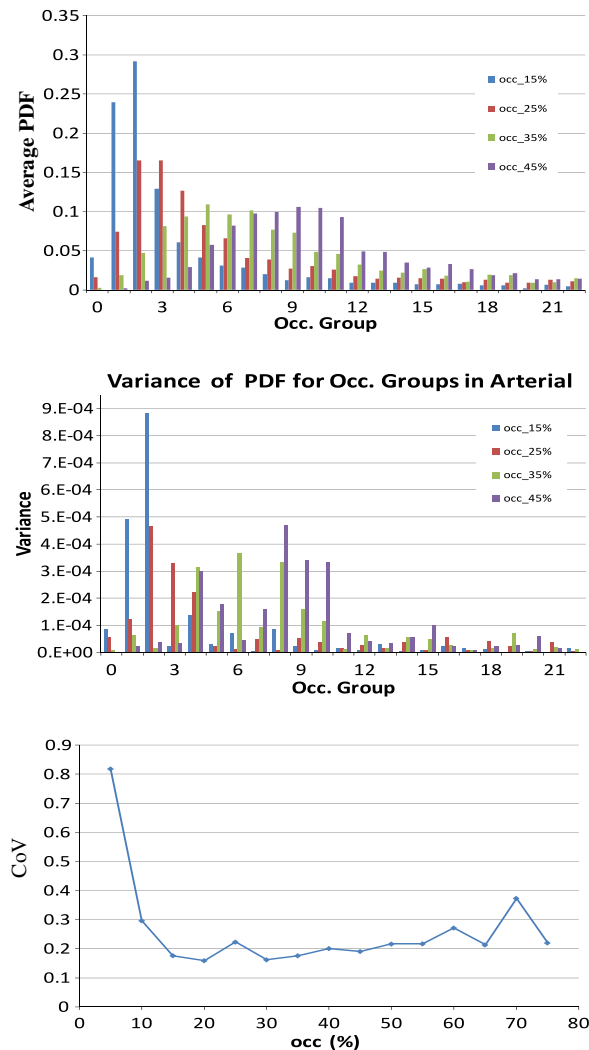


Fig. 9. (a) Average pdf of detector occupancy, $\bar{p}(o_i)$, for different values of O ; (b) variance of the same pdf, $V\bar{p}(o_i)$, for different values; (c) coefficient of variation of individual detector flow, q , for different occupancy for a sample of detectors.

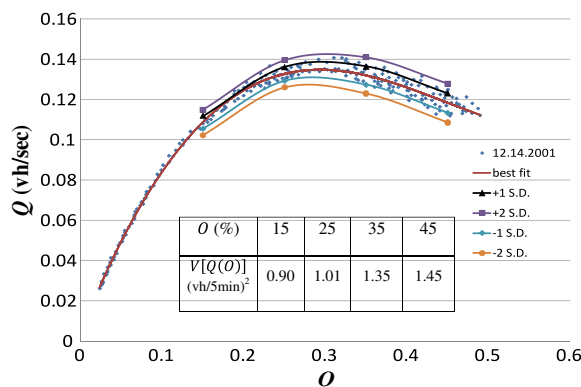


Fig. 10. Yokohama's estimated 1- and 2-standard deviation bands for the MFD.

of congestion). This is not necessarily the case in freeways because of hysteresis phenomena, which have been reported by many researchers (e.g. Treiterer and Myers, 1974).

Fig. 10 shows the results of variance estimation for Yokohama. Individual dots are Q vs. O points for every 5 min for the weekday of December 14, 2001. The lines show the 1- and 2-standard deviation bands arising from Eq. (10) on each side of a fitted curve. Fig. 10 also summarizes the estimated variance. Our simplified theoretical approach describes well the variance increase as the level of congestion increases. Note that the CoV of the MFD is about 3–4%. This value is 5–6 times smaller than the CoV of an individual link, but much higher than it would have been estimated if the central limit theorem was applied for $M = 500$ detectors.

5. Discussion

We have studied how the spatial variability of vehicle density can affect the shape, the scatter and the existence of a well-defined macroscopic fundamental diagram. Most studies until now have looked at macroscopic relationships between the average flow and density, but did not explore the effect of inhomogeneous distribution of vehicles in space and time. The results of this paper show a sufficient condition for the existence of an MFD with low scatter. This condition states that if the spatial distribution of link density is the same for two different time intervals with the same number of vehicles in the network, then these two time intervals should have the same average flows. While MFDs should be expected for arterial networks that satisfy the above condition, freeway networks have topological or control characteristics that are different (non-redundant, no traffic signals, hysteresis phenomena) and may not be well defined as we showed for the Twin Cities freeway network. Work in progress further analyses freeway data and investigates the main reasons of high scatter in Aggregated Traffic Relationships.

In Section 3 of the paper we derived an analytical approximation of the spatial distribution of vehicles in the network, which considered correlation between adjacent links by using a probability mixture model. We also showed that the variability of vehicles in an arterial network is much higher than the one assuming a random distribution of vehicles. In Section 4 we defined upper bounds for the errors of an MFD based on the errors of individual link FDs and errors in the probability density function of link occupancy.

While the results of this paper provide a clearer understanding of macroscopic modeling of traffic in cities, additional research is needed in different types of networks to understand how variations in the topology/structure of the network can affect the shape and scatter of an ATR or MFD. A proper parameterization and description of heterogeneity that would yield good estimates of macroscopic representation of the network is of critical importance. Several questions related to this issue should be addressed: (a) How a heterogeneous urban network can be partitioned in neighborhoods with well-defined MFDs? (b) Should this partitioning be time-dependent? (c) How are the traffic dynamics of multi-partition systems? (d) How can we control inter-transfers between successive neighborhoods to increase cities' mobility and accessibility measures? and (e) how to allocate/locate available sensing resources between different types of sensors (fixed, mobile, etc.) to maximize system observability at the city level?

Acknowledgements

The authors are grateful to Prof. Masao Kuwahara (University of Tokyo) for his continued support with the Yokohama data and for his many comments.

References

- Ardekani, S., Herman, R., 1987. Urban network-wide traffic variables and their relations. *Transportation Science* 21 (1), 1–16.
- Buisson, C., Ladier, C., 2009. Exploring the impact of the homogeneity of traffic measurements on the existence of macroscopic fundamental diagrams. *Transportation Research Record* 2124, 127–136.
- Daganzo, C.F., 2007. Urban gridlock: macroscopic modeling and mitigation approaches. *Transportation Research Part B* 41 (1), 49–62.
- Daganzo, C.F., Geroliminis, N., 2008. An analytical approximation for the macroscopic fundamental diagram of urban traffic. *Transportation Research Part B* 42 (9), 771–781.
- Daganzo, C.F., Gayah, V., Gonzales, E., 2011. Macroscopic relations of urban traffic variables: bifurcations, multivaluedness and instability. *Transportation Research Part B* 41 (1), 278–288.
- Edie, L.C., 1963. Discussion of traffic stream measurements and definitions. In: Almond, J. (Ed.), *Proceedings of the 2nd International Symposium on the Theory of Traffic Flow*. OECD, Paris, France, pp. 139–154.
- Geroliminis, N., Daganzo, C.F., 2007. Macroscopic Modeling of Traffic in Cities. The 86th Transportation Research Board Annual Meeting, Paper No. 07-0413, Washington, DC.
- Geroliminis, N., Daganzo, C.F., 2008. Existence of urban-scale macroscopic fundamental diagrams: some experimental findings. *Transportation Research Part B* 42 (9), 759–770.
- Godfrey, J.W., 1969. The mechanism of a road network. *Traffic Engineering and Control* 11 (7), 323–327.
- Goodman, L., 1960. On the exact variance of products. *Journal of the American Statistical Association* 55 (292), 708–713.
- Helbing, D., 2009. Derivation of a fundamental diagram for urban traffic flow. *The European Physical Journal B* 70 (2), 229–241.
- Mahmassani, H.S., Peeta, S., 1993. Network performance under system optimal and user equilibrium dynamic assignments: implications for ATIS. *Transportation Research Record* 1408, 83–93.
- Mahmassani, H., Williams, J.C., Herman, R., 1987. Performance of urban traffic networks. In: Gartner, N.H., Wilson, N.H.M. (Eds.), *Proceedings of the 10th International Symposium on Transportation and Traffic Theory*. Elsevier, Amsterdam, The Netherlands.
- Olszewski, P., Fan, H.S., Tan, Y.W., 1995. Area-wide traffic speed-flow model for the Singapore CBD. *Transportation Research Part A* 29 (4), 273–281.
- Treiterer, J., Myers, J.A., 1974. The hysteresis phenomenon in traffic flow. In: Buckley, D.J. (Ed.), *Proceedings of the 6th International Symposium on Transportation and Traffic Theory*, pp. 13–38.
- Williams, J.C., Mahmassani, H.S., Herman, R., 1987. Urban traffic network flow models. *Transportation Research Record* 1112, 78–88.