EJTIR

Issue 15(2), 2015
pp. 226-242
ISSN: 1567-7141
tlo.tbm.tudelft.nl/ejtir

An Area-Aggregated Dynamic Traffic Simulation Model

Victor L. Knoop[1]
Transport & Planning, Delft University of Technology, Netherlands.

Serge P. Hoogendoorn[2]
Transport & Planning, Delft University of Technology, Netherlands.

Microscopic and macroscopic dynamic traffic models not fast enough to run in an optimization loop to coordinate traffic measures over areas of twice a trip length (50x50 km). Moreover, in strategic planning there are models with a spatial high level of detail, but lacking the features of traffic dynamics. This paper introduces the Network Transmission Model (NTM), a model based on areas, exploiting the Macroscopic or Network Fundamental Diagram (NFD). For the first time, a full operational model is proposed which can be implemented in a network divided into multiple subnetworks, and the physical properties of spillback of traffic jams for subnetwork to subnetwork is ensured. The proposed model calculates the traffic flow between to cell as the minimum of the demand in the origin cell and the supply in the destination cell. The demand first increasing and then decreasing as function of the accumulation in the cell; the supply is first constant and then decreasing as function of the accumulation. Moreover, demand over the boundaries of two cells is restricted by a capacity. This system ensures that traffic characteristics move forward in free flow, congestion moves backward and the NFD is conserved. Adding the capacity gives qualitatively reasonable effects of inhomogeneity. The model applied on a test case with multiple destinations, and re-routing and perimeter control are tested as control measures.

Keywords: *traffic simulation, traffic dynamics, network dynamics, network fundamental diagram, macroscopic fundamental diagram.*

1. Introduction

Nowadays, due to increased communication techniques, traffic control measures can be coordinated over larger areas. For this, control concepts need to be developed. Moreover, the concepts need to be tested, possibly on-line, for which traffic simulation programs are being used. We argued earlier (Knoop et al., 2013b) that for the optimization the larger the area of interest, for instance a network, the larger the look-ahead period should be. Moreover, the simulation time for one run increases if one increases the number of vehicles (microscopic) or links (macroscopic) in order to optimize the control for the larger network. For model predictive control within an optimization loop, many iteration runs are needed. In fact, the number of runs is dependent on the number of control variables, the surface of the solution and the algorithm. For real networks, especially large ones, this might be in the order of thousands. This is infeasible to do in real time using traditional simulation programs. If the calculation times per iteration increase due to the larger network and the number of iterations increases due to the larger

[1] A: Stevinweg 1, 2628 CN Delft, The Netherlands T: +31 152 788 413 F: +31 152 783 179 E: v.l.knoop@tudelft.nl
[2]A: Stevinweg 1, 2628 CN Delft, The Netherlands T: +31 152 785 475 F: +31 152 783 179
E: s.p.hoogendoorn@tudelft.nl

solution space, also off-line applications run into problems. For example, 1000 iterations at 30 mins – a typical time for a macroscopic region-wide model – is approximately 3 weeks; 17,500 iterations of 30 mins is one year. This is (too) long for planning purposes, let alone for operational purposes. Hence, there is a need for a quick model which can simulate a wide area dynamically in a fast computation time.

Within dynamic models, the most suitable option up to now is the link transmission model (Yperman et al., 2006) which calculates the traffic states per link. The model is fast, in the order of real time for a regional size network. This makes it feasible for off line analysis, but running the model in an on-line analysis in an optimization loop is difficult. The model calculation time increases with the number of links included in the model. Besides CPU time, memory issues might play a role in the traffic assignment of many travellers in larger networks.

In this paper we take a different approach and use the Network Fundamental Diagram (NFD, sometimes also referred to as Macroscopic Fundamental Diagram) (Daganzo, 2007) as basis for a dynamic simulation. The NFD considers traffic operations at a higher level. Rather than the traditional microscopic (i.e. vehicle level) or macroscopic (i.e., link level) description, the NFD considers traffic on a network scale. The base element is a cell in which there are many roads in which traffic interacts. The NFD relates the number of vehicles in the area to their averaged speeds.

This concept will be exploited in this paper to form the basis for a dynamic simulation program. The advantage is that it is really fast since the number of elements (i.e., cells) can be low even for large networks. In particular, for very large networks one can consider choosing larger cells in order to keep the simulation time limited. Due to the averaging effects, the accuracy of the description is not necessarily worse than in case of smaller cells, although heterogeneity in the cells should be limited Geroliminis and Daganzo (2008).

This paper aims to develop a model describing the dynamics of aggregated traffic states, applicable to a network with multiple subnetworks and taking the boundary capacity and spillback of congestion into account. Such a model is useful for strategic purposes, as replacement for static models in planning purposes. Moreover, it can be used for on-line optimization of traffic measures. The paper also shows the application of the model, by which the basic traffic properties are being shown. It also is implemented with simple NFD-based control concepts, showing the use of the model for active traffic management.

The remainder of the contribution is organized as follows. In the next section, we first present the background on the NFD and macroscopic traffic flow simulations. Section 3 describes the model. Then, section 4 describes control scenarios. The model and control scenarios are implemented in a case study, as is presented in section 5. Section 6 comments on the results and finally, section 7 presents the conclusions.

2. Background

This section describes the background in traffic control concepts related to the NFD, as well as related traffic simulation concepts describing traffic dynamics.

2.1 Network Fundamental Diagram and its use

It has been shown that on an aggregate level there is a relation between the number of vehicles in an area and their speeds (Daganzo, 2007; Geroliminis and Daganzo, 2008), and hence between the number of vehicles in an area and the internal flows. This is called the Macroscopic Fundamental Diagram or Network Fundamental Diagram (called NFD in the sequel of the paper). The relationship in this concept of the NFD has been proven for homogeneous, recurrent conditions

(Geroliminis and Daganzo, 2008); the same paper shows the relationship between the internal flows and the exiting flows, the production.

Some dynamics have been described, explaining how traffic dynamics cause traffic jams to propagate from one area to the next, and how this creates inhomogeneity and a lower production (Knoop et al., 2013a). Theoretically, this is can be explored by an infinite grid network. Daganzo et al. (2011) present so in a simplification thereof, a two-ring network with periodic boundaries. The models describing the flows from one area to the next are often only considering two regions Geroliminis et al. (2012), but there are extensions to more regions e.g., (Zhang et al., 2013). There, however, the flow is governed by the NFD and could exceed the capacity of the links crossing the boundary between two zones. More importantly, also the concept of limited physical space to accommodate the vehicles in the destination cell is not considered. Hence, in earlier descriptions the notion of queues propagating upstream and possibly influencing other areas is not incorporated. Possibly the only exception is the paper by Yildirimoglu and Geroliminis (2014), which uses a multi-zone network where inflow is limited based on the accumulations in the neighbouring zones. However, in their formulation, no internal gridlocks in a zone are possible, which is one of the characteristics of the NFD.

2.2 *Control using NFDs*
Already several decades ago, the concept of an area-wide fundamental diagram was proposed (Godfrey, 1969). This has been tested for various types of traffic assignment (Mahmassani and Peeta, 1993). After Daganzo reintroduced the concept (Daganzo, 2007) in relation to gating, the topic has gained considerable attention. Gating means that traffic is stopped at the perimeter of an area. This delays traffic outside the "protected zone", but if the travel time gains inside are higher, this measure could be useful nonetheless. The paper by Geroliminis and Daganzo (2008) has shown empirically that the relation between accumulation and production is indeed a very crisp one. Other researchers have further studied the impact of inhomogeneity. For instance, there are simulation observations that productions decrease with inhomogeneity (Mazloumian et al., 2010), empirical observations during strikes and hence large inhomogeneities (Buisson and Ladier, 2009). Daamen et al. (2010) explained the network dynamics in a simulation network and Daganzo et al. (2011) looked at the dynamics of a simplified system. Knoop et al. (2013a) presented an equation to incorporate inhomogeneity in the production function. All in all, though, the attention shows that the research community considers it conceivable that on an network-aggregate level the accumulation and the production have a relationship.

The idea of gating has been studied extensively, for instance by Keyvan-Ekbatani et al. (2012) or Geroliminis et al. (2012). The basic idea is to keep the number of vehicles in an area under the critical level, i.e. the level where the production starts to decrease. Namely, a higher accumulation will then lead to a lower (and not even constant) outflow. This could give considerable travel time gains. Another control concept introduced in relation to the area-wide traffic description is routing based on the NFD (Knoop et al., 2012), also leading to considerable travel time gains. Both of these control concepts will be tested here.

2.3 *Related traffic simulation concepts describing traffic dynamics*
In traffic flow theory, several macroscopic models are available. One of the most intuitively understandable is the cell transmission model (Daganzo, 1994). Consistent with the ideas of a demand and a supply (see also (Lebacque, 1996)), this model describes the evolution of traffic states on a road. The road is split in cells. Up to a critical density, demand is an increasing function. For densities higher than the critical density, the demand is equal to the capacity of the road. The supply has a value of the capacity of the road up to the critical density. For higher densities, the supply decreases. The flow from one cell to the next is the minimum of the upstream demand and the downstream supply.

There are other models, leading to analytical solutions. For instance, Newell proposed one of these (Newell, 1993), which is adapted in the Link Transmission Model (Yperman et al., 2006). Similarly, there are hybrid approaches or solutions in Lagrangian coordinates (Leclercq, 2007). These models are less suited for network models since they rely on the coordinate system moving with the traffic. Whereas there are solutions for multi-class network models, the model is developed for uni-directional flow, and will not work for multiple directions within each cell, for instance with crossing flows.

The above models describe how traffic flows on links. A network model also needs to describe how traffic behaves at nodes. A good overview of node models and their requirements is given by Tampère et al. (2011).

Concluding, we propose here to combine concepts of the cell transmission model and good node models in a Network Transmission Model (NTM).

3. Multi-region aggregated macroscopic modeling

This section describes the traffic flow model. First, section 3.1 describes how traffic is described using aggregated quantities. Then, section 3.2 describes the traffic model.

3.1 Variables and traffic representation

The basic elements of the network description proposed here are subnetworks, called cells in the following description of the computational methodology. Table 1 gives an overview of the most important symbols and their meaning. The basic quantities used in this paper are accumulation *K* and performance *P*, which can be seen as weighted average density and flow, respectively. Note that performance is the flow which exits a network, rather than the internal flows. It has been

Table 1. Variables and their meaning; *A* and *B* refer to cells

Symbol	Meaning
K	Accumulation
P	Production (outflow)
ζ_s	Part of the accumulation heading to destination *s*
$\eta_{s,A,B}$	Fraction of traffic in A heading to *s* which has a route which goes from *A* to *B*
C_A^B	The capacity of the boundary *A*-*B*
S^B	Supply in cell *B*
D_A	Total demand out of *A*
$D_{A,s}$	Demand in A to destination *s*
D_A^B	Demand from *A* to *B*, not accounting for capacity restriction C_A^B
$\tilde{D}_A^B$	Demand from *A* to *B*, accounting for capacity restriction C_A^B
D^B	Total demand to cell *B*

shown that the performance is strongly correlated with the internal flow, the production. The accumulation K in each cell A is the average density K for all links Z in the cell weighted to their length L and the number of lanes l. This total weighting factor is indicated by w:

$$w_A = \sum_{Z \in A} L_Z l_Z \tag{1}$$

The accumulation is now calculated as

$$K = \frac{\sum_{Z \in A} k_Z L_Z l_Z}{w_A} \tag{2}$$

For each cell, it is registered which part of the vehicles (and thus accumulation) is heading towards which final destination s; this is called ζ_s. The routing from cell A to the destination is coded by the next neighbouring cell B in so called destination-specific splitfractions $\eta_{s,A,B}$. All neighbouring cells of A are indicated by the set B. The fact that the split fractions are destination specific ensures the method also works if queues spill back from one destination. Split fractions are chosen over routes because they proved a very large route set, they enable rapid changes of routes (only changes in one cell). Besides, in this way the route set, and hence the computations, do not grow too fast with an increasing number of OD pairs. It also is a natural step in the description of the cell-aggregated traffic model to code the routing on the cell-aggregated level. The splitfraction lies between 0 and 1, $0 \le \eta_{s,A,B} \le 1$, and all vehicles should be heading somewhere: $\sum_{B \in \mathrm{B}} \eta_{s,A,B} = 1$.

The demand is modelled as separate connecting area with infinite boundary capacity to the origin cell. Then, the same traffic operations are applied. The model is not suitable for traffic conditions where the demand to an origin cell is higher than its supply – in other words, all demand arrives at the origin cell. In our formulation, vehicles are assumed to have arrived at their destination once they arrive somewhere in the cell. This could be changed in a future version.

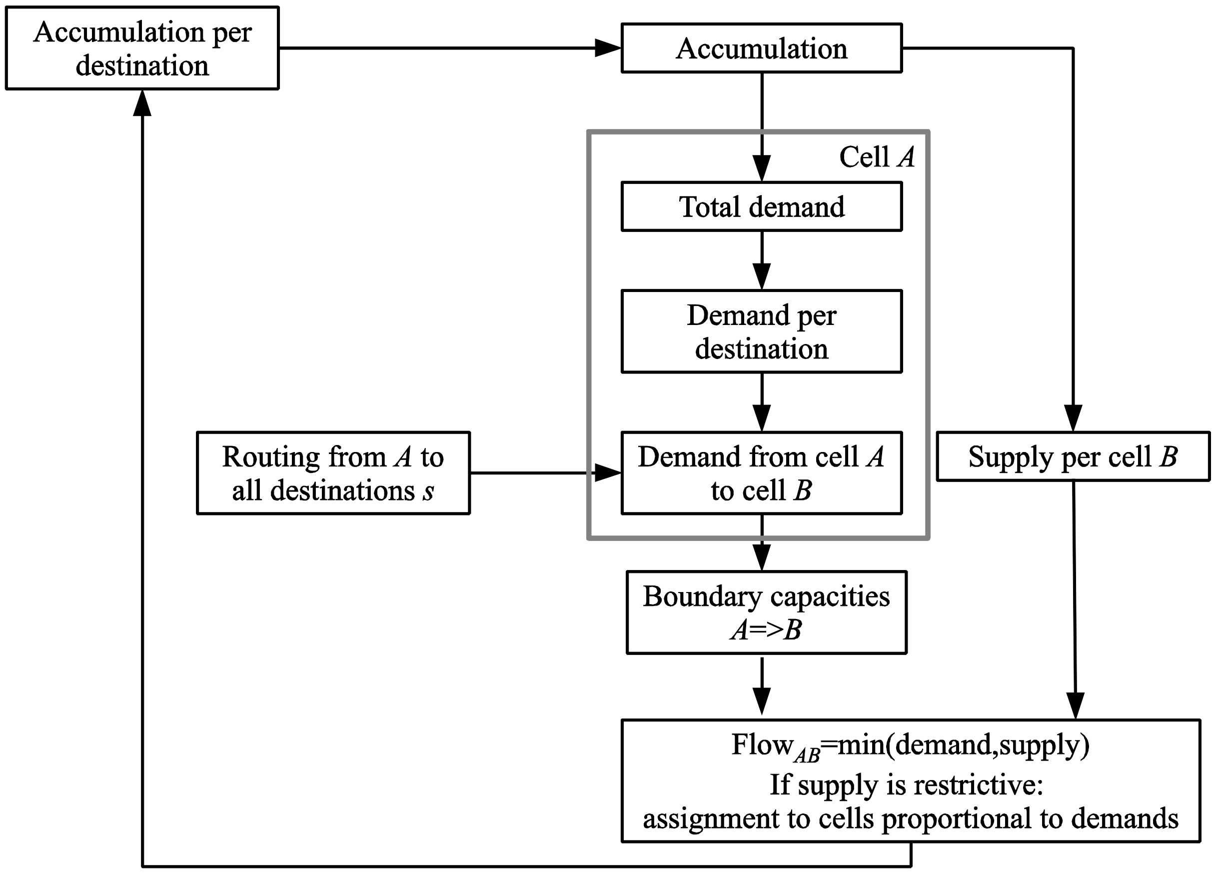
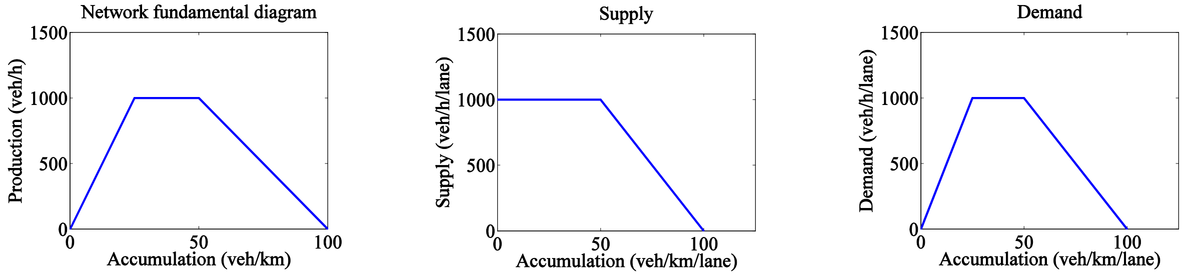


Figure 1. A graphical representation of the steps taken in the computation scheme. The demands are considered for vehicles wanting to leave A, and the supply is calculated for the receiving (neighbouring) cell B, leading to a flow from A to B.

3.2 Dynamics

This section describes the dynamic model, using text and equations. A flow diagram of the model can be found in figure 1. This section only presents the traffic flow model, and assumes the traffic routing as given exogenously. The routing will be presented in section 4.2.



(a) Network Performance Function (b) Supply (c) Demand

Figure 2. The factors determining the flow

The dynamics of traffic are simulated in these subnetworks. The flow from cell A to cell B is determined by the minimum of three elements

1. The capacity of the boundary between cell A and cell *B*, C_A^B; this is determined exogenously

2. The demand from cell A to cell B, D_A^B
3. The supply in cell B, related to the total demand to cell B

Supply
The supply from A to B is determined based a function (Network Performance Function, or NPF) which relates performance P to the accumulation K: $P = P(K)$. Since production and performance are closely related (Geroliminis and Daganzo, 2008), this function will be similar in shape to the NFD (the NPF chosen for illustration purposes in this article is shown in figure 2a). This function has to be determined exogenously, for which are several methods, empirically (Geroliminis and Daganzo, 2008) or theoretically (Leclercq and Geroliminis, 2013). In fact, we can construct a demand and supply scheme similar to the cell transmission model proposed by Daganzo (1994). The supply can be determined in the same way as in the cell transmission model, that is, it is at capacity if the accumulation in the receiving cell is lower than the critical density and equal to the NPF for higher accumulations:

$$S = \begin{cases} P_{\text{crit}} & \text{if } K \leq K_{\text{crit}} \\ P(K) & \text{if } K > K_{\text{crit}} \end{cases} \tag{3}$$

This is shown graphically in figure 2b.

An alternative would be – similar to Dynasmart (2003) – to allow a minimum speed or a minimum flow, to avoid gridlock situations. This is a good possibility, but not tested in this paper. First we want to explore the possibilities of following the MFD. The minimum flow can be introduced in a later stage, for instance when the model is calibrated for a real life case and it turns out that drivers indeed avoid gridlock themselves, for instance by adapting routes dynamically. It is not advised to try to enforce a minimum flow would even into cells with a limited supply (so adapting or overruling the supply as well). This way, the system might react unexpectedly since the jam density can be exceeded. Here, the system dynamics are also different than for a one-dimensional flow where such a minimum flow rule could be applied since the downstream cell could have the same rule, thereby allowing space for the extra number of vehicles.

Demand
Contrary to the cell transmission model, the demand in a cell decreases with an increasing accumulation at values over the critical accumulation. This is due to the impact of internal congestion in the cell, limiting the potential outflow. We thus have:

$$D = P(K) \tag{4}$$

This is graphically shown in figure 2c.

Additionally, a minimum flow can be defined. This would allow a demand even from a completely full cell.

The total demand from cell A to cell B, D_A^B is only a part of the total demand in cell A, D_A. To come to the demand to B we have first to consider the different (final) destinations s, and then combine this to the flow to B. The fraction of the demand in A towards destination s is indicated ζ_s, so the demand in A for each of the destinations is

$$D_{A,s} = \zeta_s D_A \tag{5}$$

For each of these partial demands, the fraction heading to neigbouring cell B is indicated by $\eta_{A,s}B$ The demand from cell A towards cell B hence is

$$D_A^B = \sum_{\text{all destinations}s} \eta_{A,s}^B D_{A,s} \tag{6}$$

This is now limited to the capacity of the boundary between A and B, C_A^B, giving the effective demand $\tilde{D}_A^B$:

$$\tilde{D}_A^B = \min\left\{D_A^B, C_A^B\right\} \tag{7}$$

The total demand towards cell B is determined by adding all effective demands towards cell B

$$D^B = \sum_{A \in \text{neighboring cells of B}} \tilde{D}_A^B \tag{8}$$

This is compared with the supply in cell B. The principle is that the flow from A to B (q_{AB}) is the minimum of demand and supply.

Flows from A to B
If the supply is larger, the flow is unrestricted. However, if the supply is lower, the fraction of the flow which can flow into cell B Ψ^B is calculated:

$$\Psi^B = \min\left\{\frac{S^B}{D^B}, 1\right\} \tag{9}$$

All cells B, neighbours of A, which have effective demand $\tilde{D}_A^B$ larger than zero are combined in set $\mathcal{B}$ We assume that the same restriction holds for traffic to all directions B. In this assumption we assume that the travel infrastructure is not different for the different destinations. (The discussion, section 6, will comment on possible changes on the zoning if it is.) If the infrastructure is the same, a limited outflow to one cell will reduce the speed on the links and create similar congestion, and outflow rates, for all destinations. Compare it to a link with traffic to two direction, and only 50% of the traffic to one direction can be serviced. If the traffic is mixed, traffic to the other direction is reduced by the same amount. In our model, mathematically, we express this restricting factor for cell A Ψ_A, which is hence the minimum of the restriction factors from A to B, in which B is the set of all neighbouring cells of A:

$$\Psi_A = \min_{B \in \mathrm{B}} \left\{\Psi^B\right\} \tag{10}$$

If the supply restricts the flow, the actual flow to cell B is proportional to the demands to cell B for all cells A (limited to the capacity of the boundary AB), and within cell A for traffic to all destinations which move via B. Ψ_A is applied to all demands in congestion. This way, the supply restriction is moved from the supply side to the demand, and we obtain an equation for the flow:

$$q_A^B = \Psi_A \tilde{D}_A^B \tag{11}$$

Applying the destination specific split factor $\eta_{A,s}^B$, this gives the flow from A to B with destination s:

$$q_{A,s}^{B} = \eta_{A,s}^{B} q_{A}^{B}. \tag{12}$$

This flow is assumed to be constant between two consecutive time steps.

The accumulation in any cell A towards destination s can now be updated based on the flows from B to A with destination s, denoted by $q_{B,s}^{A}$ and the flow in the opposite direction, $q_{A,s}^{B}$:

$$K_A^s(t+\tau) = K_A^s(t) + \left(\sum_{B \in \mathrm{B}} q_{B,s}^{A} - \sum_{B \in \mathrm{B}} q_{A,s}^{B} \right) \tau / w_A \tag{13}$$

In this equation, τ is the simulation time step.

4. Control concepts

The Network Transmission Model can be used for traffic control. Two example control applications are shown here, adaptive routing based on the NFD and gating. Both will be discussed in this section.

4.1 Gating

The first idea of control using the NFD, already mentioned by Daganzo (2007), is limiting the inflow into an area. In practice this can for instance be done by changing the green times of the traffic lights at the boundary. For the study at hand, we choose limit the inflow such that the accumulation will not exceed the critical accumulation K_{crit}, i.e. the accumulation from which the performance is decreasing. The number of vehicles that can be added to the cell is:

$$\left(K_{\text{crit}} - K\right) w \tag{14}$$

Under the control we limit the supply (modify eq. 3) such that this is not exceeded

$$S = \max\left\{\left(K_{\text{crit}} - K\right) w / \tau, 0\right\} \tag{15}$$

The consequence is that the vehicles are waiting in the the neighbouring cells. In this paper, we follow Knoop et al. (2012) and propose a reactive control scheme. No predictive control strategies are developed. In the simple control scheme applied here, we simply reduce the supply (eq. 3). These restrictions can be applied to any cell. In the case study (section 5) we vary the cells to which these are applied.

4.2 Routing

Another type of traffic control is routing. Routes can be suggested in-vehicle (if communication is available), or by dynamic traffic signs. We differentiate between a routing only based on the NFD and one which also takes the direction into account.

NFD-based routing

The routing is based on the travel time of vehicles. In this routing scheme, this is determined by the distance a vehicle has to drive in a particular cell and the speed in the cell. For the distance, the vehicle either has to cross the cell completely if the next cell is at the opposite side, or it should go from one cell to another cell which is located at the side. In that case, we consider a shorter distance, as shown for the distance from cell 4 to 8 in figure 3.

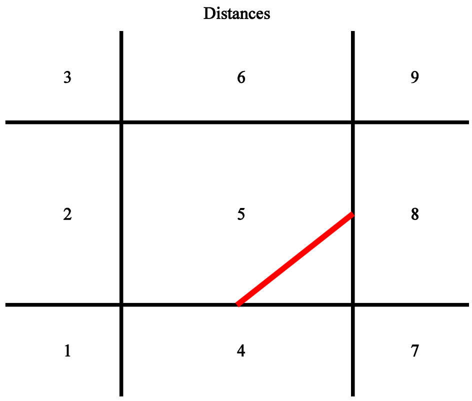


Figure 3. The distances for cells that have to be crossed

For this routing case, we simplify matters and test to which extent the NFD can really hold. In fact, the NFD assumes the cells to be homogeneous areas, so a homogeneous speed is assumed. This speed is derived from the properties of the cell using the fundamental equation $v = P/K$.

From distance and speed the time to go from one neighbouring cell of A to another can be determined. These times are disturbed with a normal error with mean 0 and a standard deviation of 10%. Within this disturbed time, the shortest paths for all cells to all destinations are found using a Floyd-Warshall algorithm. Per cell A it is determined which of the neighbouring cells B is the next cell in the fastest route to destination s. The routing is repeated for different normal random disturbances of the travel times (probit assignment). After these iterations, for each cell A it is stored which fraction of the shortest routes to destination s follows to neighbour B. This fraction, called $\eta_{s,A,B}$ in section 3, is applied in the simulation.

NFD direction based routing

The NFD provides an average speed of the vehicles in the area. Contrary to a basic idea with one area with homogeneous travellers, in reality there are several exiting directions. One could conceive a traffic flow which is limited in one direction, but not in the other (orthogonal, or the reverse direction). The NTM explicitly computes the flow from one cell to the next cell. We therefore propose a routing algorithm which takes these differences into account.

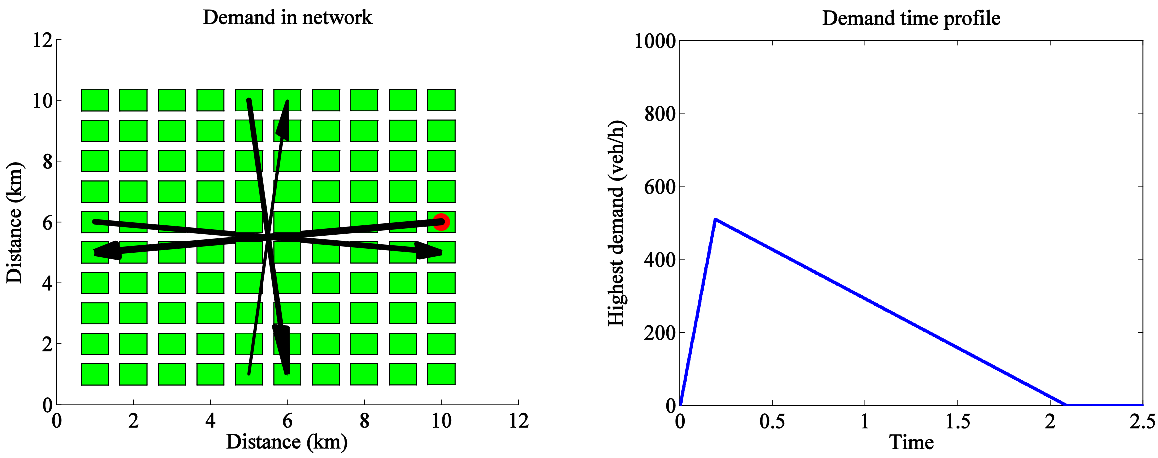
For each destination, the instantaneous average travel time from one cell to the next is determined. This is done by comparing the number of travellers in cell A heading to a neighbouring cell B and compare that with the number of travellers flowing from cell A to B in the current time step. The average expected number of time steps that travellers have to spend in cell A if they are heading to B is the number of travellers heading to B divided by the flow from A to B. The last vehicle in cell A heading to B hence leaves after T:

$$T = \frac{N_A^B}{q_A^B} \tag{16}$$

In which N is the number of vehicles in cell A heading to B.

The direction-specific NFD routing algorithms uses the expected travel time as basis for the route choice. In the routing, the cost of travelling from A to B has to be determined, which we indicate by C_A^B, which will be the travel time in this paper. For the trip from A to B for traffic towards cell s we add two elements: half of the full travel time from A to B in A (T_A^B, calculated with the above equation) and half of the maximum travel time for traffic in B towards destination s: T_B^s. Thus, C_A^B represents travel times from the "middle" of each cell to the middle of the next cell.

We use the same probit routing assignment as explained in section 4.2.1, only with different times to cross cells. The average times per cell are disturbed and using the Floyd-Warshall algorithm the fastest route is determined. This is repeated for several iterations in which the travel times are each iteration disturbed by a different random factor (10% of the travel time). After n iterations, the routing is aggregated. For each cell it is considered which fraction n fastest routes from cell A to destination s goes towards cell B. This is determined for all cells B which are neighbours of A. These numbers give the split rate of traffic for traffic to s in cell A.



(a) Demands in the network (b) Demand time profile

Figure 4. Case study setup

5. Case study

The Network Transmission Model is introduced based on principles and reasoning. In this section, the model, as well as the control principles, are applied to see their working.

5.1 Setup

Network

For the case study we set up a network with 10x10 cells, each representing an area and all having the same characteristics. The cells have a size of 1x1 km and 10 kms of roadway length. The NPF of the cells is shown in figure 2a. The capacity on the boundary between two cells is high enough that it does not restrict the flow. The time step used in the case study is 15 seconds.

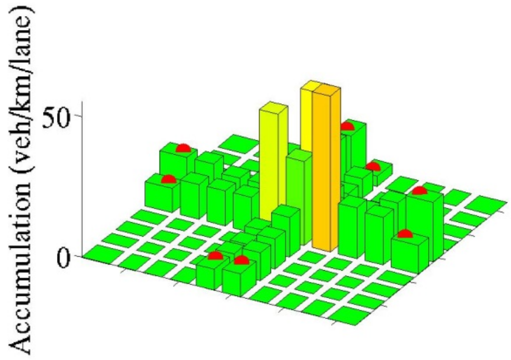
Demand

A cross-network demand is loaded onto the network, shown graphically in figure 4a. The arrow width indicates the size of the demand. The demand at the start of the simulation for directions top-down and left-right is 625 veh/h, to left is 833 veh/h and the demand bottom-up is 312 veh/h. The demand changes piecewise linearly over time, first increasing and decreasing over time. figure 4b shows this demand for the direction right-left, the highest demand. The other demands have a demand profile scaled to their maximum demands. After the demand has decreased to zero, the simulation continues to empty the network. In case of no adaptive routing the traffic might end up in a grid lock situation, in which case the simulation is ended after 7500 time steps.

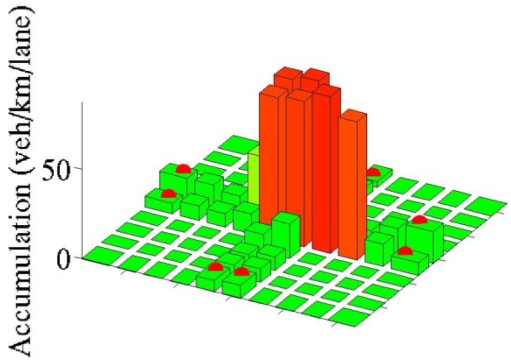
Control

For gating, we test four scenarios: (1) not limit any inflow, (2) limit the inflow in the four centre cells, (3) limit the inflow in the centre cells and the destination cells (4) limit the inflow in the centre cells and their neighbouring cells. For the rerouting several route update times are tested: 10, 50 and 100 time steps, equalling 150, 750, 1500 seconds.

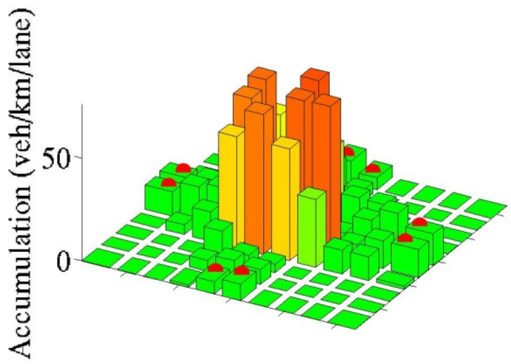
5.2 Results



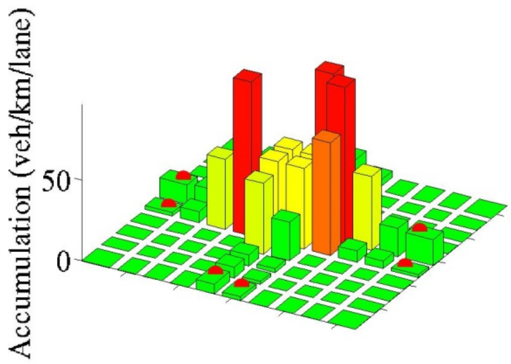
(a) Start up of congestion



(b) Jamming in middle cells, and spillback



(c) Rerouting to neighbouring cells



(d) perimeter control: the middle cells do not get to over critical accumulations

Figure 5. Snap shots of the traffic operations – colour indicates speed, and bar height accumulation.

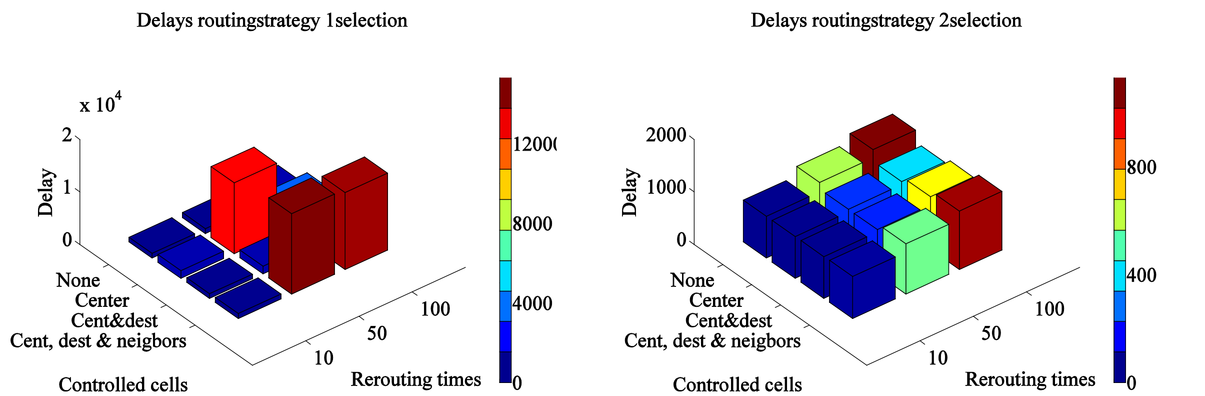
Figure 5 shows snapshots of the model during the simulation. The model shows a clear build-up of congestion, and spill back to the upstream cells. This is shown in the progress of traffic states in figure 5a and figure 5b. This leads to very congested middle cells, and as is visible in figure 5b the upstream cells also get congested. This is what is expected based on traffic flow theory. Also other aspects are present. The dynamics show that traffic is moving forward, and that congested patterns are moved in the opposite direction of the traffic: they grow at the "tail" of the queue (in quotes since there is not a single queue).

By assuming homogeneity within the cells, information can travel one cell per time step. This speed can exceed the speed of information in real world, which is limited by the free flow speed or the wave speed. Note however, that these effects are low. Consider for instance vehicles entering a cell. If only a few vehicles enter a cell, and then because of homogenization, they are spread over the cell, the average density is very low. In the next cell, this will be even lower. A similar reasoning can be given for jams spilling back at the downstream end of a cell, of which the influence at next cell in the next time step (so for the fast traveling wave) will be diluted by the homogenization. Other discretization methods or coordinate systems have been helped to overcome these problems in in directional traffic streams. It goes beyond the scope of this paper to further investigate the numerical effects of the solution scheme.

If rerouting is applied, see figure 5a, once the congestion starts to occur in the middle cells, the traffic is rerouted to the cells around these middle cells. Still, congestion in the middle cells occurs. With gating, figure 5d, the middle cells do not exceed the critical accumulation, but as a consequence the cells around these cells can get congested. Note that if a cell gets completely congested, due to the decreasing outflow in the NFD, this might take a long time to recover from – a typical phenomenon related to gridlock effects.

The total delays in the network under different routing strategies are shown in figure 6a. It seems to show a bimodal distribution for the delays: for some settings the delay is high, and for others it is low. This is caused by one situation in which the traffic control is able to prevent gridlock, and another in which it is not. Generally, it is noted that by gating one can prevent gridlocks. It is remarkable that for longer update times, the application of gating to more cells increases the gridlock effects. This could be caused by the fact that there are less vehicles allowed in a particular cell, which means that by the time of a routing update, the cell is full and no other vehicles can enter. This increases delay substantially.

The results for the NFD destination specific routing (section 4.2.2, figure 6b, gives lower delays than if only NFD routing is applied. In fact, with fast updates, the delay is as low as 500 vehicle hours. This increases to higher values with less frequent route updates.



(a) NFD based routing, see section 4.2.1

(b) NFD direction based routing, see section 4.2.2

Figure 6. The delays for different routing strategies. Note the scale for the z-axis and colour is different for both figures.

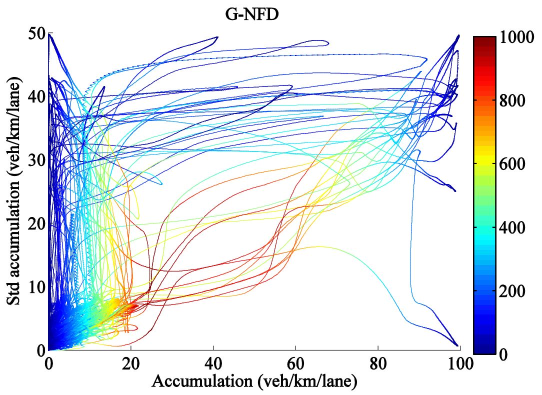


Figure 7. The performance (veh/h/lane) as function of the accumulation and the spatial spread of density, the Generalised NFD

Earlier it has been found (Knoop and Hoogendoorn, 2013) that the flow can be considered a function of the accumulation (concave) and of the standard deviation of the accumulation (linearly decreasing). This is called the Generalised Network Fundamental Diagram (G-NFD). In the G-NFD, the performance can be plotted as colour graph from the top in the accumulation-accumulation variation plane. In this case, we approximate the accumulation variation by the standard deviation of the accumulations of the neighbouring cells for each cell in each time step. We are aware that the standard deviation of accumulation between cells is not the same as the standard deviation of the densities in the cell. However, it can be a good approximation, as preliminary work shows (Knoop et al., 2011). The simulation also gives the accumulation and the production. Figure 7 shows this in the G-NFD.

As theoretically and empirically shown (Knoop and Hoogendoorn, 2013), also in the Network Transmission Model the performance increases and decreases with accumulation (due to the NFD). Moreover, it decreases with increasing inhomogeneity. The modelling principles causing this effect are the following. If a neighbouring cell is congested, the flow in the cell can be reduced even though according to the NFD and the demand the flow can be higher. Note that this is a qualitative similarity, and the amount of decrease is a matter of calibration of the fundamental diagram. What is important in this stage is that the general pattern (concave in accumulation, decreasing in standard deviation in accumulation) is similar.

6. Discussion

The Network Transmission Model introduced in this contribution presents a traffic model on an aggregated level. This section discusses some properties of the model, its use and the further scope.

Basic requirements we expect from the model to call it face valid are (i) if accumulation in a cell increases, the outflow decreases (ii) congestion propagates upstream, (iii) traffic conditions in free flow propagates downstream. The first requirement is met due to the fact that the model is implementing an NFD, and the demand is in overcritical situations a decreasing function of density. The other two requirements are fulfilled by this model due to the demand-supply structure. We hence call it face valid. There are, however, some further steps to consider.

First of all, the model is only suited to handle multi-direction traffic cells. Extensions or alternative models can be developed of to handle mono-direction traffic cells. In particular, one could model per cell the directional links separately. If that is done, the jam density would only be reached in the links towards the downstream end and not in the other links (which indeed must be present). Since there are then also links with low densities, the accumulation is not at a level where the demand is reduced to zero. In general, this shows that the application for the model for strongly directional traffic is limited. If one wants to do so, a minimum demand for high accumulations should be guaranteed, as suggested in section 3.2.2. Moreover, the model in its current form is not developed to handle traffic from zone A to zone A. A possible extension would be to include such self-demands to the cell, by adding those to the accumulation of A.

Secondly, it has been shown that the spread of congestion has an effect on density. This can be analysed Mazloumian et al. (2010), explained Daganzo et al. (2011) and formalized in a equation Knoop et al. (2013a). Some of the features of this performance decrease in case of less homogeneous networks are captured by the model because of the limited capacity over the cell boundaries, as shown in figure 7. It is unclear whether this captures all these effects. A comparison with a real network, homogeneously and inhomogeneously loaded, should show so.

As third remark, the NTM is introduced based on theoretical considerations. Whereas the model shows a reasonable propagation of traffic flows and traffic jams, only a full test of the model can show its correctness. The model should be tested against a real network – either simulated or preferably measured real life. Characteristics which should be correct are the traffic flows, densities, and the dynamics of the congested areas. There are many – observed, but also unobserved – variables which need calibration, more than for stretch of road. The combination of calibration of all these parameters (NFDs, OD, route choice) is a challenging task on its own, which is why it is not presented in this paper along with the fundamentals of the model. Once developed, the same techniques used for calibration can be used to carry out a validation for a different day or a different area. This falls outside the scope of the current paper.

Improvements of the model can lie in the separation of roads of different hierarchy or direction into different NFDs within the same spatial scope. Moreover, the travel times to each direction might differ, which is not yet taken into account in this model. These are all follow-up improvements of this model, which will be tested later along with the real data.

7. Conclusions

This paper introduced the Network Transmission Model describing the traffic flow dynamics on an aggregate level. The network was split into cells and for the traffic flow dynamics a numerical approach based on the MFD was introduced. This model overcomes the problems with temporal congestion that one faces if using static models. The model is face valid, but further studies should test the model and calibrate and validate it against real data or more often used traffic simulation programs. Also the value of several improvements upon this base concept like multi-destination and roads at different hierarchical levels should be tested.

The model was applied in a dynamic test case in which we analysed to which extent the model could be used to predict the impact of traffic control. The model was capable of reducing the delay based on adaptive routing. Also, the concept of gating was implemented and showed to have an impact on the travel times. Its impact was found to be smaller than optimizing the routing. For further research, in a calibrated system the marginal effects of each of the control concepts needs to be studied. Moreover, future research will develop more ingenious control schemes than the simple rule-based control schemes shown here. A model predictive control scheme seems to fit very well with the Network Transmission Model.

Acknowledgements

This research was sponsored the Netherlands Organisation for Scientific Research (NWO) in the project "There is plenty of room in the other lane".

References

Buisson, C. and Ladier, C. (2009). Exploring the impact of the homogeneity of traffic measurements on the existence of the macroscopic fundamental diagram. *Tranportation Research Records, Journal of the Transportation Research Board.*

Daamen, W., Loot, M. E. J., and Hoogendoorn, S. P. (2010). Empirical analysis of merging behavior. In Proceedings of the *89th Annual Meeting of the Transportation Research Board.*

Daganzo, C. (2007). Urban gridlock: Macroscopic modeling and mitigation approaches. *Transportation Research Part B: Methodologica*l, 41(1):49–62.

Daganzo, C., Gayah, V., and Gonzales, E. (2011). Macroscopic relations of urban traffic variables: Bifurcations, multivaluedness and instability. *Transportation Research Part B: Methodological*, 45(1):278–288.

Daganzo, C. F. (1994). The Cell Transmission Model: a Dynamic Representation of Highway Traffic Consistent With the Hydrodynamic Theory. *Transportation research part B*, 28B(4):269–287.

Dynasmart (2003). DYNAMSMART-X – User's Guide. Technical report, University of Maryland.

Geroliminis, N. and Daganzo, C. F. (2008). Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transportation Research Part B: Methodological*, 42(9):759–770.

Geroliminis, N., Haddad, J., and Ramezani, M. (2012). Optimal perimeter control for two urban regions with macroscopic fundamental diagrams: A model predictive approach. *IEEE Transactions on Intelligent Transportation Systems*, 14(1):348–359. ISSN 1524-9050.

Godfrey, J. (1969). The mechanism of a road network. *Traffic Engineering and Control*, 11(7):323–327.

Keyvan-Ekbatani, M., Kouvelas, A., Papamichail, I., and Papageorgiou, M. (2012). Exploiting the fundamental diagram of urban networks for feedback-based gating. *Transportation Research Part B: Methodological*, 46(10):1393–1403.

Knoop, V. L. and Hoogendoorn, S. P. (2013). Empirics of a generalised macroscopic fundamental diagram for urban freeways. . *Transportation Research Record: Journal of the Transportation Research Board*, 2391, p. 133-141

Knoop, V. L., Hoogendoorn, S. P., and Van Lint, J. W. C. (2012). Routing strategies based on macroscopic fundamental diagram. *Transportation Research Record: Journal of the Transportation Research Board*, 2315(1):1–10.

Knoop, V. L., Hoogendoorn, S. P., and van Lint, J. W. C. (2013a). The impact of traffic dynamics on the macroscopic fundamental diagram. *In Proceedings of the 92nd Annual Meeting of the Transportation Research Board*.

Knoop, V. L., Van Lint, J. W. C., and Hoogendoorn, S. P. (2011). Data requirements for traffic control on a macroscopic level. In Proceedings of *2nd International Workshop on Traffic Data Collection & its Standardisation*.

Knoop, V. L., van Lint, J. W. C., Vries, J., Kester, L., and Passchier, I. (2013b). Relationship between application scale and maximum time latency in intelligent transport solutions. *Transportation Research Record: Journal of the Transportation Research Board*, 2380, p 1-9

Lebacque, J. P. (1996). 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.

Leclercq, L. (2007). Hybrid approaches to the solutions of the lighthillwhithamrichards model. *Transportation Research Part B: Methodological*, 41(7):701 – 709. ISSN 0191-2615.

Leclercq, L. and Geroliminis, N. (2013). Estimating mfds in simple networks with route choice. In Hoogendoorn, S. P., Knoop, V. L., and Van Lint, H., (Ed), *Proceedings of the 20th International Symposium on Transportation and Traffic Theory*.

Mahmassani, H. and Peeta, S. (1993). Network performance under system optimal and user equilibrium dynamic assignments: implications for advanced traveler information systems. *Transportation Research Record: Journal of the Transportation Research Board*, (1408):83–93.

Mazloumian, A., Geroliminis, N., and Helbing, D. (2010). The spatial variability of vehicle densities as determinant of urban network capacity. *Philosophical Transactions of the Royal Society A*, 368:4627–4647.

Newell, G. F. (1993). A simplified theory of kinematic waves in highway traffic, part ii: Queueing at freeway bottlenecks. *Transportation Research Part B: Methodological*, 27(4):289 – 303.

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

Yildirimoglu, M. and Geroliminis, N. (2014). Approximating dynamic equilibrium conditions with macroscopic fundamental diagrams. In Proceedings of *the 93rd Transportation Research Board Annual Meeting (TRB). Washington: Transportation Research Board*.

Yperman, I., Logghe, S., Tampère, C. M. J., and Immers, B. (2006). Multicommodity link transmission model for dynamic network loading. In Proceedings of *the 85th Annual Meeting of the Transportation Research Board*.

Zhang, L., Garoni, T. M., and de Gier, J. (2013). A comparative study of macroscopic fundamental diagrams of arterial road networks governed by adaptive traffic signal systems. *Transportation Research Part B: Methodological*, 49:1–23.