

Delft University of Technology

Traffic dynamics Its impact on the macroscopic fundamental diagram

Knoop, VL; van Lint, JWC; Hoogendoorn, SP

DOI 10.1016/j.physa.2015.06.016

Publication date 2015 Document Version

Accepted author manuscript **Published in** Physica A: Statistical Mechanics and its Applications

Citation (APA)

Knoop, VL., van Lint, JWC., & Hoogendoorn, SP. (2015). Traffic dynamics: Its impact on the macroscopic fundamental diagram. *Physica A: Statistical Mechanics and its Applications*, *438*(November), 236-250. https://doi.org/10.1016/j.physa.2015.06.016

Important note

To cite this publication, please use the final published version (if applicable). Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

Traffic Dynamics: its impact on the Macropscopic Fundamental Diagram

Victor L. Knoop, Hans van Lint, Serge P. Hoogendoorn

Delft University of Technology, Transport & Planning, Stevinweg 1, 2628 CN Delft, the Netherlands, v.l.knoop@tudelft.nl

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, of which the MFD is a projection. Using the GMFD, we can describe the mentioned hysteresis patterns in the MFD. 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.

Keywords: Macropscopic Fundamental Diagram, Network Fundamental Diagram, Traffic inhomogeneity, Traffic dynamics

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.

The concept of describing the traffic dynamics for a zone, consisting of several roads is appealing and possibly a very powerful tool. The idea that the speed of the vehicles depends on the number of vehicles dates back to the sixties [1]. Similar findings are presented later [2, 3], and the research gained momentum after Daganzo reintroduced the concept [4] and Geroliminis and Daganzo showed the application using real world data [5]. They call it the Macrosopic Fundamental Diagram (MFD), and show that 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 after [5] (e.g., [6, 7]) suggest that such a crisp relationship exists only for homogeneously loaded networks.

The papers handling the inhomogeneity usually do not comment on the cause of the inhomogeneity, which lies in the traffic dynamics in networks. In fact, traffic dynamics in congested networks inherently leads to traffic inhomogeneity, because, under increasing traffic volume loaded onto the network at some point congestion will set in at one of the (potentially many) bottlenecks in a network. This will lead to an inhomogeneous distribution of vehicular density in the network, with congestion in one part of the network and still freely flowing traffic in the remaining parts.

This process influences the network performance for the same accumulation. To still have a description of the traffic performance in a zone, we propose a generalization of the MFD, the Generalized Fundamental Diagram (GMFD), which describes the traffic production as a continuous function of both the average density (accumulation) and the spatial spread 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, as a function of vehicle accumulation only, is a one-dimensional projection of this plane.

The use of this paper is twofold. At one hand, its 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 traffic control concepts, based on the MFD, e.g. [8, 9, 10, 11].

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. The macroscopic fundamental diagram and inhomogeneity

Original ideas on the MFD[1, 3] described that traffic speeds decrease with increasing accumulation. The recent increase of attention came after Daganzo [4] reintroduced the concept and showed how the production could decrease if accumulation increases. The name Macroscopic Fundamental Diagram has been proposed by Geroliminis and Daganzo [5]. In their paper they showed, based on data, that the traffic flow decreases after the accumulation increased over a critical threshold. These papers give a snapshot situation of traffic and do not account for the dynamics or heterogeneity.

Heterogeneity is explicitly studied by Buisson and Ladier [6], creating the MFD in heterogeneous situations. They found, using data from three different days in the French city of Toulouse, that the MFD does not hold in heterogeneous conditions.

The explanation for the reducing production is also described. First, the effect of inhomogeneity is further discussed by Mazloumian et al. [12] and Geroliminis and Ji [7]. 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 decrease the network performance, 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 then confirmed by an empirical analysis [7], 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 [13] showing that if clustered in bins with similar variation of density, the MFD is a well defined curve. Daganzo et al. [14] 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.

A theoretical explanation for the phenomenon of the influence of the spatial inhomogeneity of density on the accumulation is given by Daganzo et al. [14]. They shows that turning at intersections is the key reason for the drop in production under spatially inhomogeneous conditions. Gayah and Daganzo [15] 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 traffic production found at the same accumulation for more homogeneous conditions, and in the accumulation-production plot this traffic state must thus be located graphically located under the MFD. This way, there are hysteresis loops in the MFD. Note that hysteresis in the MFD is a result of macroscopic queueing and spillback processes. Saberi et al. [16] discuss this in more detail, how the averaging of traffic states leads to a lower prediction. Furthermore, this is confirmed with real data. Mahmassani et al. [17] study the MFD for the Chicago city business district in simulation and finds strong hysteresis effects. Those are assigned to the loading pattern, which is in line with the studies by Gayah and Daganzo [15]. Zhang et al [18]

also discuss traffic dynamics in a network. Their focus is on the effect of traffic signals, and the hysterisis which is related to the adaptiveness of the signals. A more theoretical and fundamental approach of the network settings, showing the way traffic states can evolve for two link networks, is given by Gayah et al. [19]. They also comment on the way adaptive signal settings can be used (up to a certain density) to avoid the clustering of traffic.

Since then, the hysteresis patterns are observed more often, but no attention is given to the physical traffic flow phenomena causing it. Instead, modelling approaches have been proposed with links to control [20]. Our article continues on tracking the dynamics of the aggregated traffic states. Contrary to the other approaches, we use a macroscopic approach, and see the whether the effects can be reproduced if we include traffic dynamics but not individual vehicles. This way, it gives an explanation to the hysteresis based on the heterogeneity in the traffic conditions caused by traffic dynamics. 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 the spatial inhomogeneity of density.

3. Definitions and variables

In this section we define the variables used in this paper, which is summarized in table 1.

Standard traffic flow variables are flow, q, defined as the vehicle distance covered in a unit of time, and density, k, defined as 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:

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

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

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

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. The rate at which drivers arrive at their destination is called performance, and is indicated by R. Earlier, Geroliminis and Daganzo [5] showed that there is a strong correlation between production and performance.

This article also studies the inhomogeneity of density expressed by the standard deviation of the cell density. This is calculated by considering all cell densities in an area at one moment in time, and determine the standard deviation of these numbers. We hence define the inhomogeneity, γ , by:

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

Symbol	Table 1: The variables used
	Nodo
T	Note
x	Cell in the discretised traffic now 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
kj	Jam density
ϕ_{ij}	Flux from link <i>i</i> to link <i>j</i>
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 accumu-
	lation
σ	Standard deviation

Table 1: The variables used

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

The MFD links the accumulation A to a production (P): $P_{\text{MFD}} = P_{\text{MFD}}(A)$. The generalised macroscopic fundamental diagram, as will be proposed in this paper, describes production as function of accumulation and inhomogeneity of density: $P_{\text{GMFD}} = P_{\text{GMFD}}(A, \gamma)$. These relationships are being tested in the remainder of the paper.

4. The effect of spatial inhomogeneity of density on network performance: no dynamics

In section 3 we have defined the MFD as a function that maps the accumulation, i.e. the weighed vehicular density, to the production, i.e. the weighted average flow in a network. This section will show what the effect is of a simple averaging, without taking typical traffic phenomena and dynamics into account. Saberi et al [16] describe the the time-evolution of two links described analytically. This paper adds in the sense that a stochastic analysis is applied on many links. 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 paragraph we consider the general case, and later we will present a specific case. In general terms, we can consider a fundamental diagram $q = q^*(k)$, and a distribution of densities over all links indicated by the probability density function $\mathcal{P}(k)$. In this continuous case, the production can then be calculated by determining the expectation value of the flow, in which q^* indicates the flow at a certain density according to the fundamental diagram:

$$P = \int_0^{kj} \left(\mathcal{P}(k) q^*(k) \right) dk \tag{4}$$

The spatial inhomogeneity can be calculated from the standard deviation of the probability density function

$$\gamma = \sqrt{\int_0^{kj} \left(\mathcal{P}(k) \left(q^*(k) - P \right)^2 \right) dk}$$
(5)

To evaluate these values, we need to specify the fundamental diagram and the probability density function. The specific case we will consider is a uniform distribution (denoted U). This is chosen because it is the most simple distribution, and it is possible to analytically solve the problem. The densities of the cells are bound to a minimum $k_{\text{mean}} - p$ and maximum $k_{\text{mean}} + p$, with $p \ge 0$ a variable indicating the spread of the density:

$$\mathcal{K} \sim U\left(k_{\text{mean}} - p, k_{\text{mean}} + p\right) \tag{6}$$

We will show later on (section 5) a more realistic approach. The goal of this paper is to show the effects of network dynamics on the production, of which we will show that is due to the distribution of densities in the network. Section 6 discusses the differences between the two approaches (with and without dynamics) and thereby verifies the assumption of this uniform distribution.

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

$$k_{\text{mean}} - p \ge 0 \tag{7}$$

$$k_{\text{mean}} + p \leq k_{\text{jam}}$$
 (8)

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

$$0 \le p \le \min\{k_{\text{mean}}, k_{\text{jam}} - k_{\text{mean}}\}\tag{9}$$

Next, we assume that all N links have the same triangular fundamental diagram [21], which maps the density uniquely to the flow:

$$q = \begin{cases} \frac{k}{k_c} q_{\max} & \text{if } k \le k_c \\ \left(1 - \left(\frac{k - k_c}{k_{jam} - k_c}\right)\right) q_{\max} = w \left(k - k_{jam}\right) & \text{otherwise} \end{cases}$$
(10)

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

$$w = \frac{q_{\max}}{k_c - k_{jam}} \tag{11}$$

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:

$$A = \overline{k} = k_{\text{mean}} \tag{12}$$

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

$$\gamma = \sigma(k) = \frac{1}{\sqrt{3}}p\tag{13}$$

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:

$$P = \overline{q(k)} = \frac{\int_{k_{\text{mean}}-p}^{k_{\text{mean}}-p} q(k)dk}{2p}$$
(14)

The production is obtained by substituting the fundamental diagram (equation 10) in equation 14. 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 $k_{\text{mean}} + p$:

$$P = \frac{\int_{k_{\text{mean}}-p}^{k_{\text{mean}}-p} q(k)dk}{2p} = \frac{1}{2p} \left(\int_{k_{\text{mean}}-p}^{k_c} q(k)dk \right) + \left(\int_{k_c}^{k_{\text{mean}}+p} q(k)dk \right)$$
$$= \frac{1}{2p} \left(\int_{k_{\text{mean}}-p}^{k_c} \frac{k}{k_c} q_{\text{max}}dk \right) + \left(\int_{k_c}^{k_{\text{mean}}+p} w \left(k - k_{\text{jam}}\right)dk \right)$$
(15)

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

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

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

4.2. Numerical results and implications for the GMFD

To show numerical results, use the following setup, which match the real network we will present later in section 5.1. We choose the number of links N=100, critical density $k_c=25$ veh/km/lane, cam density $k_{iam} = 150$ 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 16. 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 1(a). Also the cross-sections of this Generalized Macroscopic Fundamental Diagram are given in figure 1(b). We observe the following. First, the basic shape the MFD as found in [5] is clearly visible: an increase of production with



To improve visibility of the surface, the graph is rotated els of spatial inhomogeneity of density. such that the axis of density increases from right to left.

(a) The Generalized Macroscopic Fundamental Diagram. (b) Macroscopic fundamental diagrams with constant lev-

Figure 1: The generalized macroscopic fundamental diagram based on independent road stretches

an increase of the density for values lower than the critical density, and a decrease later on. Second, the effect that a mixture of congested and uncongested traffic states gives a lower production than a state with the same production but all links are either uncongested or congested. This effect has also been presented by [22]. 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.

Analysing the results gives the insight that that for the very high accumulations all links are probably congested, especially for the lower values for the inhomogeneity of density. As consequence, all links are in a state at the right hand side of the triangular fundamental diagram, and the actual spread inhomogeneity does not matter (the effect from [22] as summarized above). Only when states from free flow states and congestion are mixed, there is a mixed state. This leads to a state with a lower production than the fundamental diagram would predict for the same density. A MFD is hence only is valid for a spatially homogeneous area, i.e. if all link densities are equal, or at least are in the same branch (which needs to be a straight line). The consequence is that the 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



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

simulation model used to describe the traffic dynamics. Then the outcomes of the simulations are presented: first, the traffic flow phenomena are qualitatively described (section 5.3). Section 5.5 sheds some light at the relation between production and performance. Finally, section 5.4 shows the resulting Generalized Macroscopic Fundamental Diagram and comments on its shape based on the traffic patterns.

5.1. Experimental set-up

In this paper an urban network is simulated, since this is the main area where MFDs have been tested. We choose is the simplest network for which the principle of network effects are visible, of which the remainder of the paper shows the importance. Inspired by [12], we follow [23] and choose a square grid network with one way roads and a periodic boundary conditions. This implies that the nodes are located at a regular grid with periodic boundary conditions. This means 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.

Then, one-way links connect nodes in the grid network. 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. This could be any grid size; for instance, [24] choose a similar approach with a 1x1 grid. This is too small for this study to aims to find out the dynamics of congestion over the network. The network should be large enough that congestion can propagate without immediately overlapping with other areas of congestion. For reasons of symmetry, we prefer an even number of roads in order not to break symmetry at the boundary. In this case, we choose 20x20 size. The links have 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.

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. Symmetry in the network is broken by selecting the destination nodes. If all nodes were destination nodes, and all traffic was split over all destinations equally, all turning movements at all intersections would be equal (due to the symmetry in the network) and no links would become congested. At the other hand, a single destination would lead to all traffic going there, leading to a centralised congestion. A realistic number of destinations should hence be in between. We choose to have 19 destinations (1 less than the number of rows/columns, to ensure symmetry is broken), which are randomly chosen.

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 [25]. Lebacque [26] showed that this is basically an analytical, continuum LWR-model proposed earlier [27, 28] that is discretized in space and time and solved with a Godunov scheme [29].

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):

$$\phi_{x,x+1} = \min\{D_x, S_{x+1}\};$$
(16)

At the nodes, we do not choose for a detailed traffic light setting, because the traffic phases are in the order of the model time step. Instead, we have an aggregated model which accounts for the restricted capacity at the nodes. That is, for a node r there could be sufficient capacity out of the node, but an extra restriction is set to the total flow over the node. In practice, green times can be distributed over different phases ("directions"). A good setting for adaptive signalling would be to have the green time, hence flow, per direction proportional to the demand. Exactly these flows are set in the macroscopic model, as is explained below. How other signal settings work out in a MFD is shown by Zhang et al. [18], and this can even be actively influenced (see [20]).

Computationally, the node model works as follows. At each node r there are 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.

$$\alpha_r = \operatorname*{argmin}_{[j \text{ leading away from } r]} \left\{ \frac{S_j}{D_j} \right\}$$
(17)

This model formulation [30] proposes 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 [31]). 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:

$$\beta_r = \frac{C_r}{\sum_{\forall i \text{to } r} D_i} \tag{18}$$

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

$$\zeta = \min\left\{\alpha_r, \beta_r, 1\right\} \tag{19}$$

Similar to [30], 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:

$$\phi_{ij} = \zeta D_{ij} \tag{20}$$

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 4(a). 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 4(b). 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 4(b) 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), most cells have 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 4(c). 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 of density, as shown in [22].

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,



(a) start of the simulation





(c) 1 hour



(d) 1.5 hour



(e) 2 hour



(f) 3 hour



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 – time evolution anticlockwise



(c) Production and the standard deviation of density – time evolution clockwise

Figure 4: The evolution of the network characteristics over time

Evolution of density distribution -- demand 0.7



Figure 5: Evolution of the distribution of densities

the production will increase, because less vehicles are blocked.

Section 4 showed the computation of the production as function of the densities, assuming a uniform distribution. Figure 5 shows the evolution over time of the distribution of densities for the simulated network, and hence with a realistic density distribution. At the beginning all densities have the same value. It then gradually changes into a bimodal distribution, with most cells being empty, and some cells having densities close to the jam density. If the same simulations are done with a higher initial density, the fraction of cells ending up in jam conditions is larger. In the graph, times are aggregated in bins of 20 minutes, therefore even in the first row there is some spread in density.

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 6(a) shows the GMFD. Figure 6(b) shows the same surface, but now in isoproduction lines. The production decreases once the inhomogeneity increases, and that this holds for every accumulation. Especially in congestion the effect of the inhomogeneity is remarkable, since for high accumulations, the production directly decreases with a slight inhomogeneity. As will be discussed in section 6, this is a major difference with the averaging of random traffic states.

The figure shows the results from the simulation. For the no inhomogeneity case ($\gamma = 0$), the fundamentantal diagram is used since this situation is only theoretically conceivable, and does not occur in practice. 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 does not occur. These theoretical states, completely homogeneous, are added, and the values for the production are taken from the fundamental diagram. This can be done, because all roads have the same characteristics and a homogeneous traffic state means the same flow, hence the same production.



(a) The Generalized Macroscopic Fundamental Diagram (b) Iso-production lines in the accumulationinhomogeneity of density plane

Figure 6: The Generalized Macroscopic Fundamental Diagram

The distribution density as plotted in figure 5 can also be used to compute the accumulation and the spread following equations 4 and 5 respectively. The equations needs to be changed to a discrete probability for each density. This will return the equation 2 (for cell lengths L_x the same for all cells). Similarly, equation 3 is a discrete version of equation 5. Using these equations on the density distribution set will yield the production and spread in density.

5.5. Performance versus production

Geroliminis and Daganzo [5] indicate that the production, i.e. the average flow, is correlated to the performance, i.e. the exit flow of the network. Figure 7(a) shows the figure of time series, with data extracted from [5]. 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 7(b) 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.



Figure 7: 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 7(b) at the beginning of the simulation time are hence due to the loading technique.

6. Discussion of the effect of traffic dynamics on the MFD

Let us first discuss the network effects. 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 4(c) 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 [22]. 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 4(c) with a high traffic demand are less steep than those with a low demand.

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,



Figure 8: Evolution of traffic states for the Amsterdam ring road

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.

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 the network layout might cause the fundamental diagram to be different [32].

This paper shows that the production can be expressed as a continuous function of accumulation and inhomogeneity: GMFD, $P_{\text{GMFD}}(A, \gamma)$. If, for the sake of argument, the inhomomogeneity can be considered a function of the accumulation, $\gamma = \gamma(A)$, then the MFD would be a cross section of the GMFD. In equations, the production according to the MFD, $P_{\text{MFD}}(A)$, can be expressed as a value from the production according to the GMFD, $P_{\text{GMFD}}(A, \gamma)$.

$$P_{\rm MFD}(A) = P_{\rm GMFD}(A, \gamma(A)) \tag{21}$$

In practice, there is not this strict relationship, which is shown by the results of this paper. In particular, we observe that at the start of a peak hour the traffic is dense and it will breakdown when links are heavily used. Then, traffic will cluster in the queues upstream of the bottleneck, leading a higher standard deviation of density. At the end of the peak hour, the demand reduces, but the queues remain and will solve from one end. Since the density in the queues will remain more or less stable, this means that the standard deviation of density increases. Hence, for the same accumulation (as before the breakdown and clustering), the standard deviation of density is higher, and thus, according to the GMFD, the production is lower. This is the hysteresis pattern that is found in many studies.

In the simulation, the accumulation is constant until the demand drops after 3 hours. For a real life situation, we expect the increase of the demand creating a bottleneck and increasing at the same time the inhomogeneity. Urban traffic data was not available, but we can re-use the data from an urban freeway we used earlier, see [33], also for the description of the data. Figure 8 presents the data of the A10 ring road of Amsterdam, between km 10 and 32 (clockwise) on 13 October 2011. Just like in the simulated data, the decrease of performance is caused by the increase of the standard deviation of density, at the same accumulation. If we look into detail of the density distribution of the sections at the road, we see indeed the increase in standard deviation from approximately 8 am, leading to some very congested road sections, whereas the number of road sections which is loaded at a medium level is decreasing. This confirms that the homogeneous loads used in section 4 are not realistic.

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 this can be described using aggregate traffic dynamics.

A study of the real life data confirms that the hysteresis loops as found in data are coinciding with a change in the standard deviation of density. These can be explained from a traffic flow theory standpoint.

The understanding of traffic dynamics on a large scale can be utilized to improve the traffic situation. Earlier it has been shown how 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 also can change the shape of the GMFD [18]. 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 [8]. This paper suggests that the routing algorithm can be improved by not only using the accumulation but also the spatial spread of density. Moreover, preliminary results show that a change of inhomogeneity can indicate a change in traffic state earlier than a change in accumulation [34]. The response in traffic control can improve using these concepts.

Acknowledgement

This research is supported by the grant of NWO "there is plenty of room in the other lane". The authors are grateful for the reviewer comments which improved the paper substantially.

References

- [1] J. Godfrey, The mechanism of a road network, Traffic Engineering and Control 11 (7) (1969) 323–327.
- [2] R. Herman, I. Prigogine, A two-fluid approach to town trafc, Science 204 (1979) 148–151.
- [3] H. Mahmassani, J. C. Williams, R. Herman, Investigation of network-level traffic flow relationships: some simulation results, Transportation Research Record 971 (1984) 121–130.
- [4] C. Daganzo, Urban gridlock: Macroscopic modeling and mitigation approaches, Transportation Research Part B: Methodological 41 (1) (2007) 49–62.
- [5] N. Geroliminis, C. F. Daganzo, Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings, Transportation Research Part B: Methodological 42 (9) (2008) 759–770.
- [6] C. Buisson, 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.
- [7] N. Geroliminis, Y. Ji, Spatial and temporal analysis of congestion in urban transportation networks, in: Proceedings of 90th Annual Meeting of the Transportation Research Board, 2011.
- [8] V. L. Knoop, J. W. C. Van Lint, S. P. Hoogendoorn, Route advice and its effect on the marcoscopic fundamental diagram, Transportation Research Records 2315 (2012) 1–10.
- [9] N. Geroliminis, J. Haddad, M. Ramezani, Optimal perimeter control for two urban regions with macroscopic fundamental diagrams: A model predictive approach, Intelligent Transportation Systems, IEEE Transactions on 14 (1) (2012) 348–359. doi:10.1109/TITS.2012.2216877.
- [10] M. Keyvan-Ekbatani, A. Kouvelas, I. Papamichail, M. Papageorgiou, Exploiting the fundamental diagram of urban networks for feedback-based gating, Transportation Research Part B: Methodological 46 (10) (2012) 1393–1403.

- [11] K. Aboudolas, N. Geroliminis, Perimeter and boundary flow control in multi-reservoir heterogeneous networks, Transportation Research Part B: Methodological 55 (2013) 265–281.
- [12] A. Mazloumian, N. Geroliminis, D. Helbing, The spatial variability of vehicle densities as determinant of urban network capacity, Philosophical Transactions of the Royal Society A 368 (2010) 4627–4647.
- [13] N. Geroliminis, J. Sun, Properties of a well-defined macroscopic fundamental diagram for urban traffic, Transportation Research Part B: Methodological 45 (3) (2011) 605 – 617.
- [14] C. Daganzo, V. Gayah, E. Gonzales, Macroscopic relations of urban traffic variables: Bifurcations, multivaluedness and instability, Transportation Research Part B: Methodological 45 (1) (2011) 278–288.
- [15] V. V. Gayah, C. F. Daganzo, Clockwise hysteresis loops in the macroscopic fundamental diagram: An effect of network instability, Transportation Research Part B: Methodological 45 (4) (2011) 643 – 655.
- [16] M. Saberi, 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 2315 (2012) 153163.
- [17] M. S. Hani Mahmassani, A. Zockaie, Urban network gridlock: Characteristics, dynamics and control, in: Proceedings of the 20th International Symposium of Transporation and Traffic Theory, 2013.
- [18] L. Zhang, T. M. Garoni, J. de Gier, A comparative study of macroscopic fundamental diagrams of arterial road networks governed by adaptive traffic signal systems, Transportation Research Part B: Methodological 49 (2013) 1–23.
- [19] V. V. Gayah, X. S. Gao, A. S. Nagle, On the impacts of locally adaptive signal control on urban network stability and the macroscopic fundamental diagram, Transportation Research Part B: Methodological 70 (2014) 255–268.
- [20] M. Ramezani, J. Haddad, N. Geroliminis, Integrating the dynamics of heterogeneity in aggregated network modeling and control, in: Transportation Research Board 93rd Annual Meeting, no. 14-0710, 2014.
- [21] C. F. Daganzo, Fundamentals of Transportation and Traffic Operations, Pergamon, 1997.
- [22] M. Cassidy, K. Jang, C. Daganzo, Macroscopic fundamental diagram for freeway networks: Theory and observation, in: Proceedings of the 90th Annual Meeting of the Transportation Research Board, 2011.
- [23] V. L. Knoop, S. P. Hoogendoorn, Two-Variable Macroscopic Fundamental Diagrams for Traffic Networks, Springer, 2013, pp. 351–360.

- [24] V. V. Gayah, C. F. Daganzo, Effects of turning maneuvers and route choice on a simple network, Transportation Research Record: Journal of the Transportation Research Board 2249 (1) (2011) 15–19.
- [25] C. F. Daganzo, The Cell Transmission Model: a Dynamic Representation of Highway Traffic Consistent With the Hydrodynamic Theory, Transportation research part B 28B (4) (1994) 269–287.
- [26] J. P. Lebacque, 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.
- [27] M. J. Lighthill, 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 229 (1178) (1955) 317 – 345.
- [28] P. I. Richards, Shock waves on the highway, Operations Research 44(1956)42 51.
- [29] S. K. Godunov, A difference scheme for numerical computation of of discontinuous solutions of equations of fluid dynamics, Math. Sb. 47 (1959) 271 – 290.
- [30] W. L. Jin, H. M. Zhang, On the distribution schemes for determining flows through a merge, Transportation Research Part B: Methodological 37 (6) (2003) 521–540.
- [31] C. Tampère, R. Corthout, D. Cattrysse, L. Immers, A generic class of first order node models for dynamic macroscopic simulation of traffic flows, Transportation Research Part B 45 (2011) 289–309.
- [32] V. Knoop, D. De Jong, S. Hoogendoorn, The influence of the road layout on the network fundamental diagram, Transportation Research Records, journal of the Transportation Research Board.
- [33] V. L. Knoop, S. P. Hoogendoorn, Empirics of a generalized macroscopic fundamental diagram for urban freeways, Transportation Research Record: Journal of the Transportation Research Board 2391 (1) (2013) 133–141.
- [34] S. P. Hoogendoorn, V. L. Knoop, H. van Lint, H. L. Vu, Applications of the generalized macroscopic fundamental diagram, in: Traffic and Granular Flow'13, Springer, 2015, pp. 577–583.