The Impact of Traffic Dynamics on the Macroscopic Fundamental Diagram

V. L. Knoop PhD
Delft University of Technology
Transport & Planning
Stevinweg 1
Delft, The Netherlands
+31 15 278 8413
v.l.knoop@tudelft.nl

S.P. Hoogendoorn PhD
Delft University of Technology
Transport & Planning
Stevinweg 1
Delft, The Netherlands
s.p.hoogendoorn@tudelft.nl

J.W.C. Van Lint PhD
Delft University of Technology
Transport & Planning
Stevinweg 1
Delft, The Netherlands
j.w.c.vanlint@tudelft.nl

October 25, 2012

Word count:

nr of words in abstract	199
nr of words	5480
nr of figures& tables	8* 250 = 2000
total	7480

Presented at the 92th Annual Meeting of the Transportation Research Board

ABSTRACT

Literature shows that – under specific conditions – the macroscopic fundamental diagram (MFD) describes a crisp relationship between the average flow (production) and the average density in an entire network. The limiting condition is that traffic conditions must be homogeneous over the whole network. Recent works describe hysteresis effects: systematic deviations from the MFD as result of loading and unloading.

This article proposes a two dimensional generalization of the MFD, the so-called Generalized Macroscopic Fundamental Diagram (GMFD), which relates the average flow to both the average density and the (spatial) inhomogeneity of density. The most important contribution is that we show this is a continuous function. Using this function, we can describe the mentioned hysteresis patterns. The underlying traffic phenomenon explaining the two dimensional surface described by the GMFD is that congestion concentrates (and subsequently spreads out) around the bottlenecks that oversaturate first. We call this the nucleation effect. Due to this effect, the network flow is not constant for a fixed number of vehicles as predicted by the MFD, but decreases due to local queuing and spill back processes around the congestion ”nuclei”. During this build up of congestion, the production hence decreases, which gives the hysteresis effects.

1 INTRODUCTION

The dynamics of traffic differs considerably from many other transport phenomena. Most striking is that if the load (i.e., number of vehicles) exceeds a critical number, the traffic performance, for instance measured by arrival rates, decreases with an increasing traffic load. This differs for instance from fluid dynamics where a higher load leads to higher flows. In this article we will study how traffic dynamics evolve in a traffic network, and how this influences macroscopic traffic variables.

Geroliminis and Daganzo (*1*) investigate the Macroscopic Fundamental Diagram (MFD). They show that if aggregated over an area, the relationship between accumulation (i.e., the average density of roadway length) and production (i.e., the average flow of vehicles per unit of time) is quite crisp. This can be an aspect of the law of large numbers: the more data is aggregated, the less influence the differences in drivers characteristics have. For control purposes, it is very useful to have a strict relationship on the basis of which control can be applied. However, several research efforts (e.g., (2, 3)) since Geroliminis and Daganzo (*1*) suggest that such a crisp relationship exists only for homogeneously loaded networks. In practice, congestion is seldom spread homogeneously over the network.

Recent literature studies the effects of inhomogeneity of densities over the network (e.g., (2, 4, 5, 6). However, most of these papers study the effect of inhomogeneity without taking into account that the congestion spreads over the network. The effects of traffic dynamics and inhomogeneous congestion on the MFD are currently being described as hysteresis, but its explanation and quantification is currently a gap in the literature, which this article aims to fill. In this article we propose a generalization of the MFD, the *Generalized Fundamental Diagram* (GMFD), giving the production as function of the accumulation and spatial inhomogeneity of density. This latter quantity can be interpreted as an average measure for inhomogeneity of traffic conditions. We will demonstrate that this GMFD describes a well-defined (crisp) two dimensional plane. The classic MFD is a one-dimensional projection of this plane.

The use of this paper is twofold. At one hand, it scientific use is that it makes clear the that production is a *continuous* function of accumulation and spatial inhomogeneity of density. It also, after the analysis of simulated traffic dynamics in a network, describes hysteresis effects found in the MFD using the GMFD. At the other hand, for practical use, this newly found GMFD can be used in traffic control schemes. For instance, it can be used for estimating speed in a (sub-)network, and traffic can be guided over the faster routes. This way, GMFD can improve the traffic routing algorithm based on the MFD(*7*).

The remainder of the article is set-up as follows. First, a review is provided of the literature of the MFD. Then, section 3 introduces the variables used in this article. Section 4 then presents how GMFDs look when traffic dynamics are not incorporated. We do so by simply averaging randomly chosen traffic states. Section 5 presents the simulation experiment. Using a macroscopic traffic simulation, a simplified traffic network is simulated, and the queuing dynamics are analyzed. Furthermore the effect of the spatial inhomogeneity of density on the GMFD of density is studied in detail. Section 6 explicitly compares the GMFD of the traffic simulation with GMFD obtained by randomly chosen traffic states, and explains the differences in results. Thus, we are able to comment on the effect of traffic dynamics on the GMFD. Finally, section 7 presents the conclusions and an outlook on further research and applications.

2 LITERATURE OVERVIEW OF MACROSCOPIC FUNDAMENTAL DIAGRAMS

The original idea of a Macroscopic Fundamental Diagram (MFD) dates back to the late sixties (8), but only after Daganzo's 2007 reintroduction of the concept, research into the MFD has accelerated, often focussing on potential for control and management of network traffic. An overview of the most important studies with the MFD are given in table 1.

The best-known studies are the ones by Daganzo (*9*) and Geroliminis and Daganzo (*1*). Geroliminis and Daganzo (*1*) show the relationship between the number of completed trips and the production function

TABLE 1 Overview of the papers discussing the macroscopic fundamental diagram

Paper	data	network	insight
Daganzo (*9*)	theory	none	Overcrowded networks lead to a performance degradation – the start of the MFD.
Geroliminis and Daganzo (*1*)	real	Yokohama	MFDs work in practice, and there is a relation between the average flow (production) and the arrival rate (performance).
Daganzo and Geroliminis (*10*)	data & simulation	Yokohama & San Francisco	The shape of MFDs can be theoretically explained.
Buisson and Ladier (*2*)	real	urban + urban freeway	There is scatter on the FD if the detectors are not ideally located or if there is inhomogeneous congestion.
Ji et al. (*11*)	simulation	urban + freeway	Hybrid networks give a scattered MFD; inhomogeneous congestion reduces flow, and should therefore be considered in network control.
Cassidy et al. (*4*)	real	3 km freeway	MFDs on freeways only hold if stretch is completely congested or not; otherwise, there are points within the diagram.
Mazloumian et al. (*12*)	simulation	urban grid – periodic boundary	spatial inhomgeneity of density is important in deriving the production.
Geroliminis and Ji (*3*)	real	Yokohama	Spatial inhomogeneity of density is important in deriving the production.
Knoop and Hoogendoorn (*13*)	simulated	grid	Traffic congestion attracts congestion; production is a continuous function of accumulation and inhomogeneity of density.
Wu et al. (*14*)	real	900m arterial	There is an arterial fundamental diagram, influenced by traffic light settings.
Knoop et al. (*7*)	simulated	grid	Traffic performance can be improved by routing based on an average accumulation in an area. Possibly spatially correlated densities in a network.
Daganzo et al. (*15*)	simulated	grid	Equilibrium states in a network are either free flow, or heavily congested. Rerouting increases the critical density for the congested states considerably.
Gayah and Daganzo (*5*)	simulated	grid/bin	Hysteresis loops exist in MFDs due to a quicker recovery of the uncongested parts; this is reduced with rerouting.
Saberi and Mahmassani (*6*)	real	Portland region	Spatial spread of a network cause the traffic production to reduce; this leads to hysteresis loops.

which is defined as a weighted average of the flow on all links. Their results indicate that the network production can be used as a good approximation of the utility of the users for the network, i.e., it is related to their estimated travel time. Furthermore, after some theoretical work, Geroliminis and Daganzo (*1*) were the first to show that MFDs are not just theoretical concepts, but also exist in practice. With pioneering work using data from the Yokohama metropolitan area, an MFD was constructed with showed a crisp relationship between the network production and the accumulation.

Also, theoretical insights have be gained over the past years. Daganzo and Geroliminis (*10*) have shown that rather than to find the shape of the MFD in practice or by simulation, one can theoretically predict its shape. This gives a tool to calculate the highest production of the network, which then can be compared with the actual network production.

One of the requirements for the crisp relationship is that the congestion should be homogeneous over the network. Buisson and Ladier (*2*) were the first to test the how the MFDs change if the congestion is not homogeneously distributed over the network. They showed a reasonably good MFD for the French town Toulouse in normal conditions. However, one day there were strikes of truck drivers, driving slowly on the freeways, leading to traffic jams. The researchers concluded that this leads to a serious deviation from the MFD for normal conditions. The inhomogeneous conditions were recreated by Ji et al. (*11*) in a traffic simulation of a urban freeway with several on-ramps (several kilometers). They found that inhomogeneous congestion leads to a reduction of flow. Moreover, they advised on the control strategy to be followed, using ramp metering to create homogeneous traffic states. Cassidy et al. (*4*) studied the MFD for a freeway road stretch. They conclude, based on real data, that the MFD only holds in case the whole stretch is either congested or in free flow. In case there is a mix of these conditions on the studied stretch the production is lower than the production which would be predicted by the MFD. This is due to the convex shape of the fundamental diagram.

The effect of inhomogeneity is further discussed by Mazloumian et al. (*12*) and Geroliminis and Ji (*3*). Contrary to Ji et al. (*11*), both papers focus on urban networks. First, Mazloumian et al. (*12*) show with simulation that the spatial inhomogeneity of density over different locations is an important aspect to determine the total network production. So not only too many vehicles in the network in total, but also if they are located at some shorter jams at parts of the networks. The reasoning they provide is that "an inhomogeneity in the spatial distribution of car density increases the probability of spillover, which substantially decreases the network flow." Furthermore, they conclude that it is essential to model flow quantization to obtain the effects of reducing performance with an increasing variability. This finding from simulation and reasoning is confirmed by an empirical analysis by Geroliminis and Ji (*3*), using the data from the Yokohama metropolitan area. The main cause for this effect is claimed to be the turning movement of the individual vehicles. The same data is used by Geroliminis and Sun (*16*) showing that if clustered in bins with similar variation of denisity, the MFD is a well defined curve. Daganzo et al. (*15*) simplify the setup and approach the traffic flows and MFDs analytically. They show there are several equilibria to which the network will relax if loaded. Where they do show the equilibrium points, they do not show the direct influence of density variability on network production. In preliminary work (13) we already studied the effect of inhomogeneity of density in a traffic network. We then concluded that the macroscopic fundamental diagram was a continuous function of both the accumulation and the inhomogeneity of the density in an area. This is further analysed and quantified in this paper.

Recently, we showed (7) that the MFD could be used for control purposes as well. Because speeds on neighboring roads are correlated, the traffic state in a whole area could be approximated by an average, the MFD. Using the speeds derived from the MFD, we showed that traffic could be rerouted based on the accumulation in a subnetwork and the MFD which gives the speed. In this way, the traffic situation improved. This routing strategy can be improved if the prediction of the speed in a subnetwork is improved, which is one of the aims of this paper.

A theoretical explanation for the phenomenon of the influence of the spatial inhomogeneity of density on the accumulation is given by Daganzo et al. (*15*). He shows that turning at intersections is the key reason for the drop in production under spatially inhomogeneous conditions. Gayah and Daganzo (*5*) then use this information by adding dynamics to the MFD. If congestion solves, it will not solve instantaneous over all locations. Rather, it will solve completely from one side of the queue. Therefore, reducing congestion will increase the spatial inhomogeneity of density and thus (relatively) decrease the production. This means that the production for a system of dissolving traffic jams is lower than the the equilibrium state, and thus graphically located under the MFD. This way, there are hysteresis loops in the MFD, as also noted by Ji et al. (*11*). Note that these loops are an effect by themselves and are not associated with the capacity drop (17, 18). Instead, hysteresis in the MFD case is a result of macroscopic queueing and spillback processes. Saberi and Mahmassani (*6*) discusses this in more detail, how the averaging of traffic states leads to a lower prediction. Furthermore, this is confirmed with real data.

Our article continues this line of research and investigates the effect of network dynamics on the MFD. It does do by first considering the MFD as a simple average of randomly chosen traffic states. This will be compared with a macroscopic traffic simulation, simplified to the essence, by which we reveal the effect of network dynamics and using which we will find the production as function of accumulation and

TABLE 2 The variables used

Symbol	meaning
r	Node
x	Cell in the discretised traffic flow simulation
L_x	Length of the road in cell x
q_x	Flow in cell x
k_x	Density in cell x
k_c	Critical density
ϕ_{ij}	Flux from link i to link j
S	The supply of cell x
D	The demand from cell x
i	The links towards node r
j	The links from node r
N	The number of links
C	The capacity of node r in veh/unit time
α	The fraction of traffic that can flow according to the supply and demand
β	The fraction of traffic that can flow according to the demand and the node capacity
ζ	The fraction of the demand that can flow over node r
X	An area
A_X	Accumulation of vehicles in area X
P_X	Production in area X
R_X	Performance (arrival rate) in area X
γ_X	Spatial inhomogeneity of the density in area X
l	Accumulation in the network at the start of the simulation, expressed as fraction of the critical accumulation
σ	Standard deviation

the spatial inhomogeneity of density.

3 DEFINITIONS AND VARIABLES

In this article, several traffic flow variables are used to describe the network traffic flow operations. This section explains them and shows the way to calculate them. The variables are listed in table 2.

Standard traffic flow variables are flow, q, being the vehicle distance covered in a unit of time, and density, k, the number of vehicles per unit road length. The network is divided into cells, which we denote by x, which have a length L_x. Flow and density in cells are denoted by q_x and k_x.

Furthermore, the accumulation A in an area X is the weighted average density:

$$A_X = \frac{\sum_{x \in X} k_x * L_x}{\sum_{x \in X} L_x} \tag{1}$$

Similarly, the production P in an area X is the weighted average flow:

$$P_X = \frac{\sum_{x \in X} q_x * L_x}{\sum_{x \in X} L_x} \tag{2}$$

Since in the examples of this manuscript the cell length are the same for all links in the network, the accumulation and production are average densities and flows. Recall that there is a strong relationship between the production and the arrival rate, as shown by Geroliminis and Daganzo (*1*). This arrival rate is called performance, and is indicated by R.

This article also studies the inhomogeneity of density expressed by the standard deviation of the cell density. This is found by considering all cell densities in an area at one moment in time, and calculate the standard deviation of these numbers, indicated by γ:

$$\gamma_X = \sqrt{\sum_{x \in X} \frac{(k_x - k_{\text{mean}})^2}{N}} \tag{3}$$

In this equation, k_{mean} is the mean density of all cells.

4 THE EFFECT OF SPATIAL INHOMOGENEITY OF DENSITY ON NETWORK PERFORMANCE: NO DYNAMICS

The macroscopic fundamental diagram can be interpreted as the connection between the average density and the average flow in the area for which it is created. This section will show what the effect is of a simple averaging, without taking typical traffic phenomena and dynamics into account. This adds to the contribution of Saberi and Mahmassani (*6*) in the sense that it adds the stochastic analysis of many links to the time-evolution of two links described analytically in that paper. Section 5 will include traffic dynamics and show the effects thereof, which will be compared to the results obtained in this section. Now, section 4.1 first explains the experiment of averaging random traffic states. Section 4.2 then presents the results thereof.

4.1 Mathematical analysis

In this section we study the effect of averaging without incorporating traffic dynamics. For each link in a set of N links, a traffic state is chosen according to the fundamental diagram. The traffic states on this set of links are averaged, and the accumulation, flow and spatial inhomogeneity of density are calculated. The idea is to find the performance function as function of the average density and the spatial inhomogeneity of density. In this section, we consider these variables independent. In section 5 the inhomogeneity of density is calculated endogenously in a traffic simulation program.

In this section, we consider link density as a random variable $\mathcal{K}$. For reasons of simplicity, it is chosen from a uniform distribution (denoted U) with minimum $k_{\text{mean}} - p$ and maximum $k_{\text{mean}} + p$, with $p \geq 0$ a variable indicating the spread of the density:

$$\mathcal{K} \sim U\left(k_{\text{mean}} - p, k_{\text{mean}} + p\right) \tag{4}$$

This is chosen because it is the most simple distribution, and an analytical is possible. We will show later on (section 5) choose a more realistic approach and comment on the differences in section 6.

In this distribution, parameters k_{mean} and p can be chosen independently, but all values of $\mathcal{K}$ have to lie within the admissible range for densities, ranging from 0 to jam density. Hence, it is required that

$$k_{\text{mean}} - p \geq 0 \tag{5}$$

$$k_{\text{mean}} + p \leq k_{\text{jam}} \tag{6}$$

This means we can freely choose a value for the average density between 0 and k_{jam}. Now the value of p is restricted:

$$0 \leq p \leq \min\{k_{\text{mean}}, k_{\text{jam}} - k_{\text{mean}}\} \tag{7}$$

Next, we assume that all N links have the same triangular fundamental diagram (19), which maps the density uniquely to the flow:

$$q = \begin{cases} \dfrac{k}{k_c} q_{\max} & \text{if } k \leq k_c \\ \left(1 - \left(\dfrac{k - k_c}{k_{\text{jam}} - k_c}\right)\right) q_{\max} = w\left(k - k_{\text{jam}}\right) & \text{otherwise} \end{cases} \tag{8}$$

In the second equation, w is the slope of the congested branch of the fundamental diagram, defined by

$$w = \frac{q_{\text{max}}}{k_c - k_{\text{jam}}} \tag{9}$$

The average density, the standard deviation of the density and the average flow are needed to construct the Generalized Fundamental Diagram. Since we use a continuous distribution for the traffic states, the averages are replaced by the expectation value; throughout the paper, the expectation value is indicated by a bar over the variable. For the expectation value of the density we find by definition:

$$A = \overline{k} = k_{\text{mean}} \tag{10}$$

The spatial inhomogeneity γ can also be directly derived from the known standard deviation of the uniform distribution function

$$\gamma = \sigma(k) = \frac{1}{\sqrt{3}} p \tag{11}$$

For the production, we have to find the expectation value of the flow (equation 2), indicated by $\overline{q(k)}$. To this end, we need to evaluate this integral:

$$P = \overline{q(k)} = \frac{\int_{k_{\text{mean}}-p}^{k_{\text{mean}}+p} q(k)dk}{2p} \tag{12}$$

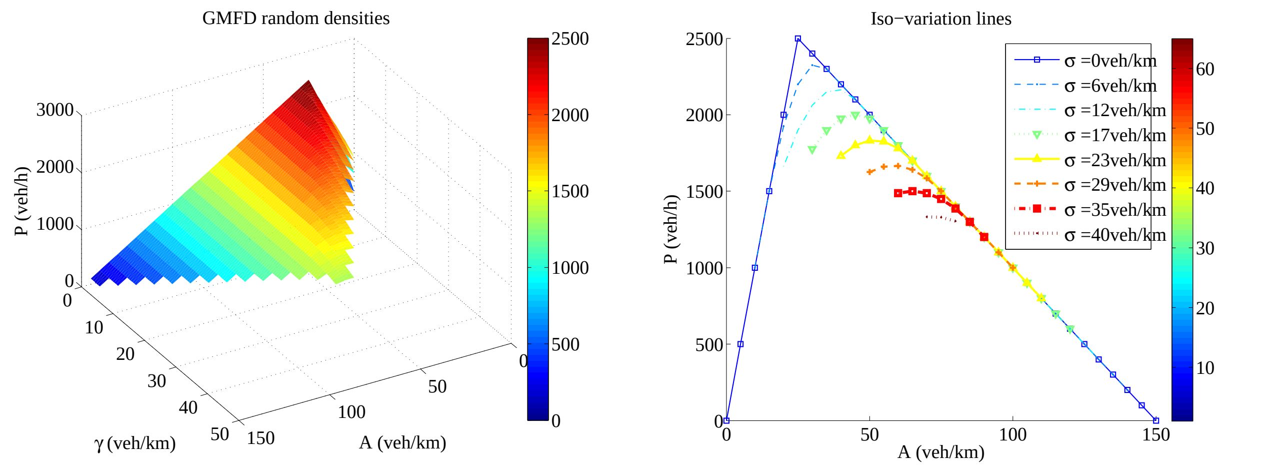
The production is obtained by substituting the fundamental diagram (equation 8) in equation 12. This integral can be studied for three cases: all links are in free flow conditions ($k_{\text{mean}} + p < k_c$), all links are congested ($k_{\text{mean}} - p > k_c$), or a combination of congested and non-congested links. For this case of mixed traffic states, the integral is split over 2 parts, the free part for densities $k = k_{\text{mean}} - p$ to the critical density k_c and the congested part for $k = k_c$ to $k_{\text{mean}} + p$:

$$\begin{aligned} P = \frac{\int_{k_{\text{mean}}-p}^{k_{\text{mean}}+p} q(k)dk}{2p} &= \frac{1}{2p}\left(\int_{k_{\text{mean}}-p}^{k_c} q(k)dk\right) + \left(\int_{k_c}^{k_{\text{mean}}+p} q(k)dk\right) \\ &= \frac{1}{2p}\left(\int_{k_{\text{mean}}-p}^{k_c} \frac{k}{k_c} q_{\text{max}} dk\right) + \left(\int_{k_c}^{k_{\text{mean}}+p} w\left(k - k_{\text{jam}}\right) dk\right) \end{aligned} \tag{13}$$

The integral, and thus the production, then is a straightforward integration:

$$P = \frac{\int_{k_{\text{mean}}-p}^{k_{\text{mean}}+p} q(k)dk}{2p} = \begin{cases} \frac{q_{\text{max}}}{k_c} k_{\text{mean}}, & (k_{\text{mean}} + p) < k_c \\ w\left(k_{\text{mean}} - k_{\text{jam}}\right) & (k_{\text{mean}} - p) > k_c \\ \frac{1}{2p}\left\{\frac{1}{2}\frac{q_{\text{max}}}{k_c}\left(k_c^2 - (k_{\text{mean}} - p)^2\right) + \dots\right. & \\ \left.\frac{1}{2} w\left((k_{\text{mean}} + p)^2 - k_c^2\right) + w k_{\text{jam}}\left((k_{\text{mean}} + p) - k_c\right)\right\} & \text{otherwise} \end{cases} \tag{14}$$

These equations will be used to show the Generalized Fundamental Diagram for a case without traffic dynamics in section 4.2.



(a) The Generalized Macroscopic Fundamental Diagram. To improve visibility of the surface, the graph is rotated such that the axis of density increases from right to left.

(b) Macroscopic fundamental diagrams with constant levels of spatial inhomogeneity of density.

FIGURE 1 The generalized macroscopic fundamental diagram based on independent road stretches

4.2 Numerical results and implications for the GMFD

To show numerical results, use the following values in our experimental setup: number of links N= 100, critical density k_c=25 veh/km/lane, cam density k_{jam} = 150 veh/km/lane, capacity $q_{\max}$ = 2500 veh/h, length of a link L = 1 km.

A grid is made of values for the mean density k_{mean} and the spatial inhomogeneity of density γ. For each combination the resulting production is calculated using equation 14. This is plotted as function of the mean density and the spatial inhomogeneity of density, and thus, a Generalized Macroscopic Fundamental Diagram is constructed, see figure 1a. Also the cross-sections of this Generalized Macroscopic Fundamental Diagram are given in figure 1b. The basic shape of the MFD as found by Geroliminis and Daganzo (*1*) is clearly visible: an increase of production with an increase of the density for values lower than the critical density, and a decrease later on. Also, the effect explained by Cassidy et al. (*4*) is visible. If the the spread of density is 0, all links have the same density, and therefore the average density is equal to the density at any of the links. Consequently, we observe the triangular fundamental diagram which has been put in.

Note too, that for the very high accumulations all links are probably congested, especially for the lower values for the inhomogeneity of density. That means that all links are in the right hand side of the triangular fundamental diagram, and the actual spread inhomogeneity does not matter (see Cassidy et al. (*4*) for a detailed reasoning). Only when free flow states and congested states are mixed, there is a mixed state, leading to a lower production than the triangular MFD, valid for a spatially homogeneous area. Hence, a decrease in the direction of increasing density is most pronounced at accumulation values around capacity, where congested states and non-congested states are equally likely to occur.

5 THE EFFECT OF SPATIAL INHOMOGENEITY OF DENSITY ON NETWORK PERFORMANCE INCLUDING TRAFFIC DYNAMICS

Where the previous section described an average of randomly drawn traffic states, this section will study the MFD with traffic dynamics taken into account. The section first describes what will be simulated in terms of network and demands. Then, section 5.2 describes the macroscopic traffic simulation model used to describe the traffic dynamics. Then the outcomes of the simulations are presented: first, the traffic flow

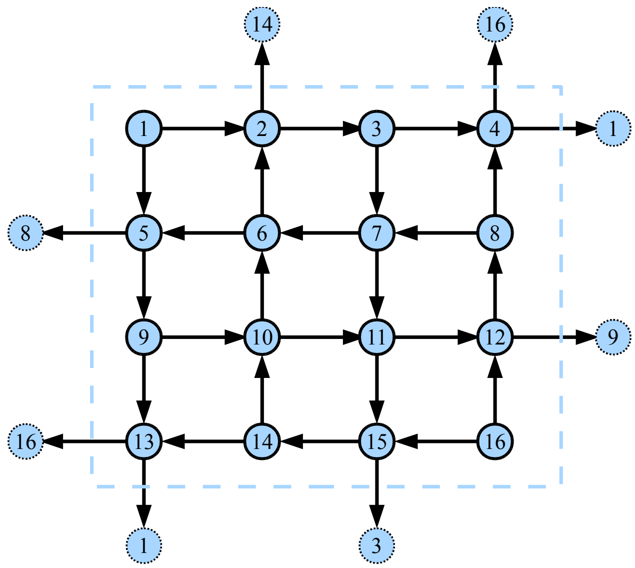


FIGURE 2 Illustration of a 4x4 grid network with periodic boundary conditions

phenomena are qualitatively described (section 5.3). Section 5.4 shows the resulting Generalized Macroscopic Fundamental Diagram, and relates this to the averaging experiment in the previous section. Finally, section 5.5 sheds some light at the relation between production and performance.

5.1 Experimental set-up

In the paper an urban network is simulated, since this is the main area where MFDs have been tested. Inspired by Mazloumian et al. (*12*), we follow Knoop and Hoogendoorn (*13*) and choose a Manhattan-like grid network with square blocks and periodic boundary conditions. This implies that the nodes are located at a regular grid, for which we choose a 20x20 size. Then, one-way links connect these nodes. The direction of the links changes from block to block, i.e. if at $x = 2$ the traffic is allowed to drive in the positive y direction, at $x = 1$ and at $x = 3$ there are one-way roads for traffic to drive in the negative y direction. We assume 2 lanes per link, a 1 km block length, a triangular fundamental diagram with a free speed of 60 km/h, a capacity of 1500 veh/h/lane and a jam density of 150 veh/km/lane.

Furthermore, periodic boundary conditions are used, meaning that a link will not end at the edge of the network. Instead, it will continue over the edge at the other side of the network. An example of such a network is given in figure 2. Traffic can continue in a direct link from node 13 to node 1 or from node 5 to node 8. This way, all nodes have two incoming and to outgoing links and network boundaries have no effect.

The destinations are randomly chosen from all points in the network. In the network, there are 19 nodes chosen as destination nodes. There are no origin nodes. Instead, at the beginning of the simulation, traffic is put on the links. Vehicles are assigned to a destination, and for this distribution is equal over all destinations.

The simulated time is 4.5 hours. During the first three hours of the simulation, the cars that reach their destination will not leave the network, but instead they are assigned a new destination. The arriving vehicles are split equally over the 18 other destinations; this is possible because we use a macroscopic model (see section 5.2). As a result of the new destination labeling, the number of cars in the network is constant for the first three hours and as such can be set as parameter setting in the simulation. This demand level is expressed as the density on all links at the start of the simulation, as fraction of the critical density; this is indicated by l.

The reason for this redestination is that most people will go the busier places (e.g., CBD), and this is also the area where many people leave. At specific periods of the day (morning peak) where this strict balance does not hold, but at other moments of the day (say, around noon), the chosen approach of vehicle regeneration at the same place as where they arrive is a good first approximation.

Figure 3a shows the network used under initial conditions. After three hours, the arriving vehicles will not be reassigned to a new destination, but they will be removed from the simulation. Hence, the number of vehicles in the network will gradually decrease after three hours.

5.2 Traffic flow simulation

This section describes the traffic flow model. For the traffic flow modelling we use a first order traffic model, the Cell Transmission Model (20). Lebacque (*21*) showed that this is basically an analytical, continuum LWR-model proposed by Lighthill and Whitham (*22*) and Richards (*23*) that is discretized in space and time and solved with a Godunov scheme (24).

We choose to split links into cells of 250 meters in length (i.e., 4 cells per link). The flux ϕ over a node, from one link to the next, is basically restricted by either the demand from the upstream node (free flow) or by the supply from the downstream node (congestion):

$$\phi_{x,x+1} = \min\left\{D_x, S_{x+1}\right\}; \tag{15}$$

At a node r we have inlinks, denoted by i which lead the traffic towards node r and outlinks, denoted by j which lead the traffic away from r. At each node r, the demand D to each of the outlinks of the nodes is calculated, and all demand to one link from all inlinks is added. This is compared with the supply S of the cell in the outlink. In case this is insufficient, a factor, α, is calculated which shows which part of the demand can continue.

$$\alpha_r = \underset{[j \text{ leading away from } r]}{\operatorname{argmin}} \left\{\frac{S_j}{D_j}\right\} \tag{16}$$

This is the model developed by Jin and Zhang (*25*). They propose that all demands towards the node are multiplied with the factor α, which gives the flow over the node.

This node model is slightly adapted for the case at hand here. Also the node itself can restrict the capacity. In our case, there are two links with a capacity of 1500 veh/h as inlinks and two links with a capacity of 1500 veh/h as outlinks. Since there are crossing flows, it is not possible to have a flow of 1500 veh/h in one direction *and* a flow of 1500 veh/h in the other direction. To overcome this problem, we introduce a node capacity (see also for instance Tampère et al. (*26*)). The node capacity is the maximum of the capacities of the outgoing links. This means that in our network, at maximum 1500 veh/h can travel over a node. The fraction of the traffic that can continue over node r, indicated by β, is calculated as follows:

$$\beta_r = \frac{C_r}{\sum_{\forall i \text{to } r} D_i} \tag{17}$$

The demand factor ζ is now the minimum of the demand factor calculated by the nodes and the demand factor due to the supply:

$$\zeta = \min\left\{\alpha_r, \beta_r, 1\right\} \tag{18}$$

Similar to Jin and Zhang (*25*), we take this as multiplicative factor for all demands to get to the flux ϕ_{ij}, i.e. the number of cars from one cell to the next over the node:

$$\phi_{ij} = \zeta D_{ij} \tag{19}$$

Note further that once the node capacity is insufficient, upstream of the node congestion will form. Since the cells in the network are small, the ratio between of flows from the inlinks will be divided according to their respective capacity, since the demand in congested conditions is capacity.

The path choice is static, and determined based on distance to the destination. Traffic will take the shortest path towards the destination. For intersections where both directions will give the same path length towards a destination, the split of traffic to that direction is 50-50. We choose this simple routing because a fixed routing allows to pinpoint the problems in the network. With more advanced routing, as for instance equilibrium, drivers can start deviating before congestion starts. For this paper it is important to understand the network effects and the best way to separate them from route choice effects is by keeping one constant.

5.3 Discussion and implications of the results: the nucleation effect

This section first describes the traffic flow over time. Figure 3 shows the outcomes of the simulation, in snapshots of the density and speed over time. At the start of the simulation (see figure 3a), traffic is evenly distributed over all links, since this was the initial situation as it was regulated externally.

When the traffic starts to run, various distributed bottlenecks become active. This is shown in figure 3b. After some time (figure 3d-f), traffic problems concentrate more and more around one location. The number of vehicles in the rest of the network reduces, ensuring free flow conditions there. This complete evolution can be found in figure 3a-f. The network has periodic boundary conditions, which means that the network edges do not have any effect. Any deviations from a symmetry are due to random effects and thus to the location of the destinations, since the traffic simulation is deterministic.

Finally, the situation stabilizes. After 3 hours vehicles which arrive at their destination will be taken from the network, instead of being assigned to another destination as in the beginning. This reduces the number of vehicles in the network. This evolution of the number of vehicles over time is shown in figure 4a. Note that the number decreases after 3 hours.

Also, for each time step the accumulation and the standard deviation of the density over all cells in the simulation can be determined. The evolution of the network in this plane is shown in figure 4b. In general, traffic starts at an accumulation but with no spread in density (i.e., at the right bottom of the figure). Then, traffic becomes less homogeneous, but the accumulation stays constant (in the figure: the line goes up) In fact, it can be seen that as soon as a queue starts, the outflow will be reduced due to the reduced speed. Since the inflow will not change, the queue will grow. So, congestion attracts more congestion. The spatiotemporal dynamics of traffic jams are thus ruled by individual points where congestion starts; this phenomenon we will call the *nucleation effect* of traffic jams due to its similarities in physics in state transitions.

The curve from the top to the lower left corner in figure 4b start when the vehicles are set to reach their destination and hence the accumulation reduces (in the end to zero, the origin in the figure). If the network is relatively empty (undercritical), the largest part of the cells has a low density. Once the traffic jams dissolve, a larger fraction of the cells will be in lower density and the inhomogeneity of density hence decreases. However, if the network is relatively full, most cells will have a high density. Removing vehicles from the simulation will increase the number of cells with a relatively low density, and hence the inhomogeneity of density increases. Consider for example the extreme case that at the beginning all cells are congested. Then, removing some vehicles will give some cells a lower density and hence inhomogeneity of density will increase.

It is interesting to see what the traffic production will do in these situations; this is shown in figure 4c. First the production will decrease with an increasing inhomogeneity in density. Contrary to the randomly drawn densities, due to the nucleation effect, there is an immediate decrease in performance. Congestion will start at nodes, and this will create congested cells and uncongested cells. Hence, two traffic states are simultaneously present and the production is lower than with the same accumulation without inhomogeneity

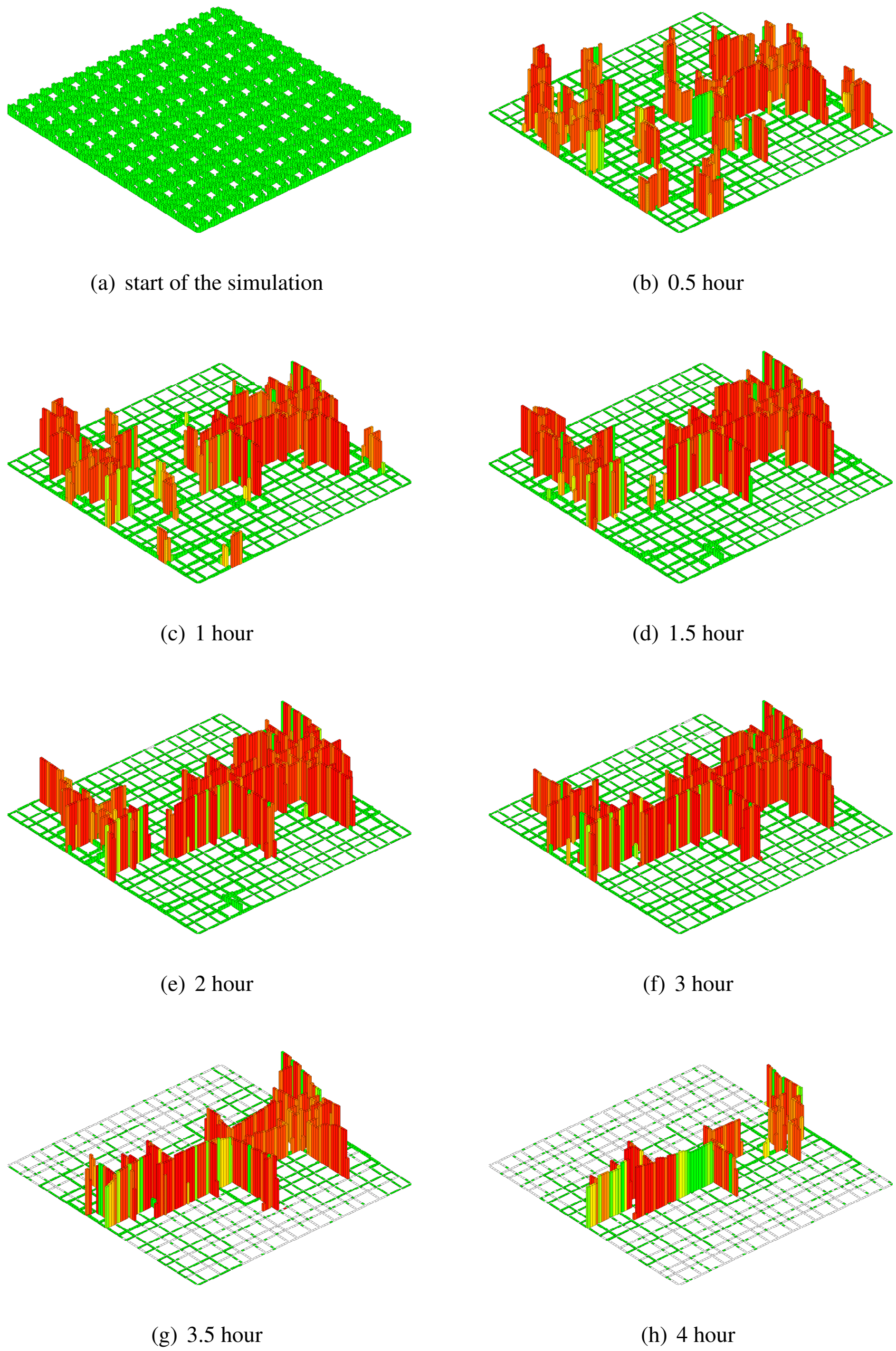
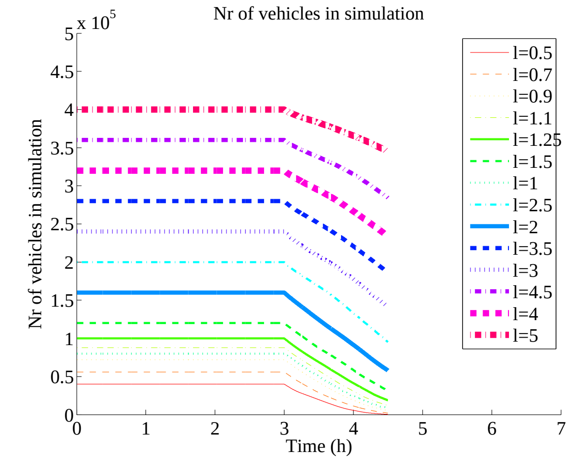
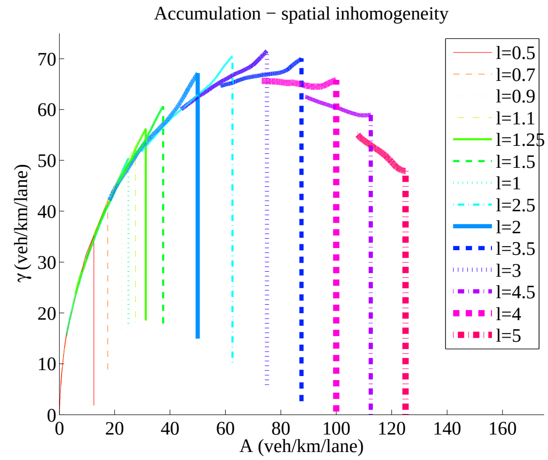


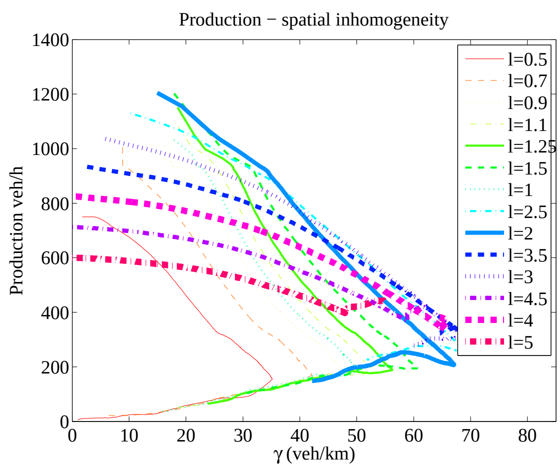
FIGURE 3 Evolution of the densities (bar heights) and speeds (colours) in the network



(a) Evolution of the number of vehicles in the network



(b) Accumulation and the standard deviation of density



(c) Production and the standard deviation of density

FIGURE 4 The evolution of the network characteristics over time

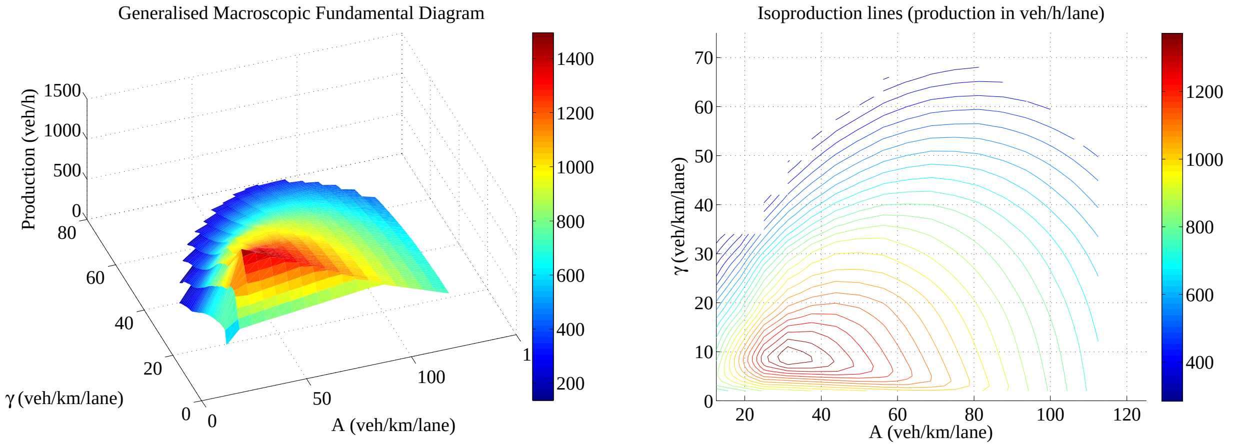
of density (4).

Let's now consider what happens when the traffic accumulation decreases. In undercritical conditions, the production will decrease due to this lower density. However, in overcritical conditions, the production will increase, because less vehicles are blocked.

5.4 Generalized Macroscopic Fundamental Diagram

As shown above, the traffic production varies as function of both the accumulation and the inhomogeneity in density. Figure 5a shows the GMFD. Figure 5b shows the same surface, but now in isoproduction lines. The production decreases once the inhomogeneity increases, and that this holds for every accumulation. As will be discussed in section 6, this is a major difference with the averaging of random traffic states.

The figures show that the maximum production can be achieved at a density of approximately 30 veh/km/lane, and at a standard deviation of density of approximately 10 veh/km/lane. For lower standard deviations of density, the production flattens and reduces. This contradicts with the reasoning in section 4 and with common sense: if there is no spatial spread in density, the production should show the same pattern as the flow in the fundamental diagram, which was chosen triangular in the simulations. Therefore,



(a) The Generalized Macroscopic Fundamental Diagram

(b) Iso-production lines in the accumulation-inhomogeneity of density plane

FIGURE 5 The Generalized Macroscopic Fundamental Diagram

we attribute this flattening is to interpolation effects, as is argued in the next paragraph.

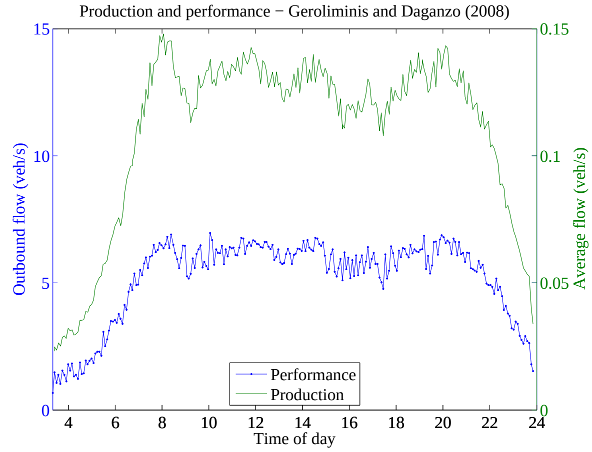
As argued, due to the nucleation effect, as soon as congestion starts, this attracts more congestion. Therefore, the situation with accumulation but no spread in density hardly ever occurs. The situations with zero standard deviation of density which can occur are (1) an empty network with no production and (2) a completely congested network with no production. In the interpolation, any point on the line of $\gamma = 0$ will be influenced by the points with a similar accumulation and a higher standard deviation of density, but also by the two aforementioned points. Furthermore, the lower the standard deviation of density, the higher will be the influence of these two points with no production. We therefore see a reduction of production with decreasing standard deviation of density for the very low values, which is an artefact of the interpolation method.

5.5 Performance versus production

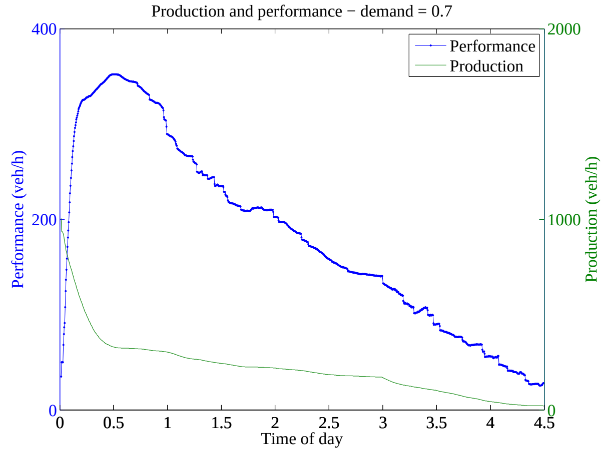
Geroliminis and Daganzo (*1*) indicates that the production, i.e. the average flow, is correlated to the performance, i.e. the exit flow of the network. Figure 6a shows the figure of time series, with data extracted from Geroliminis and Daganzo (*1*). The production and performance show a remarkable correlation.

In our simulations this is different. This mainly has to do with the network transition towards an equilibrium state. At the beginning, all vehicles are spread over the network, regardless of their destination. They then all drive towards their destination, and get closer to it. Hence close to a destination the fraction of vehicles heading to that destination is generally higher. The flow reaching the destination will be considered as new demand for all different destinations. In equilibrium if no traffic with another destination were to pass the considered area, one would close to a destination hence find as much traffic heading towards the destination as from it, where the outbound flow is divided over all 18 other directions. In the beginning, the ratio between flow to a destination and flow to another destination is 1:18 over the whole network. Although it is not true that other traffic will not pass the destination, the example makes clear why in an equilibrium situation one expects more traffic closer to the destination than randomly spread.

For the above reasoning, all traffic situations where the equilibrium had not been reached yet, will distort the correlation between traffic production and performance. Figure 6b shows the time series of the production and performance. It shows that the performance starts low and needs some time to climb to the value.



(a) The San Francisco network – data from Geroliminis and Daganzo (*1*)



(b) The simulation correlations

FIGURE 6 The relation between production and performance

After this initialization phase, equilibrium sets in (at approximately 0,5h after the start). From that moment, performance degrades gradually due to the increasing congestion. More importantly, this performance pattern follows the production pattern from that moment in time. At 3h after the start, the vehicles start to arrive, and the arrival pattern is still remarkably similar to the performance pattern. The difference between the lines in figure 6b at the beginning of the simulation time are hence due to the loading technique.

6 DISCUSSION ON THE EFFECT OF TRAFFIC DYNAMICS

It is interesting to compare the results and the GMFDs for the case with and without traffic dynamics to understand the effect of traffic dynamics. With network dynamics included, the GMFD is more then a simple average of the fundamental diagram because correlation of the densities in an area is inherent to traffic processes. Contrary to the randomly chosen states (section 4), in the simulation the production decreases with the slightest increase of inhomogeneity of density. This is because already at the start, the restricted capacity at the nodes will create congestion in some cells. This significantly reduces the traffic production, and hence the lines in figure 4c are steep. For the high accumulations, the nucleation effect also plays a role. However, an increase of inhomogeneity of density means that traffic moves towards the clusters of congestion. If all cells were in congestion, this new, mixed, traffic state must have a lower performance (4). However, since there was no high performance (no high speeds or capacity flows), the performance reduction is less than in case free flowing traffic changes into standing traffic. Hence, the lines with a downslope in figure 4c with a high traffic demand are less steep than those with a low demand.

The final patterns depend on the exact characteristics of the traffic and the network, including the network type and layout, the position of the destinations, the demand pattern, and the shape of the fundamental diagrams. It is believed that a continuous function also holds for other networks, but that the shape might differ. This can be checked in two ways, which both are suggested as line of further research: (1) A sensitivity study, how the results compare in real life, with heterogeneous drivers and different road types, including different fundamental diagrams; (2) empirical studies showing the GMFD for real networks. The studied the GMFD for a real life urban freeway network (27).

Also, the relation between production and performance only holds in quasi-equilibrium situations.

This equilibrium is defined by the spatial spread of drivers heading for a specific destinations, related to the drivers heading towards other destinations. In particular, during network loading the production and the performance differed. In the studied situation this effect was dominant due to the artificial loading situation and the static route choice. In real life, the situation is not likely to deviate from the equilibrium state as much as in the test case. However, it would be interesting to study the effects of network loading and route choice on the relation between production and performance.

7 CONCLUSIONS AND OUTLOOK

In this manuscript we studied network traffic dynamics, and their effect on the macroscopic fundamental diagram (MFD). In the traffic dynamics, we identified a nucleation effect. Congestion starts in a network at one point, which we call the nucleation point. By definition, traffic moves slower there, and therefore, the flow is reduced, and the traffic jam will grow at the tail. If the tail reaches an intersection, the congestion spreads to different links, thus inducing even more congestion. Congestion thus attracts more congestion. A simulation with periodic boundaries has shown how a network evolves from freely flowing via a state with several local bottlenecks into a final state where all traffic is queuing to pass one bottleneck.

We conclude that a continuous Generalised MFD exists, which expresses the production as function of the accumulation and the spatial inhomogeneity of density. In the analysed case, the total number of travellers in the network remains constant, but the network production, i.e. the average flow, reduces. This observation contradict with the existence of a single macroscopic fundamental diagram which links the production in a network with the accumulation of vehicles in the network. This paper proposed a Generalized Macroscopic Fundamental Diagram (GMFD), a two dimensional function giving the production in the network as function of the accumulation in the network and the spatial inhomogeneity of density over the network, in this case expressed as the standard deviation of the densities in different cells in the network. The production is shown to be a continuous function of accumulation and inhomogeneity of density, for the simulation as studied in the paper.

It is also found that due to nucleation the GMFD is different from a GMFD created by averaging randomly chosen traffic states. In the latter one, no correlation between traffic states is present, and the production is, only decreasing with large inhomogeneity or near capacity. Due to the spatial correlation of congestion and the above-mentioned nucleation effect, inhomogeneity in densities almost always are caused by some parts of the network being congested and others being not congested. Then, the production is less than with the same accumulation, but without spread. Therefore, in the GMFD the production decreases as a function of the inhomogeneity in density.

The decrease of traffic performance with the increase of inhomogeneity of density, and hence the GMFD, is also found in a continuous macroscopic traffic representation. This means that the effect is not due to individual traffic movements, or gaps, but the nucleation effect, and hence the asymmetric loading of the network is the main cause. This study furthermore showed that the production decreases as function of the spread of density, and that the causes thereof does not lie in the traffic light settings or microscopic merging, since these are not explicitly included in the simulation used.

The understanding of traffic dynamics on a large scale can be utilized to improve the traffic situation. It is well possible that traffic control, for instance traffic light settings or routing, can not only change the independent variables, i.e., the accumulation and the inhomogeneity of density, but possibly it can also change the shape of the GMFD. For instance, production at the same accumulation and the same inhomogeneity of density might be higher if all travelers are guided towards free flow traffic conditions. This is an example of optimizing within a (sub)network. Alternatively, it can be used to optimize traffic guidance through various subnetworks. In previous work we showed that guidance of travelers over subnetworks with low accumulation improved the traffic state (*7*). This paper suggests that the routing algorithm can be improved by not only using the accumulation but also the spatial spread of density. Preliminary results look promising (*28*), and further improvements on the details of the algorithm are subject for further research.

Acknowledgement This research was sponsored by a ICT Innovation Platform Cooperation Challange subsidy from ICTregie/NWO in the project *SI4MS*, Sensor Intelligence for Mobility Systems, and by the foundation Next Generation Infrastructures

REFERENCES

[1] Geroliminis, N. and C. F. Daganzo, Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transportation Research Part B: Methodological*, Vol. 42, No. 9, 2008, pp. 759–770.

[2] Buisson, C. and C. Ladier, Exploring the Impact of the Homogeneity of Traffic Measurements on the Existance of the Macroscopic Fundamental Diagram. *Tranportation Research Records, Journal of the Transportation Research Board*, 2009.

[3] Geroliminis, N. and Y. Ji, Spatial and Temporal Analysis of Congestion in Urban Transportation Networks. In *Proceedings of 90th Annual Meeting of the Transportation Research Board*, 2011.

[4] Cassidy, M., K. Jang, and C. Daganzo, Macroscopic Fundamental Diagram for Freeway Networks: Theory and Observation. In *Proceedings of the 90th Annual Meeting of the Transportation Research Board*, 2011.

[5] Gayah, V. and C. Daganzo, Clockwise hysteresis loops in the macroscopic fundamental diagram: An effect of network instability. *Transportation Research Part B*, 2011.

[6] Saberi, M. and H. S. Mahmassani, Exploring the Properties of Network-wide Flow-Density Relations in a Freeway Network. *Tranportation Research Records, Journal of the Transportation Research Board*, Vol. in print, 2012.

[7] Knoop, V., J. Van Lint, and S. Hoogendoorn, Route Advice and its Effect on the Marcoscopic Fundamental Diagram. *Transportation Research Records*, in print.

[8] Godfrey, J., The mechanism of a road network. *Traffic Engineering and Control*, Vol. 11, No. 7, 1969, pp. 323–327.

[9] Daganzo, C., Urban gridlock: Macroscopic modeling and mitigation approaches. *Transportation Research Part B: Methodological*, Vol. 41, No. 1, 2007, pp. 49–62.

[10] Daganzo, C. and N. Geroliminis, An analytical approximation for the macroscopic fundamental diagram of urban traffic. *Transportation Research Part B: Methodological*, Vol. 42, No. 9, 2008, pp. 771 – 781.

[11] Ji, Y., W. Daamen, S. Hoogendoorn, S. Hoogendoorn-Lanser, and X. Qian, Macroscopic fundamental diagram: investigating its shape using simulation data. *Transportation Research Record, Journal of the Transporation Research Board*, Vol. 2161, 2010, pp. 42–48.

[12] Mazloumian, A., N. Geroliminis, and D. Helbing, The spatial variability of vehicle densities as determinant of urban network capacity. *Philosophical Transactions of the Royal Society A*, Vol. 368, 2010, pp. 4627–4647.

[13] Knoop, V. L. and S. P. Hoogendoorn, *Proceedings of Traffic and Granular Flow 2011*, chap. Two-Variable Macropscopic Fundamental Diagrams for Traffic Networks, in print.

[14] Wu, X., H. Liu, and N. Geroliminis, An empirical analysis on the arterial fundamental diagram. *Transportation Research Part B: Methodological*, Vol. 45, No. 1, 2011, pp. 255 – 266.

[15] Daganzo, C., V. Gayah, and E. Gonzales, Macroscopic relations of urban traffic variables: Bifurcations, multivaluedness and instability. *Transportation Research Part B: Methodological*, Vol. 45, No. 1, 2011, pp. 278–288.

[16] Geroliminis, N. and J. Sun, Properties of a well-defined macroscopic fundamental diagram for urban traffic. *Transportation Research Part B: Methodological*, Vol. 45, No. 3, 2011, pp. 605 – 617.

[17] Hall, F. L. and K. Agyemang-Duah, Freeway Capacity Drop and the Definition of Capacity. *Transportation Research Record: Journal of the Transportation Research Board No.1320*, 1991, pp. 91–98.

[18] Cassidy, M. and R. Bertini, Some traffic features at freeway bottlenecks. *Transportation Research Part B: Methodological*, Vol. 33, No. 1, 1999, pp. 25 – 42.

[19] Daganzo, C. F., *Fundamentals of Transportation and Traffic Operations*. Pergamon, 1997.

[20] Daganzo, C. F., The Cell Transmission Model: a Dynamic Representation of Highway Traffic Consistent With the Hydrodynamic Theory. *Transportation research part B*, Vol. 28B, No. 4, 1994, pp. 269–287.

[21] Lebacque, J. P., The Godunov Scheme and What it Means for First Order Traffic Flow Models. In *Proceedings of the 13th International Symposium on Transportation and Traffic Theory*, 1996.

[22] Lighthill, M. J. and G. B. Whitham, On Kinematic Waves. II. A Theory of Traffic Flow on Long Crowded Roads,. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, Vol. 229, No. 1178, 1955, pp. 317 – 345.

[23] Richards, P. I., Shock waves on the highway. *Operations Research 4*, Vol. 4, 1956, pp. 42 – 51.

[24] Godunov, S. K., A difference scheme for numerical computation of of discontinuous solutions of equations of fluid dynamics. *Math. Sb.*, Vol. 47, 1959, pp. 271 – 290.

[25] Jin, W. L. and H. M. Zhang, On the distribution schemes for determining flows through a merge. *Transportation Research Part B: Methodological*, Vol. 37, No. 6, 2003, pp. 521–540.

[26] Tampère, C., R. Corthout, D. Cattrysse, and L. Immers, A generic class of first order node models for dynamic macroscopic simulation of traffic flows. *Transportation Research Part B*, Vol. 45, 2011, pp. 289–309.

[27] Knoop, V. and S. Hoogendoorn, Empirics of a Generalised Macroscopic Fundamental Diagram for Urban Freeways. In *Proceedings of the 92nd Annual Meeting of the Transportation Research Board*, Washington D.C., 2013.

[28] Knoop, V., J. Van Lint, and H. S. P., Traffic Network Guidance using Area Accumulation and Spatial Variation in Density. In *Proceedings of the 1st European Symposium on Quantitative Methods in Transportation*, Lausanne (Switzerland), 2012.